

Welfare economics and bounded rationality: The case for model-based approaches*

Paola Manzini

Marco Mariotti[†]

University of St. Andrews and IZA

Queen Mary University of London

This version: March 2014

Abstract

In this paper we examine the problems facing a policy maker who observes inconsistent choices made by agents who are boundedly rational. We contrast a model-less and a model-based approach to welfare economics. We make the case for the model-based approach and examine its advantages as well as some problematic issues associated with it.

J.E.L. codes: B4, D03, D6.

Keywords: Bounded Rationality, Welfare, Choice Theory.

*We have greatly benefitted from exchanges with Douglas Bernheim and the comments of the Editors, for which we are grateful. We also thank the audience at the San Sebastian Summer School in Welfare Economics at which we vented some preliminary ideas contained in this paper. Mariotti gratefully acknowledges financial support through a Leverhulme Fellowship. Both authors acknowledge financial support from the ESRC through grant RES-000-22-3474.

[†]Corresponding address: Paola Manzini, School of Economics & Finance, Castlecliffe, The Scores, University of St Andrews KY16 9AL, UK. Email: paola.manzini@st-andrew.ac.uk, m.mariotti@qmul.ac.uk.

1 Introduction

A world in which agents are boundedly rational and thus may make inconsistent choices is problematic for the welfare economist. In the standard approach to welfare economics, choice behaviour guides welfare considerations. For a standard preference maximiser there is no conflict between any sequence of choices: for example, if x is chosen and y is rejected in some circumstance, it never happens that y is chosen in the presence of z and z is chosen in the presence of x . This feature makes it possible to associate in a direct manner observed choices with a transitive and complete *binary relation R that is being maximised* by choices. The relation R connects x with y if x is chosen when y is available. The binary relation R is usually, but not necessarily interpreted as the agent's 'revealed' preference. Whatever the interpretation of R , it is R that is postulated to guide a planner's welfare rankings in standard welfare economics. The planner, when acting on behalf of the agent, will also seek to maximise R (whether or not the agent is 'seeking' to maximise R). Or at least the planner, when choosing policies that affect the agent, will take R into account. For example, if Pareto optimality is a criterion for the planner, the R of each agent will be the one inputed in the Pareto optimality check.

If instead the choices of the agent are generated by some boundedly rational procedure, they may not provide any consistent guidance about welfare rankings. Choices may fail to maximise a binary relation that is transitive and complete. Choices may indeed fail to maximise any binary relation at all. This destroys the traditional connection between choice data and welfare analysis. If one wants to perform welfare economics, some change in approach from the traditional one is needed. In this article we discuss this problem.

We adopt the working hypothesis that, as in traditional welfare economics, choice data are at the heart of economic welfare analysis and that they can be exploited to make welfare comparisons between various situations. In particular, we do not assume the direct observability of 'happiness' or, more in general, 'mental states'. Therefore we will not consider assertions of the type 'alternative x should be implemented rather than alternative y because it will make the agent happier'.¹ The justification for implementing an alternative rather than another will always lead back to the agent's choices. However, as we shall see, we will also support committing to a hypothesis on the cognitive process that underlies choices. In other words, we advocate the use of explanations that are not 'as if'. For this purpose, choice data alone may not be enough. Theories of bounded rationality are typically weaker (that is, compatible with more patterns of choice data) than the standard model. In addition, it may be difficult to tell one theory of bounded rationality from another on the basis of choice data alone. So we do not dismiss as irrelevant data different from choices, such as verbal reports or direct information on the

¹See Bernheim [5] for a critical discussion of the alternative hypothesis that happiness is observable.

cognitive processes of decision makers. We argue that such data may be useful in an ‘auxiliary’ role: they help the observer to make educated guesses about the *reasons* for the agent’s choice, reasons that may be welfare-relevant. Such ‘reasons’ may or may not take the form of standard preference maximisation - typically, preferences are only one of the components of the process that drives choices. The crucial distinguishing feature between the ‘reasons’ we consider and other similar concepts like ‘happiness’ or mental states is that by reason we mean some cognitive process (as opposed to a given state) that, while unobservable, has observable correlates in choice that are amenable to testing. Sen’s polite guest, who is supposed to never choose the uniquely largest item (see Sen [38]), will reveal himself as a fraud as soon as he picks the last remaining piece of cake on the tray, or, in a subtler way, when he fails the restrictions on observable choice data that a politeness-motivated process of choice implies (Baigent and Gaertner [3]).

Within these boundaries, we propose the following broad categorization into two idealised types of strategies to deal with the welfare problem:

1. model-less approach ;
2. model-based approach.

The model-less approach is identified with the influential work of Bernheim and Rangel [6] (henceforth BeRa) and Bernheim [5], though others have contributed in this vein (e.g. Nishimura [33]). This approach is characterised by the fact that it does not rely on an explicit model of decision making. The model-based approach relies instead on some specific model of decision making. Just like standard preference maximisation, normally such a model still identifies some restrictions on choice data: that is, some data are compatible with the postulated model, and others are not. So it is possible to reject models on the basis of observable data. But, as argued above, the whole set of relevant data may include both choice data (in which case revealed-preference types of empirical techniques that are standard in economics can be employed) and auxiliary ones. The welfare conclusions that are derived in this way are model-dependent, in that the same choice data can have different welfare implications according to the cognitive process that has generated them. To put it bluntly, choosing x over y cannot, in this approach, be automatically interpreted as evidence of the welfare superiority of x over y : such an interpretation needs to be mediated by a specific model and by the empirical evidence supporting that model. For example, if it is hypothesised that the agent suffers from imperfect attention (as in Masatlioglu, Nakajima and Ozbay [29] and Manzini and Mariotti [25]), the choice of x over y might simply mean that the agent has not paid attention to y , not that x is in any sense ‘better’. The task in this case will be, first, to validate a specific model of imperfect attention empirically (that is, to check whether

the data are compatible with such a model) and, secondly, if the model receives empirical support, to attempt to infer from the data what drove the choice of x over y (inattention or ‘preference’).

The model-less approach is explained and defended at length in BeRa and Bernheim [5]. Here (section 2) we limit ourselves to a brief discussion of it, and to reporting some critical developments that we formulate more technically elsewhere (Manzini and Mariotti [26]). Our main purpose in this paper is instead to articulate the possible advantages of a model-based approach, discuss some of the problems associated with it and possible ways to ameliorate them.

2 The model-less approach: which welfare relation?

BeRa have proposed a theory of choice-based welfare analysis that applies irrespective of whether choice data come from fully rational or from boundedly rational agents. BeRa aim to *define* welfare as choice, and avowedly not as the realisation of objectives inferred from choices.

They write for example:

‘choices provide appropriate guidance because they are choices, not because they reflect something else’ (BeRa p. 2).

‘non-instrumental justifications maintain that choices are normatively compelling simply because they are choices; hence it is possible to define welfare in terms of choice without implicitly invoking other objectives.’ Bernheim [5] (p. 2)

‘we propose a generalized welfare criterion that *respects choice* directly, without requiring any rationalization involving potentially unverifiable assumptions concerning underlying objectives and their relationships to choice.’ (Bernheim p. 2, our emphasis).

BeRa’s main methodological device is that of ‘choice with ancillary conditions’ in which a *choice situation* is not simply described as a menu S from which the agent chooses, but as a pair (S, d) , where S is a menu and d is a description of ‘ancillary conditions’ that are (or should be if the agent were rational) extraneous to the evaluation of choice objects. This framework is essentially the same as Salant and Rubinstein [36]’s ‘choice with frames’: what the latter call a ‘frame’ BeRa call an ‘ancillary condition’. For example, a frame or ancillary condition d might indicate a default option, or whether the agent, when stating his willingness to pay for an object, has been prompted beforehand to state his National

Insurance Number,² or the order in which alternatives are listed in the actual description of S , or the position of the alternatives on a supermarket shelf.

BeRa introduce two binary relations that, they argue, are natural extensions of the standard complete and transitive relation R to this more general setup. In particular they suggest that the following binary relation P^* is both a ‘correct’ welfare criterion to use in domain of choice problems with ancillary conditions and one that will always yield an answer. An alternative x is in the relation P^* with y if x is sometimes chosen when y is available (that is, there is at least one choice situation (S, d) for which S contains y and x is chosen), while y is never chosen when x is available.³ They define an alternative x to be a strict welfare improvement over y whenever it is the case that xP^*y . An alternative x is a weak welfare optimum in a menu S if and only if there is no alternative in S that weakly improves on x . The advantage of using P^* as a welfare criterion is that it is acyclic, and as a consequence it never fails to pick welfare optima from any choice situation.

The problematic aspect of this approach is the issue of what ‘respecting choice’ really means.

As explained in the introduction, if welfare is defined as choice, the ideal situation is one where the choice behaviour of the agent identifies a binary welfare relation that is *de facto* being maximised by the agent. This relation might be required to satisfy additional properties, but the basic preliminary question is: When is choice with ancillary conditions consistent with the maximisation of a binary relation? Unfortunately this consistency requirement turns out to be extremely demanding: it is in a sense tantamount to requiring the agent’s behaviour to be invariant with respect to ancillary conditions. It is in fact possible to show (Manzini and Mariotti [26]) that choices with ancillary conditions maximise a binary relation if and only if *any alternative which is chosen, in some choice situation, in the presence of any other alternative in a set (where the choice situation may vary depending on the alternative), must in fact be chosen from that set whatever the ancillary conditions that accompany it.* For example, if apricot jam is chosen over some of the other jams when it is on the left of the supermarket shelf, and it is chosen over the rest of the available jams when it is in the middle of the shelf, it must also be chosen (possibly together with other jams) when it is displayed on the right of the shelf.

This is a very strong consistency requirement on behaviour, because it essentially prevents ancillary conditions from playing *any* substantive role. It is logically impossible to have a theory of welfare which is valid also for situations in which the agent exhibits some even very mild form of inconsistency in choice *across* ancillary conditions, and in

²See e.g. Ariely, Loewenstein and Prelec [2].

³Formally, xP^*y if and only if there exists no set in which y is chosen in the presence of x . Given that BeRa’s choice of domain includes binary sets, then it necessarily follows that if xP^*y then x is chosen at least in the binary set for choice to be non empty.

which the agent is never ‘overruled’. At least some of the agent’s choices will have to be overruled, as a matter of logic. There is no escape from this. So the real question is: how should choices be overruled?

2.1 Padding and trimming

BeRa have chosen one particular route to overcome this problem, one that they call ‘inclusive libertarian’. It is inclusive in that it never overrules the agent by eliminating from the set of welfare optima alternatives that the agent himself would choose. And it is libertarian in the sense that it uses only the agent’s choices to determine the set of welfare optima. This approach falls into a category that we might call that of ‘padding strategies’ in which the set of choices that the agent actually makes are automatically declared welfare optimal in that situation, and then they are ‘padded’ with other choices, by sometimes overruling the agent only in the sense of declaring some alternatives welfare optima (together with the choices of the agent) *even when the agent would not choose them*. In BeRa’s approach this happens whenever the agent has chosen that ‘padding’ alternative in some *other* choice situation. For example, if apricot jam was once chosen over cherry jam when it was on the left of the shelf, it should be declared welfare optimal in the set of jams even if it is not chosen when it is on the right of the shelf. This grounding of the planner’s selection on the set of choices that the agent sometimes makes is the libertarian aspect of the approach, but this should not obscure the fact that some of the agent’s choices are overruled (which, as explained, is a logical necessity). There are some choices that the agent would not make *in that situation* but that the planner would instead declare acceptable (this leaves open the question of which choice the planner would ultimately select among the acceptable ones, but we don’t dwell on this issue here).

For a padding solution different from that chosen by BeRa, consider a *paternalistic* view, in which the planner would pad the set of welfare optima on considerations external to the agent’s own choices. This solution is ‘inclusive’ but not ‘libertarian’. For solutions that are not padding, consider ‘trimming solutions’, in which the agent is only overruled by declaring some alternatives *not* welfare optima, even in principle, even in some choice situations in which he would choose them (while in the padding approach all choices made by the agent in some situation are acceptable at least *in principle*). This is in fact the route more often taken in alternative, model-based, theories. These theories lay a claim to identifying the cognitive process that underlies choice and thus to identifying which choices genuinely have a welfare value.⁴

⁴Modern experimental techniques allow for instance the collection of ‘intermediate data’ on ‘intermediate’ stages of decision making, both in experimental settings and market settings. These can be used to identify partially or completely the unobserved procedural components, see e.g. Caplin and Dean

To summarise, there exists no theory that can have an absolute claim to respecting choice when choice is nontrivially dependent on ancillary conditions. We argue that the focus should thus be on *the best way to disrespect choice*. Any such choice will presumably be made, even in the model-less approach, on the basis of a meta-theory about the agent’s motivation, even if that theory is not spelled out in the form of a model.

It is this kind of considerations that leads us to study the alternative, model-based approach.

3 Model-based, falsifiable explanations of behavior: advantages

The core of the approach discussed in this section is the premise that, if some choices are merely the expression of the bounded rationality of the agent, for the purposes of welfare analysis it is better not to use directly those choices as a basis for welfare rankings. Rather, a model should be used to sift carefully through choices. We will first illustrate some advantages of a model-based approach. Later we will study three classes of potential failures that may affect model-based welfare analysis, providing some indications on how to overcome these difficulties.

Let’s start with an illustration of the contrast between model-based and model-less welfare analysis in an example offered by BeRa themselves (p. 13).⁵ Suppose that a choice set $A = \{x_1, \dots, x_{30}, n\}$ consists of thirty types of jams (the x_i), plus the option of choosing nothing (n), and we observe that n is chosen from A . However, suppose we also observe that x_1 (strawberry jam) is chosen when the only other alternatives are x_i and n , for all $i = 2, \dots, 30$. In this case both x_1 and n are the strict welfare optima in A , but this ambiguity in welfare recommendations seems disturbing. Can it really be a matter of indifference for the planner to give the agent strawberry jam or nothing? We suggest that an appropriate model of decision making should be sought that formalises (and allows to uncover) the agents’ *reasons* for his conflicted behaviour. The model should allow one to take a stance on the welfare superiority of either n or x_1 . For example, one might plausibly hypothesise that the agent suffers from ‘cognitive overload’, as argued by Iyengar and Lepper [21]. Such a model would need to be validated by the rest of the available evidence, both choice based and non-choice based. After such a validation, it would then

[9]; Eliaz, Richter and Rubinstein [16]; Costa-Gomes, and Crawford [12]; and Costa-Gomes, Crawford and Broseta [13]. Furthermore, the development of econometric techniques allows the identification of e.g. consideration sets from market data, as in Chiang et al. [11]; Manski [28]; and van Nierop, Paap, Bronnenberg, Franses and Wedel [32].

⁵Based on Iyengar and Lepper [21]’s celebrated ‘jam study’.

be natural to ignore the choices from large sets for welfare purposes, because under the assumption that the model is true, those choices express the cognitive limitations of the agent rather than anything associated with welfare.

3.1 Which model?

Although we will refer to various specific models to exemplify the issue raised (just a small subset of the recent, rapidly expanding literature in this field), we are of course focussing on economic modeling as a methodology, where the ex-ante choice of model will be suggested by the specific situation at hand, and the ex-post choice between competing models/explanations will be settled by testing. In this respect, the class of models we consider falls squarely within the classic economic tradition, in that we only consider models that are testable. All the boundedly rational frameworks we consider posit that the economic agent follows some procedure/heuristic, which can be characterised by means of properties on observed choices. In other words, all these models show the equivalence between a postulated, unobservable choice procedure and observable patterns in the agent's choices. This is a crucial feature that sets these models apart from the literature on 'fast and frugal' heuristics program pioneered by Gerd Gigerenzer and Peter Todd and their adaptive toolbox. While the 'fit' of these latter models to the problem at hand can be measured, the absence of an axiomatization makes it extremely hard to falsify them.⁶

Let's now examine more systematically what we see as the major benefits of a model-based approach.

3.2 Advantage 1: information efficiency

In the model-based approach, the economist's task is to provide the policy maker with a 'box' of positive modelling tools from which he can pick according to the scientific evidence in support of the model and his best contextual judgement. BeRa explicitly reject this view when they argue for an inclusive libertarian relation. In their approach, the only means available to disqualify a chosen alternative from the set of welfare optima is thus to *disqualify the choice situation itself* in which it is chosen. But this technique may sometimes be wasteful. The reason is that some choice situations which are 'suspect' may nevertheless provide information about the decision mechanisms used by the agent when 'crossed' with non-dubious choices. The observations of behaviour in choice situations have two uses in particular:

1. to infer the cognitive variables that influence choice; and

⁶Of course, the fact that fast and frugal types of procedure tend not to be axiomatised does not mean that they *cannot* be axiomatised. See e.g. Manzini and Mariotti [23].

2. to test the theory.

For example, concerning point (1) consider the Revealed Attention model introduced by Masatlioglu, Nakajima and Ozbay [29], according to which an agent only pays attention to a potentially strict subset of the available alternatives, and only maximises his preferences on that subset (the ‘consideration set’). The crucial hypothesis of this model is that the removal from a menu of an alternative which is overlooked cannot detract attention from an alternative which is paid attention to: for example, if you normally only pay attention to the first page of a Google search, removing items from the second page does not make you cease to pay attention to the items on the first page. Similarly, if alternatives are listed and you usually only look at the first three items in a list, removing the tenth item does not alter the set of items to which you pay attention. Suppose now that, when choosing from a menu $\{x, y, z\}$, an agent is observed to select x , whereas when only x and z are available, the agent chooses z . Under the assumption that the Revealed Attention model is true, these choices contain the information that y is in the consideration set of the agent in the larger menu even if it is not chosen from it, for otherwise it could not have affected the agent’s choice when removed. An observer can thus infer that the choice of x when y is available must be determined by his preference for x over y , rather than by some cognitive limitation. The choice from $\{x, z\}$ may well be ‘spurious’ (i.e. determined by attention rather than preference) but eliminating from the domain the choice situation that generates it, as BeRa would recommend, would be tantamount to throwing information away. The spurious choice situation can be *used* to derive insights on the cognitive process.

On point (2), a spurious choice situation can also be used to test the model non-parametrically. Consider for example the ‘Categorize Then Choose’ model (Manzini and Mariotti [24]). A simple revealed preference axiom characterises the following procedure for choice: the agent first ‘simplifies’ the problem by eliminating some subsets (categories) of the choice situation he is facing; then he picks the maximal element according to his preference among the surviving alternatives. The Categorise Then Choose framework views only choices from binary sets as definitely expressing preference, whereas it views choices from larger sets as expressing a mixture of preference and a categorisation strategy to cope with complexity (e.g., rather than comparing pairwise all jams available in the supermarket, you first decide on a brand, and then pick your favourite taste produced by that brand). The model is characterised by a property of a choice function called Weak WARP (both the model and the axiom will be described more precisely in section 4.1). This property involves comparing choices from large sets with choices from binary sets. A violation of Weak WARP allows an observer of choice data to reject the model, and an observation of such a violation is a composite event involving choices not reflecting pref-

ferences, so that the a priori elimination of choice situations where the agent is ‘confused’ by complexity would make the model unfalsifiable.

3.3 Advantage 2: Precision

The crucial advantage of having a fully fledged model of behaviour is that if we understand the mechanism that leads an agent to make certain choices, we have a better chance to understand the true connection between those choices and welfare. If, for instance, we have a validated decision model and a corresponding notion of welfare which predicts that the agent systematically minimises his welfare in some contexts, we can take his choices as a guide to enhance his welfare by reversing them or by generating environments where the agent cannot harm himself.

Since we are not looking for ‘as if’ explanations but rather for realistic decision mechanisms, the process of model selection is greatly enhanced by the possibility of rejecting a model when it does not tally with the choice data. As choice data are easily available in economics, we view this testability feature as crucial.

Notice that Bernheim’s [5] assertion that additional evidence (in the form of choice data or other data validating a model) could also be useful within the BeRa approach may prove to be slightly impractical,⁷ in the sense that it is only a specific model that can provide specific guidance as to the class of data needed to test or falsify the theory. In other words, what constitutes useful evidence is model dependent: we look for specific types of evidence, among all conceivable evidence, *only because we have a model that guides us to do so*. For example, if we attempt to explain the apparently inconsistent behaviour of consumers at the supermarket through a model of salience, we may suspect that choices are affected by the position of the items on the shelf. Then we may look for evidence of salience-driven behaviour, such as a tendency to focus visual attention only on a specific subset of the alternatives, such as the items on the left of each shelf or those on the shelf that is at eye-level. Or perhaps, as in Krajbich and Rangel [20] (with neuroeconomics evidence) and Reutskaya *et al.* [34] (in a laboratory experiment), we may be interested in possible correlations between the probability of an item’s salience and its value. The point is that a model-less approach leaves us in a complete void about what type of evidence to look for.

On the Bernheim logic above, suppose that evidence comes forth to ‘prove’ that a model is correct (not just the positive implications, but also its normative interpretation). In that case, it is argued, the same evidence would justify a restriction on the welfare-

⁷Bernheim [5], p. 293, states: ‘For example, if (contrary to my expectations) adequate evidence is assembled to defend a normatively substantive process model, within our framework the same evidence would typically justify a choice-domain restriction that leads to the same welfare criterion’.

relevant domain in the BeRa setting. For instance, if (as seems reasonable) the individual never uses satisficing strategies⁸ within sets consisting of only two elements, then one could justify restricting attention to binary sets, and the ‘correct’ welfare criterion would emerge.⁹ The problem with this reasoning, in our view, is that we need a model to show that (and why) only binary sets are relevant. Once it has been accepted that this model is correct, the work is done: there is no further scope for a model-less criterion.

As a concluding caveat, we observe that making welfare inferences under the assumption that the model is known is not necessarily a mechanical exercise, which consists simply of singling out one of the primitive binary relations of the model and declaring them ‘utility’ (although in many examples we do so for convenience). It may be the case that several of the agent’s objectives have welfare relevant features, and it may also be the case that such objectives are only partially revealed. For example, if the agent ‘categorises then chooses’, it is possible that the categorisation process itself contains welfare relevant aspects (if for example a jam brand X is favoured over another brand Y it may be the case that on average brand X jams prove to be better than brand Y jams). In these situations, some process of judicious aggregation and completion, which does not emerge from the choice data, will be necessary. In general, the working hypothesis here is that making welfare judgments involves some knowledge of what is important for the agent, and this in turn can be gleaned from information on his decision making process.

3.4 Advantage 3: Scope

While quite general, a model-less approach still rests on some founding assumptions which limit its scope. In particular, inbuilt in the framework is an hypothesis on how the data present themselves, namely in a deterministic way. The primitive is a deterministic choice function c that indicates the selection that the agent makes in each choice situation. Yet real choice data are typically noisy. For this reason, in recent theoretical works (as well as some classical ones), the primitive used is a *stochastic* choice function. The observed data are assumed to be probabilities of choice (as expressed in frequencies of choice). This type of data seems particularly apt in the case of modelling of boundedly rational agent, since a form of bounded rationality is that the agent makes errors with respect to rational behaviour consisting of the maximisation of a preference relation. It is very natural to model these errors as *random* errors. For example, if it is postulated that bounded rationality takes the form of the agent overlooking some of the alternatives (as in Masatlioglu, Nakajima and Ozbay [29]), it seems more natural to assume that there is some probability that an agent overlooks a given alternative in a menu, rather than to

⁸See Simon [39].

⁹We thank Douglas Bernheim for suggesting this example.

assume that he overlooks the alternative for sure. In other words, while the model-based approach seeks to identify the reasons for behaviour, such reasons often are best thought of as having a random component. This gives rise to stochastic choices (see e.g. Gul, Natenzon and Pesendorfer [19] and Manzini and Mariotti [25] for recent examples of this type of models).¹⁰ By definition, the model-based approach poses no limit to the type of models that can be considered. On the contrary, any model-less approach (and the BeRa approach in particular) is limited by the assumptions on the underlying type of choice theory that is being considered. Later in section 4.3 we will argue that a stochastic modelling of choices may offer some advantages.

4 Model based, falsifiable explanations of behavior: problems

Positive decision models are falsifiable, but there may be more than one model that is not falsified by the data. The usefulness of the model-based approach depends on the power (in a statistical sense) of the conditions that characterise those models to *accept* the right model among the potentially many which are not falsified, or in other words to which degree the conditions that characterise the models are invulnerable to type II errors.¹¹ As we shall see, this is a critical point in the methodology.

More precisely, we try to systematise the study of failures by classifying them into three types (two of which already highlighted by Bernheim [5]), that is

- Type 1 failure: two or more plausible positive models are observationally undistinguishable in terms of choice;
- Type 2 failure: only one plausible model is compatible with the data, but there is more than one possible interpretation of the model's primitives;

¹⁰Apesteguia and Ballester [1], too, assume that decision makers have an underlying true preference relation guiding their choices, and that any inconsistency in choice is attributable to 'mistakes'. Using as the basic primitive the frequency with which each alternative is chosen from a set, these authors introduce the 'swaps index' which measures deviations ('swaps') from rationality, and associate to it the deviation minimising preference relation(s). Although Apesteguia and Ballester do not specify any particular choice procedure that may cause these mistakes, they show that two general classes of models used to rationalise stochastic choice, namely random utility models and trembles models, have underlying preferences that minimise the swaps index.

¹¹A type II error occurs when a test fails to reject an incorrect null hypothesis (false negative). The power of a statistical test is the probability of *not* committing a Type II error, that is, it is the probability of correctly rejecting the null hypothesis when the alternative is true.

- Type 3 failure: there is only one accepted interpretation of the primitives of the only plausible model compatible with the data, but there are several specifications of the primitives that are compatible with the data (e.g. the preference relation cannot be pinned down).

As we discuss below, there may be ways of addressing these failures successfully.

4.1 Type 1 failure: Choice data are compatible with multiple model structures

As noted by Bernheim [5], some choice data that do not refute a specific model of the cognitive process underlying choice are also compatible with other plausible models, and therefore possibly with several welfare relations depending on which model is used. So, if even there is no choice-based falsification of a particular model, we may still not be justified in using that model (rather than some other model also compatible with the choice data) as a basis for normative judgments. Let's call this **type 1 failure**.

We illustrate this failure concretely with a slightly more technical example. We need some additional notation and definitions. Let \succ^* denote an asymmetric binary relation on the *sets* of alternatives, and for each menu A let $M^*(A, \succ^*)$ denote the set of all the alternatives in A that are not elements of any subset of A that is dominated by another subset of A according to the relation \succ^* . The way we will interpret this is that sets of alternatives are *categories* in which the agent mentally subdivides all alternatives (e.g. the set of all restaurants), and \succ^* is a *shading* relation whereby a category (such as that of Mexican restaurants) may trump another category (such as Italian restaurants). Choice behaviour is called *Categorise Then Choose (CTC)* (Manzini and Mariotti [24]) if there exists an asymmetric binary relation \succ on the alternatives (interpreted as preference) and a shading relation \succ^* on the sets of alternatives, such that on each menu S the choice is precisely $M^*(A, \succ^*)$.

In contrast, choice behaviour is a *Rationalisation choice* (Cherepanov, Feddersen and Sandroni [10]) if there exist $K + 1$ asymmetric binary relations \succ_i , $i = 1, \dots, K$, on the set of alternatives (the first K interpreted as psychological 'rationales' and the $(K + 1)^{th}$ interpreted as a preference) such that on each menu A the choice comprises precisely those alternatives that are maximal in the preference relation among all the alternatives in A that can be 'rationalised', that is that are maximal according to *at least* one rationale.

The CTC model and the rationalisation model are structurally dissimilar and they formalise very different psychological processes underlying choice. In the CTC model it is imagined that the agent first simplifies his choice task by categorising alternatives and ignoring alternatives that belong to losing (shaded) categories, and then makes a choice

by maximising preference. In the Rationalisation model, on the other hand, it is imagined that the agent first selects all alternatives that he can psychologically justify to himself on the basis of some rationale (e.g. alternatives that are ‘morally acceptable’), and then maximises his preference on the set of rationalised alternatives.

The following result is a consequence of the characterisations in Manzini and Mariotti [24] and in Cherepanov, Feddersen and Sandroni [10]:

Proposition 1 *The following statements are equivalent:*

- (1) *Choice behaviour is Categorise Then Choose.*
- (2) *Choice behaviour is a Rationalisation choice.*
- (3) *Choice behaviour satisfies the following property (Weak WARP): suppose that an alternative x is chosen from a binary menu in which only x and another alternative y are available, and from a larger menu A that contains y . Then y is not chosen from any menu ‘intermediate’ between $\{x, y\}$ and A that contains x .*

Weak WARP is introduced in Manzini and Mariotti [23]. It is a weakening of the standard Weak Axiom of Revealed Preference and it can be used to test theories of bounded rationality just as WARP is used to test the theory of rational choice. So here we have two models which, in spite of being structurally different and having very different interpretations, generate exactly the same restrictions on choice data, namely that satisfaction of Weak WARP. Moreover, the first stage selections and the second stage preference relation will be typically different in the two models. This is an example of Type 1 failure.

How serious is Type 1 failure? There are already several distinct models of boundedly rational choice in the literature, which are mostly characterised by different axioms on observed choice. This suggests that gathering choice data offers good prospects of falsification in the best scientific tradition. Axioms that restrict choice data are in any case useful to provide a ‘shortlist’ of suitable models: for each model, the characterising axioms do have statistical power against several other plausible alternatives. But, as Proposition 1 makes clear, no one can exclude that a ‘true’ unknown model lurks behind a model which has not been falsified by observations. That is, because the two models in the statement of the proposition are equivalent, a researcher might be ‘tricked’ into using one model to explain the data, whereas the driver of choices is in fact the other model. This situation, however, does not appear to be significantly different from that in any other scientific endeavour, notably that of ordinary economic modelling (we return to this point below). So it is not clear that models of boundedly rational choice should be subjected to a particular treatment or should raise special concern.

Moreover, there are effective strategies to increase the identifiability of such models by means of choice data beyond the first coarse ‘shortlisting’ stage. One strategy consists of the use of *contextual information and auxiliary data*, as we explain below in relation with failures of Types 2 and 3. Such additional data may help for example understand whether the agent is categorising or psychologically rationalising his behaviours.

A second (considerably under-used) strategy relies upon ‘*parsimony*’ as a criterion to rank competing explanations of choice. The concept of parsimony can be operationalised and made precise, for example, through Selten’s Measure of Predictive Success (Selten [37]). Selten’s measure was specifically designed to evaluate ‘area theories’ like those characterised by revealed preference type of axioms, such as Weak WARP. These theories exclude deterministically a subset of the possible outcomes, those incompatible with the model. The measure takes into account not only the ‘descriptive power’ of the model (measured by the proportion of ‘hits’, the observed outcomes consistent with the model), but also its ‘parsimony’. The lower the proportion of theoretically possible outcomes consistent with the model, the more parsimonious the model. More precisely, the measure, denoted s , is expressed as

$$s = r - a$$

where r is the descriptive power (number of actually observed outcomes compatible with the model divided by the number of possible outcomes) and a is the ‘relative area’ of the model, namely the number of outcomes in principle compatible with the model divided by the number of all possible outcomes. For a vacuous theory we would have $s = 0$.

While Selten [37] shows that his measure is the only one satisfying a small number of appealing axioms, there may of course be other plausible measures of parsimony. Some measures might be appropriate in some contexts and others in different contexts. The point here, however, is that, as we mentioned already, we believe that parsimony criteria are both severely underused in choice theory research (one notable use in the context of revealed preference theory is Beatty and Crawford [4]) and potentially tremendously effective for model selection.

So while Type 1 failure must be taken seriously, it looks far from being damning for model-based welfare analysis. To complement this point, we conclude by observing that *procedural models can generate behaviour that is consistent with utility maximization*. This makes evident how even standard non-parametric validations of ordinary utility maximisation are subject to type 1 failure in exactly the same way as models of boundedly rational choice are. An example is the ‘Checklist’ model of choice (Mandler Manzini Mariotti [22]). In this model the agent has in mind a sequence of desirable properties. For example, when buying a car, the properties may be ‘having an automatic gear’, meeting a given budget’, ‘coming in a desirable colour’, etcetera. When faced with a menu A ,

the agent progressively whittles down the set of eligible alternatives by first discarding those that do not possess the first property, then those that do not possess the second important property, and so on until only one alternative remains, which is chosen. Not all choice behaviour can be described in this way. But if it *can* be described in this way, we say that *it has a checklist*. The following result is proved in Mandler, Manzini and Mariotti [22]:

Proposition 2 (Mandler, Manzini and Mariotti [22]) *Choice behaviour maximises a utility function if and only if it has a checklist.*

So the Strong Axiom of Revealed Preference (SARP), which is normally taken to validate welfare rankings based on the revealed preference relation, is actually compatible with the consumer being boundedly rational. If it is actually the checklist procedure that generates an agent’s behaviour, the resulting revealed utility is not necessarily a good representation of that agent’s welfare. For example, an agent can declare alternative x better than alternative y (in the sense that he never chooses y when x is also available) on the basis of the fact that x possesses the first property P_1 while y does not. This makes sense for an agent who needs to limit the time and computational effort employed in choosing. But if y possesses *all* the other properties P_i with $i > 1$ while x does not, it is dubious whether x can be declared welfare superior to y by a planner who has the time and resources to pay careful considerations to *all* aspects relevant for choice. For a proper welfare inference a considered judgement, involving an evaluation of the properties and their importance to the agent, is needed. This is an illustration of the fact that mechanical recipes for inferring welfare from choice are unlikely to succeed.

4.2 Type 2 failure: Multiple interpretations of the primitives

Even if choice data were compatible with only one plausible model structure, the multiplicity of preference relations can be generated simply by giving different interpretations to the parameters of the model. As observed by Bernheim [5]: ‘for any process theory, Γ , it is usually possible to formulate other theories simply by reinterpreting and/or relabeling elements of Γ (most notably, the elements that are labeled as preferences)’ (p. 280). Let’s call this **type 2 failure**.

A clear illustration of type 2 failure is given by Tyson’s [41] two-stage ‘satisficing-plus’ model, which grafts a second stage on a satisficing model *à la* Simon. In this model alternatives are assumed to possess both a utility value and a secondary non-utility characteristic, such as visual salience or fit with a positive-action legal requirement. The agent chooses, among the alternatives that are deemed to be ‘satisficing’ on the utility criterion, the one with the highest level of the secondary characteristic. So, in our usual

jam at the supermarket example, imagine that the agent considers as satisfactory both a cherry jam and an apricot jam at their given prices. He is unwilling to exert the cognitive effort needed to break this utility pseudo-indifference (in Tyson’s terminology), for example by examining further the small print on the labels to check the exact list of ingredients, calorie content, etcetera. Instead, he simply picks, of the two satisfactory jams, the one that has a brighter label, a characteristic that may be unrelated to utility. Formally, the agent maximises a function s (representing the level of the secondary characteristic) subject to the alternatives passing a utility satisfaction threshold $\theta(A)$ for menu A , that is $u(x) \geq \theta(A)$ for some utility function u . The one just described is the intended interpretation of the functions s and u . Observe how the model leaves ample scope for the interpretation of the second stage characteristic. But even with this leeway, and even granting the assumption that the model is ‘structurally correct’ (that is, choices are compatible with the maximisation of a function s subject to a function u exceeding a critical level $\theta(A)$, where the critical level satisfies the additional cross-menu restriction that Tyson imposes), there is a difficulty: *how can we know* whether the functions s and u should be interpreted, respectively, as a secondary characteristic and utility as asserted, or instead as utility and a *primary* non-utility characteristic, or indeed whether utility has anything to do with choice at all? So, in the jam at the supermarket example, perhaps the consumer is first focusing on the visually salient jams on the shelves, and then he picks a satisficing one among these. Or perhaps he is simply responding first to visual cues and then making a final selection on the basis of brand conditioning.

Type 2 failure is indeed a serious problem. Unlike the other types of failures it is peculiar to models of boundedly rational choice; however it can be mitigated in several ways.

Solution 1: Manipulation of the external environment. Different interpretations of a model will typically lead to different predictions about choice responses after changes to the external environment. For example, in the satisficing-plus example discussed before, if s is interpreted as salience then manipulations of the salience of the objects (e.g. through an advertising campaign or alterations of the location of the objects in the visual field) will have different choice implications than if s was interpreted as utility, or as anything else. To the extent that such implications can be tested, a specific interpretation can be singled out by the data.

Solution 2: Use of auxiliary data. Additional information can usually be obtained in the form of auxiliary data to help understand which parameter of the model, if any, is most suitable to be interpreted as preference. Such auxiliary data include all data which, while obtained during, or in lieu of, or in addition to, the process of choice, are themselves not physical and voluntary choices within the task of interest. Leading examples of auxiliary

data are physiological and neurophysiological data recorded during a choice task, or survey responses and verbal evidence on internal states and personal variables related to choice tasks. Indeed, in the experimental literature the answer to survey questions are routinely used as regressors, much like ‘hard’ evidence. For example, for the satisficing-plus model, one could control in an experimental setting the environmental conditions affecting the visual salience of x and y , to show for example that if y is not chosen in the presence of x , then this cannot be due to x ’s greater salience. Or, with the usual caveat on incentive compatibility, one can simply *ask* the agent to state how aware he was of the feasible alternatives (voters in an election are routinely asked how aware they are of the various candidates/parties). Psychological and neurophysiological evidence can serve the purpose of detecting whether the agent is operating on the basis of dual or multiple selves (‘long-run’ and ‘short-run’) or not. Evidence from fMRI studies (e.g. McClure, Laibson, Loewenstein, and Cohen [31]) suggests that impulsive behavior is associated with the activation of different brain areas of the brain than planned behavior. While this type of evidence is in our opinion usually not conclusive when used *by itself*, what we are suggesting is that its use in an auxiliary role, beyond standard choice data, can be of help in the interpretation of positive models of choice.

Solution 3: Use of contextual information. The ancillary conditions themselves can form part of the evidence used to select a positive model of choice (this point refers to type 1 failure, too). For example, if ancillary conditions that make the problem more complex tend to produce more violations of WARP, this is *prima facie* evidence that a model based on complexity may be correct. BeRa argue that this type of evidence should only guide the *selection of choice situations* that are included in the domain of choice. They give the example of the choices of a blind person made in choice situations where information is presented visually, which should be discarded. But the same argument that Bernheim [5] makes regarding the use of ancillary evidence to exclude suspect choice situations can be tuned to *select between positive models of choice*. For example, ancillary conditions that affect the salience of alternatives may help interpret the primitives in the satisficing-plus model. Or, a documented choice fact is the tendency of employees to place large part of their savings in pension fund of the company for which they work (eg. Beshears et al. [7]), thus increasing the risk to their wealth to a level that is hard to justify rationally. Here, there is strong contextual evidence that the workers’ choice of pension plan expresses status-quo bias rather than ‘reasoned’ choice. Then, rather than discarding choice situations, one could use instead positive choice models of status-quo bias, e.g. Masatlioglu and Ok [30].

Of course, we have only pointed to possibilities that apply to specific contexts. One may be pessimistic on the effectiveness of contextual, or generally auxiliary, evidence to

help select between observationally equivalent models: As Bernheim (private communication) expresses it, ‘Will non-choice evidence help us settle on the right normative model? Maybe, and maybe not. With very limited exceptions...I don’t think we currently know how to use that data to test between theories, particularly if the theories in question are normatively distinct but have identical positive implications for choice. Consequently, a framework that requires such information is (in contrast to the Bernheim-Rangel approach) not ready for implementation, and may never be.’ We do not agree that the lack of a *systematic* method of using contextual evidence justifies a policy maker in ignoring it. Any contextual evidence which is good enough to exclude some choice situations from the domain seems to us good enough to help select between observationally equivalent positive models (in a more efficient way). So the positive approach is in this respect exactly as implementation-ready as the model-less approach.

4.3 Type 3 failure: Multiple preferences compatible with an identified model structure

There are cases in which it seems impossible to make choice-based welfare inferences even when (a) the structure of the model is exactly known and (b) the interpretation of the primitives is also known. **Type 3** failure consists in the inability of choice data to identify preferences when (a) and (b) hold.

It is important to understand that this problem holds in an exactly analogous form for rational agents. We give an example from the theory of rational choice over time. Suppose an agent is shown to satisfy the Fishburn–Rubinstein [17] axioms on preferences over date-outcome pairs (x, t) . Then there are continuous real-valued functions u on outcomes and δ on dates such that (x, t) is preferred to (y, s) if and only if $\delta(t)u(x) \geq \delta(s)u(y)$ (‘exponential discounting’). However, and this is not generally understood, one may fix the ‘discount factor’ δ arbitrarily to represent a given preference relation that satisfies the axioms, provided the ‘utility function’ u is calibrated accordingly.¹² In other words, for any two discount factors δ and δ' , there exist two utility functions u and v such that (u, δ) preferences are identical to (v, δ') preferences. The way out to pinpoint δ as a uniquely determined parameter expressing impatience is to devise an *external* method to fix u . We suggest the same method may be applied for boundedly rational agents.

We illustrate type 3 failure in detail within the context of one specific model. Consider the particular version of Masatlioglu, Nakajima and Ozbay’s [29] attention model in which the consideration set has a threshold structure, that is in each menu A the agent considers only the alternatives x for which $a(x) \geq \tau(A)$ where a is a real-valued attention function on the alternatives and τ a real-valued threshold function on the menus, and

¹²See e.g. Fishburn and Rubinstein [17].

then maximises a utility function u among those alternatives which exceed the attention threshold.¹³

As we discussed before in relation to the satisficing-plus model there is a first interpretational hurdle: that even if one cannot reject the mathematical structure of the model, still one cannot identify from choice data whether the model can be interpreted as described or otherwise. For example one cannot tell whether instead a is a utility function and u an attention function, so that the agent chooses the most attention grabbing alternative among those alternatives which are satisficing. That is, choice data cannot distinguish between this specification of the attention model and (a special interpretation of) the satisficing-plus model. Nevertheless, set aside this difficulty (which as argued before could be overcome by auxiliary evidence), and assume instead that *not only the constrained maximization structure of the model is correct, but also that the interpretation of the functions a and u is correct*. Does this mean that choices alone ‘reveal’ welfare?

The answer appears, at first sight, resoundingly negative. It is in fact possible to show formally (details available from the authors) that if a threshold representation exists with a given utility function u , then *any* other utility function, which respects the indifference classes of u but arbitrarily reorders them, could also work as an alternative utility function! So, even with all information about both structure and interpretation of the model, the choice data leave the utility function almost entirely undetermined. The ‘almost’ refers to the fact that at least the indifference classes can be identified (through the observation of common membership of the set of chosen alternatives out of some choice set).

As we argued before for the exponential discounting model, we suggest that one powerful way to address type 3 failure is to gather richer data on the choice process, notably:

Solution: The use of stochastic choice data. The collection of choice *frequencies* instead of choice *events*, or the estimation of choice probabilities may improve dramatically the identifiability of the primitives. The attention model of Masatlioglu, Nakajima and Ozbay [29] provides a particularly sharp example of how this methodology can be used. We have seen already that this model presents severe type 3 identification problems even in its threshold version, in that the utility function remains almost entirely arbitrary. Moreover, in the general version, it may not always be possible to determine whether a choice has been made on the basis of attention or on the basis of preference. For example, if x is selected from $\{x, y\}$ and $\{x, y, z\}$, y from $\{y, z\}$ and z from $\{x, z\}$, we cannot tell whether x is chosen over y because it is preferred to y or because y is not paid attention to in the menu $\{x, y\}$. In Manzini and Mariotti [25], we study a stochastic version of the attention model and we show that if we had access to stochastic choice data, we could

¹³This type of model is structurally identical to the Tyson [41] satisficing-plus model except that the threshold function here is unrestricted. We study it in detail in [27].

tell exactly both the amount of attention paid to each alternative and the preference relation over them. In particular, we could tell this by observing whether y is chosen more frequently when x is not available. While we refer to [25] for the technical details, the intuition is simple. In the stochastic model, each alternative is noticed in a menu with a given probability, and then the agent maximises utility over the set of the alternatives that he notices. If x is preferred to y , then the presence of x in a menu will diminish the frequency with which y is chosen, because whenever x is noticed, y cannot be chosen even if it is noticed (since in any case the agent would pick x rather than y).

This is just one ‘success story’ for a specific model, but we are fairly confident that this approach can be extended to help address other instances type 3 failures, which are so endemic in discrete and deterministic choice models.

5 Conclusions

Bernheim and Rangel [6] have offered an extension of classical welfare analysis to a context of bounded rationality. We have argued that a promising approach that should complement their analysis is to try and understand the decision process that underlies choice, a move that is explicitly rejected by them.¹⁴ This raises the need for a validated model of individual decision-making. The validation of such a model can be executed first and foremost, though not exclusively, via a standard revealed-preference type of methodology, that is by examining the restrictions that the model poses on observable choice data and checking whether such restrictions are satisfied by the available data. Caplin [8] goes further than that, and boldly proposes to extend the entire standard revealed preference approach to the new types of data which technology makes available (e.g. the collection of data on dopamine response). Caplin’s own work makes ample strides in this direction (see references in [8]). We agree that in the future this type of extension of the revealed preference methodology might become prevalent. Yet in this paper we have used the alternative and more conservative working hypothesis that non-choice data can be used in a merely auxiliary role within a process of model selection, giving priority to choice data. This can be justified on the basis of the advantages from discipline specialisation and

¹⁴Space consideration prevent us from doing justice to several other works which, while related to the theme developed in this paper, are outside the main focus of our analysis. The most relevant is Sugden [40] whose approach is choice-based and model-less but who, unlike BR, does not advocate any active role for policy makers, leaving it to the market to determine outcomes. Other authors, e.g. Green and Hojman [18], have proposed models of choice, with preferences as primitives, but which are not falsifiable.

Closer to our methodology are Dalton and Ghosal ([14], [15]), while Rubinstein and Salant [35] offer a general, model-based framework to deal with ‘behavioural datasets’, (collections of possibly conflicting preference observations elicited from the same decision maker in different, preference irrelevant contexts).

of imperfections in current technology for collecting non-standard data. In this respect our working hypothesis is more similar to the position adopted by Bernheim and Rangel. Nevertheless, we stress that ours was just a working hypothesis: we hope to have shown that even within its confines the model-based approach to welfare offers much promise. We have examined some of its problems and have suggested ways of addressing them. The new and growing literature we have discussed exemplifies this process and witnesses to exciting research at the frontier but, of course, most of the work remains to done.

References

- [1] Apestequia, J. and M. Ballester, 2013, "A Measure of Rationality and Welfare", mimeo, Universitat Pompeu Fabra and Universitat Autònoma de Barcelona
- [2] Ariely, D., G. Loewenstein and D. Prelec, 2003, "Coherent Arbitrariness": Stable Demand Curves Without Stable Preferences", *Quarterly Journal of Economics*, 118 (1): 73-106.
- [3] Baigent, N. and W. Gaertner, 1996, "Never Choose the UNiquely Largest", *Economic Theory* 8: 239-49.
- [4] Beatty, Timothy K. M., and Ian A. Crawford. 2011. "How Demanding Is the Revealed Preference Approach to Demand?" *American Economic Review*, 101(6): 2782-95.
- [5] Bernheim, B. D., 2009, "Behavioral Welfare Economics", *Journal of the European Economic Association*, 7: 267–319.
- [6] Bernheim, B. D. and A. Rangel, 2009, "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics", *Quarterly Journal of Economics*, 124: 51-104.
- [7] Beshears, J., J. J. Choi, D. Laibson, B. C. Madrian, 2008, "How are Preferences Revealed?", *Journal of Public Economics* 92 (8-9): 1787-1794..
- [8] Caplin, A., 2008, "Economic Theory and Psychological Data: Bridging the Divide", in: *The Foundations of Positive and Normative Economics: A Handbook*, A. Caplin and Andrew Schotter (eds.), Oxford University Press.
- [9] Caplin, A., and M. Dean, 2011, "Search, Choice and Revealed Preference", *Theoretical Economics*, 6 (1): 19–48.
- [10] Cherepanov, V., T. Feddersen and A. Sandroni, 2013, "Rationalization", *Theoretical Economics*, 8 (3): 775–800.

- [11] Chiang, J., S. Chib, and C. Narasimhan, 1999, "Markov Chain Monte Carlo and Models of Consideration Set and Parameter Heterogeneity", *Journal of Econometrics*, 9: 223-248.
- [12] Costa-Gomes, M. A. and V. P. Crawford, 2006, "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study," *American Economic Review* 96: 1737-1768
- [13] Costa-Gomes, M. A., V. P. Crawford and B. Broseta, 2001, "Cognition and Behavior in Normal-Form Games: An Experimental Study," *Econometrica* 69: 1193-1235.
- [14] Dalton, P. and S. Ghosal, 2008, "Behavioral Decisions and Welfare", mimeo, University of Warwick.
- [15] Dalton, P. and S. Ghosal, 2012, "Decisions with Endogenous Preference Parameters", *Social Choice and Welfare*, 38 (4): 585-600.
- [16] Eliaz, K., M. Richter and A. Rubinstein, 2011, "Choosing the two finalists" *Economic Theory* 46(2): 211-219.
- [17] Fishburn, P. C., and A. Rubinstein, 1982, "Time Preference", *International Economic Review*, 23: 677-694.
- [18] Green, J. and D. Hojman, 2007, "Choice, Rationality and Welfare Measurement", mimeo, Harvard Institute of Economic Research Discussion Paper No. 2144.
- [19] Gul, F., P. Natenzon, and W. Pesendorfer, 2012, "Random Choice as Behavioral Optimization". mimeo, Princeton University.
- [20] Krajbich, I. and A. Rangel, 2011, "Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions", *Proceedings of the National Academy of Sciences*, 108, 13852-13857.
- [21] Iyengar, S. S. and M. R. Lepper, 2000, "When choice is demotivating: Can one desire too much of a good thing?", *Journal of Personality and Social Psychology*, 79(6): 995-1006.
- [22] Mandler, M., P. Manzini and M. Mariotti, 2012, "A Million Answers to Twenty Questions: Choosing by Checklist", *Journal of Economic Theory* 147 (1): 71-92.
- [23] Manzini, P. and M. Mariotti, 2007, "Sequentially Rationalizable Choice", *American Economic Review*, 97: 1824-1839.

- [24] Manzini, P. and M. Mariotti, 2012, "Categorize Then Choose: Boundedly Rational Choice and Welfare", *Journal of the European Economic Association*, 10 (5): 939–1213.
- [25] Manzini, P. and M. Mariotti, 2014, "Stochastic Choice and Consideration Sets", *Econometrica* 82 (3): 1153-1176.
- [26] Manzini and Mariotti, 2013b, "Welfare economics and bounded rationality: the 'welfare as choice' approach", mimeo.
- [27] Manzini, P. , M. Mariotti and C.J. Tyson, 2013, "Two-stage Threshold Models of Choice", *Theoretical Economics*, 8 (3): 875–882.
- [28] Manski, C. F., 1977, "The Structure of Random Utility Models", *Theory and Decision*, 8: 229-254.
- [29] Masatlioglu, Y., D. Nakajima and E. Ozbay, 2012, "Revealed Attention", *American Economic Review*, 102 (5): 2183-2205.
- [30] Masatlioglu, Y. and E. Ok, 2009, "A Canonical Model of Choice with Initial Endowments", forthcoming in *Review of Economic Studies*.
- [31] McClure, S. M., D. I. Laibson, G. Loewenstein and J. D. Cohen, 2004, "Separate Neural Systems Value Immediate and Delayed Monetary Rewards", *Science* 306: 503 - 507.
- [32] van Nierop, E., B. Bronnenberg,, R. Paap, M. Wedel & P.H. Franses, 2010, "Retrieving Unobserved Consideration Sets from Household Panel Data", *Journal of Marketing Research*, 47 (1): 63-74.
- [33] Nishimura, H., 2013, "The transitive core: inference of welfare from nontransitive preference relations", mimeo, Caltech.
- [34] Reutskaya, E., R. Nagel, C. F. Camerer, and A. Rangel (2011) "Search Dynamics in Consumer Choice under Time Pressure: An Eye-Tracking Study", *The American Economic Review*, 101, 900-926.
- [35] Rubinstein, A. and Y. Salant, 2012, "Eliciting Welfare Preferences from Behavioral Datasets", *Review of Economic Studies*, 79 (1): 375-387.
- [36] Salant, Y. and A. Rubinstein, 2008, "(A, f): Choice with Frames", *Review of Economic Studies*, 75: 1287–1296.

- [37] Selten, R., 1991, "Properties of a measure of predictive success", *Mathematical Social Sciences*, 21: 153-167.
- [38] Sen, A. K., 1993, "Internal Consistency of Choice", *Econometrica*, 61(3): 495-521.
- [39] Herbert A. Simon, 1982/1997, *Models of Bounded Rationality*, MIT Press (three volumes)
- [40] Sugden, R., 2004, "The Opportunity Criterion: Consumer Sovereignty Without the Assumption of Coherent Preferences", *American Economic Review* 94: 1014-1033.
- [41] Tyson, C.J., 2013, "Satisficing behavior with a secondary criterion.", working paper, Queen Mary University of London.