Choice Deferral, Status Quo Bias, and Matching

Choice Deferral, Status Quo Bias, and Matching

Gökhan Buturak

Dissertation for the Degree of Doctor of Philosophy, Ph.D.
Stockholm School of Economics 2011.

Keywords: Choice deferral, choice avoidance, preference for flexibility, preference for commitment, subjective states, anticipated regret, status quo bias, choice avoidance, incomplete preferences, one-sided many-to-many matching, network formation, setwise stable set, bargaining set, individually rational core, weak separability, implementation, strategy-proofness, Pareto efficiency.

Choice Deferral, Status Quo Bias, and Matching

ISBN 978-91-7258-842-4

Printed in Sweden by:
Intellecta Infolog, Göteborg 2011

Distributed by:
The Research Secretariat
Stockholm School of Economics
Box 6501, SE-113 83 Stockholm, Sweden
www.hhs.se

Preface

This report is a result of a research project carried out at the Department of Economics at the Stockholm School of Economics (SSE).

This volume is submitted as a doctor's thesis at SSE. The author has been entirely free to conduct and present his/her research in his/her own ways as an expression of his/her own ideas.

SSE is grateful for the financial support provided by the Jan Wallander and Tom Hedelius Foundation, which has made it possible to fulfill the project.

Filip Wijkström
Associate Professor
SSE Director of Research

Acknowledgement

I would like to thank my advisors, Tore Ellingsen and Mark Voorneveld, for patiently taking me through the difficult task of economics research. Their advice, guidance and support were essential for this thesis to come out.

Jens Josephson, from Stockholm University, has been the opponent for my licentiate defense. His suggestions were important to better shape this thesis. He was also very supportive during the job market period.

In the last stage of my PhD studies, I had the opportunity to write a paper with Özgür Evren. I am grateful to him for encouraging me to do decision theory. He has always been a good friend, from whom I have learned a lot.

My gratitude also extends to the administrative staff of the Department of Economics at SSE. They were always friendly and willing to help.

Stockholm, February 2011.

Milda'ya

-Doktora tezini bitirdiniz mi?
-Dün akşam bitirdim dedi, ve deminki çocukça utanmasını daha çocukça bir neşe ile tamamladı; dün akşam son sahifenin altına kırmızı kalemle kocaman bir çizgi çizdim. Altına birinciden daha kalın bir çizgi, bir tane daha, en altına dört mayıs, saat 23'ü beş geçe diye yazdım. Bir imza attım. Sonra kalktım; balkona çıktım. İsveç usulü üç dört derin nefes aldım.

A.H. Tanpınar,
Huzur (s. 75).

Contents

Introduction 1

Paper 1: Rational Choice Deferral 7

1. Introduction 8
2. The Model of Choice Deferral 14
3. The Basic Axioms and the "Anomalies" Regarding Deferral 15
4. Representation Theorems 17
5. Explaining the "Anomalies" Regarding Deferral 23
6. More on Choice Avoidance and Anticipated Regret 26
7. Identifying the Relevant States and the Aggregator 27
8. Maximizing an Incomplete Preference Relation: A Comparison 30
9. Concluding Remarks 30

Apendix 32

References 46

Paper 2: Status Quo Bias with Choice Avoidance 51

1. Introduction 52
2. The Model 55
3. The Axioms and the Representation Result 56
4. Characterization of Choice Avoidance 59
5. Representation of Non-Averse Choice Correspondences 60
6. Relating Status-Quo Bias to Choice Avoidance 60
7. Related Literature 62

Apendix 64

References 70

Paper 3: One-Sided Many-to-Many Matching: An Ordinal Theory of Network Formation 75

1. Introduction 76
2. The Model 80
3. Examples 84
4. Relationships among Solution Sets and Nonemptiness 86
5. Implementation and Some Properties of the Solution Sets 90
6. Concluding Remarks 96

Apendix 97

References 108

Introduction

This thesis consists of three independent papers. They are put in reverse chronological order according to when they were initiated.

The first paper, which is a joint work with Özgür Evren, extends the standard rational choice framework with the option to postpone the act of selecting an alternative. In that paper, we propose an axiomatic model of choice over risky prospects that restricts the classical rationality axioms solely to those instances in which the decision maker does not defer. The cardinal approach we follow allows us to identify the preference relation of the decision maker over lotteries, even if the choice data is very scarce due to deferral. Moreover, we also derive the value of deferring choice from a given set of options, which turns out to be an affine utility function over choice sets. At each choice situation, the decision maker compares the utility of each available alternative with that of deferral so as to decide on opting for an alternative immediately.

The second paper is a model of status quo bias with choice avoidance. It describes the choice behavior of an otherwise standard decision maker whose choices are affected by the presence of a status quo alternative. The status quo emerges as a temporary choice, which may be reversed upon arrival of new (introspective or objective) information, or upon finding new alternatives.

The third paper considers the network formation problem from a matching perspective. In that paper, agents want to link with each other and each has preferences over the subsets of others. We consider various solution concepts regarding the stability of a matching between the agents, establish relations between these concepts under several preference restrictions, and provide sufficient conditions for these solutions to be nonempty.

Below, I provide a more detailed summary of each of these three papers.

1. Rational Choice Deferral

Choice theory studies how people select an alternative from a given set. In reality, however, the act of choosing may not be immediate; there is often the option of deferral. This generates a more fundamental choice situation: whether to choose now or later. Starting with Corbin (1976, 1980) and the subsequent influential work of Tversky and Shafir (1992), experimental findings cast doubt on the explanatory power of the standard rational choice theory for such situations. The trouble is that "when each option has significant advantages and disadvantages, people often experience conflict that makes choice aversive and compels them to delay decision and seek additional information or options" (Tversky and Shafir, p. 358). This reflects itself with an increase in subjects' tendency to defer choice as additional alternatives become available. In contrast, from the standpoint of the standard theory "deferring choice is just another option to be selected whenever its (subjective) value exceeds that of the available alternatives" (Tversky and Shafir, p. 358). Consequently, the standard theory predicts that a decision maker who defers choice from a given set of options would also defer when faced with a subset of those options.

On the other hand, because of intertemporal trade-offs that might be involved, it seems only natural to view the act of deferring as a special option. Motivated by these observations, we propose an axiomatic model of choice over risky prospects that restricts the classical rationality axioms solely to those instances in which the decision maker does not defer. In this setup, the independence axiom allows us to identify the preference relation of the decision maker over lotteries, even if the choice data is very scarce due to deferral. Moreover, the decision maker behaves as if the value of deferring choice from a given set of options is determined by an affine utility function W over choice sets. Other than the independence axiom, this basic model does not require a demanding assumption on behavior across choice problems where the decision maker might defer. Therefore, our theory is compatible with the observed anomalies and various interpretations that might be suitable in a given situation. The standard theory is subsumed as the particular case in which W is constant (across choice sets).

We also investigate the cases in which W can be written as the aggregate, anticipated utility associated with how the decision maker might behave upon deferral, depending on the realization of endogenously determined (i.e., subjective) states. We relate the forms of the aggregator to the particular forms of anomalies that one may observe, and characterize those cases in which the aggregator can be considered as an expectation operator across "non-trivial" states (which may be thought of as the future (unknown) tastes of the decision maker). In such cases, the anomaly noted by Tversky and Shafir (1992) occurs by necessity. A further merit of the subjective state space approach is that it facilitates the "anticipated regret" interpretation in the context of choice deferral.

We also find that the aggregator and the set of relevant states are almost fully identified (upon normalization). The only difficulty is that the decision maker's current preference relation over lotteries remains unidentified in the subjective state space representation of the value of deferral.

2. Status Quo Bias with Choice Avoidance

People often avoid making a choice and end up with default options (for experimental evidence, see Anderson (2003) and the references therein). When choice induces conflict and compels a decision maker to contemplate more about the available alternatives or search for more alternatives, the default option plays the role of a temporary choice. In such situations, people may keep their default option even when they find it inferior to some of the available alternatives. This appears to be a challenge to rational choice theory and raises the question if and how one can rationalize the observed behavior in such choice situations.

The main purpose of this paper is to reconcile the notion of status quo bias with choice avoidance within an abstract choice framework. We propose an axiomatic model of status quo bias that allows for choice avoidance. More specifically, we view choice avoidance as a particular form of status quo bias: By keeping her default option, a decision maker (temporarily) gains the flexibility of postponing the task of making an active choice. Later on, the decision maker may switch to another alternative. On the other hand, the absence of a default option supposedly forces the decision maker to select an option. Accordingly, our representation result describes the choice behavior of an otherwise standard decision maker who mentally constrains her choice problems with respect to a status quo and selects it whenever she is indecisive about which alternative is the best among the constrained choice set induced by the presence of the status quo.

According to our representation result, the decision maker's choice behavior is identified with three objects: (i) a complete preference relation which reflects that the decision maker has a complete ranking of the available alternatives and makes her choices accordingly in the absence of a default option; and, for each alternative x, (ii) a set $\mathcal{Q}(x)$ containing the alternatives the decision maker finds at least as good as the default option x, and (iii) a (possibly incomplete) preference relation corresponding to the decision maker's choices over the alternatives in $\mathcal{Q}(x)$. With regard to the latter, the decision maker tries to identify the best element(s) in $\mathcal{Q}(x)$ and sticks to her default option whenever she cannot find one. By contrast, in the standard models of status quo bias (see Masatlioglu and Ok, 2005, 2009; Sagi, 2006; and Apesteguia and Ballester, 2009), though the decision maker may not be able to compare all the alternatives with the default option, she is able to rank those that are at least as good as the default. Therefore, existing models are not compatible with the notion of choice avoidance as they respect the classical rationality postulates, holding the status quo fixed. Accordingly, we show that the representation in Masatlioglu and Ok (2009) is a special case of ours.

3. One-Sided Many-to-Many Matching: An Ordinal Theory of Network Formation

In this paper, we consider the network formation problem from a matching perspective. In the existing models of network formation, agents are usually assumed to be identical. Therefore, each agent's linking decisions are independent of the identities of the others. By contrast, the matching approach we take does not assume that the agents are identical. Each agent has a ranking of the subsets of the others.

Also, our model generalizes a class of matching problems. We investigate to what extent the results for these matching problems carry over to our framework. We establish relations between several solution sets, and provide sufficient conditions for these sets to be nonempty. We show that, under a weak separability condition, the solution concepts we study coincide and select a unique matching. We also study some properties of the stable matchings and investigate whether the stable matching rule satisfies some appealing properties such as strategy-proofness and Pareto efficiency. We provide a game in strategic form and show that, under weak separability, the unique stable matching coincides with the pairwise Nash equilibrium outcomes of this game.

References

C.J. Anderson, The psychology of doing nothing: forms of decision avoidance result from reason and emotion, *Psychological Bulletin* **129** (2003), 139–167.

J. Apesteguia, M.A. Ballester, A theory of reference-dependent behavior, *Economic Theory* **40** (2009), 427–455.

R.M. Corbin, *No-choice option in decision-making*, Ph.D. Dissertation, McGill University, 1976.

R.M. Corbin, Decisions that might not get made, in T.S. Wallsten (Ed.), *Cognitive Processes in choice and decision behavior*, Hillsdale, Lawrence Erlbaum Associates Inc., 1980.

Y. Masatlioglu, E.A. Ok, Rational choice with status quo bias, *Journal of Economic Theory* **121** (2005), 1–29.

Y. Masatlioglu, E.A. Ok, A canonical model of choice with initial endowments, New York University (2009), mimeo.

J.S. Sagi, Anchored preference relations, *Journal of Economic Theory* **130** (2006), 283–295.

A. Tversky, E. Shafir, Choice under conflict: the dynamics of deferred decision, *Psychological Science* **3** (1992), 358–361.

Paper 1

Rational Choice Deferral

Gökhan Buturak and Özgür Evren

Abstract

A variety of empirical findings suggest that in choice problems where deferral is also an option, the structure of the set of feasible alternatives may well influence the value that the decision maker attaches to the act of deferring choice, thereby giving rise to seemingly "irrational" mode of choice behavior. To accommodate this phenomenon, we propose a model of choice over risky prospects that restricts the classical rationality axioms solely to those instances in which the decision maker does not defer. We show that under weak assumptions, the implied behavior can be represented as if the value of deferring choice from a given set of options is determined by an affine utility function over choice sets which can be written as the aggregate, anticipated utility associated with how the agent might behave upon deferral, depending on the realization of subjective states. We relate the associated "anomalies" to the forms of the aggregator, characterize those cases in which the aggregator can be considered as an expectation operator (across subjective states), and examine the extent of uniqueness of the representations. Our theory is also applicable to choice problems with an outside option. In this context, the model explains the notion of "choice avoidance" by relating it to anticipated regret, and subsumes the classical model as the particular case in which the decision maker does not expect to experience regret.

JEL Classification: D11, D81.
Keywords: Choice deferral, choice avoidance, preference for flexibility, preference for commitment, subjective states, anticipated regret.

For their useful comments and suggestions, we are grateful to Kfir Eliaz, Tore Ellingsen, Guillaume Fréchette, Georgios Gerasimou, Jens Josephson, Asen Kochov, Efe Ok, David Pearce, Debraj Ray, Ariel Rubinstein, Avital Shamir, Ennio Stacchetti, Mark Voorneveld, and to the seminar participants in New York University and Stockholm School of Economics. All remaining errors and deficiencies belong to us.

1. Introduction

Choice theory studies how people *select* an alternative from a given set. In reality, however, the act of choosing may not be immediate; there is often the option of deferral. This generates a more fundamental choice situation: whether to choose now or later. For example, a policymaker may either commence one of the several available energy projects immediately, or may decide that research is needed to determine their potential effects on the environment. An employee may choose a 401(k) retirement savings plan now, or may wait to see how her private life evolves, which, in turn, may affect her consumption-saving decisions. A worker may either decide to accept one of the current job offers, or keep searching, possibly at the cost of losing some of the available offers.

Starting with Corbin (1976, 1980) and the subsequent influential work of Tversky and Shafir (1992), experimental findings cast doubt on the explanatory power of rational choice theory for such situations. The primary problem is that "when each option has significant advantages and disadvantages, people often experience conflict that makes choice aversive and compels them to delay decision and seek additional information or options" (Tversky and Shafir, p. 358). This reflects itself with an increase in subjects' tendency to defer choice as additional alternatives become available. But, as Tversky and Shafir put it, "in the rational theory of choice ... each option x has a value $v(x)$ such that, given an offered set, the decision maker selects the option with the highest value" (p. 358). Moreover, "deferring choice is just another option to be selected whenever its (subjective) value exceeds that of the available alternatives" (p. 358). When viewed this way, rational choice theory predicts that a decision maker who defers choice from a given set of options would also defer when faced with a subset of those options. To challenge this hypothesis, in their Study 2, Tversky and Shafir conduct an experiment which shows that "adding a new alternative to a given choice set can increase conflict and enhance the tendency to defer decision, contrary to value maximization" (p. 358). Similar findings are reported by Dhar (1992, 1997) and White and Hoffrage (2009), among others.

However, as we see it, the problem lies not in the failure of the standard axioms but in the view that the act of deferring choice is "just another option." For instance, in Study 2 of Tversky and Shafir (1992), the subjects are asked whether they would like to wait and "learn more about the various models" if they happen to enter a store that offers two attractive CD players which are marked by their relative advantages in the dimensions of quality and price. While subjects tend to "wait" in this context, they declare preference for immediate purchase when they are asked to imagine entering a store that offers only one type of an attractive CD player. In line with the "conflict" interpretation of Tversky and Shafir, one explanation of this behavior is that subjects see as pointless deferring an obvious or necessary choice, whereas in a more difficult choice problem deferral may be perceived as a valuable option because of the need for additional information.

More generally, since deferral is simply a means of keeping one's options open, it appears that its value may very well depend on the structure of the set of available options at the moment of the decision regarding deferral. A major purpose of the present paper is to extend the standard choice model to lay axiomatic foundations of rational behavior that is compatible with this phenomenon. In this regard, our most structured representation theorem shows that under quite general assumptions, the value of deferring choice from a given set can be considered as anticipated utility associated with how the decision maker might behave upon deferral, depending on the realization of endogenously determined (i.e., subjective) states.

This is not to mean, however, that all modes of related behavior can be rationalized. Rather, in the context of choice problems with deferral, our theory provides a framework in which rationalizable modes of behavior can be distinguished from "irrational" ones. For instance, in response to the experimental observations, recently Gerasímou (2010) has proposed a model which predicts that the decision maker would defer unless one of the available alternatives dominates all others with respect to an incomplete preference relation. By well-known representation results, this is equivalent to saying that the decision maker (behaves as if she) has a set of criteria in her mind, and she defers unless she is absolutely sure that one of the available alternatives is the best in all dimensions. This appears to be a boundedly rational mode of behavior, as it ignores the intertemporal trade-offs that come about when deferring choice is costly. It turns out, however, that Gerasímou's model satisfies all of the essential axioms of our main representation theorem which describes a decision maker who behaves as if she is a sophisticated dynamic programmer. (For more on this, see Section 8 below.) By contrast, as we shall discuss in Section 3, a decision maker who finds it difficult to search for the best options in a large choice set may very well violate the rationality axioms demanded by our model which regulate behavior across choice problems in which the agent does *not* defer.

Another challenge to the rational choice theory comes from a class of related experiments in which deferring choice is not viable, although the subjects can still decide not to make an *active* choice in return for an outside option. The findings in such choice problems are quite similar to those in the context of choice deferral: subjects tend to avoid choice when they face "too many" alternatives, or a difficult choice problem, even when the outside option is revealed to be inferior to some of the feasible alternatives. For example, in their Study 1, Tversky and Shafir (1992) show that replacing an inferior lottery in a choice set with another lottery that is not so easily comparable with the other available lotteries increases subjects' propensity to pay for an outside option which is randomly drawn from a known set.[1] As noted by several researchers, such behavior, which may aptly be called choice *avoidance*, can be related to anticipated regret which may come about if the decision maker fears that any of the available (ordinary) alternatives may turn out to be inferior (see, e.g., Anderson, 2003; Dhar and Simonson, 2003).

Our theory is also applicable in choice problems with a default option, which may take the form of an outside option or an initial endowment.[2] In particular, upon a suitable transformation, our subjective state space representation can be read as a model of choice avoidance that is compatible with the "anticipated regret" interpretation. The implied decision maker is almost rational in the sense that when comparing ordinary alternatives, she never violates the standard rationality postulates. However, when assessing the value of the default option, the decision maker does not account for the possibility that she may experience regret ex-post, leading to violations of the weak axiom of revealed preferences across choice problems in which the default option is (passively) selected.

1.1 Overview of Results

For choice problems with deferral, the most general version of the decision maker that we wish to model compares the utility of the best available alternative with the value of deferring choice which, in turn, depends on the set of feasible options. If the former value exceeds the latter, the agent selects the best available alternative(s); if the converse inequality holds she decides to defer; and in case of a tie, we understand that the agent is indifferent between immediate choice and deferral. Therefore, in this context, the primitives that we have in mind consist of a utility function φ on the grand set of alternatives, an object $\ominus$ that represents "choice deferral," a choice correspondence $\mathbf{c}$, and a value function W on the collection of choice sets. Formally, the behavior that we have just described reads as follows: For a generic choice set x and an ordinary alternative $p \in x$,

$$p \in \mathbf{c}(x) \quad \text{if, and only if,} \quad \varphi(p) = \max\left\{W(x), \max_{q \in x} \varphi(q)\right\}, \tag{1}$$

[1] A short list of related studies includes Iyengar and Lepper (2000), Boatwright and Nunes (2001), Shah and Wolford (2007), and Reutskaja and Hogart (2009). Perhaps the most important finding in this strand of literature is due to Iyengar et al. (2004) which establishes a negative relationship between the number of 401(k) plans offered by an employer and the participation rate of the employees. However, the authors do not discuss whether the said observation may be attributed to choice deferral, as opposed to choice avoidance.

[2] Rational choice models with status-quo bias are not compatible with the notion of choice avoidance as they respect the weak axiom of revealed preferences, holding the initial endowment fixed (see Masatlioglu and Ok, 2005, 2009; Sagi, 2006; and Apesteguia and Ballester, 2009).

and

$$\ominus \in \mathbf{c}(x) \quad \text{if, and only if,} \quad W(x) \geq \max_{q\in x} \varphi(q). \tag{2}$$

We attack this representation problem in the framework of choice under risk, for in the ordinal framework, with the sort of choice data that our model demands, typically neither the structure of the value function W nor the agent's preference relation over the set of alternatives can be determined in a meaningful way. The present cardinal approach allows us to identify φ up to positive affine transformations by coupling a fairly standard independence axiom with the classical rationality and continuity axioms, despite the fact that the choice data may be quite scarce due to deferral. It should be emphasized, however, that our theory restricts the rationality axioms solely to those instances in which the agent does not defer, since we view deferral as a special object that only has a context dependent meaning.[3]

By a further independence axiom that is akin to those of the menu choice literature, we are also able to deliver an affine value function W.[4] Except for this independence axiom, the basic model that we have just described does not require a demanding assumption on behavior across choice problems in which the decision maker defers. Consequently, our theory is compatible with various forms of interesting behavior related to deferral and several interpretations that might be suitable in a given situation. For instance, the agent may defer to find new alternatives in view of the choice set that she faces, or to make a more informed decision on which of the available alternatives to choose. (See Section 9 for a discussion of how our model can be structured to accommodate the former case.)

Our focus will be those cases which are compatible with the latter scenario that allows us to import the machinery developed in the menu choice literature.[5] Specifically, we will show that upon addition of a further continuity axiom to our basic model, one can find a state space $\mathcal{S}$, a state contingent expected utility function $U(\cdot,\cdot)$, and a (signed) measure μ on the set $\mathcal{S}$ such that, for each choice set x,

$$W(x) = \int_{\mathcal{S}} \max_{q\in x} U(s,q)\,\mu(ds). \tag{3}$$

[3]In fact, because of excessive deferral, in the ordinal framework our rationality axioms may not suffice for rationalizability (see Section 3).

[4]The menu choice literature studies how people select a *set* (from a collection of sets) which determines the options that will be available to the decision maker in a future date. The most prominent work in this strand of literature includes Kreps (1979), Dekel et al. (2001), and Gul and Pesendorfer (2001).

[5]Let us note, however, that in the present framework, the axioms of the menu choice literature are not readily applicable, for in our model, the value of deferral across different choice problems cannot be compared directly. (A detailed discussion of a related difficulty can be found in Section 4.3 below.)

Next, we relate the forms of the aggregator μ to the particular forms of the associated behavior of interest, and characterize those cases in which μ can be considered as a probability measure over "non-trivial" states, which may be thought of as the future (unknown) tastes of the decision maker. In such cases, the instances of behavior that Tversky and Shafir (1992) note in their Study 2 occurs by *necessity*; that is, the decision maker defers choice from a given set of options, but she makes a selection when faced with a subset of those options (see Section 5).

We also find here that the aggregator and the set of relevant states are almost completely identified (upon normalization). The only difficulty is that the decision maker's current preference relation over lotteries remains unidentified in the subjective state space representation of the value of deferral. Since the choice data available in our model is much smaller than that demanded by the menu choice models, this level of identification seems to be surprisingly good.

A few remarks are in order to explain why we do not directly assume a decision maker who has a complete understanding of all contingencies that might affect the consequences of her decision to defer choice. First of all, as Dekel, Lipman and Rustichini (2001) (henceforth, DLR) and Epstein et al. (2007) argue, assuming such a complete list of contingencies at the outset would seem unrealistic in most decision problems, and the experiments on choice deferral are not exceptional, nor are the examples of real economic problems that we mentioned earlier. Perhaps more importantly, like any other choice theoretic exercise, our purpose in the present paper is to comment on the structure of the decision making process that may underlie a given choice data, and naturally, the set of contingencies that the decision maker may have in her mind is an integral part of this structure.

In the context of choice problems with a default option, our theory can be read as a model of choice avoidance, but is general enough to subsume the standard model as the particular case in which W is constant. A further merit of the subjective state space approach is that it facilitates the "anticipated regret" interpretation in this context. To clarify this transformation, let us first note that the set of maximizers of $\varphi(\cdot)$ on a choice set x coincides with that of $\Phi(\cdot, x) := \varphi(\cdot) - W(x)$. Thus, (1) and (2) can equivalently be written as

$$p \in \mathbf{c}(x) \quad \text{if, and only if,} \quad \Phi(p, x) = \max\left\{0, \max_{q \in x} \Phi(q, x)\right\}, \tag{4}$$

and

$$\ominus \in \mathbf{c}(x) \quad \text{if, and only if,} \quad 0 \geq \max_{q \in x} \Phi(q, x). \tag{5}$$

If W takes the form (3), the number $\Phi(p, x)$ can be viewed as the anticipated utility of choosing the alternative p from the set x. Indeed, upon setting $v(p) := \varphi(p) - W(\{p\})$, it follows that

$$\Phi(p, x) = v(p) - \int_{\mathcal{S}} \left(\max_{q \in x} U(s, q) \ - U(s, p) \right) \mu(ds). \tag{6}$$

Following Sarver (2008), the term $\max_{q \in x} U(s, q) \ - U(s, p)$ can be considered as a measure of the strength of the sense of regret that the agent would experience upon realization of the state s, if, ex-ante, she has chosen p from x. In turn, $v(p)$ may be viewed as the context independent, material utility attached to the alternative p. Thus, overall, we can read $\Phi(p, x)$ as the utility of choosing p from x, net of the mental cost associated with the possibility of regretting this choice. Finally, 0 represents the context independent utility entailed by selecting $\ominus$ (passively), which now stands for the default option.

1.2 Related Literature

To the best of our knowledge, in the theoretical literature, the only other papers that touch the problem of choice deferral are Kochov (2009) and Kopylov (2009). The latter paper is mainly concerned with relating indecisiveness driven choice deferral to ambiguity aversion. The primitive of Kopylov's model is an incomplete preference relation, and the decision maker in question is assumed to defer whenever she cannot identify a dominant alternative, as in the aforementioned model of Gerasímou (2010). Kochov (2009), on the other hand, examines the epistemic value of a set of options in the menu choice framework. While he assumes that deferral is never an inferior option, in our model, deferring choice may incur costs such as foregone consumption or having "too much" flexibility in the second period of the decision problem. In particular, we relate a particular form of interesting behavior regarding deferral to those cases in which flexibility may be undesirable (see Sections 3 and 5). Kochov's epistemic analysis focuses on the information that a choice set might reveal about the modes of behavior that the agent may wish to adopt upon deferral. In Section 9, we will comment on a possible extension of our model in this direction.

In the context of choice problems with a default option, Dean (2008) had previously suggested an alternate formulation of Gerasímou's (2010) model on choice deferral. The comparison of our theory of choice avoidance with that of Dean is analogous: the preference relation that we model is obtained upon a probabilistic aggregation of how the decision maker might feel ex-post, whereas Dean's model describes an agent who opts for the default whenever she has second thoughts on which of the available alternatives might be the best, irrespective of how all other alternatives compare with the default option.

The plan of the remainder of the paper is as follows. In Section 2, we introduce our model of choice deferral formally. In Section 3, we discuss the basic rationality axioms that we will utilize and the forms of seemingly anomalous behavior in the context of choice deferral. We report our representation theorems in Section 4, and relate them to the "anomalies" in Section 5. Choice problems with a default option are discussed in Section 6, while Section 7 deals with the identification issue. In Section 8, we formally compare our model with Gerasímou (2010). Possible directions for future research are discussed in Section 9, while Appendix contains the omitted proofs.

2. The Model of Choice Deferral

We denote by B a finite set of riskless prizes. Δ stands for the set of all lotteries on B (equipped with the Euclidean norm), and $\mathcal{X}$ for the collection of all nonempty closed subsets of Δ endowed with the Hausdorff metric, d_h (induced by the Euclidean norm). We interpret Δ as the set of all (ordinary) **alternatives.** A generic element x of $\mathcal{X}$ will be referred to as a **choice set**.

We consider an agent who has two periods to select an alternative. In period 1, the agent faces a choice set x. She can either select an alternative from x immediately, or may decide not to do so. In the latter case, we understand that the agent defers choice. To formalize the agent's observed behavior, we identify "choice deferral" with an arbitrary object $\circleddash$ that does not belong to Δ. Therefore, the agent's behavior is modeled by a *nonempty* valued correspondence $\mathbf{c}$ from $\mathcal{X}$ into $\Delta \cup \{\circleddash\}$ such that, for every $x \in \mathcal{X}$,

$$\mathbf{c}(x) \subseteq x \cup \{\circleddash\}.$$

Following the usual interpretation of the notion of a choice correspondence, if an alternative p belongs to $\mathbf{c}(x)$, we understand that the agent *may* select p from x (immediately). On the other hand, by $\circleddash \in \mathbf{c}(x)$ we mean that the agent may postpone the task of making a choice. In particular, if $\mathbf{c}(x)$ contains both $\circleddash$ and an alternative p, we understand that (given the choice set x) "opting for p immediately" and "deferring choice" are both likely actions that the agent may take.[6]

We do not explicitly model the agent's behavior upon deferral in the second period of the choice problem. Therefore, the representation theorems that we report below will merely provide interpretations of why the agent may be deferring at a given instance. Our most structured representation theorems focus on a particular interpretation that is based on the assumption that the set of feasible alternatives remains unchanged across

[6] Conceptually, this interpretation is not different than the usual interpretation of situations in which two or more mutually exclusive alternatives might be chosen.

the two periods. However, as we noted earlier, our general representation (Theorem 1) is also compatible with those cases in which the set of feasible alternatives may change upon deferral (see Section 9). In this general result, we provide necessary sufficient conditions that allow one to find an expected utility function $\varphi : \Delta \to \mathbb{R}$ and a continuous, affine[7] function $W : \mathcal{X} \to \mathbb{R}$ such that, for each $x \in \mathcal{X}$ and $p \in x$, the conditions in (1) and (2) hold. In what follows, such a pair (φ, W) will be referred to as an **affine representation**. To simplify our notation, we will write $\max \varphi(x)$ instead of $\max_{q \in x} \varphi(q)$.

3. The Basic Axioms and the "Anomalies" Regarding Deferral

We start with adapting classical axioms of rational choice theory (see, e.g., Kreps, 1988, pp. 11-15).

Property α. Let $x, y \in \mathcal{X}$ be such that $x \subseteq y$, $\mathbf{c}(x) \subseteq \Delta$ and $\mathbf{c}(y) \subseteq \Delta$. Then, $p \in \mathbf{c}(y) \cap x$ implies $p \in \mathbf{c}(x)$.

Property β. Let $x, y \in \mathcal{X}$ be such that $x \subseteq y$, $\mathbf{c}(x) \subseteq \Delta$ and $\mathbf{c}(y) \subseteq \Delta$. Then, $p, q \in \mathbf{c}(x)$ and $q \in \mathbf{c}(y)$ imply $p \in \mathbf{c}(y)$.

The only difference of these axioms from their traditional versions is that the scope of α and β, as formulated above, is restricted to those situations in which the agent does not defer choice. In particular, to account for the related behavior noted by Tversky and Shafir (1992), we allow instances of the following form:

$$\mathbf{c}(y) = \{\ominus\} \quad \text{and} \quad \mathbf{c}(x) \neq \{\ominus\} \text{ for some } x \subseteq y. \tag{7}$$

On the other hand, together, properties α and β are equivalent to the next property, which is none but *a restricted version of the weak axiom of revealed preferences*:

For any $x, y \in \mathcal{X}$ with $x \subseteq y$ and $\mathbf{c}(x), \mathbf{c}(y) \subseteq \Delta$, we have $\mathbf{c}(y) \cap x = \mathbf{c}(x)$ if $\mathbf{c}(y) \cap x \neq \varnothing$.

This asymmetry in the model enables our theory to capture the context dependent nature of the value of deferral. When viewed from this perspective, (7) amounts to saying that flexibility gained by deferring choice may be more valuable when one faces a larger choice set.

There is also some experimental evidence which indicates that subjects sometimes

[7] The function $W : \mathcal{X} \to \mathbb{R}$ is **affine** if $W(\lambda x + (1-\lambda)y) = \lambda W(x) + (1-\lambda)W(y)$, for any $x, y \in \mathcal{X}$ and $\lambda \in (0,1)$. (As usual, the convex combination $\lambda x + (1-\lambda)y$ is defined as $\{\lambda p + (1-\lambda)q : p \in x, q \in y\}$. Convex combinations involving more than two sets are defined analogously.)

tend to defer upon removal of some of the feasible alternatives:[8]

$$\mathbf{c}(y) \cap \Delta \neq \varnothing \quad \text{and} \quad \mathbf{c}(x) = \{\ominus\} \text{ for some } x \subseteq y. \tag{8}$$

It should be noted, however, that instances of this form are also consistent with the behavior of an agent who deems the act of deferring "just like any other option," as the removal of a superior alternative p from a choice set y may trigger deferral in a rather trivial way. However, the following stronger form of (8) cannot be explained without incorporating context dependence:

$$\mathbf{c}(y) \cap x \neq \varnothing \quad \text{and} \quad \mathbf{c}(x) = \{\ominus\} \text{ for some } x \subseteq y. \tag{9}$$

Intuitively, such behavior may be considered as a manifestation of a decrease in the value of deferral when additional alternatives become available, even if these additional alternatives are inferior to those which were available at the outset. If deferring choice when facing a larger set of feasible alternatives is associated with having more flexibility in period 2, such decrease in the value of deferral may stem from a desire for less flexibility in the second period of the decision problem. For instance, the agent may believe that in period 2 she will be under the influence of temptation which might force her to exercise costly self-control when many alternatives are available. (For more on this, see Section 5 below.)

In the next section, we will couple properties α and β with some independence and continuity axioms. Since α and β do not impose any restriction on behavior across situations in which the agent defers, the obtained theory will be compatible with instances of the form (7) and (9). It should be noted, however, that our model is *not* meant to cover all modes of related behavior. Consider, for instance, an agent who finds it difficult to search for the best options in a large choice set. Such an agent would presumably violate our property α, for when faced with a large choice set, she could opt for an inferior alternative that she wouldn't select from a smaller set, where some of the superior alternatives can easily be identified. Of course, this form of behavior is also compatible with instances of the form (7), as the agent may simply defer when the set of feasible options is so large that she cannot identify any alternative which is worth selecting. While this is not a rationalizable mode of behavior because of the failure of property α, in Section 7 we will see that Gerasímou's (2010) aforementioned model satisfies all of the essential axioms that we will utilize in this paper.

In passing, let us note that in the present framework, properties α and β (or the restricted form of the weak axiom of revealed preferences) are not sufficient for rational-

[8] See, for instance, Dhar (1997) and White and Hoffrage (2009).

izability. For example, suppose that the agent defers with certainty unless the choice set that she faces equals $x := \{p, q, r\}$ or $y := \{p, q, r'\}$. Also assume that $\mathbf{c}(x) = \{p\}$ while $\mathbf{c}(y) = \{q\}$. Then, since x and y are not nested with respect to set inclusion, properties α and β hold voidly. However, it appears that the agent prefers p over q when r is available, while she prefers q over p when r' is available. Here, the trouble stems from the fact that the choice data is so scarce that the properties α and β do not impose a meaningful restriction on behavior. In the next section, we will see that any choice correspondence that respects some independence and continuity axioms is rich enough to render α and β functional.

4. Representation Theorems

Our general representation requires the following standard continuity axiom.

Continuity. Let (x_n) be a sequence in $\mathcal{X}$ that converges to a choice set x.
(i) If $p_n \in \mathbf{c}(x_n) \cap \Delta$ for each n and $p_n \to p$, then $p \in \mathbf{c}(x)$.
(ii) If $\ominus \in \mathbf{c}(x_n)$ for each n, then $\ominus \in \mathbf{c}(x)$.

Next, we introduce a nontriviality axiom that requires the presence of an instance in which the agent defers with certainty, and another one in which she avoids deferral. Intuitively, this implies that the decision to defer entails serious trade-offs.

Regularity. There exist $y_0, y_\ominus \in \mathcal{X}$ such that $\mathbf{c}(y_0) \subseteq \Delta$ and $\mathbf{c}(y_\ominus) = \{\ominus\}$.

Our first independence axiom consists of two parts which serve for quite different purposes:

Independence-C. Pick any $\lambda \in (0, 1)$, and let $x, y \in \mathcal{X}$ be such that $\mathbf{c}(x) \cap \Delta \neq \varnothing$ and $\mathbf{c}(y) \subseteq \Delta$. Then:
(i) For every $p \in x$ and $q \in y$, we have

$$\lambda p + (1 - \lambda)q \in \mathbf{c}(\lambda x + (1 - \lambda)y) \quad \text{iff} \quad p \in \mathbf{c}(x) \text{ and } q \in \mathbf{c}(y).$$

(ii) $\ominus$ does not belong to $\mathbf{c}(\lambda x + (1 - \lambda)y)$.

The first part of independence-c is satisfied in the standard model of choice under risk. However, in axiomatic treatments of the standard model, one would often assume a weaker form of independence that requires the compound lottery $\lambda p + (1 - \lambda)r$ to be selected from $\lambda x + (1 - \lambda)\{r\}$ whenever p is selected from x. We find the stronger form above useful because the structure of the present exercise does not allow us to elicit the preference relation of the agent in the standard way, by declaring p to be revealed preferred to q if, at least in one instance, the agent selects p while q is available. In the present setup, it may be impossible to determine the rank of a given pair of lotteries with the standard method. Indeed, at the outset all we know is that, by the regularity axiom, there exists a choice set $y_0 \in \mathcal{X}$ with $\mathbf{c}(y_0) \subseteq \Delta$. By using the second part of the continuity axiom, one can then show that actually for any choice set y that is sufficiently close to y_0 we have $\mathbf{c}(y) \subseteq \Delta$. In particular, for any pair of lotteries p, q, one can find a sufficiently small number $\lambda \in (0,1)$ such that $\mathbf{c}(\lambda\{p,q\} + (1-\lambda)y_0) \subseteq \Delta$. Then, it becomes natural to view p to be revealed preferred to q whenever $\lambda p + (1-\lambda)p_0 \in \mathbf{c}(\lambda\{p,q\} + (1-\lambda)y_0)$ for some $p_0 \in y_0$. This discussion also clarifies how the strength of the properties α and β is restored in the present cardinal approach.[9]

On the other hand, the second part of independence-c is not relevant to the problem of eliciting the agent's preference relation over lotteries. Rather, it regulates decisions regarding deferral. Specifically, it requires that the agent should not defer whenever a compound lottery is available that is obtained by mixing two lotteries one of which is "strictly preferred to deferral" while the other one is "weakly preferred to deferral." The reader will note that this reasoning is actually based on the assumption that with regard to the independence property, the act of deferring choice can be considered just like any other option. While this idea seems reasonable by itself, our approach can also be justified by viewing the mixture of two choice sets as a lottery over those sets, as in the literature on menu choice (see, e.g., Dekel et al., 2001; and Gul and Pesendorfer, 2001). We now introduce a further axiom of similar nature:

Independence-D. Let $x, y \in \mathcal{X}$ be such that $\ominus \in \mathbf{c}(x)$ and $\mathbf{c}(y) = \{\ominus\}$. Then, $\mathbf{c}(\lambda x + (1-\lambda)y) = \{\ominus\}$ for every $\lambda \in (0,1)$.

Independence-d is simply the dual of the second part of independence-c, and hence, it does not require further motivation. Let us note, however, that one might be able to model interesting behavior by relaxing independence-d as well as the second part of independence-c, just as it is the case with the setwise independence axioms of the literature on menu choice. We will comment on this issue in Section 9.

In what follows, we say that an affine representation (φ, W) for a choice correspondence $\mathbf{c}$ is **regular** if there exist $y_0, y_\ominus \in \mathcal{X}$ such that $\max \varphi(y_0) > W(y_0)$ and $\max \varphi(y_\ominus) < W(y_\ominus)$. (Similarly, we will say that $\mathbf{c}$ is **regular** if it satisfies the regularity axiom.) The next theorem shows that the axioms that we have introduced so far are necessary and sufficient for the existence of a regular affine representation.

Theorem 1. *A choice correspondence* $\mathbf{c}$ *on* $\mathcal{X}$ *satisfies properties* α, β, *independence-c, independence-d, continuity and regularity if, and only if, it admits a regular affine*

[9] Another paper that deals with the problem of eliciting preference from scarce data is due to Stoye (2010), but he assumes that all convex sets are in the domain of the choice correspondence, as he is concerned with agents who randomize their choices.

representation.

Theorem 1 is the most general representation result that we report in this paper. In what follows, our focus will be the subclass of subjective state space representations. An alternate form that is covered by Theorem 1 will be discussed in Section 9. First, we wish to examine the degree of uniqueness of affine representations.

4.1 Uniqueness of Affine Representations

Notice that, given a set of feasible alternatives x, a choice correspondence $\mathbf{c}$ gives three pieces of information about the agent's possible behavior: (i) whether she may decide to defer choice; (ii) whether she may select some of the available alternatives immediately; (iii) if so, which alternatives are likely to be selected.

Suppose now that the choice correspondence $\mathbf{c}$ admits an affine representation (φ, W). Then, by definitions, the sign of $W(x) - \max \varphi(x)$ encodes all information embedded in $\mathbf{c}$ with regard to points (i) and (ii) above. Moreover, in this case, maximization of φ on x determines the content of $\mathbf{c}$ with regard to (iii). It thus follows that given any $(\varphi', W') \in \mathbb{R}^{\Delta} \times \mathbb{R}^{\mathcal{X}}$, if $W'(\cdot) - \max \varphi'(\cdot)$ is a positive multiple of $W(\cdot) - \max \varphi(\cdot)$, and if φ' is a positive affine transformation of φ, then the pair (φ', W') is also an affine representation for $\mathbf{c}$. The next lemma shows that the converse of this observation also holds.

Lemma 1. *Let $\mathbf{c}$ be a regular choice correspondence on $\mathcal{X}$ that admits an affine representation (φ, W). Then, a pair $(\varphi', W') \in \mathbb{R}^{\Delta} \times \mathbb{R}^{\mathcal{X}}$ is another affine representation for $\mathbf{c}$ if, and only if, φ' is a positive affine transformation of φ, and there exists a $\gamma > 0$ such that*

$$W'(x) - \max \varphi'(x) = \gamma(W(x) - \max \varphi(x)),$$

for each $x \in \mathcal{X}$.

Lemma 1 will be our main tool when discussing the identification of subjective state space representations in Section 7. We now formally introduce this representation notion.

4.2 Subjective State Space Representations

Given a (countably additive and signed) measure μ on the set

$$\mathcal{S} := \{s \in \mathbb{R}^B : \sum_{b \in B} s(b) = 0, \ \|s\| = 1\},$$

for each $x \in \mathcal{X}$ we define

$$W_\mu(x) := \int_{\mathcal{S}} \max_{q \in x} U(s, q) \, \mu(ds), \tag{10}$$

where $U(s,q)$ stands for the expected value of s with respect to q, that is, $U(s,q) := \sum_{b\in B} s(b)q(b)$. A **subjective state space representation** (henceforth, **S-representation**) refers to an affine representation of the form (φ, W_μ). We will often write (φ, μ) instead of (φ, W_μ).

Following the literature on menu choice, we interpret the set $\mathcal{S}$ as a (normalized) subjective state space that consists of all nontrivial preference relations over Δ which satisfy the hypotheses of the classical expected utility theory. In turn, if the set of feasible alternatives remains unchanged upon deferral, the measure μ in an S-representation can be considered as an additive aggregator that translates the statewise maximum expected utilities on a given choice set to the anticipated second period utility associated with that set. As usual, in our model of preference for flexibility, we will require μ to be positive so that its support can be interpreted as the set of probable modes of behavior that the agent may adopt in period 2.

As shown by Dekel, Lipman, Rustichini and Sarver (2007) (henceforth, DLRS), an affine function W on $\mathcal{X}$ can be written in the form (10) only if it is Lipschitz continuous.[10] We now impose a technical axiom that ensures this property in the present framework, which is quite similar to the corresponding axiom introduced by DLRS.

L-Continuity. For any $y_\ominus \in \mathcal{X}$ with $\mathbf{c}(y_\ominus) = \{\ominus\}$ and $y_0 \in \mathcal{X}$ with $\mathbf{c}(y_0) \subseteq \Delta$, there exists $n > 0$ such that for every $\varepsilon \in (0, 1/n)$ and every $x, y \in \mathcal{X}$ with $d_h(x,y) \leq \varepsilon$,

$$\ominus \in \mathbf{c}((1-n\varepsilon)y + n\varepsilon y_0) \quad \text{implies} \quad \ominus \in \mathbf{c}((1-n\varepsilon)x + n\varepsilon y_\ominus).$$

To understand the content of the axiom, let y_0 and $y_\ominus$ be as above, and consider $x, y \in \mathcal{X}$ with $\mathbf{c}(y) = \{\ominus\}$ and $\mathbf{c}(x) \subseteq \Delta$. Then, the first part of the continuity axiom implies that $\mathbf{c}\left(\lambda y + (1-\lambda)y_0\right) = \{\ominus\}$ for all sufficiently large $\lambda \in (0,1)$. However, in general, it is indeterminate if for any such λ we will have $\ominus \in \mathbf{c}\left(\lambda x + (1-\lambda)y_\ominus\right)$. The L-continuity axiom implies that as x and y get arbitrarily close to each other, there exists a range of λ over which the first property implies the second, and the upper bound of this range converges to 1 smoothly.

The next lemma shows that this axiom characterizes the class of affine representations that take the particular form in question.

Lemma 2. *Let* $\mathbf{c}$ *be a regular choice correspondence on* $\mathcal{X}$ *that has an affine representation. Then,* $\mathbf{c}$ *satisfies the L-continuity axiom if, and only if, it also admits an S-representation.*

[10] A real function g on a set $\mathcal{X}' \subseteq \mathcal{X}$ is **Lipschitz continuous** if there exists $\bar{n} > 0$ such that $g(x) - g(y) \leq \bar{n} d_h(x,y)$ for all $x, y \in \mathcal{X}'$. The number $\bar{n}$ is referred to as the **Lipschitz coefficient** of g.

4.3 Representation of Preference for Flexibility

As we have seen in Section 3, the instances of the form (9) may be considered as a manifestation of preference for commitment motivated by potential harms of having many options in the second period of the choice problem. The next axiom rules out this sort of behavior, and thereby, implies that the decision maker has preference for flexibility:

Monotonicity. Let $x, y \in \mathcal{X}$ with $x \subseteq y$. Then, $p \in \mathbf{c}(y) \cap x$ implies $p \in \mathbf{c}(x)$.

We now show that the monotonicity axiom is equivalent to requiring an S-representation with a nonnegative aggregator.

Theorem 2. *Let* $\mathbf{c}$ *be a regular choice correspondence on* $\mathcal{X}$ *that has an S-representation. Then,* $\mathbf{c}$ *satisfies the monotonicity axiom if, and only if, it admits an S-representation* (φ, μ) *with* $\mu \geq 0$.

It is readily seen that if a choice correspondence $\mathbf{c}$ admits an affine representation, then, in fact, the monotonicity axiom is *equivalent* to ruling out the "anomaly" (9).[11] Therefore, Theorem 2 can be restated as follows:

Theorem 2'. *Let* $\mathbf{c}$ *be a regular choice correspondence on* $\mathcal{X}$ *that has an S-representation. Then, the following two conditions are equivalent.*
(i) *There do not exist* $x, y \in \mathcal{X}$ *which satisfy* (9).
(ii) $\mathbf{c}$ *admits an S-representation* (φ, μ) *with* $\mu \geq 0$.

It should be noted that unlike the corresponding theorem of DLR on representation of preference for flexibility, in Theorem 2, the existence of an S-representation (and hence, the L-continuity axiom) is assumed at the outset. This modeling choice is motivated by the lesser degree of identification in our representation theorems. Specifically, if (φ, W) is an affine representation for a choice correspondence, then so is, say, $(\frac{1}{2}\varphi, W - \frac{1}{2}\max\varphi)$. This implies that a choice correspondence that satisfies the monotonicity axiom may admit an affine representation (φ, W) that does *not* respect the usual setwise monotonicity property ($W(y) \geq W(x)$ if $y \supseteq x$). Consequently, in the proof of Theorem 2, we are forced to select a "well-behaved" representation that respects this property. If $\mathbf{c}$ admits an S-representation, this is a relatively simple task. As a first step, in the proof of Theorem 2, we show that the monotonicity axiom implies that for any S-representation (φ, μ), the restriction of μ to $\mathcal{S} \setminus \{s^*\}$ is nonnegative, where s^* stands for the unique element of $\mathcal{S}$ that represents the same preference relation over lotteries

[11] Of course, here "affinity" of the representation is irrelevant. All one needs to observe is that such a correspondence satisfies the following version of the weak axiom: For any $x, y \in \mathcal{X}$ with $x \subseteq y$, $\mathbf{c}(x) \cap \Delta \neq \varnothing$ and $\mathbf{c}(y) \cap \Delta \neq \varnothing$, we have $\mathbf{c}(y) \cap x = \mathbf{c}(x) \cap \Delta$ whenever $\mathbf{c}(y) \cap x \neq \varnothing$. (The continuity axiom is the main force behind the validity of this slightly stronger version.)

as φ (whenever the latter is a nonconstant function).[12] Then, we deal with the case $\mu(\{s^*\}) < 0$ by shifting φ and μ simultaneously in such a way that the measure μ' obtained upon this transformation satisfies $\mu'(\{s^*\}) = 0$. At present, we do not know if an analogous selection argument would be feasible for an arbitrary affine representation.

Another point to note is that in Theorem 2, the aggregator μ is allowed to be the zero measure, which corresponds to the case $W_\mu = 0$. In turn, this implies a decision maker who views deferral as any other alternative, which is not compatible with the interesting modes of behavior under investigation. On the other hand, by itself, letting $\mu > 0$ is not very helpful either, for s^* may happen to be the only probable state. For example, let μ be the point mass of s^*, which we denote as δ_{s^*}, and suppose that $\varphi(\cdot) = \gamma U(s^*, \cdot) + \theta$ for some $\gamma > 1$ and $\theta \in \mathbb{R}$. Then, by Lemma 1, the function $\varphi'(\cdot) := (\gamma - 1)U(s^*, \cdot) + \theta$ together with the zero measure on $\mathcal{S}$ give us another S-representation for the choice correspondence in question, which takes us back to the uninteresting case. Therefore, it seems to be in order to focus on those cases in which s^* is not the only the probable state. This is equivalent to requiring the measure $\mu_{-*} := \mu - \mu(\{s^*\})\delta_{s^*}$ to be nonzero. The next technical axiom characterizes this property in the more general case of a signed measure μ.

Nontriviality-D. There exist $x, y \in \mathcal{X}$ such that $\mathbf{c}(x) \cap \mathbf{c}(y) \cap \Delta \neq \varnothing$ and $\ominus \in \mathbf{c}(x) \setminus \mathbf{c}(y)$.

In terms of an S-representation (φ, μ), this axiom requires the existence of a pair of choice sets x, y such that $W_\mu(x) = \max \varphi(x) = \max \varphi(y) > W_\mu(y)$. In this case, W_μ is not only nonconstant, but it is also ordinally distinct from both $\max \varphi$ and $-\max \varphi$. This, in turn, obviously implies that μ is a nonzero measure which is not equal to any multiple of δ_{s^*}, as we seek. Before showing that the converse also holds, we formally exclude the cases in which φ is constant:

Nontriviality-C. There exists an $x \in \mathcal{X}$ such that $\mathbf{c}(x) \cap \Delta$ is a nonempty, proper subset of x.

The next lemma characterizes the class of nontrivial representations that we seek.

Lemma 3. *Let* $\mathbf{c}$ *be a regular choice correspondence on* $\mathcal{X}$ *that has an S-representation. Then,* $\mathbf{c}$ *satisfies nontriviality-c and nontriviality-d if, and only if, it admits an S-representation* (φ, μ) *such that* μ_{-*} *is nonzero and* φ *is nonconstant.*

By combining this observation with Theorem 2, we obtain a characterization of those cases in which μ can be chosen to be a "nontrivial" probability measure:

[12] It is important to note that s^* does not depend on the choice of a particular S-representation, for the function φ is unique up to positive affine transformations.

Corollary 1. *Let* **c** *be a regular choice correspondence on* $\mathcal{X}$ *that has an S-representation. Then,* **c** *satisfies nontriviality-c, nontriviality-d and monotonicity if, and only if, it admits an S-representation* (φ, μ) *such that* φ *is nonconstant and* μ *is a probability measure with* $\mu(\mathcal{S} \setminus \{s^*\}) > 0$.

Next, we examine potential forms of φ that are compatible with the interpretation that underlies an S-representation.

4.4 On Possible Forms of φ

If $W(\{p\})$ is interpreted as the anticipated utility of consuming p in period 2, we may consider the following specific form for φ:

$$\varphi(p) = W(\{p\}). \tag{11}$$

The specification (11) identifies the situation when choices in period 1 simply determine what will be consumed in period 2. Alternatively, without the restriction in (11), $\varphi(p)$ may correspond to either the case where an alternative that is chosen in period 1 is consumed in both periods, or the case in which consumption takes place only in one period and what is chosen in period 1 has to be consumed immediately. Of course, without demanding any further structure from φ, these two cases cannot be distinguished.

It is crucial to note that the model of preference for flexibility that we presented above is not consistent with the regularity axiom under the specification (11), since when $\mu \geq 0$, we have $W_\mu(x) \geq \max_{p \in x} W_\mu(\{p\})$ for any choice set x, implying that the agent would never consider deferral as an inferior option. By contrast, under the other specifications for φ involving consumption in period 1, the preference for flexibility model is consistent with the regularity axiom, since in the associated cases, deferral may incur significant costs because of the foregone consumption in period 1.

5. Explaining the "Anomalies" Regarding Deferral

Consider a regular choice correspondence **c** *that admits an S-representation* (φ, μ). To avoid trivialities, we will presume that φ is nonconstant and, as before, denote with s^* the state in $\mathcal{S}$ that represents the same preference relation over Δ as φ.[13] Let μ^+ and μ^- be the positive and negative parts of μ with respect to the Jordan decomposition (which are uniquely defined), so that $\mu = \mu^+ - \mu^-$. Roughly speaking, the support of μ^+ can be viewed as the set of states in which flexibility will be valuable in period 2,

[13]When φ is constant, one can let s^* be the origin of $\mathbb{R}^B$. With this convention, the results that we report below would remain valid.

and the support of μ^- as the set of states in which flexibility will be harmful. We recall, however, that the sign of $\mu(s^*)$ is not very informative about the agent's behavior, as one can represent the same choice correspondence with more than one aggregators that attach different signs to the weight of the state s^*. The next proposition shows that whenever s^* is not the only "positive state," the behavior described in (7) occurs by necessity.

Proposition 1. *Suppose* $\mu^+ (\mathcal{S} \setminus \{s^*\}) > 0$. *Then, there exist* $x, y \in \mathcal{X}$ *that satisfy* (7).

In the context of preference for flexibility, Proposition 1 simply says that instances of the form (7) will occur unless the agent knows for sure that her tastes will not change at all. In particular, any choice correspondence as in Corollary 1 exhibits such instances. In turn, the existence of a "negative state" that is distinct from s^* is necessary and sufficient for the occurrence of the second type of behavior under investigation:

Proposition 2. $\mu^- (\mathcal{S} \setminus \{s^*\}) > 0$ *if and only if there exist* $x, y \in \mathcal{X}$ *that satisfy* (9).[14]

The main idea behind Propositions 1 and 2 is embedded in the proof of Theorem 2, which is closely related to the analysis of DLR concerning the uniqueness of subjective state space representations in their framework. Specifically, in the proof of Theorem 2, we show that if $\mu^- (\mathcal{S} \setminus \{s^*\}) > 0$, then one can perturb a sphere $z \subseteq \Delta$ to obtain a set z' such that

(i) $z \subseteq z'$; (ii) $\max_{q \in z} U(s^*, q) = \max_{q \in z'} U(s^*, q)$; and (iii) $W_\mu(z) > W_\mu(z')$.[15,16]

Clearly, condition (ii) amounts to saying that $\max \varphi(z) = \max \varphi(z')$. Moreover, if we consider a pair of sets $y_0, y_\ominus \in \mathcal{X}$ as in the regularity axiom, for suitably chosen $y_* \in \{y_0, y_\ominus\}$ and $\lambda \in (0, 1]$, we will have $W_\mu(\lambda z + (1-\lambda) y_*) = \max \varphi(\lambda z + (1-\lambda) y_*)$. Then, since $\max \varphi(\lambda z + (1-\lambda)y_*) = \max \varphi(\lambda z' + (1-\lambda)y_*)$ and $W_\mu(\lambda z + (1-\lambda)y_*) > W_\mu(\lambda z' + (1-\lambda)y_*)$, it also follows that $\max \varphi(\lambda z' + (1-\lambda)y_*) > W_\mu(\lambda z' + (1-\lambda)y_*)$. Hence, for a sufficiently small $\gamma \in (0, 1)$, the sets $x := \gamma y_\ominus + (1-\gamma)(\lambda z + (1-\lambda)y_*)$ and $y := \gamma y_\ominus + (1-\gamma)(\lambda z' + (1-\lambda)y_*)$ must satisfy the following:

$$W_\mu(x) > \max \varphi(x) = \max \varphi(y) > W_\mu(y),$$

[14] By the shifting argument that we noted in Section 4.3, when $\mu^- (\mathcal{S} \setminus \{s^*\}) = 0$ we may as well assume $\mu \geq 0$, and thereby, rule out instances of the form (9). Hence the "if" part of Proposition 2. However, the analogue of this conclusion does not hold in the context of Proposition 1, for when $\mu^+ (\mathcal{S} \setminus \{s^*\}) = 0$, depending on the size of $\mu(\{s^*\})$ relative to the "strength" of first stage preferences we may or may not observe instances of the form (7).

[15] Dekel et al. (2009) utilize the structure of spheres in a similar way in an environment with finitely many relevant states.

[16] Using the terminology of DLR, from (ii) and (iii) it follows, in particular, that the state s^* is not "sufficient."

which imply that $\mathbf{c}(x) = \{\ominus\}$ and $\mathbf{c}(y) \cap x = \arg\max_{q \in x} \varphi(q) \neq \varnothing$, as in (9). This proves the "only if" part of Proposition 2. (For the "if" part, see Footnote 14.)

Similarly, the proof of Proposition 1 builds upon the existence of sets z, z' with $W_\mu(z) < W_\mu(z')$, which also satisfy the properties (i) and (ii) above. (We omit the details.)

As a concrete example that is consistent with both types of behavior under investigation, let us consider an agent who is on a diet and who may have a friend at dinner. In the morning, she contemplates going for shopping now, or later in the evening when she will know if her friend will really be able to join her. If she will have to eat alone, she thinks she should cook vegetarian food. On the other hand, if her friend will join her, she thinks it would be better to cook fish, an enjoyable option that goes against her diet. Moreover, she anticipates that if she goes for shopping in the evening, she will be hungry and therefore be tempted by steak, and to a lesser extent, by fish. Thus, postponing the task of shopping allows the agent to make a more informed decision, which may come at the cost of having to exercise self-control. In turn, committing to an alternative in the morning has the benefit of avoiding temptation, but the alternative in question may turn out to be inferior. So, should the agent shop in the morning without exactly knowing if her friend will come or, in the evening under temptation, possibly after learning that her friend will not be able to come?

In this example, the agent's normative preferences (about what she should do) depend on if her friend will join her. More generally, we may consider the following representation that covers this story:

$$\begin{aligned} W_\mu(x) &= \sum_{i=1}^{I} \lambda_i \left\{ \max\nolimits_{p \in x} \left(U(s_i, p) - \left(\max\nolimits_{q \in x} U(s^-, q) - U(s^-, p) \right) \right) \right\} \\ &= \sum_{i=1}^{I} \lambda_i \left\{ \max\nolimits_{p \in x} U(s_i + s^-, p) - \max\nolimits_{q \in x} U(s^-, q) \right\}. \end{aligned} \tag{12}$$

Here, s_i's and s^- are distinct elements of $\mathcal{S}$ that represent the agent's normative and temptation rankings, respectively, while the term $\max_{q \in x} U(s^-, q) - U(s^-, p)$ captures the anticipated self-control cost that the agent incurs by choosing p instead of the most tempting, feasible alternative in period 2. The agent assigns the probability $\lambda_i > 0$ to each normative ranking, and plans to maximize the compromise utility $U(s_i + s^-, \cdot)$ if she decides that s_i is the "right" ranking. Using our previous terminology, then, the support of the positive part of the associated aggregator is $\mathcal{S}^+ := \{\frac{s_i + s^-}{\|s_i + s^-\|} : s_i + s^- \neq 0$, $i = 1, ..., I\}$, while the only negative state is s^-. (Notice that $|\mathcal{S}^+| \geq I - 1$ and $s^- \notin \mathcal{S}^+$ since $s_1, ..., s_I$ and s^- are assumed to be distinct.) In turn, with the specification (11), the agent's utility from committing to an alternative p in period 1 is $\varphi(p) = W_\mu(\{p\}) =$

$\sum_{i=1}^{I} \lambda_i U(s_i, p)$. In this setup, s^* will be typically different than s^-, and hence, the conclusion of Proposition 2 will apply. Moreover, if $I \geq 2$, Proposition 1 is certainly applicable, for then at least one element of $\mathcal{S}^+$ is distinct from s^*.[17]

The model that we have just sketched is certainly interesting as it is general enough to produce instances of both types of the seemingly anomalous behavior in a relatively simple framework. However, laying axiomatic foundations of this type of an S-representation is beyond the scope of the present paper, for not much is known about value functions that take the form (12). Rather, to our knowledge, in the menu choice literature the related models on temptation and self-control focus on those cases in which there is only one possible normative ranking.[18] In turn, if consumption takes place only in period 2, in that setup the agent would have no reason to defer, and hence, to model interesting behavior one would have to consider additional intertemporal trade-offs.

5.1 The Standard Model as a Particular Case

The following standard axioms describe a decision maker who views deferral like any other option.

Unrestricted α. Let $x, y \in \mathcal{X}$ be such that $x \subseteq y$. Then, $p \in \mathbf{c}(y) \cap x$ implies $p \in \mathbf{c}(x)$, and $\ominus \in \mathbf{c}(y)$ implies $\ominus \in \mathbf{c}(x)$.

Unrestricted β. Let $x, y \in \mathcal{X}$ be such that $x \subseteq y$. Then, $\rho, \rho' \in \mathbf{c}(x)$ and $\rho' \in \mathbf{c}(y)$ imply $\rho \in \mathbf{c}(y)$.

The reader will note that unrestricted α rules out (9) immediately. As for (7), consider a pair of choice sets x, y with $x \subseteq y$ and $\mathbf{c}(y) = \{\ominus\}$. Then, unrestricted α implies $\ominus \in \mathbf{c}(x)$. Hence, if we were to have $\mathbf{c}(x) \neq \{\ominus\}$, unrestricted β would imply $\mathbf{c}(y) \neq \{\ominus\}$, a contradiction. It follows that under these axioms, neither (7) nor (9) can occur. Hence, if $\mathbf{c}$ is a choice correspondence that admits an S-representation (φ, μ), by Propositions 1 and 2, we must have $\mu(\mathcal{S} \setminus \{s^*\}) = 0$. It is also not difficult to see that without loss of generality we can in fact assume $\mu(\{s^*\}) = 0$. Thus, the standard model is characterized as the particular case that allows one to set $W_\mu = 0$.

6. More on Choice Avoidance and Anticipated Regret

In many real-life situations, deferring the selection of an alternative is impossible, yet the decision maker has the option of not making an active choice, as in the case of a

[17] When $|\mathcal{S}^+| \geq 2$ this is trivially true. Thus, the only case that requires caution is of the form $I = 2$, $s^- \neq s_1 \neq s_2 = -s^-$. But then, $\lambda_1 s_1 + \lambda_2 s_2$ and $s_1 + s^-$ can be collinear only if so are s_1 and s^-, which is not the case.

[18] See, e.g., Gul and Pesendorfer (2001), Dekel et al. (2009) and Stovall (2010).

consumer who faces the problem of whether to buy a new dress for an upcoming party. If we relabel $\ominus$ so as to denote the associated outside option or initial endowment, our theory can be used to analyze such choice problems.

In this context, independence-d and the second part of independence-c are motivated more strongly, for it becomes more natural to consider $\ominus$ like any other option with regard to the independence properties. Yet, as we discussed earlier, experimental evidence points to biased behavior in favor of the default option in difficult choice problems. To model this phenomenon, it is natural to rule out instances of the form (9), as the removal of inferior alternatives should only simplify the choice problem, and thereby, promote an active choice. Thus, Theorem 2 is readily applicable in this context. As we have seen in the introduction, the corresponding S-representation can be transformed to facilitate an interpretation based on the notion of anticipated regret. It is also worth noting that, by the discussion in Section 5.1, in this context unrestricted α and β simply describe a classical decision maker who attaches the normalized value 0 to the default option $\ominus$.

Let us now consider a triplet (Φ, v, μ) that represents a choice correspondence $\mathbf{c}$ in the sense of (4) and (6), where μ is a nonnegative measure on our normalized state space, v is an expected utility function on Δ, and $\Phi(\cdot, \cdot)$ is defined by μ and v as in (6). Then, the corresponding S-representation of $\mathbf{c}$ takes the form (φ, μ) where $\varphi(p) := v(p) + W_\mu(\{p\})$ for $p \in \Delta$. The regularity axiom requires v to attach positive utility to some alternatives, for otherwise the decision maker would never make an active choice. In turn, if $v(p)$ is thought of as the aggregate, material utility associated with ex-ante and ex-post benefits of choosing p, the behavior entailed by maximization of φ can be viewed as a further compromise in favor of ex-post preferences, which may come about as a consequence of the agent's urge to avoid ex-post regret. It is equally conceivable that ex-post consumption is the only source of utility, and $v(p)$ takes the form $KW_\mu(\{p\})$ for some $K > 0$. In this case, the rank of lotteries with respect to φ and W_μ would coincide. Hence, the agent's choices among lotteries would be governed by maximization of the restriction of W_μ to singleton sets. Moreover, as can be seen from the definition of $\Phi(\cdot, \cdot)$ in (6), one may consider $1/K$ as a measure of the agent's sensitivity to anticipated regret when comparing ordinary alternatives with the default option. In any case, whenever μ is nondegenerate, that is, whenever the agent is genuinely uncertain about her tastes, instances of the form (7) occurs by Proposition 1.

7. Identifying the Relevant States and the Aggregator

In view of our previous discussions, it is clear that without further assumptions, the state s^* that governs the agent's first-stage choices among alternatives cannot be identified in an S-representation. We shall now show that excluding the state s^*, the support of the aggregator associated with an S-representation is uniquely determined by the choice data. Moreover, for any state $s \neq s^*$, the importance of s relative to all other states which are distinct from s^* can also be identified. Put formally, the main finding of this section shows that the measure $\mu_{-*} := \mu - \mu(\{s^*\})\delta_{s^*}$ induced by the aggregator μ associated with an S-representation is unique up to positive multiplication:

Theorem 3. *Let (φ, μ) and (φ', μ') be S-representations for a choice correspondence* **c** *that satisfies regularity and nontriviality-c. Then, there exists a number $\gamma > 0$ such that $\mu'_{-*} = \gamma\mu_{-*}$.*

Remark. If nontriviality-c fails, an S-representation is identified in a stronger sense, for then the measure μ itself is unique up to positive multiplication. (This can be verified by letting s^* be the origin of $\mathbb{R}^B$ so that $\mu(\{s^*\}) = 0$ and $\mu_{-*} = \mu$ by definitions.)

It follows from Theorem 3 that if one fixes the (total variation) norm of μ_{-*}, this measure can be determined uniquely. We now state a version of Theorem 3 that builds upon this method of normalization. In what follows, given a number $k > 0$, we say that an S-representation (φ, μ) for a choice correspondence is **semi k-normalized** if $\|\mu_{-*}\| = k$ and φ is nonconstant.

Theorem 3'. *Let* **c** *be a regular choice correspondence on $\mathcal{X}$ that admits an S-representation. Then:*
(i) **c** *satisfies nontriviality-c and nontriviality-d if, and only if, it admits a semi k-normalized S-representation for some $k > 0$.*
(ii) *In fact, if* **c** *satisfies nontriviality-c and nontriviality-d, for any $k > 0$, there exists a semi k-normalized S-representation.*
(iii) *Given any $k > 0$, if (φ, μ) and (φ', μ') are semi k-normalized S-representations for* **c***, then we have $\mu_{-*} = \mu'_{-*}$.*

We next turn to the identification problem related to the state s^*, which is caused by the freedom of shifting φ and μ simultaneously. There are two methods of normalization that one can utilize to eliminate this identification problem per force. First, one can specify a number $k^* > 0$ and require that $\varphi(p^*) - \varphi(p_*) = k^*$ for a pair of predetermined alternatives p^*, p_*. The next lemma establishes the existence and uniqueness of an S-representation that satisfies this additional property.

Lemma 4. *Let* **c** *be a choice correspondence on $\mathcal{X}$ that admits an S-representation and that satisfies regularity, nontriviality-c and nontriviality-d. Consider any pair of alternatives p^*, p_* with $U(s^*, p^*) > U(s^*, p_*)$, and any pair of positive numbers k^*, k. Then,* **c** *admits a unique semi k-normalized S-representation (φ, μ) such that $\varphi(p^*) - \varphi(p_*) = k^*$.*

While the normalization method in Lemma 4 is likely to be useful in many instances, on occasion, conceptual difficulties might arise, as the choice of $\varphi(p^*) - \varphi(p_*)$ may in fact have normative implications. For instance, when discussing the proof of Theorem 2, we have seen that in order to obtain an S-representation (φ, μ) with $\mu(\{s^*\}) \geq 0$, one may be forced to choose a large value for $\varphi(p^*) - \varphi(p_*)$. In turn, one may wish to avoid relating the strength of first stage preferences to the normative properties of the representation of the anticipated second period utility.

On the other hand, in many cases of interest, it may be possible to determine a natural specification for $\mu(\{s^*\})$ at the outset. For instance, if one is willing to obtain a nonnegative aggregator, it may be desirable to focus on the case $\mu(\{s^*\}) = 0$, whenever possible. Hence, in the next lemma, we consider an alternate method of normalization based on the choice of $\mu(\{s^*\})$.

Lemma 5. *Let* $\mathbf{c}$ *be a choice correspondence on* $\mathcal{X}$ *as in Lemma 4. Consider any* $k > 0$ *and* $k_* \in \mathbb{R}$. *Then* $\mathbf{c}$ *admits at most one semi* k*-normalized S-representation* (φ, μ) *such that* $\mu(\{s^*\}) = k_*$. *Moreover, holding* k *constant, if such a representation exists for* k_*, *then it also exists for any* $k'_* > k_*$.

Of course, there is a duality relation between Lemmas 4 and 5. In fact, the infimum of k_*'s that are compatible with a normalized representation in the sense of the latter result equals the limit of $\mu(\{s^*\})$ in the former result as $\varphi(p^*) - \varphi(p_*)$ tends to 0. Our final result shows that these two methods of normalization can be used in a complementary fashion in representation of preference for flexibility:

Lemma 6. *Let* $\mathbf{c}$ *be a choice correspondence on* $\mathcal{X}$ *as in Lemma 4 that also satisfies the monotonicity axiom. Consider any* $k, k^* > 0$ *and* $p^*, p_* \in \Delta$ *with* $U(s^*, p^*) > U(s^*, p_*)$. *Let* (φ^*, μ^*) *denote the semi* k*-normalized S-representation of* $\mathbf{c}$ *such that* $\varphi^*(p^*) - \varphi^*(p_*) = k^*$. *Then, either* $\mu^* \geq 0$ *for every such* (k, k^*, p^*, p_*)*; or, for any* $k > 0$, *the correspondence* $\mathbf{c}$ *admits a (unique) semi* k*-normalized S-representation* (φ, μ) *such that* $\mu(\{s^*\}) = 0$ *and* $\mu \geq 0$.

Proof. Suppose that we do not have $\mu^* \geq 0$ for some (k, k^*, p^*, p_*). By Theorems 2 and 3, we must have $\mu^*_{-*} \geq 0$ and $\mu^*(\{s^*\}) < 0$. Then, by Lemma 5, there exists a semi k-normalized S-representation (φ', μ') such that $\mu'(\{s^*\}) = 0$. Clearly, we must in fact have $\mu' \geq 0$. Hence, for any $\gamma > 0$, the pair $(\varphi, \mu) := (\gamma\varphi', \gamma\mu')$ is a semi γk-normalized S-representation such that $\mu(\{s^*\}) = 0$ and $\mu \geq 0$. $\square$

From Lemma 6, it immediately follows that in the context of Corollary 1, the choice correspondence $\mathbf{c}$ either admits a unique S-representation (φ, μ) such that μ is a probability measure with $\mu(s^*) = 0$, or for any k, k^*, p^*, p_* as above, we can find a unique semi k-normalized S-representation (φ^*, μ^*) with $\mu^* > 0$ and $\varphi^*(p^*) - \varphi^*(p_*) = k^*$. In the latter case, by taking a joint multiple of φ^* and μ^* we can transform μ^* into a probability measure, and the class of all such probability measures would only differ in the relative weights of the events $\mathcal{S} \setminus \{s^*\}$ and $\{s^*\}$.

8. Maximizing an Incomplete Preference Relation: A Comparison

Let $\succeq$ be a preorder (a transitive and reflexive binary relation) on Δ that satisfies the standard independence axiom ($p \succeq q$ iff $\lambda p + (1-\lambda)r \succeq \lambda q + (1-\lambda)r$ for $\lambda \in (0,1)$) and that is closed ($\lim p_n \succeq \lim q_n$ whenever $p_n \succeq q_n$ for each n). We assume that $\succeq$ is **incomplete,** i.e., there exist $p, q \in \Delta$ such that neither $p \succeq q$ nor $q \succeq p$ hold; and write $p \bowtie q$ for any such p, q. We wish to examine a choice correspondence $\widetilde{\mathbf{c}}$ which is defined as follows: For every $x \in \mathcal{X}$,

$$\widetilde{\mathbf{c}}(x) = \begin{cases} \mathbf{m}(x) & \text{if } \mathbf{m}(x) \neq \varnothing, \\ \{\ominus\} & \text{otherwise,} \end{cases}$$

where $\mathbf{m}(x) := \{p \in x : p \succeq q \text{ for each } q \in x\}$.

This is a direct adaptation of the corresponding model of Gerasímou (2010, Proposition 3) to the present framework of choice under risk. In the context of choice deferral, the model predicts that the decision maker would defer whenever she faces two incomparable alternatives, regardless of the significance of their advantages and disadvantages relative to each other, and regardless of the costs that deferral may incur. For instance, \$1000 extra compensation per month for an otherwise less desirable job may be a good enough reason to give up one month's salary for contemplation, irrespective of whether the alternate offer is \$30,000 or \$2,000 per month.

Despite the boundedly rational flavor of this model, it is not difficult to verify that, in fact, it satisfies the monotonicity axiom as well as all of the axioms demanded by Theorem 1 except for the second part of the continuity axiom. (We omit the details.) Thus, it appears that the loss of generality entailed by the sophisticated, dynamic programming approach that we propose in this paper is not significant, as it is limited by the content of the continuity axiom(s). Moreover, our model is also functional without the monotonicity axiom.

9. Concluding Remarks

In many situations, people defer their choices with the hope for better alternatives, which may also entail the risk of losing some available alternatives. Theorem 1 is compatible with such cases. For instance, let H be a self map on $\mathcal{X}$ where $H(x)$ represents the set of alternatives that the agent expects to observe in period 2, upon facing the choice set x in period 1. Also assume that the timing of the choice of an alternative is immaterial, and that there is no uncertainty regarding the agent's tastes. Then, given the choice set x, we could write the value of deferral as $W_{\varphi,H}(x) := \max \varphi(H(x))$, where $\varphi(p)$ measures the time invariant value of alternative p. Notice that the class of choice correspondences that are compatible with such representations is quite rich. In fact, for any $\mathbf{c}$ that admits an affine representation (φ, W) we can find a suitable H with $W_{\varphi,H} = W$ provided that $\max_{x\in\mathcal{X}} W(x) = \max \varphi(\Delta)$ and $\min_{x\in\mathcal{X}} W(x) = \min \varphi(\Delta)$.[19] Thus, in principle, instances of the form (7) and (9) can also be explained in this setup. The real difficulty is to find a particular structure that H may assume which could easily be related to such behavior (as we have done in Section 5 by focusing on the structure of μ), and that can be identified from the choice data up to a meaningful level of uniqueness. We leave this as an open problem for future research.

As we discussed in the introduction, Kochov (2009) models a decision maker whose belief about the modes behavior that she may wish to adopt upon deferral depends on the choice set that she faces. In Kochov's theory, the rationality axioms as well as the setwise independence assumptions of the menu choice literature are restricted so as to regulate behavior only across epistemically equivalent choice problems. If one assumes that the epistemic value of a set of lotteries ultimately depends on the set of riskless prizes associated with those lotteries, one can model a similar learning process in the present framework by restricting our theory to each subcollection $\mathcal{X}_{B'}$ of choice sets x which satisfy the following two properties for a given nonempty set $B' \subseteq B$: (i) $p(B \setminus B') = 0$ for each $p \in x$; and (ii) for each $b \in B'$, there exists $q \in x$ such that $q(b) > 0$. Since any such collection $\mathcal{X}_{B'}$ is closed under the setwise mixture operations, this approach might allow one to obtain a subjective state space representation for each collection $\mathcal{X}_{B'}$. One potential difficulty may be the lack of compactness of $\mathcal{X}_{B'}$, as we utilize compactness of $\mathcal{X}$ when proving Lemma 2.[20]

Following Epstein et al. (2007), one can argue that coarseness of a decision maker's conceptualization of what might happen upon deferral can be considered as a form of ambiguity that may create an urge for hedging. This could reflect itself with a tendency to defer more often when the set of available options is a mixture of more extreme choice sets. In particular, the second part of independence-c could fail. An important direction for further research may be to explore how this form of ambiguity aversion can be incorporated into our model.

Perhaps more importantly, our independence axioms regarding the instances of de-

[19] Indeed, we can simply set $H(x) := \{p \in \Delta : \varphi(p) \leq W(x)\}$ for each $x \in \mathcal{X}$.

[20] Kochov's general theory also permits learning from probabilities that lotteries in a given set assign to certain prizes, but this version would not have much predictive power in the present context, for it actually views as epistemically distinct any two sets of lotteries.

ferral do not appear to be so plausible when deferring choice actually corresponds to selecting a flexible and reasonably good alternative $\ominus$ which may or may not be abandoned in period 2. For example, suppose that the set of feasible alternatives remains unchanged across the two periods. Then, effectively, the agent faces choice sets of the form $x \cup \{\ominus\}$ in both periods. But, in general, a 50-50 lottery over a pair of sets $x \cup \{\ominus\}$, $y \cup \{\ominus\}$ is much different than $(\frac{1}{2}x + \frac{1}{2}y) \cup \{\ominus\}$, even when the latter is viewed as a 50-50 lottery over x and y with the additional option to choose $\ominus$.[21] Thus, we can invoke the usual arguments in favor of the setwise independence axioms only if the decision maker is reasonably sure that she would not select $\ominus$ in period 2. Our model of choice deferral is immune to this problem simply because we have assumed that the agent has to choose an ordinary alternative in period 2. In a sense, our model of choice avoidance corresponds to the opposite case in which the choice of $\ominus$ in period 1 is irreversible. By focusing on these particular cases, we have been able to explain the experimental observations in a minimal framework.

Appendix

A1. Proof of Theorem 1

We omit the "if" part of the proof which is a routine exercise. For the "only if" part, let $\mathbf{c}$ be a choice correspondence on $\mathcal{X}$ that satisfies properties α, β, independence-c, independence-d, continuity and regularity.

Put $\mathcal{X}_0 := \{x \in \mathcal{X} : \mathbf{c}(x) \subseteq \Delta\}$ and $\mathcal{X}_\ominus := \{x \in \mathcal{X} : \mathbf{c}(x) = \{\ominus\}\}$. Notice that by independence-c, $\mathcal{X}_0$ is convex in the sense that $x, y \in \mathcal{X}_0$ imply $\lambda x + (1-\lambda)y \in \mathcal{X}_0$ for every $\lambda \in (0,1)$. Similarly, $\mathcal{X}_\ominus$ is also convex by independence-d.

We proceed with a few claims that will prove useful in the subsequent arguments.

Claim A1. *$\mathcal{X}_0$ and $\mathcal{X}_\ominus$ are open subsets of $\mathcal{X}$.*

Proof. The second part of continuity implies that $\{x \in \mathcal{X} : \ominus \in \mathbf{c}(x)\}$ is a closed subset of $\mathcal{X}$. Hence, $\{x \in \mathcal{X} : \mathbf{c}(x) \subseteq \Delta\}$ is open. Using compactness of Δ and the first part of continuity, it is also easy to see that $\{x \in \mathcal{X} : \mathbf{c}(x) \cap \Delta \neq \varnothing\}$ is a closed subset of $\mathcal{X}$ as well, which implies $\{x \in \mathcal{X} : \mathbf{c}(x) = \{\ominus\}\}$ is open. $\square$

Claim A2. **(i)** *For any $x, y \in \mathcal{X}$, $p \in \mathbf{c}(x) \cap \Delta$, $q \in \mathbf{c}(y) \cap \Delta$ and $\lambda \in (0,1)$, we have $\lambda p + (1-\lambda)q \in \mathbf{c}(\lambda x + (1-\lambda)y)$.*
(ii) *For any $\lambda \in (0,1)$ and $x, y \in \mathcal{X}$ with $\ominus \in \mathbf{c}(x) \cap \mathbf{c}(y)$, we have $\ominus \in \mathbf{c}(\lambda x + (1-\lambda)y)$.*

Proof. We start with the proof of (i). Applying the regularity axiom, pick a set $y_0 \in \mathcal{X}$ with $\mathbf{c}(y_0) \subseteq \Delta$ and a point $q_0 \in \mathbf{c}(y_0)$. Let $x, y \in \mathcal{X}$, $p \in \mathbf{c}(x) \cap \Delta$ and $q \in \mathbf{c}(y) \cap \Delta$.

[21] DLR present a thorough discussion of why a convex combination of choice sets may be considered as a lottery over those sets.

Fix any $\lambda, \gamma \in (0,1)$ and put $q_\gamma := \gamma q + (1-\gamma) q_0$, $y_\gamma := \gamma y + (1-\gamma) y_0$. Then, as $\mathbf{c}(y) \cap \Delta \neq \varnothing$ and $\mathbf{c}(y_0) \subseteq \Delta$, independence-c implies $q_\gamma \in \mathbf{c}(y_\gamma)$ and $\mathbf{c}(y_\gamma) \subseteq \Delta$. Thus, by applying independence-c to the sets x and y_γ, we also see that $\lambda p + (1-\lambda) q_\gamma \in \mathbf{c}(\lambda x + (1-\lambda) y_\gamma)$. Since γ is an arbitrary number in $(0,1)$, passing to limit as $\gamma \to 1$ gives $\lim_{\gamma \to 1} \lambda p + (1-\lambda) q_\gamma \in \mathbf{c}(\lim_{\gamma \to 1} \lambda x + (1-\lambda) y_\gamma)$ by continuity. As $\lim_{\gamma \to 1} \lambda p + (1-\lambda) q_\gamma = \lambda p + (1-\lambda) q$ and $\lim_{\gamma \to 1} \lambda x + (1-\lambda) y_\gamma = \lambda x + (1-\lambda) y$, we obtain the desired conclusion: $\lambda p + (1-\lambda) q \in \mathbf{c}(\lambda x + (1-\lambda) y)$.

Part (ii) is readily verified by the symmetric argument obtained by replacing independence-c with independence-d and y_0 with a set $y_\ominus \in \mathcal{X}_\ominus$. $\square$

The following lemma is well-known (see, e.g., Schneider, 1993, Theorem 1.1.2, p.2).

Lemma A1. *For any $x, y \in \mathcal{X}$ and $\lambda \in (0,1)$,*

$$\operatorname{conv}(\lambda x + (1-\lambda) y) = \lambda \operatorname{conv}(x) + (1-\lambda) \operatorname{conv}(y).$$

While we have not been able to find it elsewhere, the following lemma is also likely to be known. We therefore omit its proof which is available from the authors upon request.

Lemma A2. *Take any $x \in \mathcal{X}$ and define $x^* := \bigcup_{n=0}^{\infty} x_n$ where $x_0 := x$ and $x_{n+1} := \frac{1}{2} x_n + \frac{1}{2} x_n$ for $n = 0, 1, ...$ Then, $\operatorname{cl}(x^*) = \operatorname{conv}(x)$.*

Claim A3. *For any $x \in \mathcal{X}$, we have $\mathbf{c}(x) \cap \Delta = \mathbf{c}(\operatorname{conv}(x)) \cap x$. Moreover, $\ominus \in \mathbf{c}(x)$ if, and only if, $\ominus \in \mathbf{c}(\operatorname{conv}(x))$.*

Proof. Take any $x \in \mathcal{X}$. Let x^* and x_n $(n = 0, 1, ...)$ be as in Lemma A2. It is clear that $x_{n+1} \supseteq x_n$ for every n. Hence, as shown by DLR (Lemma 5), x_n converges to $\operatorname{cl}(x^*)$ in the Hausdorff metric. From Lemma A2, it therefore follows that $x_n \to \operatorname{conv}(x)$. Moreover, by applying Claim A2(i) inductively, we see that $p \in \mathbf{c}(x_n)$ for every n, whenever $p \in \mathbf{c}(x) \cap \Delta$. By continuity, we therefore see that $\mathbf{c}(x) \cap \Delta \subseteq \mathbf{c}(\operatorname{conv}(x))$. That $\ominus \in \mathbf{c}(x)$ implies $\ominus \in \mathbf{c}(\operatorname{conv}(x))$ is verified by the symmetric argument in which Claim A2(ii) takes the role of Claim A2(i).

To prove the remaining assertions, we first need to show that there exists a convex set $y_0 \in \mathcal{X}$ such that $\mathbf{c}(y_0) \subseteq \Delta$. Let y_0' be any choice set with $\mathbf{c}(y_0') \subseteq \Delta$. As is well-known, for any $y \in \mathcal{X}$ there exists a sequence of finite subsets of y that converges to y in the Hausdorff metric. By Claim A1, $\mathcal{X}_0$ is open. It therefore follows that there is a finite set $y_0'' \subseteq y_0'$ such that $\mathbf{c}(y_0'') \subseteq \Delta$. Put $y_0 := \operatorname{conv}(y_0'')$. Since y_0'' is finite, there exists a sufficiently small $\lambda \in (0,1)$ such that $\lambda y_0'' + (1-\lambda) y_0 = y_0$ (see the proof of Lemma 1 in DLR). Moreover, as we have seen in the first part of the proof, we have $\mathbf{c}(y_0'') \subseteq \mathbf{c}(y_0)$, and hence, $\mathbf{c}(y_0) \cap \Delta \neq \varnothing$. By independence-c, it therefore follows that $\mathbf{c}(\lambda y_0'' + (1-\lambda) y_0) \subseteq \Delta$. That is, $\mathbf{c}(y_0) \subseteq \Delta$, as we sought. It is similarly verified that there exists a convex set $y_\ominus \in \mathcal{X}$ such that $\mathbf{c}(y_\ominus) = \{\ominus\}$.

Now, let $\ominus \in \mathbf{c}(\text{conv}(x))$ and fix any $\gamma \in (0,1)$. Then, independence-d implies $\mathbf{c}(\gamma \,\text{conv}(x) + (1-\gamma)y_\ominus) = \{\ominus\}$. Set $x^\gamma := \gamma x + (1-\gamma)y_\ominus$ so that $\text{conv}(x^\gamma) = \gamma \,\text{conv}(x) + (1-\gamma)y_\ominus$ by Lemma A1 and convexity of $y_\ominus$. It follows that $\mathbf{c}(\text{conv}(x^\gamma)) = \{\ominus\}$. Moreover, as we have seen in the first part of the proof, $\mathbf{c}(x^\gamma) \subseteq \mathbf{c}(\text{conv}(x^\gamma))$. Since $\mathbf{c}(x^\gamma)$ is non-empty, we must therefore have $\mathbf{c}(x^\gamma) = \{\ominus\}$. Thus, as $\gamma \in (0,1)$ is arbitrary, passing to limit as $\gamma \to 1$ yields $\ominus \in \mathbf{c}(x)$ by continuity.

It remains to show that $\mathbf{c}(\text{conv}(x)) \cap x \subseteq \mathbf{c}(x)$. Pick any $p \in \mathbf{c}(\text{conv}(x)) \cap x$, $\gamma \in (0,1)$ and $q_0 \in \mathbf{c}(y_0)$. An obvious modification of our previous arguments yields

$$\mathbf{c}(\gamma x + (1-\gamma)y_0) \subseteq \mathbf{c}(\gamma \,\text{conv}(x) + (1-\gamma)y_0) \subseteq \Delta. \tag{13}$$

By independence-c, we also have $\gamma p + (1-\gamma)q_0 \in \mathbf{c}(\gamma \,\text{conv}(x) + (1-\gamma)y_0)$. Since (13) allows us to apply property α, we therefore get $\gamma p + (1-\gamma)q_0 \in \mathbf{c}(\gamma x + (1-\gamma)y_0)$. Finally, by passing to limit as $\gamma \to 1$, we obtain the desired conclusion: $p \in \mathbf{c}(x)$. $\square$

Throughout the remainder of the proof of Theorem 1, we will denote by y_0 and $y_\ominus$ convex sets in $\mathcal{X}$ such that $\mathbf{c}(y_0) \subseteq \Delta$ and $\mathbf{c}(y_\ominus) = \{\ominus\}$. (By the regularity axiom and Claim A3, there exist such convex sets.) As $\mathcal{X}_0$ is open, it is clear that there exists a sufficiently small $\gamma_0 \in (0,1)$ such that $\gamma_0 x + (1-\gamma_0)y_0 \in \mathcal{X}_0$ for every $x \in \mathcal{X}$. We now define a binary relation $\succsim$ on Δ as, for every $p, q \in \Delta$,

$$p \succsim q \quad \text{iff} \quad \gamma_0 p + (1-\gamma_0)p_0 \in \mathbf{c}(\gamma_0\{p,q\} + (1-\gamma_0)y_0) \text{ for some } p_0 \in y_0.$$

By the choice of γ_0, the relation $\succsim$ is complete.

Although it is a fairly standard exercise, for the sake of completeness, we now show that $\succsim$ is transitive. Take any $p, q, r \in \Delta$ with $p \succsim q$ and $q \succsim r$. Then, there exist $p_0, q_0 \in y_0$ such that

$$\gamma_0 p + (1-\gamma_0)p_0 \in \mathbf{c}(\gamma_0\{p,q\} + (1-\gamma_0)y_0), \tag{14}$$

$$\gamma_0 q + (1-\gamma_0)q_0 \in \mathbf{c}(\gamma_0\{q,r\} + (1-\gamma_0)y_0). \tag{15}$$

Put $\tilde{x} := \gamma_0\{p,q,r\} + (1-\gamma_0)y_0$ and $\tilde{z} := \{\eta \in \{p,q,r\} : \gamma_0\eta + (1-\gamma_0)\eta_0 \in \mathbf{c}(\tilde{x})\ \exists \eta_0 \in y_0\}$. Note that if $p \in \tilde{z}$ so that $\gamma_0 p + (1-\gamma_0)\eta_0 \in \mathbf{c}(\tilde{x})$ for some $\eta_0 \in y_0$, as $\gamma_0\{p,r\} + (1-\gamma_0)y_0$ is a subset of $\tilde{x}$ that contains $\gamma_0 p + (1-\gamma_0)\eta_0$, property α implies $\gamma_0 p + (1-\gamma_0)\eta_0 \in \mathbf{c}(\gamma_0\{p,r\} + (1-\gamma_0)y_0)$, as we seek. By a symmetric argument, we also see that if $q \in \tilde{z}$, then $\gamma_0 q + (1-\gamma_0)\eta_0' \in \mathbf{c}(\tilde{x}) \cap \mathbf{c}(\gamma_0\{p,q\} + (1-\gamma_0)y_0)$ for some $\eta_0' \in y_0$. But in this case, from (14) and property β, it follows that $p \in \tilde{z}$. Similarly,

$r \in \tilde{z}$ implies $q \in \tilde{z}$ by (15). Since $\tilde{z}$ is non-empty by definitions, it follows that $p \in \tilde{z}$ in all contingencies, and hence, $p \succsim r$.

We shall now show that $\succsim$ satisfies the classical independence axiom. To this end, let $p, q, r \in \Delta$, and $\lambda \in (0,1)$. Suppose $p \succsim q$ so that, for a $p_0 \in y_0$, the point $p' := \gamma_0 p + (1-\gamma_0)p_0$ belongs to $\mathbf{c}\,(\gamma_0\{p,q\} + (1-\gamma_0)y_0)$. Pick any $r' \in \mathbf{c}\,(\gamma_0\{r\} + (1-\gamma_0)y_0)$. By definitions, there exists an $r_0 \in y_0$ such that $r' = \gamma_0 r + (1-\gamma_0)r_0$. Moreover, since y_0 is convex, we see that

$$\begin{aligned}
&\lambda\,(\gamma_0\{p,q\} + (1-\gamma_0)y_0) + (1-\lambda)\,(\gamma_0\{r\} + (1-\gamma_0)y_0) \\
&= \gamma_0\,(\lambda\{p,q\} + (1-\lambda)\{r\}) + (1-\gamma_0)\,(\lambda y_0 + (1-\lambda)y_0) \\
&= \gamma_0\,\{\lambda p + (1-\lambda)r, \lambda q + (1-\lambda)r\} + (1-\gamma_0)y_0.
\end{aligned}$$

Therefore, independence-c implies

$$\lambda p' + (1-\lambda)r' \in \mathbf{c}\,(\gamma_0\,\{\lambda p + (1-\lambda)r, \lambda q + (1-\lambda)r\} + (1-\gamma_0)y_0)\,.$$

Now, note that $\lambda p' + (1-\lambda)r' = \gamma_0\,(\lambda p + (1-\lambda)r) + (1-\gamma_0)(\lambda p_0 + (1-\lambda)r_0)$. Since $\lambda p_0 + (1-\lambda)r_0$ belongs to y_0, we therefore see that $\lambda p + (1-\lambda)r \succsim \lambda q + (1-\lambda)r$, as we sought.

To see that $\succsim$ is continuous, let $(p_n), (q_n)$ be convergent sequences in Δ such that $p_n \succsim q_n$ for every n. Put $p := \lim p_n$ and $q := \lim q_n$. Let (p'_n) be a sequence in y_0 such that $\gamma_0 p_n + (1-\gamma_0)p'_n \in \mathbf{c}\,(\gamma_0\{p_n, q_n\} + (1-\gamma_0)y_0)$ for every n. Since y_0 is compact, by passing to a convergent subsequence if necessary, we can assume that (p'_n) converges to a point $p' \in y_0$. Then, clearly, the continuity axiom implies $\gamma_0 p + (1-\gamma_0)p' \in \mathbf{c}\,(\gamma_0\{p,q\} + (1-\gamma_0)y_0)$. That is, $p \succsim q$, as desired.

The properties of $\succsim$ that we have established above implies that there exists an expected utility function φ on Δ that represents $\succsim$. The next step is to show that, for any $x \in \mathcal{X}$,

$$\mathbf{c}\,(x) \cap \Delta \neq \varnothing \quad \text{implies} \quad \mathbf{c}\,(x) \cap \Delta = \arg\max_x \varphi. \tag{16}$$

Fix an arbitrary $x \in \mathcal{X}$ and let $p \in \arg\max_x \varphi$. Recall that $\mathbf{c}\,(\gamma_0 x + (1-\gamma_0)y_0) \subseteq \Delta$. Hence, there exist $q \in x$, $q_0 \in y_0$ such that $\gamma_0 q + (1-\gamma_0)q_0 \in \mathbf{c}\,(\gamma_0 x + (1-\gamma_0)y_0)$. Then, property α yields $\gamma_0 q + (1-\gamma_0)q_0 \in \mathbf{c}\,(\gamma_0\{p,q\} + (1-\gamma_0)y_0)$. But as $p \in \arg\max_x \varphi$, we have $p \succsim q$, that is $\gamma_0 p + (1-\gamma_0)p_0 \in \mathbf{c}\,(\gamma_0\{p,q\} + (1-\gamma_0)y_0)$ for some $p_0 \in y_0$. Then, property β implies $\gamma_0 p + (1-\gamma_0)p_0 \in \mathbf{c}\,(\gamma_0 x + (1-\gamma_0)y_0)$. If $\mathbf{c}\,(x) \cap \Delta \neq \varnothing$, it therefore follows that $p \in \mathbf{c}\,(x)$ by independence-c.

To establish the converse inclusion, let $p' \in \mathbf{c}\,(x) \cap \Delta$. Pick a point $p'_0 \in \mathbf{c}\,(y_0)$. Then, independence-c yields $\gamma_0 p' + (1-\gamma_0)p'_0 \in \mathbf{c}\,(\gamma_0 x + (1-\gamma_0)y_0)$. Hence, by property α, it obviously follows that $p' \in \arg\max_x \varphi$. This completes the proof of (16).

Let X be the collection of convex sets in $\mathcal{X}$. Now, we will construct a function $W : X \to \mathbb{R}$ which is affine and continuous, and which satisfies the following additional properties: For every $x \in X$,

$$\begin{aligned} W(x) > \max\varphi(x) \quad &\text{iff} \quad \mathbf{c}(x) = \{\ominus\}, \\ W(x) = \max\varphi(x) \quad &\text{iff} \quad \ominus \in \mathbf{c}(x) \neq \{\ominus\}, \\ W(x) < \max\varphi(x) \quad &\text{iff} \quad \mathbf{c}(x) \subseteq \Delta. \end{aligned} \tag{$*$}$$

Put $X_\ominus := X \cap \mathcal{X}_\ominus$ and $X_0 := X \cap \mathcal{X}_0$. Then, by Claim A1, $X_\ominus$ and X_0 are relatively open subsets of X. Thus, for any $x \in X_\ominus$, the sets $\{\lambda \in [0,1] : \mathbf{c}(\lambda x + (1-\lambda)y_0) = \{\ominus\}\}$ and $\{\lambda \in [0,1] : \mathbf{c}(\lambda x + (1-\lambda)y_0) \subseteq \Delta\}$ are relatively open subsets of $[0,1]$, which are clearly also disjoint and nonempty. By connectedness of $[0,1]$, it follows that the union of these two sets cannot be equal to $[0,1]$. That is, there exists $\lambda \in (0,1)$ such that $\ominus \in \mathbf{c}(\lambda x + (1-\lambda)y_0) \neq \{\ominus\}$. In fact this number, which we denote as $\lambda^*(x)$, is the unique number in $[0,1]$ that satisfies the latter two properties, for $\lambda > \lambda^*(x)$ implies $\mathbf{c}(\lambda x + (1-\lambda)y_0) = \{\ominus\}$ by independence-d, and $\lambda < \lambda^*(x)$ implies $\mathbf{c}(\lambda x + (1-\lambda)y_0) \subseteq \Delta$ by independence-c.

We now show that $\lambda^*(\cdot)$ is continuous on $X_\ominus$. Pick a sequence (x_n) in $X_\ominus$ that converges to some $x \in X_\ominus$. It suffices to find a subsequence (x_{n_k}) such that $\lim_k \lambda^*(x_{n_k}) = \lambda^*(x)$. For each n, put $\lambda_n^* := \lambda^*(x_n)$ and $z_n := \lambda_n^* x_n + (1-\lambda_n^*)y_0$. Then, $\ominus \in \mathbf{c}(z_n) \neq \{\ominus\}$ for each n. In particular, we can pick a sequence (q_n) such that $q_n \in \mathbf{c}(z_n)$ for each n. Since $[0,1] \times \Delta$ is compact, there exists a subsequence $(\lambda_{n_k}^*, q_{n_k})$ that converges to some $(\lambda, q) \in [0,1] \times \Delta$. Then, we also have $\lim_k z_{n_k} = \lambda x + (1-\lambda)y_0$. Hence, the continuity axiom implies $q \in \mathbf{c}(\lambda x + (1-\lambda)y_0)$ and $\ominus \in \mathbf{c}(\lambda x + (1-\lambda)y_0)$ so that $\ominus \in \mathbf{c}(\lambda x + (1-\lambda)y_0) \neq \{\ominus\}$. But we know that $\lambda^*(x)$ is the unique number in $[0,1]$ that satisfies the latter two properties. Thus, we must have $\lambda = \lambda^*(x)$, as we need. (In particular, $\lambda \in (0,1)$.)

Since $\lambda^*(\cdot)$ is continuous on $X_\ominus$, so is f which we define by $f(x) := \frac{1}{\lambda^*(x)}$ for $x \in X_\ominus$. We next show that f is affine on $X_\ominus$. Let $x, y \in X_\ominus$ and $\gamma \in (0,1)$. Put $\delta := \frac{\lambda^*(y)\gamma}{\lambda^*(y)\gamma + \lambda^*(x)(1-\gamma)}$ so that $\gamma = \frac{\delta\lambda^*(x)}{\delta\lambda^*(x) + (1-\delta)\lambda^*(y)}$ and $1-\gamma = \frac{(1-\delta)\lambda^*(y)}{\delta\lambda^*(x)+(1-\delta)\lambda^*(y)}$. Then

$$\begin{aligned} &\delta\left(\lambda^*(x)x + (1-\lambda^*(x))y_0\right) + (1-\delta)\left(\lambda^*(y)y + (1-\lambda^*(y))y_0\right) \\ &= \left(\delta\lambda^*(x) + (1-\delta)\lambda^*(y)\right)\left(\frac{\delta\lambda^*(x)}{\delta\lambda^*(x)+(1-\delta)\lambda^*(y)}x + \frac{(1-\delta)\lambda^*(y)}{\delta\lambda^*(x)+(1-\delta)\lambda^*(y)}y\right) \\ &\qquad + \left(1 - \left(\delta\lambda^*(x) + (1-\delta)\lambda^*(y)\right)\right)y_0 \\ &= \left(\delta\lambda^*(x) + (1-\delta)\lambda^*(y)\right)\left(\gamma x + (1-\gamma)y\right) + \left(1 - \left(\delta\lambda^*(x) + (1-\delta)\lambda^*(y)\right)\right)y_0. \end{aligned}$$

By Claim A2, we have

$$\circleddash \in \mathbf{c}\left(\delta\left(\lambda^*(x)x + (1-\lambda^*(x))y_0\right) + (1-\delta)\left(\lambda^*(y)y + (1-\lambda^*(y))y_0\right)\right) \neq \{\circleddash\}.$$

It therefore follows that $\lambda^*(\gamma x + (1-\gamma)y) = \delta\lambda^*(x) + (1-\delta)\lambda^*(y)$, that is, $f(\gamma x + (1-\gamma)y) = (\delta\lambda^*(x) + (1-\delta)\lambda^*(y))^{-1}$. As the latter number equals both $\theta := \frac{\gamma}{\delta\lambda^*(x)}$ and $\theta' := \frac{1-\gamma}{(1-\delta)\lambda^*(y)}$, we see that $f(\gamma x + (1-\gamma)y) = \delta\theta + (1-\delta)\theta' = \frac{\gamma}{\lambda^*(x)} + \frac{1-\gamma}{\lambda^*(y)} = \gamma f(x) + (1-\gamma)f(y)$, as we sought.

Since $X_\circleddash$ is relatively open in X, there is a small enough $\gamma_\circleddash \in (0,1)$ such that $\gamma_\circleddash x + (1-\gamma_\circleddash)\, y_\circleddash \in X_\circleddash$ for every $x \in X$. Define $f_1(x) := \frac{f(\gamma_\circleddash x + (1-\gamma_\circleddash)y_\circleddash) - (1-\gamma_\circleddash)f(y_\circleddash)}{\gamma_\circleddash}$ for every $x \in X$. By affinity of f, we have $f_1(x) = f(x)$ for every $x \in X_\circleddash$; that is, f_1 extends f. Moreover, since $x \to \gamma_\circleddash x + (1-\gamma_\circleddash)y_\circleddash$ is an affine and continuous map from X into $X_\circleddash$, the function f_1 is affine and continuous on X by the corresponding properties of f over $X_\circleddash$.

Now, we fix a number $\theta_0 < \max\varphi(y_0)$ and set, for each $x \in X$,

$$W(x) := \max\varphi(x) + (f_1(x) - 1)\left(\max\varphi(y_0) - \theta_0\right).$$

Then $W(\cdot)$ is continuous and affine on X by the corresponding properties of $\max\varphi(\cdot)$ and $f_1(\cdot)$. Pick any $x \in X$. To verify $(*)$, it suffices to establish the following:

(i) $f_1(x) > 1$ if $x \in X_\circleddash$; (ii) $f_1(x) = 1$ if $\circleddash \in \mathbf{c}(x) \neq \{\circleddash\}$; and (iii) $f_1(x) < 1$ if $x \in X_0$.

Since f_1 extends f, property (i) immediately follows from the definitions. To prove (ii), suppose $\circleddash \in \mathbf{c}(x) \neq \{\circleddash\}$. Let (γ_n) be a sequence in $(0,1)$ that converges to 1 and put $z'_n := \gamma_n x + (1-\gamma_n)y_\circleddash$ for each n. We claim

$$\lim_n \lambda^*(z'_n) = 1. \tag{17}$$

Otherwise, there exists $\varepsilon \in (0,1)$ and a subsequence (n_k) such that $\lambda^*_{n_k} := \lambda^*(z'_{n_k}) \leq 1-\varepsilon$ for each k. Then, by passing to a further subsequence if necessary, we can assume that $(\lambda^*_{n_k})$ converges to some $\lambda \in [0, 1-\varepsilon]$. But the continuity axiom and definition of $\lambda^*(\cdot)$ then imply that $\circleddash \in \mathbf{c}(\lim_k \lambda^*_{n_k} z'_{n_k} + (1-\lambda^*_{n_k})y_0) = \mathbf{c}(\lambda x + (1-\lambda)y_0)$ which contradicts independence-c. This proves (17). By definition of f and continuity of f_1 on X, we then see that $f_1(x) = \lim_n f_1(z'_n) = \lim_n f(z'_n) = 1$, as we need. Finally, to prove (iii), suppose by contradiction that $\mathbf{c}(x) \subseteq \Delta$ and $f_1(x) \geq 1$. Then, by affinity of f_1, we have $f_1(\gamma y_\circleddash + (1-\gamma)x) > 1$ for every $\gamma \in (0,1]$. But, as we have seen when defining $\lambda^*(\cdot)$, there is a number $\gamma^* \in (0,1)$ such that $\circleddash \in \mathbf{c}(\gamma^* y_\circleddash + (1-\gamma^*)x) \neq \{\circleddash\}$. This contradicts the property (ii) that we have just established.

We now extend W to $\mathcal{X}$ as $W(x) := W(\text{conv}(x))$ for each $x \in \mathcal{X}$. Using the corresponding properties over X, it is easily seen that W is affine over $\mathcal{X}$ (by Lemma A1), and continuous over $\mathcal{X}$ (since $d_h(\text{conv}(x), \text{conv}(y)) \leq d_h(x, y)$ for every $x, y \in \mathcal{X}$). Moreover, by Claim A3, properties $(*)$ actually hold for every $x \in \mathcal{X}$. This clearly completes the proof of Theorem 1. $\square$

A2. Proof of Lemma 1

The "if" part is explained in text. For the "only if" part, let (φ, W) and (φ', W') be affine representations for a regular choice correspondence $\mathbf{c}$. Then, clearly, both φ and φ' represent the preference relation $\succsim$ as defined in the proof of Theorem 1. Hence, by the uniqueness result of the classical expected utility theory, φ' must be a positive affine transformation of φ.

For any $x \in \mathcal{X}$, put $\Psi(x) := W(x) - \max \varphi(x)$ and $\Psi'(x) := W'(x) - \max \varphi'(x)$. The next step is to show that, for any $x, y \in \mathcal{X}$,

$$\Psi(x) \geq \Psi(y) \quad \text{implies} \quad \Psi'(x) \geq \Psi'(y). \tag{18}$$

Assume by contradiction that $\Psi(x) \geq \Psi(y)$ and $\Psi'(x) < \Psi'(y)$ for some $x, y \in \mathcal{X}$. By regularity of $\mathbf{c}$, it is clear that there exist $z \in \mathcal{X}$ and $\lambda \in (0, 1]$ such that $\Psi(\lambda x + (1-\lambda)z) = 0$. Then, $\ominus \in \mathbf{c}(\lambda x + (1-\lambda)z) \neq \{\ominus\}$ implying that $\Psi'(\lambda x + (1-\lambda)z) = 0$. By affinity of Ψ and Ψ', it thus follows that $\Psi(\lambda y + (1-\lambda)z) \leq 0 < \Psi'(\lambda y + (1-\lambda)z)$. But here, the weak inequality implies $\mathbf{c}(\lambda y + (1-\lambda)z) \cap \Delta \neq \varnothing$ while the strict inequality implies $\mathbf{c}(\lambda y + (1-\lambda)z) = \{\ominus\}$, a contradiction. This proves (18).

Since X (i.e., the collection of convex sets in $\mathcal{X}$) is a mixture set, by the classical uniqueness result, (18) implies that there exist $\gamma > 0$ and $\theta \in \mathbb{R}$ such that $\Psi'(x) = \gamma\Psi(x) + \theta$ for every $x \in X$. Moreover, as Ψ' and Ψ are affine and continuous functions on $\mathcal{X}$, we also have $\Psi'(x) = \Psi'(\text{conv}(x))$ and $\Psi(x) = \Psi(\text{conv}(x))$ for every $x \in \mathcal{X}$. It thus follows that $\Psi'(x) = \gamma\Psi(x) + \theta$ for every $x \in \mathcal{X}$. Finally, we note that θ must be equal to 0, for as we argued before, by regularity of $\mathbf{c}$ there exists an $x \in \mathcal{X}$ such that $\Psi'(x) = \Psi(x) = 0$. $\square$

A3. Proof of Lemma 2

Let $\mathbf{c}$ be a regular choice correspondence on $\mathcal{X}$ that has an affine representation (φ, W). As before, set $\Psi(x) := W(x) - \max \varphi(x)$ for all $x \in \mathcal{X}$. In the remainder of the appendix, for each $s \in \mathcal{S}$ and $x \in \mathcal{X}$, we set

$$\sigma_x(s) := \max_{q \in x} U(s, q).$$

From the analysis of DLRS, it follows that the following two conditions are equivalent:

(i) There exist $\theta \in \mathbb{R}$ and a measure μ on $\mathcal{S}$ such that $W(x) = \int_{\mathcal{S}} \sigma_x(s)\, \mu(ds) + \theta$ for every $x \in \mathcal{X}$.

(ii) W is Lipschitz continuous on $\mathcal{X}$.

In turn, since $W(x) = W(\text{conv}(x))$ and $d_h(\text{conv}(x), \text{conv}(y)) \leq d_h(x, y)$ for every $x, y \in \mathcal{X}$, property (ii) is equivalent to the following weaker property:

(iii) W is Lipschitz continuous on X.[22]

The next claim shows that if $\mathbf{c}$ satisfies L-continuity, then W is Lipschitz continuous on X. This establishes the "only if" part of Lemma 2, since by applying property (i) above, we can then obtain an S-representation of the form $(\varphi(\cdot) - \theta, W(\cdot) - \theta)$ for some $\theta \in \mathbb{R}$.

Claim A4. *If* $\mathbf{c}$ *satisfies L-continuity, then* W *is Lipschitz continuous on* X*.*

Proof. Since $\max \varphi$ is Lipschitz continuous, it suffices to verify Lipschitz continuity of Ψ. Thus, our task will be to find a number $\bar{n} > 0$ such that $\Psi(y) - \Psi(x) \leq \bar{n} d_h(x, y)$ for every $x, y \in X$.

By regularity of $\mathbf{c}$, there exist $y_\ominus, y_0 \in X$ such that $\mathbf{c}(y_\ominus) = \{\ominus\}$ and $\mathbf{c}(y_0) \subseteq \Delta$. Let $n > 0$ be a number that satisfies the condition stated in the L-continuity axiom, and fix a number $d \in (0, 1/n)$.

Consider any $x, y \in X$ such that $\mathbf{c}(x) = \mathbf{c}(y) = \{\ominus\}$ and $a := d_h(x, y) \leq d$. Without loss of generality, we can assume $a > 0$. Observe that $\Psi(x), \Psi(y), \Psi(y_\ominus) > 0$ and $\Psi(y_0) < 0$. In particular, affinity of Ψ implies $\Psi((1 - n\varepsilon)x + n\varepsilon y_\ominus) > 0$ for all $\varepsilon \in (0, 1/n)$. Moreover, $n\varepsilon_y$ belongs to $(0, 1)$ where $\varepsilon_y := \frac{\Psi(y)}{n(\Psi(y) - \Psi(y_0))}$. It is also clear that $\Psi((1 - n\varepsilon_y)y + n\varepsilon_y y_0) = 0$ and $\Psi((1 - n\varepsilon)y + n\varepsilon y_0) < 0$ for $\varepsilon \in (\varepsilon_y, 1/n)$.

First, suppose $\varepsilon_y \leq a$. Then,

$$\Psi((1 - na)x + nay_\ominus) \geq 0 \geq \Psi((1 - na)y + nay_0).$$

It follows from the affinity of Ψ that $\frac{na}{1-na}(\Psi(y_\ominus) - \Psi(y_0)) \geq \Psi(y) - \Psi(x)$. Since $a \leq d$, we therefore see that, for $\varepsilon_y \leq a$,

$$a \frac{n}{1 - nd}(\Psi(y_\ominus) - \Psi(y_0)) \geq \Psi(y) - \Psi(x). \tag{19}$$

Next, suppose $\varepsilon_y > a$. Then, from the definition of ε_y it follows that $1 - na >$

[22] In the discussion of DLRS it is sometimes unclear if the Lipschitz continuity of the function at hand refers to X or to the grand domain $\mathcal{X}$. However, for an affine and continuous function this point is immaterial by the equivalence of properties (ii) and (iii) above.

$\frac{-\Psi(y_0)}{\Psi(y)-\Psi(y_0)}$, that is, $\lambda_a := \frac{-\Psi(y_0)}{(1-na)(\Psi(y)-\Psi(y_0))}$ belongs to $(0,1)$. Moreover, with some algebra it is verified that

$$\Psi((1-na)(\lambda_a y + (1-\lambda_a)y_0) + nay_0) = 0. \tag{20}$$

This implies $\ominus \in \mathbf{c}((1-na)(\lambda_a y + (1-\lambda_a)y_0) + nay_0)$. But, by convexity of the sets in question, we have

$$d_h(\lambda_a x + (1-\lambda_a)y_0, \lambda_a y + (1-\lambda_a)y_0) = \lambda_a d_h(x,y) \leq a.$$

Therefore, the L-continuity axiom implies

$$\Psi((1-na)(\lambda_a x + (1-\lambda_a)y_0) + nay_\ominus) \geq 0. \tag{21}$$

By combining (20) and (21), we get

$$\Psi((1-na)(\lambda_a x + (1-\lambda_a)y_0) + nay_\ominus) \geq \Psi((1-na)(\lambda_a y + (1-\lambda_a)y_0) + nay_0).$$

Since Ψ is affine, this amounts to saying that

$$\frac{1}{\lambda_a}\frac{na}{1-na}(\Psi(y_\ominus) - \Psi(y_0)) \geq \Psi(y) - \Psi(x).$$

Note that $\frac{1}{\lambda_a}$ is bounded from above by $\frac{(1-na)(\Psi^* - \Psi(y_0))}{-\Psi(y_0)}$, where Ψ^* is the maximum value of Ψ on the compact set X. Therefore, it follows that, for $\varepsilon_y > a$, we have

$$an\frac{(\Psi^* - \Psi(y_0))}{-\Psi(y_0)}(\Psi(y_\ominus) - \Psi(y_0)) \geq \Psi(y) - \Psi(x). \tag{22}$$

Now, put $\bar{n} := \max\left\{\frac{n(\Psi^* - \Psi(y_0))(\Psi(y_\ominus) - \Psi(y_0))}{-\Psi(y_0)}, \frac{n(\Psi(y_\ominus) - \Psi(y_0))}{1-nd}\right\}$. By (19) and (22), we conclude that $\Psi(y) - \Psi(x) \leq \bar{n}d_h(x,y)$. By an argument based on telescoping summation, this observation can easily be generalized to include the case $d_h(x,y) > d$ and $\mathbf{c}(x) = \mathbf{c}(y) = \{\ominus\}$ (see the proof Lemma 1 in DLRS). In other words, we have $\Psi(y) - \Psi(x) \leq \bar{n}d_h(x,y)$ for every $x, y \in X_\ominus$.

Finally, recall that there exists $\gamma_\ominus \in (0,1)$ such that $\gamma_\ominus x + (1-\gamma_\ominus)y_\ominus \in X_\ominus$ for each $x \in X$. Hence, for any $x, y \in X$,

$$\begin{aligned}\Psi(y) - \Psi(x) &= \frac{1}{\gamma_\ominus}\left(\Psi(\gamma_\ominus y + (1-\gamma_\ominus)y_\ominus) - \Psi(\gamma_\ominus x + (1-\gamma_\ominus)y_\ominus)\right)\\ &\leq \frac{1}{\gamma_\ominus}\bar{n}\gamma_\ominus d_h(x,y) = \bar{n}d_h(x,y).\end{aligned}$$

□

Next, we prove that if W is Lipschitz continuous on $\mathcal{X}$, then the L-continuity axiom must hold. In view of the equivalence of the properties (i) and (ii), this will complete the proof of Lemma 2.

Claim A5. *If W is Lipschitz continuous on $\mathcal{X}$, then* $\mathbf{c}$ *satisfies the L-continuity axiom.*

Proof. Fix an arbitrarily chosen $(y_0, y_\ominus) \in \mathcal{X}_0 \times \mathcal{X}_\ominus$. If W is Lipschitz continuous, so is Ψ. Let $\bar{n}$ be the Lipschitz coefficient of Ψ, and put $n := \bar{n}/(\Psi(y_\ominus) - \Psi(y_0))$. Then, $\Psi(y) - \Psi(x) \leq d_h(x, y)n(\Psi(y_\ominus) - \Psi(y_0))$ for every $x, y \in \mathcal{X}$. In particular, if $d_h(x, y) < 1/n$, then for every $\varepsilon \in [d_h(x, y), 1/n)$, we have $\Psi(y) - \Psi(x) \leq \frac{n\varepsilon}{1-n\varepsilon}(\Psi(y_\ominus) - \Psi(y_0))$. Upon rearranging, this yields $\Psi((1 - n\varepsilon)x + n\varepsilon y_\ominus) \geq \Psi((1 - n\varepsilon)y + n\varepsilon y_0)$. Thus, $\ominus \in \mathbf{c}((1 - n\varepsilon)y + n\varepsilon y_0)$ implies $\ominus \in \mathbf{c}((1 - n\varepsilon)x + n\varepsilon y_\ominus)$. □

□

A4. Proof of Theorem 2

We omit the "if" part of the proof, which is trivial. For the "only if" part, consider a regular choice correspondence $\mathbf{c}$ on $\mathcal{X}$ that admits an S-representation $(\tilde{\varphi}, \nu)$. Also suppose that the monotonicity axiom holds.

As usual, let $y_\ominus \in \mathcal{X}$ and $\gamma_\ominus \in (0, 1)$ be such that $\mathbf{c}\,(y_\ominus) = \{\ominus\}$ and, for every $x \in \mathcal{X}$, $\mathbf{c}\,(\gamma_\ominus x + (1 - \gamma_\ominus)y_\ominus) = \{\ominus\}$. Also pick a set $y_0 \in \mathcal{X}$ with $\mathbf{c}\,(y_0) \subseteq \Delta$.

Take any $x, y \in \mathcal{X}$ with $x \subseteq y$. We shall first show that

$$\max \tilde{\varphi}(x) = \max \tilde{\varphi}(y) \quad \text{implies} \quad W_\nu(x) \leq W_\nu(y). \tag{23}$$

Suppose $\max \tilde{\varphi}(x) = \max \tilde{\varphi}(y)$. Put $x' := \gamma_\ominus x + (1 - \gamma_\ominus)y_\ominus$ and $y' := \gamma_\ominus y + (1 - \gamma_\ominus)y_\ominus$. Then, as we have seen in the proof of Theorem 1, there exists a number $\lambda^*(y') \in (0, 1)$ such that the set $y'' := \lambda^*(y')y' + (1 - \lambda^*(y'))y_0$ satisfies $\ominus \in \mathbf{c}(y'') \neq \{\ominus\}$. As $(\tilde{\varphi}, \nu)$ represents $\mathbf{c}$, this amounts to saying

$$W_\nu\,(y'') = \max \tilde{\varphi}(y''). \tag{24}$$

Next, we note that, since φ is an affine function, $\max \tilde{\varphi}(y) = \max \tilde{\varphi}(x)$ implies

$$\max \tilde{\varphi}(y'') = \max \tilde{\varphi}(x''), \tag{25}$$

where $x'' := \lambda^*(y')x' + (1 - \lambda^*(y'))y_0$. Moreover, $x \subseteq y$ implies $x'' \subseteq y''$. It therefore follows that $\arg\max_{q \in x''} \tilde{\varphi}(q) \subseteq \arg\max_{q \in y''} \tilde{\varphi}(q)$.

Pick any $p \in \arg\max_{q \in x''} \tilde{\varphi}(q)$. Then, as we have just seen, p maximizes $\tilde{\varphi}$ also on y''. Equation (24) therefore implies that $p \in \mathbf{c}(y'')$. Thus, the monotonicity axiom yields

$p \in \mathbf{c}(x'')$, and hence, $\max_{q \in x''} \tilde{\varphi}(q) \geq W_\nu(x'')$. By (24) and (25), we therefore see that $W_\nu(y'') \geq W_\nu(x'')$. Since W_ν is affine, this is equivalent to saying $W_\nu(y) \geq W_\nu(x)$, as we sought.

By applying the Jordan decomposition theorem, let ν^+ and ν^- be the positive and negative parts of ν, respectively, so that $\nu = \nu^+ - \nu^-$.

First, assume $\tilde{\varphi}$ is nonconstant so that s^* is a well-defined element of $\mathcal{S}$. The next step is to show that

$$\nu^-(\mathcal{S} \setminus \{s^*\}) = 0. \tag{26}$$

Assume by contradiction that $\nu^-(\mathcal{S} \setminus \{s^*\}) > 0$. As ν^+ and ν^- are disjoint, there exists a Borel set $\mathcal{S}^- \subseteq \mathcal{S}$ such that $\nu^-(\mathcal{S}^-) = \nu^-(\mathcal{S})$ and $\nu^+(\mathcal{S}^-) = 0$. Then, $\nu^-(\mathcal{S} \setminus \{s^*\}) = \nu^-(\mathcal{S}^- \setminus \{s^*\})$, and hence, $\nu^-(\mathcal{S}^- \setminus \{s^*\}) > 0$. By countable additivity of ν^-, it thus follows that there exists a closed set $F \subseteq \mathcal{S}^- \setminus \{s^*\}$ such that $\nu^-(F) > 0$ (see Aliprantis and Border, 1999, Theorem 17.24, p. 574).

Let p_0 be the element of Δ which is defined as $p_0(b) := \frac{1}{|B|}$ for each $b \in B$, so that $U(s, p_0) = 0$ for any $s \in S$. Put $z_\theta := \left\{q \in \mathbb{R}^B : \|q - p_0\| \leq \theta, \sum_{b \in B} q(b) = 1\right\}$ for each $\theta > 0$. It is clear that for any $\theta > 0$, the boundary of z_θ (relative to its affine hull) equals $\{p_0 + \theta s : s \in \mathcal{S}\}$. As is well known, it follows that for each $s \in \mathcal{S}$ and $\theta > 0$, the point $p_0 + \theta s$ is the unique maximizer of $U(s, \cdot)$ on z_θ (where we consider $U(\cdot, \cdot)$ as the dot product). Let $\theta^* > 0$ be a sufficiently small number so that $z_\theta \subseteq \Delta$ for every $\theta \in (0, \theta^*]$. Then, in particular, for any $\theta \in (0, \theta^*]$ and $s \in \mathcal{S}$, we have

$$\sigma_{z_\theta}(s) = U(s, p_0 + \theta s) = \theta.$$

Fix a number $\theta_* \in (0, \theta^*)$, and let (θ_n) be a decreasing sequence in (θ_*, θ^*) that converges to θ_*. For each $n \in \mathbb{N}$, put $z^n := z_{\theta_*} \cup y^n$ where $y^n := \{p_0 + \theta_n s : s \in F\}$. Note that since F is closed, so are y^n and z^n for each n. Moreover, as $\theta_n \to \theta_*$, it clearly follows that y^n converges to the set $y_{\theta_*} := \{p_0 + \theta_* s : s \in F\}$ in the Hausdorff metric.

Fix any $s' \in \mathcal{S} \setminus F$. Then, as $p_0 + \theta_* s'$ is the unique maximizer of $U(s', \cdot)$ on z_{θ_*}, we must have $\sigma_{z_{\theta_*}}(s') > \sigma_{y_{\theta_*}}(s')$. As $y^n \to y_{\theta_*}$, it follows that $\sigma_{z_{\theta_*}}(s') > \sigma_{y^n}(s')$ for all sufficiently large n. For any such n, we thus have $\sigma_{z_{\theta_*}}(s') = \sigma_{z^n}(s')$. Since s' is an arbitrary point in $\mathcal{S} \setminus F$, and since $\sigma_{z^n}(s) = \theta_n > \theta_* = \sigma_{z_{\theta_*}}(s)$ for each $s \in F$ and $n \in \mathbb{N}$, we conclude that

$$\bigcap_{n=1}^{\infty} \mathcal{S}_n = F, \tag{27}$$

where $\mathcal{S}_n := \{s \in \mathcal{S} : \sigma_{z^n}(s) > \sigma_{z_{\theta_*}}(s)\}$.

Moreover, for each $n \in \mathbb{N}$, we have

$$\mathcal{S}_{n+1} \subseteq \mathcal{S}_n. \tag{28}$$

To see this, fix any $s \in \mathcal{S}$ with $\sigma_{z^{n+1}}(s) > \sigma_{z_{\theta_*}}(s)$. Then, by definition of z^{n+1}, we have $\sigma_{z^{n+1}}(s) = U(s, p_0 + \theta_{n+1}\tilde{s}) = \theta_{n+1}U(s,\tilde{s})$ for some $\tilde{s} \in F$. As $\sigma_{z_{\theta_*}}(s) > 0$, it also follows that $U(s,\tilde{s}) > 0$. Since $\sigma_{z^n}(s) \geq U(s, p_0 + \theta_n\tilde{s}) = \theta_n U(s,\tilde{s})$, and since $\theta_n > \theta_{n+1}$, we therefore see that $\sigma_{z^n}(s) > \sigma_{z^{n+1}}(s)$. Thus, $\sigma_{z^n}(s) > \sigma_{z_{\theta_*}}(s)$. This proves (28).

As ν^+ is countably additive, (27) and (28) imply $\lim_n \nu^+(\mathcal{S}_n) = \nu^+(F)$. Since $\nu^+(F) = 0 < \nu^-(F)$, we thus see that

$$\nu^+(\mathcal{S}_n) < \nu^-(F) \text{ for all sufficiently large } n. \tag{29}$$

Fix any $n \in \mathbb{N}$ and note that

$$W_\nu(z^n) - W_\nu(z_{\theta_*}) = \int_{\mathcal{S}} \left(\sigma_{z^n} - \sigma_{z_{\theta_*}}\right) \nu^+(ds) - \int_{\mathcal{S}} \left(\sigma_{z^n} - \sigma_{z_{\theta_*}}\right) \nu^-(ds). \tag{30}$$

Moreover,

$$\begin{aligned}
\int_{\mathcal{S}} \left(\sigma_{z^n} - \sigma_{z_{\theta_*}}\right) \nu^+(ds) &= \int_{\mathcal{S}_n} \left(\sigma_{z^n} - \sigma_{z_{\theta_*}}\right) \nu^+(ds) + \int_{\mathcal{S}\setminus\mathcal{S}_n} \left(\sigma_{z^n} - \sigma_{z_{\theta_*}}\right) \nu^+(ds) \\
&= \int_{\mathcal{S}_n} \left(\sigma_{z^n} - \sigma_{z_{\theta_*}}\right) \nu^+(ds) \\
&\leq \int_{\mathcal{S}_n} (\theta_n - \theta_*)\, \nu^+(ds) = (\theta_n - \theta_*)\, \nu^+(\mathcal{S}_n),
\end{aligned}$$

where the inequality is a consequence of the fact that $\sigma_{z^n} \leq \theta_n$.

Similarly,

$$\begin{aligned}
\int_{\mathcal{S}} \left(\sigma_{z^n} - \sigma_{z_{\theta_*}}\right) \nu^-(ds) &= \int_{\mathcal{S}_n} \left(\sigma_{z^n} - \sigma_{z_{\theta_*}}\right) \nu^-(ds) \\
&\geq \int_F \left(\sigma_{z^n} - \sigma_{z_{\theta_*}}\right) \nu^-(ds) \\
&= \int_F (\theta_n - \theta_*)\, \nu^-(ds) = (\theta_n - \theta_*)\, \nu^-(F).
\end{aligned}$$

By (29) and (30), we therefore see that $W_\nu(z^n) - W_\nu(z_{\theta_*}) < 0$ for all sufficiently large n. But as $s^* \in \mathcal{S} \setminus F$, for all sufficiently large n equation (27) implies $\sigma_{z^n}(s^*) = \sigma_{z_{\theta_*}}(s^*)$, that is, $\max \tilde{\varphi}(z^n) = \max \tilde{\varphi}(z_{\theta_*})$. Since $z_{\theta_*} \subset z^n$ for every n, we have therefore obtained a contradiction to (23). This proves (26).

Now, by (26), we can write $\nu^- = \gamma^-\delta_{s^*}$ where $\gamma^- := \nu^-(\{s^*\})$. Put $\mu := \nu^+$ and, for every $q \in \Delta$, $\varphi(q) := \tilde{\varphi}(q) + \gamma^- U(s^*, q)$. Then, φ is a positive affine transformation of $\tilde{\varphi}$. Moreover, for any $x \in \mathcal{X}$,

$$\begin{aligned}\int_{\mathcal{S}} \sigma_x(s)\,\mu(ds) - \max \varphi(x) &= \int_{\mathcal{S}} \sigma_x(s)\,\nu^+(ds) - \big(\max \tilde{\varphi}(x) + \gamma^- \sigma_x(s^*)\big) \\ &= \int_{\mathcal{S}} \sigma_x(s)\,\nu(ds) - \max \tilde{\varphi}(x).\end{aligned}$$

Thus, from Lemma 1, it follows that (φ, μ) is an S-representation for $\mathbf{c}$.

Finally, if $\tilde{\varphi}$ is constant, (23) implies that $W_\nu(x) \leq W_\nu(y)$ whenever $x \subseteq y$. But then, by an obvious modification of the arguments above, we see that $\nu^- = 0$. Hence, in this case, we can set $\mu := \nu$ and $\varphi := \tilde{\varphi}$. This completes the proof of Theorem 2. □

A.5 Proofs of Lemma 3 and Corollary 1

Let $\mathbf{c}$ be a regular choice correspondence on $\mathcal{X}$ that has an S-representation. For the "only if" part of Corollary 1, assume that $\mathbf{c}$ also satisfies nontriviality-c, nontriviality-d and monotonicity axioms, and let $(\tilde{\varphi}, \nu)$ be an S-representation for $\mathbf{c}$. Then, $\tilde{\varphi}$ is obviously nonconstant by nontriviality-c. When proving Theorem 2, we have also seen that $\nu_{-*} \geq 0$. Moreover, nontriviality-d immediately implies $\nu_{-*} \neq 0$ as we discussed in text. Hence, we can simply set $(\varphi, \mu) := \frac{1}{\nu(\mathcal{S})}(\tilde{\varphi}, \nu)$ if $\nu(\{s^*\}) \geq 0$. Otherwise, following the final steps of the proof of Theorem 2, we can focus on an S-representation of the form (φ', ν^+) and proceed as before. In turn, the "if" part of Corollary 1 is an immediate consequence of the corresponding part of Lemma 3, which we prove next.

Suppose that $\mathbf{c}$ admits an S-representation (φ, μ) such that $\mu_{-*} \neq 0$ and φ is nonconstant. That $\mathbf{c}$ must satisfy nontriviality-c is obvious. To verify nontriviality-d, we first note that, as the support of μ is distinct from $\{s^*\}$, following the proof of Theorem 2 we can find two sets $z, z' \in \mathcal{X}$, where z is a sphere and $z \subseteq z'$, such that $\max \varphi(z) = \max \varphi(z')$ and $W_\mu(z) \neq W_\mu(z')$. (See, in particular, the arguments that follow the equation (26).)

First, let us assume $W_\mu(z') > W_\mu(z)$. It is clear that by the regularity axiom, there exist a set $y_* \in \mathcal{X}$ and a number $\lambda \in (0, 1]$ such that $\max \varphi(\lambda z' + (1-\lambda) y_*) = W_\mu(\lambda z' + (1-\lambda) y_*)$. Then, by construction, we also have $\max \varphi(\lambda z' + (1-\lambda) y_*) = \max \varphi(\lambda z + (1-\lambda) y_*)$ and $W_\mu(\lambda z' + (1-\lambda) y_*) > W_\mu(\lambda z + (1-\lambda) y_*)$. Since the set $x := \lambda z' + (1-\lambda) y_*$ contains $y := \lambda z + (1-\lambda) y_*$, it clearly follows that x and y satisfy the properties required by the nontriviality-d axiom.

If $W_\mu(z') < W_\mu(z)$, we can simply change the roles of the sets z' and z in the above argument, and thereby, complete the proof of the "if" part of Lemma 3. By our previous arguments, the "only if" part of this result is clear. □

A.6 Omitted Proofs From Section 7

We start with a simple consequence of the findings of DLR which we prove for the sake of completeness.

Lemma A3. *Let θ be a real number and consider a pair of measures μ, μ' on $\mathcal{S}$ such that $\int_{\mathcal{S}} \sigma_x(s)\,\mu'(ds) = \int_{\mathcal{S}} \sigma_x(s)\,\mu(ds) + \theta$ for each $x \in \mathcal{X}$. Then, $\mu = \mu'$ and $\theta = 0$.*

Proof. Put $\nu := \mu' - \mu$. The equations in the lemma imply that, for each $x, y \in \mathcal{X}$ and $\gamma > 0$, $\int_{\mathcal{S}} \gamma\,(\sigma_x(s) - \sigma_y(s))\,\nu(ds) = 0$. As shown by DLR (Lemma 11, p.928), the set $\{\gamma\,(\sigma_x - \sigma_y) : x, y \in \mathcal{X},\ \gamma > 0\}$ is dense in the Banach space of continuous, real functions on $\mathcal{S}$, which we denote by $\mathbf{C}(\mathcal{S})$. Hence, it follows that $\int_{\mathcal{S}} g(s)\nu(ds) = 0$ for each $g \in \mathbf{C}(\mathcal{S})$. Thus, by standard separation and duality arguments, ν must be the zero measure; that is, we must have $\mu = \mu'$. This immediately implies $\theta = 0$. $\square$

Proof of Theorem 3. By Lemma 1, under the hypotheses of the theorem, there exist $\theta, \theta' \in \mathbb{R}$ and positive numbers $\lambda, \lambda', \gamma$ such that $\varphi(\cdot) = \lambda U(s^*, \cdot) + \theta$, $\varphi'(\cdot) = \lambda' U(s^*, \cdot) + \theta'$ and $W_{\mu'}(\cdot) - \max \varphi'(\cdot) = \gamma(W_\mu(\cdot) - \max \varphi(\cdot))$. Using the two former equalities, we can rewrite the latter equality as $\int_{\mathcal{S}} \sigma_x(s)\,\mu'(ds) = \int_{\mathcal{S}} \sigma_x(s)\,\nu(ds) + \theta' - \gamma\theta$, for each $x \in \mathcal{X}$, where $\nu := \gamma\mu + (\lambda' - \gamma\lambda)\,\delta_{s^*}$. By Lemma A3, it follows that $\mu' = \gamma\mu + (\lambda' - \gamma\lambda)\,\delta_{s^*}$ so that $\mu'_{-*} = \gamma\mu_{-*}$, as we seek. $\square$

Proof of Theorem 3'. Part (i) is simply a restatement of Lemma 3, and part (iii) follows from Theorem 3 immediately. It is also clear that part (i) implies part (ii), for if $\mathbf{c}$ admits a semi k-normalized S-representation (φ, μ), then given any $\gamma > 0$, the pair $(\gamma\varphi, \gamma\mu)$ is a semi γk-normalized S-representation for $\mathbf{c}$. $\square$

Proof of Lemma 4. If $\mathbf{c}$ satisfies the hypotheses of the lemma, for any $k > 0$, Theorem 3' implies that $\mathbf{c}$ admits a semi k-normalized S-representation $(\tilde{\varphi}, \tilde{\mu})$. As usual, we can write $\tilde{\varphi}(\cdot) = \tilde{\lambda} U(s^*, \cdot) + \theta$ for some $\tilde{\lambda} > 0$ and $\theta \in \mathbb{R}$. Denote by λ^* the unique number such that $\lambda^*\,(U(s^*, p^*) - U(s^*, p_*)) = k^*$, and define $\varphi(\cdot) := \lambda^* U(s^*, \cdot) + \theta$ so that $\varphi(p^*) - \varphi(p_*) = k^*$. Then, in view of our previous discussions, it is clear that with $\mu := \tilde{\mu} + (\lambda^* - \tilde{\lambda})\delta_{s^*}$ the pair (φ, μ) is also a semi k-normalized S-representation for $\mathbf{c}$.

In fact, the pair (φ, μ) that we constructed is the unique representation with the desired properties. To see this, consider an S-representation (φ', μ') for $\mathbf{c}$. Note that if $\varphi'(p^*) - \varphi'(p_*) = k^*$, the function φ' must be of the form $\lambda^* U(s^*, \cdot) + \theta'$ for a $\theta' \in \mathbb{R}$. Moreover, if (φ', μ') is semi k-normalized, we must also have $\mu'_{-*} = \mu_{-*}$ by Theorem 3'. Following the proof of Theorem 3, this implies $W_{\mu'}(\cdot) - \max \varphi'(\cdot) = W_\mu(\cdot) - \max \varphi(\cdot)$, and hence, $\int_{\mathcal{S}} \sigma_x(s)\,\mu'(ds) = \int_{\mathcal{S}} \sigma_x(s)\,\mu(ds) + \theta' - \theta$ for each $x \in \mathcal{X}$. By Lemma A3, we therefore see that $\mu' = \mu$ and $\theta' = \theta$. Thus, we also have $\varphi' = \varphi$. $\square$

Proof of Lemma 5. In view of the proofs of Theorem 3 and Lemma 4, given any $k > 0$, the class of semi k-normalized S-representations for $\mathbf{c}$ is of the form $\{(\varphi_\lambda, \mu_\lambda) : \lambda > 0\}$ where $\varphi_\lambda(\cdot) := \lambda U(s^*, \cdot) + \theta$, and θ is a number that is independent from λ. Moreover, $\mu_\lambda(\{s^*\})$ is a linearly increasing function of λ, as the measures μ_λ and $\mu_{\lambda'}$ are related by $\mu_{\lambda'} = \mu_\lambda + (\lambda' - \lambda)\,\delta_{s^*}$ for any $\lambda, \lambda' > 0$. Hence, $k_0 := \lim_{\lambda \to 0} \mu_\lambda(\{s^*\})$ is the infimum of k_* which are compatible with the desired form of normalization. Given the structure of k-normalized S-representations, the uniqueness part of the assertion is an obvious consequence of Lemma 4. $\square$

References

C.D. Aliprantis, K. Border, *Infinite Dimensional Analysis*, Berlin, Springer, 1999.

C.J. Anderson, The psychology of doing nothing: forms of decision avoidance result from reason and emotion, *Psychological Bulletin* **129** (2003), 139–167.

J. Apesteguia, M.A. Ballester, A theory of reference-dependent behavior, *Economic Theory* **40** (2009), 427–455.

P. Boatwright, J.C. Nunes, Reducing assortment: an attribute-based approach, *Journal of Marketing* **65** (2001), 50–63.

R.M. Corbin, *No-choice option in decision-making*, Ph.D. Dissertation, McGill University, 1976.

R.M. Corbin, Decisions that might not get made, in T.S. Wallsten (Ed.), *Cognitive Processes in choice and decision behavior*, Hillsdale, Lawrence Erlbaum Associates Inc., 1980.

M. Dean, Status quo bias in large and small choice sets, New York University (2008), mimeo.

E. Dekel, B.L. Lipman, A. Rustichini, Representing preferences with a unique subjective state space, *Econometrica* **69** (2001), 891–934.

E. Dekel, B.L. Lipman, A. Rustichini, Temptation-driven preferences, *Review of Economic Studies* **76** (2009), 937–971.

E. Dekel, B.L. Lipman, A. Rustichini, T. Sarver, Representing preferences with a unique subjective state space: a corrigendum, *Econometrica* **75** (2007), 591–600.

R. Dhar, To choose or not to choose: this is the question, *Advances in Consumer Research* **19** (1992), 735–738.

R. Dhar, Consumer preference for a no-choice option, *Journal of Consumer Research* **24** (1997), 215–231.

R. Dhar, I. Simonson, The effect of forced choice on choice, *Journal of Marketing Research* **40** (2003), 146–160.

L.G. Epstein, M. Marinacci, K. Seo, Coarse contingencies and ambiguity, *Theoretical Economics* **2** (2007), 355–394.

G. Gerasímou, Rational indecisive choice, Cambridge University (2010), mimeo.

F. Gul, W. Pesendorfer, Temptation and self-control, *Econometrica* **69** (2001), 1403–1435.

S.S. Iyengar, M.R. Lepper, When choice is demotivating: can one desire too much of a good thing?, *Journal of Personality and Social Psychology* **79** (2000), 995–1006.

S.S. Iyengar, G. Huberman, W. Jiang, How much choice is too much?, Contributions to 401(k) retirement plans, in S.O. Mitchell and P.S. Utkus (Eds.), *Pension Design and Structure, New Lessons from Behavioral Finance,* New York, Oxford University Press, 2004.

A. Kochov, The epistemic value of a menu and subjective states, University of Rochester (2009), mimeo.

I. Kopylov, Choice deferral and ambiguity aversion, *Theoretical Economics* **4** (2009), 199–225.

D.M. Kreps, A representation theorem for 'preference for flexibility', *Econometrica* **47** (1979), 565–577.

D.M. Kreps, *Notes on the Theory of Choice*, Boulder, Westview Press, 1988.

Y. Masatlioglu, E.A. Ok, Rational choice with status quo bias, *Journal of Economic Theory* **121** (2005), 1–29.

Y. Masatlioglu, E.A. Ok, A canonical model of choice with initial endowments, New York University (2009), mimeo.

E. Reutskaya, R.M. Hogarth, Satisfaction in choice as a function of the number of alternatives: when "goods satiate," *Psychology & Marketing* **26** (2009), 197–203.

J.S. Sagi, Anchored preference relations, *Journal of Economic Theory* **130** (2006), 283–295.

T. Sarver, Anticipating regret: why fewer options may be better, *Econometrica* **76** (2008), 263–305.

R. Schneider, *Convex Bodies: The Brunn-Minkowski Theory*, Cambridge, Cambridge University Press, 1993.

A.M. Shah, G. Wolford, Buying behavior as a function of parametric variation of number of choices, *Psychological Science* **18** (2007), 369–370.

J.E. Stovall, Multiple temptations, *Econometrica* **78** (2010), 349–376.

J. Stoye, Revealed preference when agents can randomize, Cornell University (2010), mimeo.

A. Tversky, E. Shafir, Choice under conflict: the dynamics of deferred decision, *Psychological Science* **3** (1992), 358–361.

C. M. White, U. Hoffrage, Testing the tyranny of too much choice against the allure of more choice, *Psychology & Marketing* **26** (2009), 280–298.

Paper 2

Status Quo Bias with Choice Avoidance

Gökhan Buturak

Abstract

People often find it difficult to make a choice out of a set of alternatives, thereby possibly resolving a choice problem in favor of a default option. Also, the presence of a default option may well have an influence on the choices of a decision maker. We here propose an abstract choice model in which choice avoidance emerges as a particular form of status quo bias. The model describes an otherwise standard decision maker who mentally constrains her choice problems with respect to a default option, the choice of which can be ascribed to either the agent's indecisiveness or there being no strictly better alternative within the constrained choice set. We show that, when the agent's choice behavior satisfies the classical rationality tenets across choice problems with an identical status quo alternative, the model reduces to that of Masatlioglu and Ok (2009).

JEL Classification: D11, D81.
Keywords: Status quo bias, choice avoidance, incomplete preferences.

This paper is a by-product of my collaboration with Özgür Evren on the first paper of this thesis. I gratefully acknowledge my intellectual debt to him. I also thank Tore Ellingsen, Jens Josephson and Mark Voorneveld for their helpful comments and suggestions. All remaining errors and deficiencies are mine.

1. Introduction

People often find it difficult to choose, and often end up with default options.[1] Explanations for such choice avoidance include anticipated regret, uncertainty about one's own preferences, search for better alternatives, and conflict induced by the structure of the choice problem.

A remarkable experimental finding is that agents are more prone to choose their default options when faced with a larger set of alternatives.[2] This finding challenges rational choice theory as it implies that agents sometimes stick to their default option even when it is revealed to be inferior to some of the available alternatives. This raises the question if and how one can rationalize the observed behavior in such choice situations.

It is also well-documented that the choice behavior of agents often exhibits reference dependence with respect to a status quo alternative or an initial endowment. For instance, Madrian and Shea (2001) empirically study the savings behavior of employees in a large U.S. corporation and report that the default contribution rate and default investment allocation chosen by the company for automatic enrollment has a strong influence on the savings behavior of 401(k) participants. A substantial fraction of 401(k) participants hired under automatic enrollment stick to both the default contribution rate and fund allocation even though very few employees hired before automatic enrollment would have picked this particular outcome. These and other findings prompted an interest in formal models of reference-dependent choice. However, existing models are not compatible with the notion of choice avoidance as they respect the classical rationality postulates, holding the status quo fixed (see Masatlioglu and Ok, 2005, 2009; Sagi, 2006; and Apesteguia and Ballester, 2009).

The main purpose of this paper is to reconcile the notion of status quo bias with choice avoidance within an axiomatic choice framework. More specifically, we view choice avoidance as a particular form of status quo bias. By keeping her default option, an agent (temporarily) gains the flexibility of postponing the task of making an active choice.[3] On the other hand, the absence of a default option supposedly forces the agent to select an option. Accordingly, our representation result describes the choice behavior of an otherwise standard agent who mentally constrains her choice problems with respect to a status quo and selects it whenever she is indecisive about which alternative is the best among the constrained choice set induced by the presence of the status quo.

[1] For experimental evidence, see Anderson (2003) and the references therein.

[2] See, e.g., Iyengar and Lepper (2000), Boatwright and Nunes (2001), Shah and Wolford (2007), and Reutskaja and Hogart (2009).

[3] By making an active choice, we refer to the choice of an alternative that is deemed at least as good as the default option whenever such options are feasible.

To accommodate the observed patterns of choice behavior related to avoidance, our model restricts the classical rationality postulates solely to choice situations where the agent makes an active choice. Following Masatlioglu and Ok (2009), we employ the status quo irrelevance axiom, which identifies the instances in which the presence of a status quo is irrelevant to the choice behavior of the agent. In addition, we introduce the conflict driven aversion axiom to pin down the choice behavior of the agent under avoidance. We couple these with a continuity axiom to get a representation which identifies the agent's decision rule with three objects: (i) a complete preference relation which reflects that the agent has a complete ranking of the available alternatives and makes her choices accordingly in the absence of a default option; and, for each alternative x, (ii) a set $\mathcal{Q}(x)$ containing the alternatives the agent finds at least as good as the default option x, and (iii) a (possibly incomplete) preference relation corresponding to the agent's choices over the alternatives in $\mathcal{Q}(x)$. With regard to the latter, the agent tries to identify the best element(s) in $\mathcal{Q}(x)$ and sticks to her default option whenever she cannot find one. By contrast, in the standard models of status quo bias, though the agent may not be able to compare all the alternatives with the default option, she is able to rank those that are at least as good as the default.[4] Accordingly, we show that the representation in Masatlioglu and Ok (2009) is a special case of ours.

The main idea that underlies our model is that the status quo is just a preliminary choice, which may be reversed upon arrival of new (introspective or objective) information that may help the agent judge the available alternatives in a better way, or upon finding new alternatives. When viewed from this perspective, we may assume that the agent's choice behavior is driven mainly by her concern about her future well-being.

Depending on how one interprets immediate choices, one can justify this assumption in two different ways. First, one can assume that making an active choice immediately means that the agent takes a binding action, such as signing a contract, that will make her face the associated consequences at a later date. It is important to observe that this interpretation requires the agent to be able to commit herself *only* to a single alternative among her all feasible options. In this case, avoiding choice would always be (weakly) desirable as in the preference for flexibility models.

Second, as in the classical choice theory, one can assume that upon choosing an alternative, the agent begins experiencing the associated material consequences immediately. Of course, according to this interpretation, temporal payoffs could be important. If, however, the delay in making an active choice is relatively short, in many cases of interest one could deem this interval immaterial. For instance, from the perspective of

[4]Needless to say, in the standard choice framework, the decision maker is always decisive and her choices are not affected by the presence of a status quo alternative.

a firm, a worker's performance during her first month of employment would presumably have no significance beyond being an indicator of her future performance. Similarly, when deciding whether to undertake a long term project, the project's anticipated performance in the upcoming few weeks would presumably be the least of the concerns of an investor.

While it is reasonable to interpret our model along the above lines, a questionable feature of the representation in the present paper is that it predicts the agent would retain her default option whenever she faces two incomparable alternatives regardless of the costs that the choice of the default option may incur. For instance, given that the default option is a not-so-desirable job with a low salary, indecisiveness on two high-paying job offers undoubtedly better than the current one may cause the agent to stay at the current inferior job for another month in order to contemplate on the two job offers, irrespective of whether the loss incurred is significant or not. However, if the contracts are binding for a long enough period of time, it would be reasonable for the agent to forgo the extra increment of a month's salary for contemplation as this amount would be immaterial given the long period of contractual obligation. The reason that our representation allows for these two contrasting interpretations is that we do not provide structure for the agent's behavior at the instances she avoids making an active choice. As a result, the representation we provide cannot distinguish rational from irrational behavior regarding choice avoidance.

Recently, Buturak and Evren (2010) proposed a model of choice deferral in which the decision maker behaves as if the value of deferring choice from a given set of options is determined by an affine utility function W over choice sets. In their model, the decision maker in question compares W with the value of making a choice immediately (determined by an affine utility function φ) so as to decide on opting for an option now or later. A similar approach would be taken here by obtaining two such functions as in Buturak and Evren (2010). Given that we derive the agent's complete ranking of alternatives when she has to make an immediate choice, what remains is to find, for each alternative x, a suitable function $W(\cdot, x)$ over choice sets consistent with any given choice data. As the current model is situated in the ordinal framework, in principle, it seems possible to derive such a function W. We leave this extension as future research. A benefit of the current approach we follow is that it enables a direct comparison of our model with the existing models of status quo bias.

We shall proceed in the next section with formally introducing the model. Then, we will present our axioms and the representation result in Section 3. In Section 4, we will give a characterization of choice avoidance. We will then provide a representation of choice correspondences without avoidance in Section 5. We will, in Section 6, investigate how the notions of status quo bias and choice aversion are related to each other in the current model. In Section 7, we will discuss two related papers. We will give a particular attention to Dean's (2008) model on choice avoidance and highlight the differences between his representation and the one we provide here. The Appendix contains the omitted proofs in the text.

2. The Model

We closely follow the notation and terminology of Masatlioglu and Ok (2009). Let X be a nonempty compact metric space which represents the set of all alternatives. The generic elements of X are denoted by x, y, w, z. Also, let $\mathcal{X}$ denote the collection of all closed, nonempty subsets of X endowed with the Hausdorff metric, the typical elements of which are denoted by S, T. The symbol $\diamond$ stands for an object that does not belong to X. A **choice problem** is a pair (S, x), where $S \in \mathcal{X}$ and either $x \in S$ or $x = \diamond$. When $x \in S$, we think of x as the status quo of the choice problem (S, x), and say that (S, x) is a **choice problem with a status quo**. On the other hand, a **choice problem without a status quo** is of the form $(S, \diamond)$ for some $S \in \mathcal{X}$.[5] We denote by $\mathcal{C}(X)$ and $\mathcal{C}_{sq}(X)$, respectively, the set of all choice problems and the set of all choice problems with a status quo.

We consider an agent who faces a choice problem $(S, x) \in \mathcal{C}(X)$. In the case where $x \in S$, she can either choose an alternative from S immediately, or keep her status quo x. In the former case, the agent's choice can be considered as binding. The latter case may be interpreted as the inertia of the agent to stick to her default option or as an instrument to keep her options open and perhaps to search for more alternatives.

A **choice correspondence** refers to a nonempty valued correspondence $\mathbf{c}$ on $\mathcal{C}(X)$ such that $\mathbf{c}(S, x) \subseteq S$ for every $(S, x) \in \mathcal{C}(X)$.

Given a binary relation $\succcurlyeq$ on a nonempty set Y, we often write $y \succcurlyeq z$ instead of $(y, z) \in \succcurlyeq$. The asymmetric and symmetric parts of $\succcurlyeq$ are denoted as $\succ$ and $\sim$, respectively. We say that x and y are **$\succcurlyeq$-incomparable** if neither $y \succcurlyeq x$ nor $x \succcurlyeq y$, and write $y \bowtie x$. Moreover, we say that a binary relation $\succcurlyeq^\bullet$ on Y **extends** $\succcurlyeq$ if, for every $y, z \in Y$,

$$y \succcurlyeq z \quad \text{implies} \quad y \succcurlyeq^\bullet z, \quad \text{and} \quad y \succ z \quad \text{implies} \quad y \succ^\bullet z.$$

For any nonempty subset S of Y, by $\mathbf{m}(S, \succcurlyeq)$ we mean the set of $\succcurlyeq$-maximum elements in S, that is, $\mathbf{m}(S, \succcurlyeq) := \{x \in S : x \succcurlyeq y \ \forall y \in S\}$. As usual, a preorder refers to a transitive and reflexive binary relation.

[5] As Masatlioglu and Ok (2009) note, the usage of the symbol $\diamond$ is redundant since it is used to indicate that a choice problem is without status quo bias. Practically, the notation $\mathbf{c}(S, \diamond)$ is not different than $\mathbf{c}(S)$. However, notation-wise it will prove useful in what follows.

3. The Axioms and the Representation Result

We start with customizing the classical rationality axioms to the needs of our model.

Property α'. For any $(S,x),(T,x) \in \mathcal{C}(X)$, if $y \in T \subseteq S$, $y \in \mathbf{c}(S,x)$ and $\mathbf{c}(S,x) \neq \{x\}$, then $y \in \mathbf{c}(T,x)$.

Property β'. For any $(S,x),(T,x) \in \mathcal{C}(X)$, if $y,z \in \mathbf{c}(T,x)$, $T \subseteq S$, $z \in \mathbf{c}(S,x)$ and $\mathbf{c}(S,x) \neq \{x\}$, then $y \in \mathbf{c}(S,x)$.

By contrast, the traditional versions of these axioms are as follows.

Property α. For any $(S,x),(T,x) \in \mathcal{C}(X)$, if $y \in T \subseteq S$ and $y \in \mathbf{c}(S,x)$, then $y \in \mathbf{c}(T,x)$.

Property β. For any $(S,x) \in \mathcal{C}(X)$, if $y,z \in \mathbf{c}(S,x)$, $S \subseteq T$ and $z \in \mathbf{c}(T,x)$, then $y \in \mathbf{c}(T,x)$.

It is important to observe that properties α and α' and properties β and β' coincide on choice problems without a status quo, though, in general, the latter properties (α' and β') are weaker than the former (α and β, respectively).

Definition 1. Given a choice problem (S,x), we say that an alternative $y \in S \setminus \{x\}$ is **aversive** at (S,x), if $\mathbf{c}(S,x) = \{x\}$ but $\mathbf{c}(S \setminus \{y\},x) \neq \{x\}$. In turn, we say that the choice correspondence $\mathbf{c}$ is **non-averse** if there does not exist a choice problem with an aversive alternative.

Notice that, as is obvious from the statements for the properties α' and β', the classical rationality postulates fully hold across choice problems without a status quo alternative. Therefore, the notion of aversion as defined above can occur only in choice problems with a status quo. A practical side of this approach is that it allows us to derive the agent's preference relation over the set of alternatives in the absence of a default option, in which case she has to make a choice. This is also consistent with the existing models of status quo bias.

In the current model, we wish to rule out choice avoidance that might be caused by the presence of obviously inferior alternatives. To this end, the next axiom requires that aversive alternatives must be "good enough" in two different meanings.

Conflict Driven Aversion (CDA). If y is aversive at $(S,x) \in \mathcal{C}_{sq}(X)$, then:
i. For any $T \in \mathcal{X}$ with $\{x,y\} \subseteq T \subseteq S$, we have $\mathbf{c}(T,x) \subseteq \{x,y\}$.
ii. $\{y\} = \mathbf{c}(\{x,y\},x)$.

Part ***i.*** of CDA pins down the choices of the agent under aversion. Intuitively, if y is aversive at a choice problem (S,x), then y must be good enough so that, at any choice problem (T,x) with $T \subseteq S$, either it is chosen for it is clearly superior to all other in T, or x is chosen for there is another alternative z that the agent finds difficult to compare with y. To make it concrete, suppose the agent opts for the default option at a choice problem (S,x), whereas she chooses $y \in S$ in the absence of an alternative z and vice versa. This indicates that, at the choice problem (S,x), she has difficulty to make a choice between the alternatives y and z: The presence of the two creates avoidance which makes the agent stick to x though she finds both y and z obviously better than the others. Therefore, at any subset T of S in which y and z are present, the agent should still have difficulty to make a choice between the two and, as a result, pick x. On the other hand, in the absence of one of them, she should not have any hesitation to pick the other.

Part ***ii.*** of CDA, on the other hand, states that any aversive alternative must be better than the status quo. Put differently, the presence/absence of the alternatives that are (weakly) worse than the default option should have no impact on the choices of the agent; i.e., such alternatives cannot be aversive. Hence, any alternative inferior to the status quo should not cause any difficulty when making a choice.

Following Masatlioglu and Ok (2009), we also employ the next axiom which identifies the instances in which the presence of a status quo alternative does not influence the choice behavior.

Status Quo Irrelevance (SQI). For any $(S,x) \in \mathcal{C}_{sq}(X)$, suppose that $\mathbf{c}(T,x) \neq \{x\}$ for every non-singleton subset T of S with $x \in T$. Then, $\mathbf{c}(S,x) = \mathbf{c}(S,\diamond)$.

Given a choice problem (S,x), if the status quo alternative x is not strictly chosen at any subset of S other than the singleton set containing x, SQI states that the presence of the status quo has no influence on the choice problem and, hence, (S,x) should reasonably be equivalent to the choice problem without a status quo at S, $(S,\diamond)$. This means that y should be chosen from $(S,\diamond)$ whenever it is chosen from (S,x).

SQI implies the following link between choice with and without status quo, which Masatlioglu and Ok (2009) aptly term the *weak axiom of status quo bias*: For any $x,y \in X$ with $x \neq y$,

$$y \in \mathbf{c}(\{x,y\},x) \quad \text{implies} \quad y \in \mathbf{c}(\{x,y\},\diamond), \quad \text{and}$$
$$y \in \mathbf{c}(\{x,y\},\diamond) \quad \text{implies} \quad y \in \mathbf{c}(\{x,y\},y).$$

Also, we note that the nonemptiness of the choice correspondence $\mathbf{c}$ implies $\mathbf{c}(\{x\},x) = \mathbf{c}(\{x\},\diamond) = \{x\}$.

The representation we will provide shortly requires the following standard continuity axiom.

Upper Hemicontinuity (UHC). Let (y_n) be a sequence in X that converges to $y \in X$, and let (S_n, x) be a sequence in $\mathcal{C}(X)$ that converges in Hausdorff metric to $(S, x) \in \mathcal{C}(X)$. Then, $y \in \mathbf{c}(S, x)$ whenever $y_n \in \mathbf{c}(S_n, x)$ for every n.

We say that a choice correspondence $\mathbf{c}$ is **avoidance consistent** if it satisfies properties α', β', CDA ***i.***, CDA ***ii.***, SQI and UHC.

Now, we are ready to state our representation theorem.

Theorem 1. *Let X be a nonempty compact metric space. A choice correspondence $\mathbf{c}$ on $\mathcal{C}(X)$ is avoidance consistent if, and only if, there exist a closed-valued self-correspondence $\mathcal{Q}$ on X, a complete preorder $\succcurlyeq^*$ on X, and for each $x \in X$, a preorder $\succcurlyeq^\bullet_x$ on the set $\mathcal{Q}(x)$ such that:*
i. *For each $x \in X$, the set $\mathcal{Q}(x)$ contains x, and $\succcurlyeq^*$ extends $\succcurlyeq^\bullet_x$ on $\mathcal{Q}(x)$. Moreover, for every $x \in X$ and $y \in \mathcal{Q}(x)$ we have $y \succcurlyeq^\bullet_x x$.*
ii. $\mathbf{c}(\cdot, \diamond) = \mathbf{m}(\cdot, \succcurlyeq^*)$.
iii. *For every $(S, x) \in \mathcal{C}_{sq}(X)$,*

$$\mathbf{c}(S, x) = \begin{cases} \{x\} & \textit{if } \mathbf{m}(S \cap \mathcal{Q}(x), \succcurlyeq^\bullet_x) = \varnothing, \\ \mathbf{m}(S \cap \mathcal{Q}(x), \succcurlyeq^\bullet_x) & \textit{otherwise.} \end{cases}$$

Since $\succcurlyeq^\bullet_x$ is allowed to be incomplete, when more alternatives become available or some existing ones become unavailable, the agent *may* find it hard to identify a $\succcurlyeq^\bullet_x$-maximum element among the alternatives which are better than her status quo x. According to ***iii.*** above, in such cases the agent sticks to her status quo, which is the basis of our explanation of choice avoidance in the current model.

We remark that, whenever $\mathbf{m}(S \cap \mathcal{Q}(x), \succcurlyeq^\bullet_x)$ and $S \cap \mathcal{Q}(x) \setminus \{x\}$ are both nonempty, ***iii.*** implies that $\mathbf{c}(S, x) = \mathbf{m}(S \cap \mathcal{Q}(x), \succcurlyeq^\bullet_x) \subseteq S \cap \mathcal{Q}(x) \setminus \{x\}$. However, since $\succcurlyeq^\bullet_x$ is allowed to be incomplete, even if $S \cap \mathcal{Q}(x) \setminus \{x\}$ is nonempty, we may have $\mathbf{m}(S \cap \mathcal{Q}(x), \succcurlyeq^\bullet_x) = \varnothing$ so that $\mathbf{c}(S, x) = \{x\}$. Depending on how one interprets the choice of status quo when better alternatives are available, the choice behavior of the agent may be considered as "irrational;" yet, it may very well be rational. As we have mentioned in the introduction, the representation in Theorem 1 is not structured to provide a decision rule in which the agent compares the value of each alternative she may select immediately with the (ex ante) evaluation of the possible choices she could make later on. Consequently, with the representation in Theorem 1 in hand, it is not possible to extricate rational from irrational behavior.

In Theorem 1, property α ensures completeness of $\succcurlyeq^\bullet_x$. Since $\succcurlyeq^*$ extends $\succcurlyeq^\bullet_x$, this, in turn, implies that $\succcurlyeq^\bullet_x$ and $\succcurlyeq^*$ coincide on $\mathcal{Q}(x)$ for each $x \in X$.

Observation 1. *If $\mathbf{c}$ satisfies property α, in Theorem 1, the preorder $\succcurlyeq^\bullet_x$ coincides with*

$\succcurlyeq^*$ *on* $\mathcal{Q}(x)$ *for each* $x \in X$ *and* $\mathbf{c}$ *satisfies property* β.

Given a choice problem $(S, x) \in \mathcal{C}_{sq}(X)$, the objects $\succcurlyeq^\bullet_x$ and $\mathcal{Q}(x)$ determine which alternatives in S the agent would focus; whether she would opt for the status quo alternative; whether she would deem choosable some alternatives other than the status quo; if so, which ones she may choose. In view of this, it would be desirable if, in the representation of the choice correspondence $\mathbf{c}$, $\succcurlyeq^\bullet_x$ and $\mathcal{Q}(x)$ are unique for each $x \in X$, which indeed is the case. We put this formally by the following proposition.

Proposition 1 (Uniqueness). *The binary relation* $\succcurlyeq^*$ *and, for each* $x \in X$, *the set* $\mathcal{Q}(x)$ *and the binary relation* $\succcurlyeq^\bullet_x$ *specified in Theorem 1 are unique.*

4. Characterization of Choice Avoidance

In this subsection, we give a brief account of choice avoidance in the current model. This short exercise will be useful in understanding how the current model deviates from the standard status quo bias models.

Proposition 2. *Let* X *be a nonempty compact metric space and* $\mathbf{c}$ *be an avoidance consistent choice correspondence. Also let, for each* $x \in X$, $\succcurlyeq^\bullet_x$ *and* $\mathcal{Q}(x)$ *be as in Theorem 1. Then, for each* $x \in X$, *the following three conditions are equivalent:*
i. *The preorder* $\succcurlyeq^\bullet_x$ *is complete on its domain,* $\mathcal{Q}(x)$.
ii. *For any* $S \in \mathcal{X}$ *that contains* x, *there does not exist an alternative which is aversive at* (S, x).
iii. *The restriction of* $\mathbf{c}$ *to* $\{(S, x) : x \in S \in \mathcal{X}\}$ *satisfies property* α.

According to Proposition 2, choice avoidance is equivalent to the incompleteness of the preorder $\succcurlyeq^\bullet_x$ for some alternative $x \in X$ and also to violation of property α at some choice problem $(S, x) \in \mathcal{C}_{sq}(X)$, which is the source of choice avoidance in the current model. By contrast, in Masatlioglu and Ok (2009) $\succcurlyeq^\bullet_x$ is complete on $\mathcal{Q}(x)$ for each $x \in X$, which is also equivalent to $\mathbf{c}$ satisfying property α.

Now, as a consequence of Proposition 2, we have the following corollary:

Corollary 1. *Let* X *be a nonempty compact metric space and* $\mathbf{c}$ *be an avoidance consistent choice correspondence. Also, let* $\succcurlyeq^*$ *and, for each* $x \in X$, $\succcurlyeq^\bullet_x$ *and* $\mathcal{Q}(x)$ *be as in Theorem 1. Then, the following three conditions are equivalent:*
i. *For each* $x \in X$, *the preorder* $\succcurlyeq^\bullet_x$ *coincides with* $\succcurlyeq^*$ *on* $\mathcal{Q}(x)$.
ii. $\mathbf{c}$ *is non-averse.*
iii. $\mathbf{c}$ *satisfies property* α.

5. Representation of Non-Averse Choice Correspondences

As we have argued previously, the current model subsumes that of Masatlioglu and Ok (2009). In view of Observation 1 and Corollary 1, if $\mathbf{c}$ satisfies property α, in Theorem 1,

- $\mathbf{c}$ becomes non-averse,
- $\mathbf{c}$ satisfies property β, and
- for each $x \in X$, the preorder $\succcurlyeq_x^\bullet$ coincides with $\succcurlyeq^*$ on $\mathcal{Q}(x)$.

Therefore, upon combining Theorem 1 with Observation 1 and Corollary 1, we obtain the following version of the representation result in Masatlioglu and Ok (2009) (which does not allow for choice avoidance).

Corollary 2 (Masatlioglu and Ok (2009)). *Let X be a nonempty compact metric space. A choice correspondence* $\mathbf{c}$ *on $\mathcal{C}(X)$ satisfies properties α, β,* SQI *and* UHC *if, and only if, there exist a closed-valued self-correspondence $\mathcal{Q}$ on X and a complete preorder $\succcurlyeq^*$ on X such that:*
i. *For each $x \in X$ and $y \in \mathcal{Q}(x)$ we have $y \succcurlyeq^* x$.*
ii. $\mathbf{c}(\cdot, \diamond) = \mathbf{m}(\cdot, \succcurlyeq^*)$.
iii. *For every $(S, x) \in \mathcal{C}_{sq}(X)$,*

$$\mathbf{c}(S, x) = \mathbf{m}(S \cap \mathcal{Q}(x), \succcurlyeq^*).$$

Proof. We omit the "if" part of the proof which is a straightforward exercise. To prove the "only if" part, we first note that when properties α and β hold, axiom CDA is voidly true, for then, there does not exist an aversive alternative. Thus, the hypotheses of the "only if" part of Corollary 2 imply those of Theorem 1. Therefore, there exist a closed-valued self-correspondence $\mathcal{Q}$ and binary relations $\succcurlyeq_x^\bullet$ for each $x \in X$ and $\succcurlyeq^*$ as in Theorem 1, and ***ii.*** in Corollary 2 holds. Moreover, by Observation 1, $\succcurlyeq_x^\bullet$ coincides with $\succcurlyeq^*$ on $\mathcal{Q}(x)$ for each $x \in X$ so that ***i.*** and ***iii.*** follow from Theorem 1. $\square$

6. Relating Status-Quo Bias to Choice Avoidance

In this section, we wish to investigate the extent to which the current model relates choice avoidance and status quo bias to each other. To this end, we need the following definitions.

Definition 2. $\mathbf{c}$ is **standard** if $\mathbf{c}(S, x) = \mathbf{c}(S, \diamond)$ for every $(S, x) \in \mathcal{C}_{sq}(X)$. When $\mathbf{c}$ is not standard we will sometimes say that $\mathbf{c}$ exhibits **weak status quo bias**. If there exists $(S, x) \in \mathcal{C}_{sq}(X)$ with $\mathbf{c}(S, x) = \{x\}$ but $x \notin \mathbf{c}(S, \diamond)$ we say that $\mathbf{c}$ exhibits **status quo bias**.

It is crucial to note that, in the current model, existence of an aversive alternative necessarily implies that $\mathbf{c}$ exhibits status quo bias.

Observation 2. *Let X be a nonempty compact metric space and $\mathbf{c}$ be an avoidance consistent choice correspondence. If $\mathbf{c}$ admits an aversive alternative, then it exhibits status quo bias.*

Proof. By assumption, there is a choice problem $(S,x) \in \mathcal{C}_{sq}(X)$ and an alternative $y \in S \setminus \{x\}$ such that $\mathbf{c}(S,x) = \{x\}$ and $\mathbf{c}(S \setminus \{y\}, x) \neq \{x\}$. Then, CDA ***ii.*** implies $\{y\} = \mathbf{c}(\{x,y\},x)$ and, hence, $y \succ^{\bullet}_{x} x$. Let $\succcurlyeq^{\bullet}_{x}$, $\mathcal{Q}(x)$ and $\succcurlyeq^{*}$ be as in Theorem 1. Then, $\{y\} = \mathbf{c}(\{x,y\},x) = \mathbf{m}(\{x,y\} \cap \mathcal{Q}(x), \succcurlyeq^{\bullet}_{x})$, so $y \succ^{\bullet}_{x} x$. Since $\succcurlyeq^{*}$ extends $\succcurlyeq^{\bullet}_{x}$, it follows that $y \succ^{*} x$. Consequently, $\mathbf{c}(S,\diamond) = \mathbf{m}(S, \succcurlyeq^{\bullet}_{x}) \not\ni x$ as we seek. $\square$

The proposition below gives characterizations of standard choice correspondences and choice correspondences that do not exhibit (weak) status quo bias. This will help us to conclude how the notions of status quo bias and choice avoidance are related in the current model.

Proposition 3. *Let X be a nonempty compact metric space. Consider an avoidance consistent choice correspondence $\mathbf{c}$ and let $\mathcal{Q}$ and $\succcurlyeq^{*}$ be as in Theorem 1.*
i. *$\mathbf{c}$ is standard if, and only if, it is non-averse and*

$$\mathcal{Q}(x) \supseteq \{y \in X : y \succcurlyeq^{*} x\} \text{ for every } x \in X. \tag{1}$$

In particular, if $\mathbf{c}$ is non-averse, it exhibits weak status quo bias if, and only if, there exist $x, y \in X$ such that $y \succcurlyeq^{*} x$ and $y \notin \mathcal{Q}(x)$.
ii. *$\mathbf{c}$ does not exhibit status quo bias if, and only if, it is non-averse and*

$$\mathcal{Q}(x) \supseteq \{y \in X : y \succ^{*} x\} \text{ for every } x \in X. \tag{2}$$

In particular, if $\mathbf{c}$ is non-averse, it exhibits status quo bias if, and only if, there exist $x, y \in X$ such that $y \succ^{*} x$ and $y \notin \mathcal{Q}(x)$.

Violations of (1) and (2) occur if the agent's possible choices at some situations differ under status quo bias from the choices she would make otherwise. According to Proposition 3, existence of such situations implies that the choice behavior of the agent exhibits (weak) status quo bias. In turn, this amounts to saying that, regardless of whether $\mathbf{c}$ admits an aversive alternative, it *may* exhibit (weak) status quo bias. More generally, choice avoidance *is not* a necessary requirement to observe status quo bias, whereas status quo bias *is* to obtain choice avoidance in the current model. This result is in line with our view that choice avoidance is a special form of status quo bias.

7. Related Literature

In this section, we will highlight the differences between the present paper and two related papers. We will start with Dean's (2008) model on choice avoidance.

Dean (2008) approaches the same problem as we do here, albeit in a different way. He proposes two choice models accommodating the observed behavior in the experiments regarding status quo bias, choice avoidance and choice overload. The first model he studies, which will be our focus here, is directly related with the current one. In that model, his approach falls short of providing a proper distinction between the notions of status quo bias and choice avoidance. To make our argument more precise, we will provide a formal treatment of Dean's notion of representation shortly and show that, under some mild regularity conditions, his model reduces to the standard choice model in the absence of avoidance.

The following is Dean's notion of representation.

Definition 3. A choice correspondence $\mathbf{c}$ admits a **type D representation** if there exist a preorder $\succcurlyeq^\bullet$ on X, a complete preorder $\succcurlyeq^*$ on X that extends $\succcurlyeq^\bullet$, and a possibly empty valued correspondence T on $\mathcal{X}$ with $T(S) \subseteq S$ for every $S \in \mathcal{X}$ such that:
i. $\mathbf{c}(\cdot, \diamond) = \mathbf{m}(\cdot, \succcurlyeq^*)$.
ii. $\mathbf{c}(S, x) = \mathbf{m}(S, \succcurlyeq^*)$ for every $(S, x) \in \mathcal{C}(X)$ with $x \in T(S)$.[6] Moreover, for each such (S, x), we also have $y \succcurlyeq^\bullet x$ for every $y \in \mathbf{c}(S, x)$.
iii. For every $(S, x) \in \mathcal{C}(X)$ with $x \in S \setminus T(S)$,

$$\mathbf{c}(S,x) = \begin{cases} \{x\} & \textit{if } \mathbf{m}(S, \succcurlyeq^\bullet) = \varnothing, \\ \mathbf{m}(S, \succcurlyeq^\bullet) & \textit{otherwise.} \end{cases}$$

In this definition, the role of the correspondence T is to identify those instances in which the decision maker disregards her status quo. More specifically, when the decision maker is endowed with $x \in T(S)$, then she selects from S precisely those alternatives that she would choose if she had not have an initial endowment. By comparison, in our model, such instances are only implicitly identified by the representation. (For example, given a choice problem $(S, x) \in \mathcal{C}_{sq}(X)$, if $S \subseteq \mathcal{Q}(x)$ and $\mathbf{m}(S, \succcurlyeq^\bullet_x)$ is nonempty, then $\mathbf{c}(S, x) = \mathbf{c}(S, \diamond)$ as in part ***ii.*** of the definition above.)

It is crucial to note that the correspondence T in a type D representation does not entirely pin down the set of ineffective endowments. In particular, when $\succcurlyeq^\bullet$ is complete, irrespective of T, this model reduces to the standard model of choice driven

[6] It may be useful to note that since $\succcurlyeq^*$ extends $\succcurlyeq^\bullet$, whenever $\mathbf{m}(S, \succcurlyeq^\bullet)$ is non-empty, the sets $\mathbf{m}(S, \succcurlyeq^*)$ and $\mathbf{m}(S, \succcurlyeq^\bullet)$ coincide. Dean's corresponding definition emphasizes this point which gives rise to a minor expositional difference with the above definition.

by maximization of a complete (and status quo independent) preference relation, which is, of course, inconsistent with the notion of status quo bias. Put differently, by contrast to our model, in a type D representation, incompleteness of the preference relation $\succcurlyeq^\bullet$ does not merely explain choice avoidance. Rather, this is a necessary requirement to observe status quo bias as well. Hence, unfortunately, Dean's notion of representation seems to be short of providing a proper analytical distinction between the notions of status quo bias and choice avoidance.

Before formalizing this point, let us note that in some cases where $\succcurlyeq^\bullet$ is severely incomplete, a choice correspondence $\mathbf{c}$ that admits a type D representation may also be representable in the sense of Theorem 1. For instance, if $\succcurlyeq^\bullet$ does not rank any two distinct alternatives, and if $T(S) = \varnothing$ for every $S \in \mathcal{X}$, then the definition of type D representation implies $\mathbf{c}(S,x) = \{x\}$ for every $(S,x) \in \mathcal{C}_{sq}(X)$. This behavior can be mimicked by setting $\mathcal{Q}(x) := \{x\}$ in the realm of Theorem 1. To rule out such exceptional cases, we shall focus on preference relations that satisfy a minimal completeness requirement which we introduce in the definition below. (In what follows, we say that x and y are **$\succcurlyeq^\bullet$-incomparable** if neither $y \succcurlyeq^\bullet x$ nor $x \succcurlyeq^\bullet y$, and write $y \bowtie^\bullet x$.)

Definition 4 (Regularity). Suppose that whenever $x \bowtie^\bullet y$ *either*
i. there exists $z \in X$ such that two of the alternatives in the set $\{x, y, z\}$ are strictly ranked by $\succcurlyeq^\bullet$ while being $\succcurlyeq^\bullet$-incomparable with the third one; *or*
ii. there exists $z \in X \setminus \{x, y\}$ such that either $z \sim^\bullet x$ or $z \sim^\bullet y$ holds.
Then we say that $\succcurlyeq^\bullet$ is **regular**.

Remark. Part ***i.*** of the definition of regularity is the regularity notion of Eliaz and Ok (2006); we also allow for the case ***ii.***

The following observation shows that type D representations do not properly distinguish between the notions of choice avoidance and status quo bias.

Observation 3. *Suppose that* $\mathbf{c}$ *admits a type* D *representation with a regular* $\succcurlyeq^\bullet$*. Then* $\succcurlyeq^\bullet$ *is complete (and* $\mathbf{c}$ *is standard) whenever* $\mathbf{c}$ *is non-averse.*

Gerasímou (2010) is another recent paper related to the current one. His model allows the decision maker in question to make no choice from a set of alternatives whenever she cannot find an alternative that dominates all others. Though in Gerasímou's model "no choice" can be replaced with a default option, it would have no direct impact on the optimal choices of the decision maker. By contrast, as in Masatlioglu and Ok (2005, 2009), in the current model the optimal choices of the decision maker in the presence of a status quo alternative may differ from the choices she would make when there is no status quo. Moreover, Gerasiomu's model allows the decision maker to make (possibly) inferior choices (or satisficing ones). By contrast, except the choice of the default option in specific cases, this is definitely not the case in the current model.

Appendix
A1. Proof of Theorem 1

We omit the "if" part of the proof which is a routine exercise. To prove the "only if" part, let **c** be an avoidance consistent choice correspondence. Following Masatlioglu and Ok (2009), we define the correspondence $\mathcal{Q}$ as, for every $x \in X$,

$$\mathcal{Q}(x) := \{y \in X : y \in \mathbf{c}(\{x, y\}, x)\}.$$

Let us now fix an arbitrary $x \in X$, and define the binary relation $\succcurlyeq_x^\bullet$ on the set $\mathcal{Q}(x)$ as, for every $y, z \in \mathcal{Q}(x)$,

$$y \succcurlyeq_x^\bullet z \quad \text{if, and only if,} \quad y \in \mathbf{c}(\{x, y, z\}, x).$$

Claim 1. *For each $x \in X$, $\mathcal{Q}(x)$ is closed-valued, the binary relation $\succcurlyeq_x^\bullet$ is a preorder on the set $\mathcal{Q}(x)$, and $y \succcurlyeq_x^\bullet x$ for every $y \in \mathcal{Q}(x)$.*

Proof. We first show that, for any $x \in X$, $\mathcal{Q}(x)$ is closed-valued. Fix any $x \in X$, and any sequence (y_n) in $\mathcal{Q}(x)$ with $y_n \to y$ for some $y \in X$. Then, $y_n \in \mathbf{c}(\{y_n, x\}, x)$ for each n. So, by UHC, we find $y \in \mathbf{c}(\{y, x\}, x)$, that is, $y \in \mathcal{Q}(x)$, as we seek.

It should be noted that, by the definitions above, the following four conditions are equivalent: *i.* $y \in \mathcal{Q}(x)$; *ii.* $y \in \mathbf{c}(\{x, y\}, x)$; *iii.* $y \succcurlyeq_x^\bullet y$; *iv.* $y \succcurlyeq_x^\bullet x$. In particular, $\succcurlyeq_x^\bullet$ is reflexive on $\mathcal{Q}(x)$, and we have $y \succcurlyeq_x^\bullet x$ for every $y \in \mathcal{Q}(x)$.

To see that $\succcurlyeq_x^\bullet$ is also transitive, let $w, y, z \in \mathcal{Q}(x)$ be distinct alternatives with

$$y \succcurlyeq_x^\bullet w \quad \text{and} \quad w \succcurlyeq_x^\bullet z. \tag{3}$$

We can assume $z \neq x$, for otherwise it would immediately follow that $y \succcurlyeq_x^\bullet z$, as we just sought. Rewriting (3) yields:

$$y \in \mathbf{c}(\{x, w, y\}, x), \quad \text{and} \tag{4}$$
$$w \in \mathbf{c}(\{x, w, z\}, x). \tag{5}$$

First assume $\mathbf{c}(\{x, w, y, z\}, x) \neq \{x\}$. Then, we must have

$$y \in \mathbf{c}(\{x, w, y, z\}, x). \tag{6}$$

Indeed, when $w \in \mathbf{c}(\{x, w, y, z\}, x)$, (6) follows from (4) and properties α' and β'. Moreover, if $z \in \mathbf{c}(\{x, w, y, z\}, x)$, by (5) and properties α' and β', we also have $w \in \mathbf{c}(\{x, w, y, z\}, x)$. This establishes (6), which, by property α', implies the desired conclusion: $y \in \mathbf{c}(\{x, y, z\}, x)$, that is, $y \succcurlyeq_x^\bullet z$.

Suppose now $\mathbf{c}(\{x, w, y, z\}, x) = \{x\}$. Then, y and x must be distinct, for otherwise $\mathbf{c}(\{x, w, y, z\}, x)$ coincides with $\mathbf{c}(\{x, w, z\}, x)$, and (5) implies $w = y$, a contradiction. But if y and x are distinct, it follows from (4) that z is aversive at the choice problem $(\{x, w, y, z\}, x)$. Hence, by CDA ***i.***, $\mathbf{c}(\{x, w, z\}, x) \subseteq \{x, z\}$. As w and z are distinct, by (5), we then see that $w = x$. Hence, rewriting (5) yields $x \in \mathbf{c}(\{x, z\}, x)$. This contradicts CDA ***ii.*** We have thus shown that $\succcurlyeq_x^\bullet$ *is a preorder on* $\mathcal{Q}(x)$. $\square$

Before we proceed, it should be noted that

$$\textit{for every } (S, x) \in \mathcal{C}_{sq}(X), \textit{ we have } \mathbf{c}(S, x) \subseteq \mathcal{Q}(x),$$

as $y \in \mathbf{c}(S, x) \setminus \mathcal{Q}(x)$ would imply, by property α', that $y \in \mathbf{c}(\{x, y\}, x)$ (i.e., $y \in \mathcal{Q}(x)$), which is an absurdity. In what follows, we will use this fact without further mention. Next, we prove the main conclusion of Theorem 1.

Proof of Theorem 1 *iii.* Let $(S, x) \in \mathcal{C}_{sq}(X)$ with S a finite subset of X. We shall first show that

$$\mathbf{c}(S, x) \neq \{x\} \quad \text{implies} \quad \mathbf{c}(S, x) = \mathbf{m}\left(S \cap \mathcal{Q}(x), \succcurlyeq_x^\bullet\right). \tag{7}$$

Fix an arbitrary $y \in \mathbf{c}(S, x)$. If $\mathbf{c}(S, x) \neq \{x\}$, for every $z \in S$, property α' implies $y \in \mathbf{c}(\{x, y, z\}, x)$. In particular, $y \succcurlyeq_x^\bullet z$ for every $z \in S \cap \mathcal{Q}(x)$; that is, $y \in \mathbf{m}\left(S \cap \mathcal{Q}(x), \succcurlyeq_x^\bullet\right)$. To prove the converse inclusion, let $z \in \mathbf{m}\left(S \cap \mathcal{Q}(x), \succcurlyeq_x^\bullet\right)$. Then, z and y both belong to $\mathbf{c}(\{x, y, z\}, x)$. Hence, by property β', we must have $z \in \mathbf{c}(S, x)$. This proves (7).

The following fact will prove useful in what follows:

$$\mathbf{c}(S \cap \mathcal{Q}(x), x) \neq \{x\} \quad \text{implies} \quad \mathbf{c}(S, x) \neq \{x\}. \tag{8}$$

To prove (8), without loss of generality assume that $S \setminus (S \cap \mathcal{Q}(x))$ is nonempty and denote this set by $\{y_1, ..., y_n\}$. Since y_1 does not belong to $\mathcal{Q}(x)$, from CDA ***ii.*** it follows that y_1 is not aversive at $((S \cap \mathcal{Q}(x)) \cup \{y_1\}, x)$. Hence, the left side of (8) implies $\mathbf{c}\left((S \cap \mathcal{Q}(x)) \cup \{y_1\}, x\right) \neq \{x\}$. By repeating these arguments inductively, we obtain the desired conclusion: $\mathbf{c}\left((S \cap \mathcal{Q}(x)) \cup \{y_1, ..., y_n\}, x\right) \neq \{x\}$.

We now show that: For any $(T,x) \in \mathcal{C}_{sq}(X)$ with T a finite subset of X,

$$T \subseteq \mathcal{Q}(x) \text{ and } \mathbf{m}\left(T, \succcurlyeq^{\bullet}_{x}\right) \neq \varnothing \quad \text{imply} \quad \mathbf{c}(T,x) = \mathbf{m}\left(T, \succcurlyeq^{\bullet}_{x}\right). \tag{9}$$

Let T be as above and note the claim is trivially true if $|T| = 1$. On the other hand, if T is of the form $\{x,y,z\}$, then we obviously have $\mathbf{m}\left(T, \succcurlyeq^{\bullet}_{x}\right) \subseteq \mathbf{c}(T,x)$ by definition of $\succcurlyeq^{\bullet}_{x}$. Moreover, $\mathbf{m}\left(T, \succcurlyeq^{\bullet}_{x}\right) \neq \{x\}$ whenever $|T| \geq 2$. Hence, if $2 \leq |T| \leq 3$, (7) implies $\mathbf{c}(T,x) = \mathbf{m}\left(T, \succcurlyeq^{\bullet}_{x}\right)$, as we seek.

To proceed inductively, let $n \geq 3$ be an integer such that (9) holds whenever $|T| \leq n$. Consider any $(T',x) \in \mathcal{C}_{sq}(X)$ such that $|T'| = n+1$, $T' \subseteq \mathcal{Q}(x)$ and $\mathbf{m}\left(T', \succcurlyeq^{\bullet}_{x}\right) \neq \varnothing$. By (7), it suffices to show that $\mathbf{c}(T',x) \neq \{x\}$. Pick any $w \in \mathbf{m}\left(T', \succcurlyeq^{\bullet}_{x}\right) \setminus \{x\}$, and assume by contradiction that $\mathbf{c}(T',x) = \{x\}$. Then, for any $y \in T' \setminus \{w,x\}$,

$$w \in \mathbf{m}\left(T' \setminus \{y\}, \succcurlyeq^{\bullet}_{x}\right) = \mathbf{c}(T' \setminus \{y\}, x) \neq \{x\},$$

where the equality follows from the induction hypothesis. Thus, any such y is aversive at (T',x). Moreover, as $|T'| \geq 4$, the set $T' \setminus \{w,x\}$ contains two distinct alternatives y_1, y_2. Then, by CDA ***i.***, we must have $\mathbf{c}(T' \setminus \{y_1\}, x) \subseteq \{y_2, x\}$. Since $w \notin \{y_2, x\}$, this is a contradiction which proves (9).

Now, take any $(S,x) \in \mathcal{C}_{sq}(X)$ with S a finite subset of X. Since $\mathbf{c}$ is nonempty valued, from (7), it immediately follows that when $\mathbf{m}\left(S \cap \mathcal{Q}(x), \succcurlyeq^{\bullet}_{x}\right) = \varnothing$, we have $\mathbf{c}(S,x) = \{x\}$.

Suppose now that $\mathbf{m}\left(S \cap \mathcal{Q}(x), \succcurlyeq^{\bullet}_{x}\right) \neq \varnothing$. If $|S \cap \mathcal{Q}(x)| = 1$, we trivially have $\mathbf{c}(S,x) = \{x\} = \mathbf{m}\left(S \cap \mathcal{Q}(x), \succcurlyeq^{\bullet}_{x}\right)$. On the other hand, if $|S \cap \mathcal{Q}(x)| \geq 2$, we have $\mathbf{m}\left(S \cap \mathcal{Q}(x), \succcurlyeq^{\bullet}_{x}\right) \neq \{x\}$, and hence, (7)-(9) imply the desired conclusion:

$$\mathbf{c}(S,x) = \mathbf{m}\left(S \cap \mathcal{Q}(x), \succcurlyeq^{\bullet}_{x}\right). \tag{10}$$

We shall extend this result to arbitrary sets in $\mathcal{X}$. Fix an arbitrary $(S,x) \in \mathcal{C}_{sq}(X)$. Let $\mathbf{c}\left(S,x\right) \neq \{x\}$ and pick any $y \in \mathbf{c}\left(S,x\right)$. By property α', for any $z \in S$, we have $y \in \mathbf{c}\left(\{x,y,z\},x\right)$, hence, $y \succcurlyeq^{\bullet}_{x} z$. In particular, $y \succcurlyeq^{\bullet}_{x} x$, hence, $y \in S \cap \mathcal{Q}(x)$. Therefore, $\mathbf{c}\left(S,x\right) \subseteq \mathbf{m}\left(S \cap \mathcal{Q}(x), \succcurlyeq^{\bullet}_{x}\right)$ whenever $\mathbf{c}\left(S,x\right) \neq \{x\}$. Now, let $y \in \mathbf{m}\left(S \cap \mathcal{Q}(x), \succcurlyeq^{\bullet}_{x}\right)$. Clearly, we can find a sequence (S_n) of finite subsets of S such that $S_n \to S$ and $y \in S_n$ for every n. Then, as $y \in \mathbf{m}\left(S_n \cap \mathcal{Q}(x), \succcurlyeq^{\bullet}_{x}\right)$, (10) implies $y \in \mathbf{c}\left(S_n, x\right)$ for every n. From UHC, it obviously follows that $y \in \mathbf{c}\left(S,x\right)$. This shows that $\mathbf{m}\left(S \cap \mathcal{Q}(x), \succcurlyeq^{\bullet}_{x}\right) \subseteq \mathbf{c}\left(S,x\right)$ for every $(S,x) \in \mathcal{C}_{sq}(X)$. We have thus shown that, for any $(S,x) \in \mathcal{C}_{sq}(X)$,

$$\mathbf{c}(S,x) \neq \{x\} \quad \text{implies} \quad \mathbf{c}(S,x) = \mathbf{m}\left(S \cap \mathcal{Q}(x), \succcurlyeq^{\bullet}_{x}\right). \tag{11}$$

Since $\mathbf{c}$ is a nonempty valued correspondence, from (11), we have $\mathbf{c}(S,x) = \{x\}$ whenever $\mathbf{m}(S \cap \mathcal{Q}(x), \succcurlyeq_x^\bullet) = \varnothing$. It follows that, for any $(S,x) \in \mathcal{C}_{sq}(X)$, $\mathbf{c}(S,x) = \{x\}$ if $\mathbf{m}(S \cap \mathcal{Q}(x), \succcurlyeq_x^\bullet) = \varnothing$, and $\mathbf{c}(S,x) = \mathbf{m}(S \cap \mathcal{Q}(x), \succcurlyeq_x^\bullet)$ otherwise, as we seek. □

Now, for every $y,z \in X$, let $y \succcurlyeq^* z$ if, and only if, $y \in \mathbf{c}(\{y,z\},\diamond)$. Since $\mathbf{c}$ satisfies properties α and β on $\mathcal{C}(X) \setminus \mathcal{C}_{sq}(X)$ and UHC on the compact set X, as is well known, $\succcurlyeq^*$ is an upper semicontinuous, complete preorder on X and $\mathbf{c}(\cdot,\diamond) = \mathbf{m}(\cdot,\succcurlyeq^*)$. This proves part ***ii.*** of Theorem 1.

We conclude with the next claim which, together with Claim 1, proves part ***i.*** of Theorem 1.

Claim 2. *For each $x \in X$, $\succcurlyeq^*$ extends $\succcurlyeq_x^\bullet$ on $\mathcal{Q}(x)$.*

Proof. Let $x \in X$, and take any $y,z \in \mathcal{Q}(x)$ with $y \succcurlyeq_x^\bullet z$ so that $y \in \mathbf{c}(\{x,y,z\},x)$. Then, clearly, for any $T \subseteq \{x,y,z\}$ which contains x, we have $\mathbf{m}(T,\succcurlyeq_x^\bullet) \neq \varnothing$, and hence, $\mathbf{c}(T,x) = \mathbf{m}(T,\succcurlyeq_x^\bullet)$. Moreover, for any such T, whenever $T \neq \{x\}$, the set $\mathbf{m}(T,\succcurlyeq_x^\bullet)$ is not equal to $\{x\}$. By SQI, it thus follows that

$$\mathbf{c}(\{x,y,z\},x) = \mathbf{c}(\{x,y,z\},\diamond). \tag{12}$$

In particular, $y \in \mathbf{c}(\{x,y,z\},\diamond)$ so that $y \succcurlyeq^* z$.

Suppose now $y \succ_x^\bullet z$. Since $z \succcurlyeq_x^\bullet x$, transitivity of $\succcurlyeq_x^\bullet$ implies $\mathbf{m}(\{x,y,z\},\succcurlyeq_x^\bullet) = \{y\}$, and hence, $\mathbf{c}(\{x,y,z\},x) = \{y\}$. From (12), it thus follows that $\{y\} = \mathbf{c}(\{x,y,z\},\diamond)$. This implies $y \succ^* z$, as we seek. □

□

A2. Proof of Observation 1

Let $\succcurlyeq^*$ and $\succcurlyeq_x^\bullet$ be as in Theorem 1 for some $x \in X$ and $\mathbf{c}$ satisfy property α. We first prove that, for each $x \in X$, $\succcurlyeq_x^\bullet$ and $\succcurlyeq^*$ coincide on $\mathcal{Q}(x)$. Since $\succcurlyeq^*$ extends $\succcurlyeq_x^\bullet$, it suffices to show that the preorder $\succcurlyeq_x^\bullet$ is complete on $\mathcal{Q}(x)$. Suppose by contradiction that there exist $y,z \in \mathcal{Q}(x)$ such that neither $y \succcurlyeq_x^\bullet z$ nor $z \succcurlyeq_x^\bullet y$ holds. As $w \succcurlyeq_x^\bullet x$ for every $w \in \mathcal{Q}(x)$, by transitivity of $\succcurlyeq_x^\bullet$, both y and z must be distinct from x, so that $y \succ_x^\bullet x$ and $z \succ_x^\bullet x$. It then follows that the set $\{x,y,z\}$ does not have a $\succcurlyeq_x^\bullet$-maximum element. Hence, Theorem 1 ***iii.*** implies $\mathbf{c}(\{x,y,z\},x) = \{x\}$. But then, by property α, we must have $x \in \mathbf{c}(\{x,y\},x)$, which is a contradiction, for $\mathbf{m}(\{x,y\},\succcurlyeq_x^\bullet) = \{y\} = \mathbf{c}(\{x,y\},x)$, where the last equality follows from Theorem 1 ***iii.***

To show that $\mathbf{c}$ satisfies property β, pick any $x,y,z \in X$ and $(S,x),(T,x) \in \mathcal{C}_{sq}(X)$ such that $T \subseteq S$, $y,z \in \mathbf{c}(T,x)$ and $z \in \mathbf{c}(S,x)$. We shall show that $y \in \mathbf{c}(S,x)$. As $\mathbf{c}$ satisfies property α, from the foregoing result, $\succcurlyeq_x^\bullet$ is complete on $\mathcal{Q}(x)$. Therefore,

$\mathbf{m}(S \cap \mathcal{Q}(x), \succcurlyeq_x^\bullet)$ and $\mathbf{m}(T \cap \mathcal{Q}(x), \succcurlyeq_x^\bullet)$ are nonempty. So, by Theorem 1 ***iii.***, $\mathbf{c}(T,x) = \mathbf{m}(T \cap \mathcal{Q}(x), \succcurlyeq_x^\bullet)$ and $\mathbf{c}(S,x) = \mathbf{m}(S \cap \mathcal{Q}(x), \succcurlyeq_x^\bullet)$. It follows that, for all $x' \in T \cap \mathcal{Q}(x)$, $y \succcurlyeq_x^\bullet x'$ and, for all $y' \in S \cap \mathcal{Q}(x)$, $z \succcurlyeq_x^\bullet y'$. In particular, $y \succcurlyeq_x^\bullet z$ so that, for all $x' \in S \cap \mathcal{Q}(x)$, $y \succcurlyeq_x^\bullet x'$. Hence, $y \in \mathbf{c}(S,x) = \mathbf{m}(S \cap \mathcal{Q}(x), \succcurlyeq_x^\bullet)$. □

A3. Proof of Proposition 1

Let $x \in X$. First, we show that, for all $y \in X$, $y \in \mathcal{Q}(x)$ if, and only if, $y \in \mathbf{c}(\{x,y\},x)$. It is clear that $\mathbf{c}(S,x) \subseteq S \cap \mathcal{Q}(x)$ for each $(S,x) \in \mathcal{C}_{sq}(X)$. In particular, $y \in \mathbf{c}(\{x,y\},x)$ implies $y \in \mathcal{Q}(x)$. Conversely, $y \in \mathcal{Q}(x)$ implies $y \succcurlyeq_x^\bullet x$ by Theorem 1 ***i.*** But then, $y \in \mathbf{m}(\{x,y\}, \succcurlyeq_x^\bullet) = \mathbf{c}(\{x,y\},x)$ by Theorem 1 ***iii***. Hence, we conclude that $y \in \mathcal{Q}(x)$ if, and only if, $y \in \mathbf{c}(\{x,y\},x)$.

Next, we show that, for all $y,z \in X$, $y \succcurlyeq_x^\bullet z$ if, and only if, $y \in \mathbf{c}(\{x,y,z\},x)$. Notice that, for any $y,z \in \mathcal{Q}(x)$ with $y \succcurlyeq_x^\bullet z$ we have $y \in \mathbf{m}(\{x,y,z\}, \succcurlyeq_x^\bullet) \neq \varnothing$, implying that $y \in \mathbf{c}(\{x,y,z\},x) = \mathbf{m}(\{x,y,z\}, \succcurlyeq_x^\bullet)$. Suppose now $y \in \mathbf{c}(\{x,y,z\},x)$. If $\mathbf{c}(\{x,y,z\},x) = \mathbf{m}(\{x,y,z\}, \succcurlyeq_x^\bullet)$ there is nothing to prove: $y \succcurlyeq_x^\bullet z$. Otherwise, $\mathbf{m}(\{x,y,z\}, \succcurlyeq_x^\bullet) = \varnothing$ and $\mathbf{c}(\{x,y,z\},x) = \{x\}$ implying that $y = x$. This is a contradiction, for then $z \in \mathbf{m}(\{x,y,z\}, \succcurlyeq_x^\bullet) \neq \varnothing$. Hence, we conclude that $y \succcurlyeq_x^\bullet z$ if, and only if, $y \in \mathbf{c}(\{x,y,z\},x)$.

We skip the proof that $\succcurlyeq^*$ is unique for it can be proved using similar arguments to the foregoing ones. □

A4. Proof of Proposition 2

Let $\mathbf{c}$ be an avoidance consistent choice correspondence. Also let, for each $x \in X$, $\succcurlyeq_x^\bullet$ and $\mathcal{Q}(x)$ be as in Theorem 1.

If $\succcurlyeq_x^\bullet$ is complete, for any $(S,x) \in \mathcal{C}_{sq}(X)$, there is a $\succcurlyeq_x^\bullet$-maximum element on the set $S \cap \mathcal{Q}(x)$ and, hence, $\mathbf{c}(S,x) = \mathbf{m}(S \cap \mathcal{Q}(x), \succcurlyeq_x^\bullet)$ by Theorem 1 ***iii***. Since $y \succcurlyeq_x^\bullet x$ for every $y \in \mathcal{Q}(x)$, it follows that $\mathbf{c}(S,x) = \{x\}$ if, and only if, $S \cap \mathcal{Q}(x) = \{x\}$. Moreover, in this case, $(S \setminus \{y\}) \cap \mathcal{Q}(x) = \{x\}$ for every $y \in S \setminus \{x\}$, implying that $\mathbf{c}(S \setminus \{y\},x) = \{x\}$. This shows that ***i.*** implies ***ii.***

To deduce property α from ***ii.***, let $(S',x),(T',x) \in \mathcal{C}(X)$ with S',T' finite, $y \in T' \subseteq S'$ and $y \in \mathbf{c}(S',x)$. Without loss of generality, suppose $S' \setminus T'$ is nonempty and denote this set by $\{y_1,...,y_n\}$. If $\mathbf{c}(S',x) \neq \{x\}$, the desired conclusion follows from property α'. We can therefore assume $\mathbf{c}(S',x) = \{x\}$, so that $y = x$. Then, by ***ii.***, y_1 is not aversive at (S',x), and hence, $\mathbf{c}(S' \setminus \{y_1\},x) = \{x\}$. Inductively, it obviously follows that $\mathbf{c}(S' \setminus \{y_1,...,y_n\},x) = \{x\}$, so that $y \in \mathbf{c}(T',x)$. Now, take any $(S,x),(T,x) \in \mathcal{C}(X)$

and any $y \in X$ such that $y \in T \subseteq S$ and $y \in \mathbf{c}(S, x)$. By passing to subsequences if needed, we can find two sequences of finite subsets (S_n) and (T_n) of S and T, respectively, and a sequence (y_n) in X such that, for all n, $y_n \in T_n \subseteq S_n$ and $y_n \in \mathbf{c}(S_n, x)$, (S_n) and (T_n) converge, respectively, to S and T in the Hausdorff metric, and (y_n) converges to $y \in X$. From the foregoing argument, we have $y_n \in \mathbf{c}(T_n, x)$ for all n. It thus follows from UHC that $y \in \mathbf{c}(T, x)$, as we seek.

Finally, suppose that the restriction of $\mathbf{c}$ to $\{(S, x) : x \in S \in \mathcal{X}\}$ satisfies property α. We shall show that, for any $x \in X$, $\succcurlyeq_x^\bullet$ is complete on $\mathcal{Q}(x)$. Fix any $x \in X$ and suppose by contradiction that there exist $y, z \in \mathcal{Q}(x)$ such that $y \bowtie_x^\bullet z$. As $w \succcurlyeq_x^\bullet x$ for every $w \in \mathcal{Q}(x)$, it follows from transitivity of $\succcurlyeq_x^\bullet$ that $y \succ_x^\bullet x$ and $z \succ_x^\bullet x$. Hence, the set $\{x, y, z\}$ does not have a $\succcurlyeq_x^\bullet$-maximum element. So, Theorem 1 ***iii.*** implies $\mathbf{c}(\{x, y, z\}, x) = \{x\}$. But then, by property α, we must have $x \in \mathbf{c}(\{x, y\}, x)$, which is a contradiction, for $\mathbf{m}(\{x, y\}, \succcurlyeq_x^\bullet) = \{y\} = \mathbf{c}(\{x, y\}, x)$, where the last equality follows from Theorem 1 ***iii.*** $\square$

A5. Proof of Proposition 3

Let $\mathbf{c}$ be an avoidance consistent choice correspondence.

i. We will prove the first claim. The second claim follows from the first one.

Suppose $\mathbf{c}$ is non-averse and (1) holds. Since $\mathbf{c}$ is non-averse, by Corollary 1, $\mathbf{c}$ satisfies property α on $\mathcal{C}(X)$. It then follows from Observation 1 that $\succcurlyeq_x^\bullet$ and $\succcurlyeq^*$ coincide on $\mathcal{Q}(x)$. Therefore, we have $\mathbf{c}(S, x) = \mathbf{m}(S \cap \mathcal{Q}(x), \succcurlyeq^*)$ for all $(S, x) \in \mathcal{C}_{sq}(X)$. Moreover, as $\succcurlyeq_x^\bullet$ and $\succcurlyeq^*$ coincide on $\mathcal{Q}(x)$ and (1) holds, Theorem 1 ***i.*** implies that $\mathcal{Q}(x) = \{y \in X : y \succcurlyeq^* x\}$. Thus, for any $y \in \mathbf{c}(S, \diamond)$ and $x \in S$, $y \succcurlyeq^* x$ so that $\mathbf{c}(S, \diamond) \subseteq \mathcal{Q}(x)$. But then, by Theorem 1 ***ii.***, this amounts to saying that $\mathbf{c}(S, \diamond) = \mathbf{c}(S \cap \mathcal{Q}(x), \diamond) = \mathbf{m}(S \cap \mathcal{Q}(x), \succcurlyeq^*)$. So, $\mathbf{c}(S, \diamond) = \mathbf{c}(S, x)$ and, therefore, $\mathbf{c}$ is standard.

Suppose now $\mathbf{c}$ is standard. Assume by contradiction that there exist $x, y \in S \subseteq X$ with $y \neq x$ such that y is aversive at (S, x). Since $\mathbf{c}$ is standard, this implies that $\mathbf{c}(S, \diamond) = \{x\}$ and $\mathbf{c}(S \setminus \{y\}, \diamond) \neq \{x\}$, a contradiction for $\mathbf{c}$ satisfies properties α and β on $\mathcal{C}(X) \setminus \mathcal{C}_{sq}(X)$. Therefore, $\mathbf{c}$ is non-averse.

Finally, we prove the contraposition of the claim that $\mathbf{c}$ being standard implies that (1) holds. Suppose that (1) does not hold. We shall show that $\mathbf{c}$ is not standard. As (1) does not hold, there exist $x, y \in X$ such that $y \succcurlyeq^* x$ and $y \notin \mathcal{Q}(x)$.Then, $y \in \mathbf{c}(\{x, y\}, \diamond)$ but $\mathbf{c}(\{x, y\}, x) = \{x\}$, as we seek.

ii. We will prove the first claim. The second claim follows from the first one.

We begin with the "if" part of the claim. Suppose $\mathbf{c}$ is non-averse and (2) holds. Since $\mathbf{c}$ is non-averse, by Corollary 1, $\mathbf{c}$ satisfies property α on $\mathcal{C}(X)$. It then follows from Observation 1 that $\succcurlyeq_x^\bullet$ and $\succcurlyeq^*$ coincide on $\mathcal{Q}(x)$. Therefore, by Theorem 1 ***iii.***,

$\mathbf{c}(S,x) = \mathbf{m}(S \cap \mathcal{Q}(x), \succcurlyeq^*)$ for all $(S,x) \in \mathcal{C}_{sq}(X)$. Now, suppose by contradiction that $\mathbf{c}$ exhibits status quo bias; i.e., there exists $(S,x) \in \mathcal{C}_{sq}(X)$ such that $\mathbf{c}(S,x) = \{x\}$ and $x \notin \mathbf{c}(S, \diamond)$. So, $\mathbf{m}(S \cap \mathcal{Q}(x), \succcurlyeq^*) = \{x\}$ and there is $y \in \mathbf{c}(S, \diamond)$ with $y \neq x$. As $\mathbf{c}$ satisfies properties α and β on $\mathcal{C}(X) \setminus \mathcal{C}_{sq}(X)$ and $x \notin \mathbf{c}(S, \diamond)$, we have $\mathbf{c}(\{x,y\}, \diamond) = \{y\}$. This implies, by Theorem 1 ***ii.***, that $\mathbf{m}(\{x,y\}, \succcurlyeq^*) = \{y\}$ and, hence, $y \succ^* x$. It follows from (2) that $y \in S \cap \mathcal{Q}(x)$. But then, $y \succ^* x$ and $y \in S \cap \mathcal{Q}(x)$ imply $y \in \mathbf{m}(S \cap \mathcal{Q}(x), \succcurlyeq^*)$, a contradiction for $\mathbf{m}(S \cap \mathcal{Q}(x), \succcurlyeq^*) = \{x\}$ and $y \neq x$.

Now, we prove the contraposition of the "only if" part of the claim. If $\mathbf{c}$ is not non-averse, by Observation 2, $\mathbf{c}$ exhibits status quo bias. So, suppose that (2) does not hold. We shall show that $\mathbf{c}$ exhibits status quo bias. That (2) does not hold implies there exist $x, y \in X$ such that $y \succ^* x$ and $y \notin \mathcal{Q}(x)$.Then, $\mathbf{c}(\{x,y\}, \diamond) = \{y\}$ but $\mathbf{c}(\{x,y\}, x) = \{x\}$, as we seek. $\square$

A6. Proof of Observation 3

Let $\mathbf{c}$ be a non-averse choice correspondence that admits a type D representation with a regular $\succcurlyeq^\bullet$. Suppose by contradiction that there exist $x, y \in X$ such that $x \bowtie^\bullet y$ and assume that part ***i.*** of regularity holds. Then, there exists $z \in X$ such that either $x \succ^\bullet z$ or $z \succ^\bullet x$, and $z \bowtie^\bullet y$. (Notice that the choice of x being $\succcurlyeq^\bullet$-comparable with z is without loss of generality.) If $x \succ^\bullet z$ holds, by the definition of type D representation, we have $\mathbf{c}(\{x,y,z\}, z) = \{z\}$ and $\mathbf{c}(\{x,z\}, z) = \{x\}$, a contradiction for $\mathbf{c}$ is non-averse. If $z \succ^\bullet x$ holds, by the definition of type D representation, we have $\mathbf{c}(\{x,y,z\}, x) = \{x\}$ and $\mathbf{c}(\{x,z\}, x) = \{z\}$, a contradiction for $\mathbf{c}$ is non-averse. Now, assume that part ***ii.*** of regularity holds. Then, suppose, without loss of generality, that there exists $z \in X \setminus \{x,y\}$ such that $z \sim^\bullet x$ and, hence, $z \bowtie^\bullet y$ by transitivity. (For the other case, $z \sim^\bullet y$ and $z \bowtie^\bullet x$, the proof follows by symmetry.) It follows from the definition of type D representation that $\mathbf{c}(\{x,y,z\}, x) = \{x\}$ and $\mathbf{c}(\{x,z\}, x) = \{x,z\}$, a contradiction for $\mathbf{c}$ is non-averse. Thus, $\succcurlyeq^\bullet$ is complete. As $\succcurlyeq^*$ is a complete extension of $\succcurlyeq^\bullet$ and $\succcurlyeq^\bullet$ is complete on X, it is straightforward to see that $\mathbf{c}(S,x) = \mathbf{c}(S, \diamond)$ for all $(S,x) \in \mathcal{C}(X)$. Therefore, $\mathbf{c}$ is standard. $\square$

References

C.J. Anderson, The psychology of doing nothing: forms of decision avoidance result from reason and emotion, *Psychological Bulletin* **129** (2003), 139–167.

J. Apesteguia, M.A. Ballester, A theory of reference-dependent behavior, *Economic Theory* **40** (2009), 427–455.

P. Boatwright, J.C. Nunes, Reducing assortment: an attribute-based approach, *Journal of Marketing* **65** (2001), 50–63.

G. Buturak, Ö. Evren, Rational choice deferral, Stockholm School of Economics and New York University (2010), mimeo.

M. Dean, Status quo bias in large and small choice sets, New York University (2008), mimeo.

K. Eliaz, E.A. Ok, Indifference or indecisiveness? Choice-theoretic foundations of incomplete preferences, *Games and Economic Behavior* **56** (2006), 61–86.

G. Gerasímou, Rational indecisive choice, Cambridge University (2010), mimeo.

S.S. Iyengar, M.R. Lepper, When choice is demotivating: can one desire too much of a good thing?, *Journal of Personality and Social Psychology* **79** (2000), 995–1006.

B.C. Madrian, D.F. Shea,The power of suggestion: inertia in 401(k) participation and savings behavior, *The Quarterly Journal of Economics* **116** (2001), 1149–1187.

Y. Masatlioglu, E.A. Ok, Rational choice with status quo bias, *Journal of Economic Theory* **121** (2005), 1–29.

Y. Masatlioglu, E.A. Ok, A canonical model of choice with initial endowments, New York University (2009), mimeo.

E. Reutskaya, R.M. Hogarth, Satisfaction in choice as a function of the number of alternatives: when "goods satiate," *Psychology & Marketing* **26** (2009), 197–203.

J.S. Sagi, Anchored preference relations, *Journal of Economic Theory* **130** (2006), 283–295.

A.M. Shah, G. Wolford, Buying behavior as a function of parametric variation of number of choices, *Psychological Science* **18** (2007), 369–370.

Paper 3

One-Sided Many-to-Many Matching: An Ordinal Theory of Network Formation

Gökhan Buturak

Abstract

We develop an ordinal theory of network formation, or — equivalently — one-sided many-to-many matching. We provide (i) relations between several solution concepts, (ii) sufficient conditions for their nonemptiness and (iii) an implementation result for the pairwise stable set. We closely follow Echenique and Oviedo (2006) and show that almost all the inclusion results for the solution sets proposed in the two-sided many-to-many matching model carry over to the one-sided many-to-many matching model under the assumption that preferences are (strongly) substitutable. Nonemptiness of solution sets will be established under a weak separability condition.

JEL Classification: C78, D85.
Keywords: One-sided many-to-many matching, network formation, setwise stable set, bargaining set, individually rational core, weak separability, implementation, strategy-proofness, Pareto efficiency.

I am grateful to Jens Josephson, Jordi Massó and Mark Voorneveld for their careful readings of earlier drafts of this paper. For their helpful comments and suggestions, I thank seminar participants at the Microeconomics and Social Choice Lunch Seminar at the Department of Economics, Universitat Autònoma de Barcelona, and at the Economics Lunch Seminar at Stockholm School of Economics. All the remaining errors and deficiencies are mine. Financial support from Jan Wallander and Tom Hedelius Research Foundation is gratefully acknowledged.

1. Introduction

In this paper, we develop an ordinal theory of network formation, which can also be interpreted as one-sided many-to-many matching. We combine two streams of literature, matching theory and network formation, which are closely tied to each other. On the one hand, we approach the network formation problem from a matching perspective so as to incorporate the preferences of each agent in the network formation process over the subsets of the others. In the existing models of network formation, individuals are typically assumed to be identical.[1] So, agents have preferences over positions at possible networks that can be formed rather than over the sets of agents that they would like to link with. More specifically, although the general models of network formation do allow players to have *arbitrary* preferences over their sets of neighbors (and even over the other connections), the only consideration for an agent linking with another is the marginal benefit pertaining to the link itself, which is typically assumed to be constant.[2] As a result, in these models, identities of the agents do not play any role in the linking decisions of the agents, which is not the case in this paper.

On the other hand, we generalize the two-sided many-to-many matching problem (2MM) to the one-sided many-to-many matching problem (1MM).[3,4] When viewed as an extension of 2MM, the current paper may seem to be a straightforward generalization of the existing matching models. However, it is quite typical in matching theory that an alteration of a specific matching problem may not yield the same results as the matching problem at hand. For instance, in their seminal paper, Gale and Shapley (1962)

[1] An exception is Persitz (2008).

[2] For instance, in the symmetric connections model of Jackson and Wolinsky (1996) and the variations of it, the value and cost of forming a link are assumed to be constant. In the spatial connections models (Johnson and Gilles, 2000; Jackson and Rogers, 2005), linking costs depend linearly on the distance between agents. However, the intrinsic value of linking with another agent is still constant. For more on this, see Persitz (2008).

[3] By a two-sided matching problem we mean that there are two sides of a market each containing a group of agents (such as workers, students, etc), where each agent on one side of the market wants to match with some other on the other side of the market. Gale and Shapley's (1962) marriage problem is an example of two-sided matching problems. By a one-sided matching problem, we mean that the market consists of just one group of agents, each agent having the option of matching with those other than herself. The roommates problem in Gale and Shapley (1962) is an example of a one-sided matching problem. Many-to-many matching refers to the matching problems where each agent is allowed to match with more than one agent (either from the same side or from the other side of the market) other than herself.

[4] Indeed, 2MM is a special case of 1MM. Every 2MM matching problem can be mapped to a 1MM matching problem by keeping each agent's preferences over subsets of the other side unchanged and letting the agents find any subset containing agents of the own side unacceptable. Moreover, any solution concept can be defined in such a way that the set of solutions to the original 2MM matching problem coincides with those of the transformed 1MM matching problem. Hence, a result that holds for each 1MM matching problem also holds for each 2MM matching problem. However, the converse is not true: we will, for instance, often need stronger assumptions to ensure nonemptiness of solution sets than what is the case in the 2MM setting.

generalize the (two-sided) marriage problem to the (one-sided) roommates problem and show that some results do not extend from marriage to roommate problems. In particular, a core stable matching may fail to exist. Alkan (1988) shows with an example that the results for the marriage market no longer hold in the three-sided one-to-one matching markets. Similarly, many properties of the marriage problem (except the existence of stable matchings) do not generalize to the college admissions problem (see Roth, 1985).[5,6] In view of these examples, the current paper investigates to what extent the results for the existing matching models carry over to the one-sided many-to-many matching problem.

1.1 Preview of Results

We provide relations between several solution concepts, including the bargaining set and the setwise stable set, two core-like concepts. We closely follow Echenique and Oviedo (2006) and show that almost all the inclusion results for the solution sets proposed in the two-sided many-to-many matching model carry over to the one-sided many-to-many matching model under the assumption that preferences are (strongly) substitutable. However, in the present framework, the nonemptiness of the solution sets fails to hold under the preference restrictions in Echenique and Oviedo (2006). This is not surprising in view of the fact that the roommates problem is a particular case of 1MM. In the roommates problem, preferences of the agents necessarily satisfy strong substitutability. However, it is very well-known that the roommates problem may not have a stable matching (see, e.g., Example 3 in Gale and Shapley, 1962), which also emerges as a problem in 1MM. To make it concrete, consider three agents a, b and c where agent a prefers b to c to any other subset of agents not containing herself, agent b prefers c to a to any other subset of agents not containing herself, and agent c prefers a to b to any other subset of agents not containing herself. Now, consider the matching μ_0 defined as $\mu_0(a) = \{b\}$, $\mu_0(c) = \varnothing$. Then, b and c together form a pairwise block. The new matching, call it μ_1, then becomes $\mu_1(a) = \varnothing$ and $\mu_1(b) = \{c\}$. But then, a and c form a pairwise block to μ_1. The new matching, call it μ_2, becomes $\mu_2(a) = \{c\}$ and $\mu_2(b) = \varnothing$. Finally, a and b form a pairwise block to μ_2 and we come up with the initial matching μ_0. So, the pairwise blocks form a cycle of matchings and a stable matching cannot be obtained.

[5] The college admissions problem is a two-sided many-to-one matching problem, where each student has preferences over colleges and can be enrolled in only one college, and each college j has preferences over students and has a quota of q_j students.

[6] Needless to say, existence of stable matchings is not straightforward in the general two-sided many-to-one and many-to-many matching problems.

Furthermore, we show with an example that even if the roommates problem is avoided by introducing a top-quota restriction (i.e., the most preferred subset of each agent contains at least two other agents) strong substitutability is not sufficient for the existence of a stable matching. In that example, the preference profile is not individually rational, i.e., for at least one preference ordering, there is a set that is preferred to the empty set but a strict subset of it is preferred to itself. We may further restrict preferences with a size monotonicity condition as follows: an agent's preference ordering is said to satisfy size monotonicity if a better set contains at least as many agents as a worse set. It is tempting to say that size monotonicity would fix the problem by ensuring individually rational preferences for each agent. However, even if we restrict the preferences with this size monotonicity condition together with strong substitutability and top-quota restriction, a counterexample can easily be provided where there is no stable matching.

Consequently, we show that for the solution sets proposed in Echenique and Oviedo (2006) to be nonempty in our setup, the preferences should, in addition to strong substitutability, satisfy a property called *desirability*, where the two together are equivalent to a *weak separability* condition.[7] With these two restrictions, we show that the solution sets we consider shrink to a singleton set and coincide with the individually rational core. We also provide a game in strategic form and show that, under weak separability, the unique stable matching coincides with the pairwise Nash equilibrium outcomes of this game. We also conclude that our results under weak separability hold for general many-to-many matching problems.

1.2 Related Literature

Following the stable marriage problem of Gale and Shapley (1962), many extensions of the two-sided one-to-one matching were proposed and studied. The stable roommates problem (Gale and Shapley, 1962) was the one-sided version of the two-sided one-to-one matching. Hedonic games were introduced by Drèze and Greenberg (1980) as a generalization of the stable roommates problem, which allows a player's utility to depend on the composition of the members of and the strategies available to her coalition. Banerjee et al. (2001) and Bogomolnaia and Jackson (2002) modify the game described in Drèze and Greenberg (1980) by discarding the strategies available to each coalition and purely focusing on the hedonic aspect of coalition formation. Pure hedonic games seem to be similar to the model described in this paper, in that they also incorporate the preferences of the agents who are to be matched over the partitions they may be

[7] Throughout the paper, when we mention (weak) separability, we mean (weak) separability without a quota.

in. But the two differ since the concern of pure hedonic games is group formation, whereas the concern of the current paper is network formation. Technically, forming a group requires the unanimous consent of the group in hedonic games, whereas bilateral consent is sufficient to form links in our framework.

Previous literature on 2MM provided results for the pairwise stable set. Roth (1984a) proved that, under substitutability, the pairwise stable set is nonempty, and there are firm- and worker-optimal pairwise stable matchings. Using different definitions of supremum and infimum, Blair (1988) and Alkan (1999) provided results on the lattice structure of pairwise stable matchings in 2MM in different preference domains. In a recent work on 2MM, Echenique and Oviedo (2006) provided results for more demanding stability notions. The reason why they study other stability notions than the usual core —e.g. the setwise stable set (due to Roth, 1984a), the individually rational core (due to Sotomayor, 1999) and the bargaining set— is twofold: (i) pairwise stability does not allow for more general coalitions; (ii) the core as a stability concept is problematic in many-to-many matching problems as it may be empty; if not, the core matchings may not be individually rational or pairwise stable. Using fixpoint methods as in Adachi (2000), they show that the solution sets they study are nonempty in the substitutable and strongly substitutable preference domains. As we have already mentioned, the matching problem we introduce in the current paper is the one-sided version of the matching model in Echenique and Oviedo (2006). Indeed, this seemingly mild modification dramatically changes the results: the sufficient conditions they propose for the nonemptiness of the solution sets are no longer sufficient to guarantee the same results in our framework.

Klaus and Walzl (2009) also study the same matching problem as in Echenique and Oviedo (2006) with contracts. Besides the stability notions appearing in Echenique and Oviedo (2006), they introduce a weaker version of the setwise stable set, obtain similar inclusion results between different stability sets, and show that they are nonempty under similar restrictions on preferences as in Echenique and Oviedo (2006). Lahiri (2004) studies a one-sided many-to-many matching problem with unilateral consent to link formation, which he calls the 'directed network problem with quotas' and shows that the core is nonempty. The model we present in this paper is of bilateral consent to link formation and, thus, differs from that in Lahiri (2004). Konishi and Ünver (2006) also study 2MM and show that pairwise stability is equivalent to credible group stability when one side has q-responsive preferences and the other side has q-categorywise-responsive preferences.[8] Further, they show under the same preference restriction that the set

[8]We use *q-responsiveness* shorthand for responsiveness with quotas, and *q-categorywise-responsiveness* shorthand for categorywise-responsiveness with quotas.

of matchings resulting from the coalition-proof Nash equilibria of the strategic-form matching game coincides with the credible group-stable set. Finally, they conclude that their results are also valid for general non-bipartite many-to-many matchings. Yet, our paper differs from theirs as they study (two-sided) many-to-many matching problem under different preference restrictions than those we employ here.

The paper is organized as follows: In the next section, we introduce the model and give some definitions. In Section 3, we provide examples demonstrating that the pairwise stable set may be empty under several preference restrictions. In Section 4, we give some inclusion results between several solution sets in different preference domains. In Section 5, we give some results regarding the properties of the solution sets, such as strategy-proofness and Pareto efficiency, and provide an implementation of the pairwise stable set by a strategic form game. We conclude in Section 6. The proofs that are omitted in the text are deferred to the Appendix.

2. The Model

Let I be the non-empty finite set of agents. The preferences of agent $i \in I$ over $2^{I \setminus \{i\}}$ are summarized by a **preference relation** $P(i)$ on $2^{I \setminus \{i\}}$.[9] We refer to $R(i)$ as the weak preference relation associated with $P(i)$, the collection of preference relations $(P(i))_{i \in I}$ indexed by the set of agents I as a **preference profile**, and $(P(i))_{i \in S}$ as the preference profile for S. We denote generic preference profiles $(P(i))_{i \in I}$ and $(P'(i))_{i \in I}$ as P and P', respectively; and generic preference profiles for $S \subseteq I$, $(P(i))_{i \in S}$ and $(P'(i))_{i \in S}$, as $P(S)$ and $P'(S)$, respectively. The set of preference relations of agent $i \in I$ is denoted as $\mathscr{P}(i)$, the set of preference profiles as $\mathscr{P}$, and the set of preference profiles for $S \subseteq I$ as $\mathscr{P}(S)$. When we write definitions for a given preference profile $P \in \mathscr{P}$, we occasionally use the phrase 'at $P \in \mathscr{P}$.' A **matching** μ is a mapping from the set I into the set of all subsets of I such that for all $i, j \in I$ with $i \neq j$, $\mu(i) \in 2^{I \setminus \{i\}}, \mu(j) \in 2^{I \setminus \{j\}}$, and $j \in \mu(i)$ if, and only if, $i \in \mu(j)$. We denote by $\mathscr{M}$ the set of all matchings for the set of agents I. The complete network on I is defined as $g^I := \{\{m, n\} \subseteq I : m \in I, n \in I, m \neq n\}$. Thus, each matching $\mu \in \mathscr{M}$ induces an undirected network $g^\mu := \{\{m, n\} \subseteq I : m \in \mu(n), n \in \mu(m)\} \subseteq g^I$.

Given a preference relation $P(i)$ of agent $i \in I$, a set of partners S preferred by i to the empty set is called **acceptable**, and all the partners who are acceptable are called the **acceptance set** of agent i, denoted by A_i. Let the choice correspondence $Ch(S, P(i))$ denote agent i's most preferred subset of $S \subseteq I \setminus \{i\}$ according to i's preference relation

[9] A preference relation P on a set X is a complete, antisymmetric and transitive binary relation on the set X. R is referred as the weak preference relation associated with P. Formally, for all $x, y \in X$, $x \; R \; y$ if, and only if, $x = y$ or $x \; P \; y$.

$P(i)$. Formally, $Ch(S, P(i)) \subseteq S$ for all $S \subseteq I$, and $Ch(S, P(i))\, P(i)\, S'$ for all $S' \subseteq S$, provided that $S' \neq Ch(S, P(i))$. The preference relation $P(i)$ of agent $i \in I$ is **individually rational** if, for each $S \in 2^{I \setminus \{i\}}$ with $S\, P(i)\, \varnothing$, $Ch(S, P(i)) = S$. A preference profile $P \in \mathscr{P}$ is said to be **individually rational** if $P(i)$ is individually rational for all $i \in I$.

We start by introducing some properties of the choice correspondence:

Observation 1. *Let $i \in I$ be an agent and let the sets of agents $S, A, B \in 2^{I \setminus \{i\}}$ be such that $A, B \subseteq S$. Given a preference profile $P \in \mathscr{P}$,*
i. *$B \subseteq A$ implies $Ch\,(A, P(i))\, R(i)\, Ch(B, P(i))$.*
Moreover, if $Ch\,(A, P(i)) \subseteq B \subseteq A$, then
ii. *$Ch\,(A, P(i)) = Ch(B, P(i))$.*
In particular,
iii. *$Ch(Ch(S, P(i)), P(i)) = Ch(S, P(i))$.*

To see if the observation holds, observe that the first claim holds by definition, as the choice correspondence $Ch(\cdot, P(i))$ selects the most preferred subset of its argument and $B \subseteq A$. Moreover, if $Ch(A, P(i)) \subseteq B \subseteq A$, then $Ch(A, P(i))\, P(i)\, C$ for all $C \in 2^A \supseteq 2^B$ with $C \neq Ch(A, P(i))$, so in particular $Ch(A, P(i)) = Ch(B, P(i))$. Now, the third claim follows from the second claim by setting $A = S$ and $B = Ch(S, P(i))$.

There are two requirements for a solution to be stable; namely, individual rationality (no individual wants to severe her links assigned by a particular solution) and no blocking requirement (roughly, no two individuals or more have incentives to deviate and form other links than those assigned by a particular solution). There are various ways of forming blocks. We give the formal definitions of several forms of blocks as well as the associated stability notions below. Before that, we start with the definition of individual rationality for a matching, which can also be interpreted as immunity to blocks of single agents.

Definition. A matching μ is **individually rational** at $P \in \mathscr{P}$ if $\mu(i)\, R(i)\, S, \forall S \subseteq \mu(i)$, for all $i \in I$; i.e., $\mu(i) = Ch(\mu(i), P(i))$ for all $i \in I$.

The very basic form of a block is the pairwise block, where two individuals want to match[10] with each other, possibly severing some of their matchings. Hence, a pairwise stable matching is immune to pairwise blocks.

Definition. Let $i, j \in I$ and let μ be a matching. The pair (i, j) is a **pairwise block** of μ at $P \in \mathscr{P}$ if $i \notin \mu(j)$, $i \in Ch(\mu(j) \cup \{i\}, P(j))$, and $j \in Ch(\mu(i) \cup \{j\}, P(i))$.

Definition. A matching μ is **pairwise stable** if it is individually rational and there is no pairwise block of μ.

[10] Throughout the paper, we use 'to match' and 'to link' interchangeably.

The set of pairwise stable matchings at $P \in \mathscr{P}$ is denoted as $S(P)$.

A generalization of pairwise block is the block* where the deviation is now extended to an agent $i \in I$ and a set of agents $D \subseteq I \setminus \{i\}$, i linking with all agents in D. So, we say that a matching is stable* if it is immune to any block*.

Definition. A pair $(D, i) \in 2^I \times I$ with $i \notin D \neq \varnothing$ is said to **block*** a matching μ at $P \in \mathscr{P}$ if $D \cap \mu(i) = \varnothing$, $D \subseteq Ch(\mu(i) \cup D, P(i))$, and $i \in Ch(\mu(j) \cup \{i\}, P(j))$ for all $j \in D$.

Definition. A matching μ is **stable*** at $P \in \mathscr{P}$ if it is individually rational and there is no pair $(D, i) \in 2^{I \setminus \{i\}} \times I$ that blocks* μ.

The set of stable* matchings at $P \in \mathscr{P}$ is denoted as $S^*(P)$.

We now come to the (weak) core, which is a very well-known stability concept. A block to a matching is a set of agents severing all their matchings and linking with each other. A core matching is then immune to this kind of deviations. It is a demanding stability concept and a problematic one in the framework of many-to-many matchings. More specifically, a core matching may not be individually rational or may not be immune to pairwise deviations (see, e.g., Echenique and Oviedo (2006) for a detailed discussion of the core matchings in many-to-many matching problems).

Definition. A **block** of a matching μ at $P \in \mathscr{P}$ is a pair $\langle S', \mu' \rangle$ with $\varnothing \neq S' \subseteq I$ and $\mu' \in \mathscr{M}$ such that $\mu'(i) \subseteq S'$, $\mu'(i)\, R(i)\, \mu(i)$ for all $i \in S'$, and $\mu'(i)\, P(i)\, \mu(i)$ for some $i \in S'$.

Definition. A matching μ is a **core** matching if there are no blocks of μ.

The set of core matchings at $P \in \mathscr{P}$ is denoted as $C(P)$.

Since the core is problematic, we will instead be interested in the individually rational core, where a core matching is now required to be individually rational and immune to individually rational blocks. It was first (implicitly) defined in Sotomayor (1999).

Definition (Individually rational block). A block $\langle S', \mu' \rangle$ of a matching μ at $P \in \mathscr{P}$ is individually rational if $\mu'(i) = Ch(\mu'(i), P(i))$ for all $i \in S'$.

Definition. A matching μ is an **individually-rational core** matching if it is individually rational and has no individually-rational blocks.

The set of individually rational core matchings at $P \in \mathscr{P}$ is denoted as $IRC(P)$.

Next, we introduce the bargaining set. A matching from the bargaining set is immune to deviations (called objections) that are free from further deviations (called counter-objections). Intuitively, the notion of bargaining set adds some sort of credibility and farsightedness into coalitional deviations. It was introduced in Aumann and Maschler (1964) for cooperative games. There are different definitions for the bargaining set in the literature depending on from which group of agents the counterobjection arises. For instance, in Zhou (1994), counterobjections are formed outside the objecting group. Klijn and Massó (2003) study Zhou's bargaining set for the one-to-one matching model. To contrast the results, we follow the definition in Echenique and Oviedo (2006).

Definition. Let μ be a matching. The pair $\langle S', \mu' \rangle$ with $\varnothing \neq S' \subseteq I$ and $\mu' \in \mathscr{M}$ is called an **objection** to μ at $P \in \mathscr{P}$ if $\mu'(i) \setminus \mu(i) \subseteq S'$ and $\mu'(i)\, P(i)\, \mu(i)$ for all $i \in S'$. Let $\langle S', \mu' \rangle$ be an objection to μ. A **counterobjection** to μ is an objection $\langle S'', \mu'' \rangle$ to μ' such that $\varnothing \neq S'' \subseteq S'$.

Definition. A matching μ is in the **bargaining set** if μ is individually rational and there are no objections without counterobjections to μ.

The bargaining set at $P \in \mathscr{P}$ is denoted as $B(P)$.

Setwise stability was proposed by Roth (1984a) as an alternative to core stability. The notion of a setwise block relaxes the requirement for the usual block that all the deviating agents should match within the deviating group by allowing the deviating agents to keep some of their existing links. Yet, the members of the deviating coalition are restricted to form new links only within themselves.

Definition. A **setwise block** to a matching μ at $P \in \mathscr{P}$ is a pair $\langle S', \mu' \rangle$ with $\varnothing \neq S' \subseteq I$ and $\mu' \in \mathscr{M}$ such that $\mu'(i) \setminus \mu(i) \subseteq S'$, $\mu'(i)\, P(i)\, \mu(i)$ and $\mu'(i) = Ch(\mu'(i), P(i))$ for all $i \in S'$.

Definition. A matching μ is in the **setwise-stable set** if μ is individually rational and there are no setwise blocks to μ.

The set of setwise-stable matchings is denoted as $SW(P)$.

Fixpoint methods to show the nonemptiness of solution sets in matching theory were introduced in Adachi (2000) for the stable marriage problem. Echenique and Oviedo (2004) also use a similar method to characterize the core many-to-one matchings as fixed points of a map T on the set of pre-matchings $\mathscr{V}$, which is a superset of $\mathscr{M}$. They also use the same method in Echenique and Oviedo (2006). In the latter paper, the map T facilitates to find a matching within the setwise stable set. They show that the fixpoint set of the mapping T is equivalent to the setwise stable set under some restrictions on preferences, namely strong substitutability; and, the fixpoint set is a nonempty lattice, hence, iterating the map T results in a matching in the setwise stable set. In our model, the construction of a lattice structure is problematic for we lose the conflict/coincidence of interests property that the two-sided matching problems possess under certain preference restrictions. Therefore, we are not able to use this machinery to prove nonemptiness results. However, when establishing inclusion results, the fixpoint set will prove useful.

So, as a starting point, we define pre-matchings:

Definition. A **pre-matching** ν *is* a mapping $\nu : I \to 2^I$ with $i \notin \nu(i)$ for all $i \in I$.

The set of pre-matchings is denoted as $\mathscr{V}$.

Observe that a pre-matching ν is a matching if ν is such that, for any $i, j \in I$ with $i \neq j$, $j \in \nu(i)$ if, and only if, $i \in \nu(j)$. In other words, a pre-matching does not necessarily have the bilateral feature of a matching.

Let ν be a pre-matching and i be an agent. Given a preference profile $P \in \mathscr{P}$, define the set $U(i, \nu)$ as

$$U(i, \nu) := \{j \in I \setminus \{i\} : i \in Ch(\nu(j) \cup \{i\}, P(j))\}.$$

The set $U(i, \nu)$ is the set of agents who are willing to link with i, possibly after severing some of the links that were assigned by ν.

We define $T : \mathscr{V} \to \mathscr{V}$ as, for all $i \in I$, $(T\nu)(i) = Ch(U(i, \nu), P(i))$. We interpret $(T\nu)(i)$ as the agent i's most desirable set of agents who want to link with agent i. Now, we define the fixpoint set.

Definition. The **fixpoint set** associated with T, $\mathscr{E}(P)$, is defined as,

$$\mathscr{E}(P) = \{\nu \in \mathscr{V} : \nu = T\nu\}.$$

3. Examples

In this section, we provide two examples where the preferences of the agents satisfy strong substitutability but no stable matching exists. By strong substitutability, we mean that if an agent is chosen when she is added to a set, then she should also be chosen when she is added to a worse set. By substitutability, we mean that if an agent is chosen when she is added to a set, then she should also be chosen when she is added to a subset of it.

The first example shows that even if the roommates problem is avoided by introducing a top-quota restriction (i.e., the most preferred subset of each agent has cardinality more than one), strong substitutability is not sufficient for the existence of a stable matching. Indeed, the preference profile in this example satisfies strong substitutability, but is not individually rational.

Example 1. Let $I = \{1, 2, 3, 4, 5\}$ be a set of agents with the following preferences:

$P(1)$	$2,3$	3	$3,4$	$2,4$	2	4	$\varnothing$
$P(2)$	$3,4$	4	$4,5$	$3,5$	3	5	$\varnothing$
$P(3)$	$4,5$	5	$1,5$	$1,4$	4	1	$\varnothing$
$P(4)$	$1,5$	1	$1,2$	$2,5$	5	2	$\varnothing$
$P(5)$	$1,2$	2	$2,3$	$1,3$	1	3	$\varnothing$

We omit set braces for convenience and write, for instance, k instead of $\{k\}$. To make the notation clear, let us consider the preference ordering of agent 1: she prefers $\{2, 3\}$ to $\{3\}$, $\{3\}$ to $\{3, 4\}$, $\{3, 4\}$ to $\{2, 4\}$, $\{2, 4\}$ to $\{2\}$, $\{2\}$ to $\{4\}$, and $\{4\}$ to remaining unmatched (to the empty set). We do not list the unacceptable subsets for each agent. It is easy to check that the preference profile satisfies strong substitutability (and hence substitutability). Although each agent has top-quota more than one (so that we rule out the roommates problem), it is easy to check that there is no stable matching. The source of instability of the matchings in this example seems to be the preference profile not being individually rational. □

It is tempting to argue that a size monotonicity condition may solve the problem in the above example. However, the second example shows that even if we restrict the preferences with a size monotonicity condition (i.e., a better set contains at least as many agents as from a worse set) along with strong substitutability and top-quota restriction, there may still be no stable matching.

Example 2. Let $I = \{1, 2, 3, 4, 5\}$ be a set of agents with the preferences as follows:

$P(1)$	$2,3,4$	$2,3,5$	$2,4,5$	$3,4,5$	$2,3$	$2,4$	$2,5$	$3,4$	$3,5$	$4,5$	2	3	4	5	$\varnothing$
$P(2)$	$3,4,5$	$1,3,4$	$1,3,5$	$1,4,5$	$3,4$	$3,5$	$1,3$	$4,5$	$1,4$	$1,5$	3	4	5	1	$\varnothing$
$P(3)$	$1,4,5$	$2,4,5$	$1,2,4$	$1,2,5$	$4,5$	$1,4$	$2,4$	$1,5$	$2,5$	$1,2$	4	5	1	2	$\varnothing$
$P(4)$	$1,2,5$	$1,3,5$	$2,3,5$	$1,2,3$	$1,5$	$2,5$	$3,5$	$1,2$	$1,3$	$2,3$	5	1	2	3	$\varnothing$
$P(5)$	$1,2,3$	$1,2,4$	$1,3,4$	$2,3,4$	$1,2$	$1,3$	$1,4$	$2,3$	$2,4$	$3,4$	1	2	3	4	$\varnothing$

The preference profile satisfies strong substitutability (and hence substitutability). Moreover, each agent has top-quota more than one and, hence, we guarantee that the matching problem is not a roommates problem. Also, the preferences satisfy size monotonicity and, hence, we ensure that the preference of each agent is individually rational. We also note that the preferences in Example 2 are q-responsive with quota 3.[11] However, it is easy, yet arduous, to check that there is no stable matching. □

[11] An agent's preference ordering is q-responsive if it is q-separable and, for any two acceptable sets (with size strictly less than the quota restriction) that only differ in one agent, she prefers the set containing the more preferred one (and the sets which have size more than the quota are deemed as 'not acceptable').

4. Relationships among Solution Sets and Nonemptiness

We present inclusion relations between different solution sets under four preference domains, which are general, substitutable, strongly substitutable, and weakly separable preference domains. We show that none of the preference restrictions that we study other than weak separability is sufficient for these solution sets to be nonempty.

4.1 Results under General Preference Domain

In this subsection, we will summarize the inclusion results in the general preference domain (i.e., without any restrictions on the preferences). The results confirm those in Echenique and Oviedo (2006). Moreover, we provide further results in the general preference domain. In particular, we show the equivalence of the fixpoint set and the stable* set, which is shown under substitutability in Echenique and Oviedo (2006). We also provide the relations among the setwise stable set, the bargaining set and the pairwise stable set in the general preference domain.

Theorem 1. *For all $P \in \mathscr{P}$, $SW(P) \subseteq S^*(P) = \mathscr{E}(P) \subseteq S(P)$ and $SW(P) \subseteq B(P) \subseteq S(P)$.*

4.2 Results under Substitutability

Substitutability was introduced in Kelso and Crawford (1982), which has then been extensively used in the matching literature. It ensures that if an agent is chosen when she is added to a set, then she is also chosen when she is added to a smaller set. Below, we give a formal definition of substitutability.

Definition. An agent i's preference relation $P(i)$ satisfies **substitutability** if, for all $S, S' \subseteq I \setminus \{i\}$ with $S \subseteq S'$ and all $j \in I \setminus \{i\}$,

$$j \in Ch(S' \cup \{j\}, P(i)) \Rightarrow j \in Ch(S \cup \{j\}, P(i)).$$

A preference profile $P = (P(i))_{i \in I}$ satisfies substitutability if, for each agent $i \in I$, $P(i)$ satisfies substitutability.

After defining substitutability, we present some properties of the choice correspondence under substitutability, which also appear in Blair (1988).

Observation 2. *If P satisfies substitutability, then for all $i \in I$ and all $A, B \in 2^{I \setminus \{i\}}$, $Ch(A \cup B, P(i)) \cap A \subseteq Ch(A, P(i))$.*

To see that Observation 2 holds, for each $j \in Ch(A \cup B, P(i)) \cap A$, we have $A \cup B \cup \{j\} = A \cup B \supseteq A \cup \{j\} = A$. By substitutability, $j \in Ch(A \cup \{j\}, P(i)) = Ch(A, P(i))$.

Observation 3. *If P satisfies substitutability, then, for all $i \in I$ and all $A, B \in 2^{I \setminus \{i\}}$, $Ch(A \cup B, P(i)) = Ch(Ch(A, P(i)) \cup B, P(i))$.*

To see if the above observation holds, by Observation 2, we write $Ch\,(A \cup B, P(i)) \cap A \subseteq Ch(A, P(i))$ and $Ch\,(A \cup B, P(i)) \cap B \subseteq Ch\,(B, P(i)) \subseteq B$. Thus, $Ch\,(A \cup B, P(i)) \subseteq Ch\,(A, P(i)) \cup B \subseteq A \cup B$, and using Observation 1, we conclude that $Ch(A \cup B, P(i)) = Ch\,(Ch(A, P(i)) \cup B, P(i))$.

In Theorem 2, we provide further inclusion results.

Theorem 2. *If $P \in \mathscr{E}$ is substitutable, then $SW(P) \subseteq B(P) \subseteq S^*(P) = S(P) = \mathscr{E}(P)$.*

After presenting Theorem 2, a follow-up question would be whether the weak inclusions in Theorem 2 bind.[12] In Example 3 in Konishi and Ünver (2006), the agents have substitutable preferences, but there is a unique pairwise stable matching that is not in the setwise stable set. (Indeed, the setwise stable set is empty.) Therefore, in Theorem 2, the inclusion between the setwise stable set and the pairwise stable set does not necessarily bind. Moreover, in that example, there is also a core matching that is individually rational. Hence, it is also in the individually rational core; yet, it is not pairwise stable. So, we conclude that the pairwise stable set is not comparable with the individually rational core under substitutable preferences. However, we note that the preferences in Example 3 in Konishi and Ünver (2006) are not strongly substitutable.

As we have demonstrated, the preference profiles in both examples in Section 2 are substitutable (this can easily be checked), but there is no pairwise stable matching. So, the pairwise stable set may be empty in the substitutable preference domain, contrary to the results in Echenique and Oviedo (2006).

4.3 Results under Strong Substitutability

Strong substitutability is introduced in Echenique and Oviedo (2006) as a strengthening of substitutability in order to ensure the setwise stable set to be nonempty. Strong substitutability is simply a restriction on preferences imposing that if an agent is chosen when added to a set, then she should also be chosen when added to a worse set. Below, we give a formal definition of strong substitutability.

Definition. An agent i's preference relation $P(i)$ satisfies **strong substitutability** if, for all $S, S' \subseteq I \setminus \{i\}$ with $S' \, P(i) \, S$ and all $j \in I \setminus \{i\}$,

$$j \in Ch(S' \cup \{j\}, P(i)) \Rightarrow j \in Ch(S \cup \{j\}, P(i)).$$

[12] In the current context, the verb *bind* refers to the equivalence of two sets in case there is weak inclusion between them: $A \subseteq B$ binds if, and only if, $A = B$.

A preference profile $P = (P(i))_{i \in I}$ satisfies strong substitutability if, for each agent $i \in I$, $P(i)$ satisfies strong substitutability.

Strong substitutability is closely related to the notion of *weak separability* that we shall introduce in Subsection 4.4. But the two are not the same since weak separability, as opposed to strong substitutability, also requires that if an agent is deemed acceptable, she should be chosen when added to any set of agents. Therefore, weak separability is a stronger restriction on preferences than strong substitutability.

As previously mentioned, strong substitutability implies substitutability.

Proposition 1. *If P satisfies strong substitutability, then it satisfies substitutability.*

Below, Theorem 3 summarizes the inclusion results under strong substitutability.

Theorem 3. *If P is strongly substitutable, then $SW(P) = B(P) = S(P) = S^*(P) = \mathscr{E}(P) \subseteq IRC(P)$.*

Recall that substitutability is not sufficient to establish comparability between the pairwise stable set and the individually rational core. Under strong substitutability, however, Theorem 3 puts that the pairwise stable set is a subset of the individually rational core. Now, we ask whether the inclusion between the pairwise stable set and the individually rational core in Theorem 3 binds, which is indeed not the case: In Example 2 in Section 2, it is possible to find matchings that are in the individually rational core, but the pairwise stable set is empty. To see that the individually rational core is nonempty in Example 2, consider, for instance, the matching μ defined as $\mu(1) = \{3, 4\}$, $\mu(2) = \{3, 4, 5\}$, $\mu(3) = \{1, 2, 5\}$, $\mu(4) = \{1, 2, 5\}$ and $\mu(5) = \{2, 3, 4\}$. Notice that the matching μ is individually rational. Moreover, any individually rational block $\langle S', \mu' \rangle$ to μ should contain at least four agents. But then, checking all those sets containing four or more agents, we see that all better sets of matchings for each agent $i \in S'$ are at the expense of at least one other agent $j \in S' \setminus \{i\}$ getting a worse set of matchings than before. Hence, μ is in the individually rational core though the pairwise stable set is empty. Therefore, the inclusion between the pairwise stable set and the individually rational core in Theorem 3 does not necessarily bind.

Accordingly, we cannot guarantee the existence of a pairwise stable matching even under strongly substitutable preferences, in contrast to the results in Echenique and Oviedo (2006). To maintain the nonemptiness of the solution sets we have presented, we require further restrictions on preferences, which we shall tackle in the following subsection.

4.4 Results under Weak Separability

So far, we have shown that the setwise stable set, the bargaining set, the fixpoint set, the pairwise stable set, and the stable* set are equivalent under strong substitutability and are included in the individually rational core. But strong substitutability does not guarantee the nonemptiness of these solution sets. As already noted, we need further restrictions on preferences to obtain nonemptiness. We propose the property called *desirability*. This property is similar to *q-desirability* in Martínez et al. (2000). In our version, we do not define it for stable matchings and we no longer require quota restrictions on preferences. In words, desirability says that an agent who is deemed acceptable should be chosen when she is added to any set. We will shortly show that strong substitutability and desirability together are equivalent to a weak separability condition that is akin to separability, a well-known property in the matching and social choice literature.

We start with the formal definitions of desirability and weak separability.

Definition. An agent i's preference relation $P(i)$ satisfies **desirability** if, for all $S \subseteq I$ and all $j \in I$,

$$i \neq j, \{j\}\, P(i)\, \varnothing \Rightarrow j \in Ch(S \cup \{j\}, P(i)).$$

A preference profile $P = (P(i))_{i \in I}$ satisfies desirability if $P(i)$ satisfies desirability for all $i \in I$.

Definition. An agent i's preference relation $P(i)$ satisfies **weak separability** if, for all $j \in I \setminus \{i\}$ and $S \subseteq I \setminus \{i\}$,

$$j \in Ch(S \cup \{j\}, P(i)) \Leftrightarrow \{j\}\, P(i)\, \varnothing.$$

A preference profile $P = (P(i))_{i \in I}$ satisfies **weak separability** if $P(i)$ satisfies weak separability for all $i \in I$.

We note that this definition is weaker than the separability condition (without quota restrictions) that is used in the social choice and matching literature. In the latter, for $P(i)$ to be separable, it is required that, for all $j \in I \setminus \{i\}$ and $S \subseteq I \setminus \{i\}$, $S \cup \{j\}\, P(i)\, S \setminus \{j\}$ if, and only if, $\{j\}\, P(i)\, \varnothing$. It is an easy exercise to show that separability implies weak separability.[13] However, the converse is not true.[14]

Also, it is important to note that weak separability is definitely a stronger restriction on preferences than strong substitutability. As said, we have the following equivalence result.

Proposition 2. *$P \in \mathscr{P}$ satisfies strong substitutability and desirability if, and only if, it satisfies weak separability.*

[13] The proof is available upon request.

[14] Let $I = \{1, 2, 3, 4, 5\}$ and agent 1's preference ordering be, from best to worst, $\{2, 5\}$, $\{2\}$, $\{2, 4\}$, $\{2, 3\}$, $\{5\}$, $\varnothing$. Then, it is straightforward to check that this preference ordering satisfies weak separability. Yet, it does not satisfy separability.

Theorem 4 summarizes the findings in this subsection.

Theorem 4. *If P satisfies weak separability, all solution concepts in Theorem 3 coincide and select the matching μ with*

$$\forall i, j \in I: \quad j \in \mu(i) \Leftrightarrow \{i\}\, P(j)\, \varnothing,\ \{j\}\, P(i)\, \varnothing.$$

Under weakly separable preferences, agents are judged by their individual merits, rather than needing to refer to their merits within different coalitions. Hence, an intuitively appealing matching is exactly the one where two people are matched if, and only if, both find the other person acceptable, i.e., the matching μ indicated in Theorem 4.

A short discussion is in order to explain why we have not used alternative preference restrictions to weak separability. A related but logically distinct restriction on preferences is q-separability. To clarify the two notions, we note that q-separability refers to separability with a quota restriction where the separability condition is imposed only on those subsets which do not exceed the quota. Sotomayor (1999) shows that the setwise stable set may be empty with q-separable preferences (see Example 3 in Sotomayor, 1999).

A stronger restriction on preferences than q-separability extensively used in matching literature is q-responsiveness. In words, an agent's preferences satisfy q-responsiveness if it is q-separable and for any two sets of agents that differ in only one agent, the agent prefers the subset containing the most-preferred agent. Example 2 in Section 2 of the current paper shows that there may be no stable matchings even with q-responsive and strongly substitutable preferences. By contrast, Theorem 4 shows that, in the weakly separable preference domain, the stable set, the individually rational core and the setwise stable set coincide, shrink to a singleton set and, hence, are nonempty.

5. Implementation and Some Properties of the Solution Sets

In this section, we will study some properties of the stable matchings under weak separability and investigate whether the stable matching rule satisfies some appealing properties. We will also give an implementation result.

5.1 Rural Hospital Theorem

Roth (1984b) notes that the vacancies at rural hospitals are usually not filled. In view of this fact, he shows that, in a two-sided many-to-one matching market where agents have q-responsive preferences on both sides, those who are not matched in a stable matching remain unmatched in all other stable matchings. This property is called as the *rural hospital theorem*,[15] a well-known result in the matching literature.

Martínez et al. (2000) show that the set of single agents may not be the same in pairwise stable matchings in many-to-one matching problems under substitutability (see Example 5 in Martínez et al., 2000). They also provide sufficient conditions under which the rural hospital theorem holds. The preference profile in Example 5 in Martínez et al. (2000) does not satisfy strong substitutability. We here modify it to show that the rural hospital theorem does not hold under strongly substitutable preferences either.

Example 3. Let $I = \{a, b, c, d, 1, 2, 3\}$ be a set of agents with the following preferences:

$P(a)$	2	1	$\varnothing$									
$P(b)$	2	1	$\varnothing$									
$P(c)$	1	2	$\varnothing$									
$P(d)$	1	3	$\varnothing$									
$P(1)$	a, b	a, c	a, d	b, c, d	b, d	c, d	a	b, c	b	c	d	$\varnothing$
$P(2)$	c	a, c	b, c	a, b	a	b	$\varnothing$					
$P(3)$	d	$\varnothing$										

It can easily be checked that this preference profile satisfies strong substitutability. Two pairwise stable matchings are $\mu_N(1) = \{a, b\}$, $\mu_N(2) = \{c\}$, $\mu_N(3) = \{d\}$, and $\mu_L(1) = \{c, d\}$, $\mu_L(2) = \{a, b\}$, $\mu_L(3) = \varnothing$. Note that $\mu_L(3) = \varnothing$ while $\mu_N(3) = \{d\}$. $\square$

We already have shown in Theorem 4 that, in addition to strong substitutability, it suffices to have the desirability property for the pairwise stable set to be unique and nonempty. As a corollary to Theorem 4, we also obtain the rural hospital theorem since the stable set shrinks to a singleton set and, hence, there is no other matching with which to compare.

5.2 Strategy-proofness

Centralized procedures in matching situations such as matching firms and workers, students and colleges, hospitals and interns, even colleagues to colleagues are common in practice. By selecting certain matchings, the goal of the centralized authority in mediating the matching process is to achieve socially desirable matching outcomes with certain aspects such as stability, Pareto efficiency, individual rationality and strategy-proofness, given the preferences of agents. We refer to such procedures of selecting certain matchings as matching rules. A **matching rule** is then a mapping from the set

[15] Also known as the *lonely wolf theorem.*

of preference profiles of the agents to the set of all possible matchings which prescribes the socially desirable matching outcomes. Formally, a matching rule is a correspondence $\varphi : \mathscr{P} \to \mathscr{M}$. Two examples of matching rules are the stable rule, which selects the set of stable matchings, and the Pareto rule, which selects the set of Pareto efficient matchings. Throughout this subsection, we will be interested in the stable rule.

Individuals usually have private information about their own preferences. When they face with such centralized procedures that are designed to reflect the preferences of all individuals as a social outcome, it may be naïve to expect them to truthfully reveal their preferences if it is in their own interest to misrepresent them. However, individuals may have the incentive to truthfully reveal their preferences under some rules. Such matching rules are called strategy-proof. Formally, a matching rule φ is **strategy-proof** if

$$\varphi(P)(i)\,P(i)\,\varphi((P(j))_{I\setminus\{i\}}, P'(i))(i),\ \forall P \in \mathscr{P}, \forall P'(i) \in \mathscr{P}(i), \forall i \in I.$$

Roth (1982) shows that the stable rule is not strategy-proof for the marriage problem. Alcalde and Barberà (1994) strengthen the result in Roth (1982). Preferences are necessarily strongly substitutable in the marriage problem. Hence, the result in Roth (1982) confirms that, in our framework, the stable rule is not strategy-proof if the preferences are restricted to the strongly substitutable domain. We will, however, shortly present that it is not so in the weakly separable domain. Before that, we need some more notation.

Let $\mathcal{S} : \mathscr{P} \to \mathscr{M}$ be the **stable rule** which selects the set of pairwise stable matchings for each preference profile $P \in \mathscr{P}$. Notice that, by definition, $\mathcal{S}$ is individually rational. Let $\mathcal{S}^s$ denote the stable rule restricted to the domain of weakly separable preference profiles. Note that, due to Theorem 4, $\mathcal{S}^s$ is a single-valued matching rule.

Now, we are prepared to state our result on strategy-proofness.

Proposition 3. *$\mathcal{S}^s$ is a strategy-proof matching rule for the one-sided many-to-many matching problem.*

Even if a matching rule is strategy-proof, a group of agents may benefit by misrepresenting their preferences. A matching rule which does not allow for such misrepresentation of preferences is called coalitionally strategy-proof. Formally, a matching rule φ is **coalitionally strategy-proof** if for all $D \subseteq I$, all $P \in \mathscr{P}$, and all $P'(D) \in \mathscr{P}(D)$, there is an agent $i \in D$ such that

$$\varphi(P)(i)\,P(i)\,\varphi((P(I \setminus D)), P'(D))(i).$$

In the matching literature, the results for coalitional strategy-proofness are restricted. Dubins and Freedman (1981) show that the deferred-acceptance algorithm in which workers make offers is coalitionally strategy-proof for the workers in the college admissions problem. Martínez et al. (2004) show that the workers-optimal stable rule is coalitionally strategy-proof for the workers for the many-to-one matching model in which firms have substitutable and q-separable preferences over the subsets of workers. Besides these limited results, we also know from Roth (1982) that, in our framework, the stable rule is not coalitionally strategy-proof if the preferences are restricted to the strongly substitutable domain. Below, we show with an example that this is also the case in the weakly separable domain.

Example 4. Let $I = \{a, b, 1, 2\}$ be a set of agents with the following preferences:

$P(a)$	1	$1,2$	$\varnothing$
$P(b)$	2	$1,2$	$\varnothing$
$P(1)$	b	a,b	$\varnothing$
$P(2)$	a	a,b	$\varnothing$

It can easily be checked that this preference profile satisfies weak separability. The unique pairwise stable matching is the empty matching; i.e., $\mu(i) = \varnothing$ for all $i \in I$. Therefore, the stable rule returns the empty matching. However, instead of their true preferences, the agents may report their preferences as follows.

$P'(a)$	$1,2$	1	2	$\varnothing$
$P'(b)$	$1,2$	2	1	$\varnothing$
$P'(1)$	a,b	b	a	$\varnothing$
$P'(2)$	a,b	a	b	$\varnothing$

This preference profile also satisfies weak separability. Now, the unique pairwise stable matching is $\mu'(a) = \{1,2\}$, $\mu'(b) = \{1,2\}$, $\mu'(1) = \{a,b\}$, $\mu'(2) = \{a,b\}$. It is obvious that $\mu'(i)\, P(i)\, \mu(i)$ for all $i \in I$. Therefore, the stable rule $\mathcal{S}^s$ is not coalitionally strategy-proof. $\square$

5.3 Pareto Efficiency

Pareto efficiency of a matching rule requires that there is no matching which is weakly better than any of the matchings from the selection of the matching rule considered for all agents and is strictly better for at least one agent.

Formally, a matching $\mu \in \mathscr{M}$ is **Pareto efficient** at P if there is no other matching $\mu' \in \mathscr{M}$ such that $\mu'(i)\, R(i)\, \mu(i)$ for all $i \in I$ and $\mu'(j)\, P(j)\, \mu(j)$ for at least one $j \in I$. The set of all Pareto efficient matchings at P is denoted as $\mathcal{P}(P)$. A matching rule φ is Pareto efficient if $\varphi(P) \subseteq \mathcal{P}(P)$ for all $P \in \mathscr{P}$. We will also say that a matching $\mu' \in \mathscr{M}$

(weakly) **Pareto dominates** $\mu \in \mathscr{M}$ if $\mu'(i)\, R(i)\, \mu(i)$ for all $i \in I$ and $\mu'(j)\, P(j)\, \mu(j)$ for at least one $j \in I$.

In our framework, as it is evident from Example 4 in Section 5.2, the stable rule $\mathcal{S}$ is not Pareto efficient even if the preference domain is restricted with weak separability. Recall that, in that example, the only stable matching is the empty matching. The matching μ' defined in that example Pareto dominates the empty matching, though it is not individually rational. If we restrict attention to the individually rational matchings, however, we obtain the following Pareto efficiency result for $\mathcal{S}$.

Proposition 4. *If $P \in \mathscr{P}$ is strongly substitutable, then there is no individually rational matching $\mu \in \mathscr{M}$ that Pareto dominates a matching assigned by $\mathcal{S}(P)$.*

5.4 Implementation

We propose a noncooperative game in strategic form that implements the pairwise stable set. The set of pairwise Nash equilibrium outcomes of the game we propose is nonempty and coincides with the pairwise stable set in the weakly separable preference domain.[16] We do not use the Nash equilibrium concept since the empty network (or in our framework, the empty matching) is always a Nash equilibrium outcome of the network formation games with bilateral consent (see, e.g., Calvó-Armengol and İlkılıç, 2009), even in the case when there is at least a pair of agents who prefer matching with each other to remaining alone. Nash equilibrium is problematic in that sense. As a result, we will use the pairwise Nash equilibrium concept. We start with the description of the game:

- The set of players is defined as $I = \{1, \ldots, n\}$, where $n \geq 2$.
- The preferences of agent $i \in I$ over $2^{I \setminus \{i\}}$ is summarized by a (strict) preference relation $P(i)$ on $2^{I \setminus \{i\}}$. Write $P = (P(i))_{i \in I}$.
- Each player $i \in I$ simultaneously proposes a set of partners $\xi_i \in 2^{I \setminus \{i\}}$, so that player i's action space is $\Xi_i = 2^{I \setminus \{i\}}$. Write $\Xi = \times_{i \in I} \Xi_i$ and, for any $J \subseteq I$, $\Xi_J = \times_{i \in J} \Xi_i$.
- Given a strategy profile $\xi = (\xi_j)_{j \in I}$, the outcome of the game is the matching μ^ξ formed by mutual consent:

$$\forall i, j \in I \ : \quad j \in \mu^\xi(i) \Leftrightarrow j \in \xi_i,\, i \in \xi_j.$$

[16]The implementation result given here shows that the set of strong Nash *equilibrium outcomes* equals the setwise stable set. It should, however, be noted that several distinct strong Nash equilibrium strategy profiles yield the same *outcome*. Therefore, the mapping between the equilibrium strategy profiles and the setwise stable set may not be a bijection, whereas a stronger implementation result would establish such a bijection. To strengthen the implementation result in the latter sense, the preferences of agents can be altered in such a way that each agent prefers not to be rejected to being rejected, which is equivalent to the introduction of a slight cost of rejection in a cardinal model of network formation, so that in equilibrium, no agent proposes to the agents who reject his proposal.

- players' preferences $(\succsim_i)_{i \in I}$ over outcomes are defined in terms of P:

$$\forall i \in I, \forall \xi, \zeta \in \Xi : \quad \xi \succsim_i \zeta \Leftrightarrow \mu^{\xi}(i)\, R(i)\, \mu^{\zeta}(i).$$

Given the preference profile P, this description defines a normal form game, $\Gamma(P)$.

Now, we will define solution concepts. We start with the definition of Nash equilibrium for the game $\Gamma(P)$.

A strategy profile $\xi^* \in \Xi$ is a **Nash equilibrium** (NE) of $\Gamma(P)$ if, for all $i \subseteq I$ and all $\xi_i \in \Xi_i$,

$$\left(\xi_i^* \cap \{\tilde{\imath} \in I \setminus \{i\} : i \in \xi_{\tilde{\imath}}^*\}\right)\; R(i)\; \left(\xi_i \cap \{\tilde{\imath} \in I \setminus \{i\} : i \in \xi_{\tilde{\imath}}^*\}\right).$$

The set of all the NE of $\Gamma(P)$ is denoted as $NE(P)$.

As seen from the definition, Nash equilibrium concept considers only individual deviations. As the empty matching is always a NE of the game $\Gamma(P)$ and NE is immune to only unilateral deviations, we could instead use an equilibrium concept that allows for any coalitional deviations such as strong Nash equilibrium, whose definition is as follows.

A strategy profile $\xi^* \in \Xi$ is a **strong Nash equilibrium** (SNE) of $\Gamma(P)$ if, for all $J \subseteq I$ and all $\xi_J \in \Xi_J$, there exists an agent $i \in J$ such that

$$\left(\xi_i^* \cap \{\tilde{\imath} \in I \setminus \{i\} : i \in \xi_{\tilde{\imath}}^*\}\right)\; R(i)\; \left(\xi_i \cap \left(\{\tilde{\imath} \in I \setminus J : i \in \xi_{\tilde{\imath}}^*\} \cup \{\tilde{\imath} \in J \setminus \{i\} : i \in \xi_{\tilde{\imath}}\}\right)\right).$$

The set of all the SNE of $\Gamma(P)$ is denoted as $SNE(P)$.

It is straightforward to verify that the outcome of any SNE is setwise stable. However, the converse is not true: A setwise stable matching may not be the outcome of any SNE. For instance, for the preference profile given in Example 4 in Section 5.2, there is no SNE. However, in that example, the empty matching is the unique setwise stable matching. This is due to the fact that SNE is a very demanding equilibrium concept and usually an SNE does not exist. Also, in SNE, the credibility of the deviations is not taken into account.

Accordingly, we will work with a weaker equilibrium notion. In order to avoid the empty network as an equilibrium outcome in network games, Calvó-Armengol and İlkılıç (2009) proposed the pairwise Nash equilibrium concept, which defines equilibrium only in terms of individual and pairwise deviations. The definition of a pairwise Nash equilibrium lies between NE and SNE. In a strategic environment with many players, it is reasonable to consider only unilateral and pairwise deviations as coordination and credibility of deviations by large coalitions are hard to attain. We adopt this equilibrium concept to our framework as follows.

A strategy profile $\xi^* \in \Xi$ is a **pairwise Nash equilibrium** (PNE) of $\Gamma(P)$ if, for all $J \subseteq I$ with $|J| \leq 2$ and all $\xi_J \in \Xi_J$, there exists an agent $i \in J$ such that

$$(\xi^*_i \cap \{\tilde{\imath} \in I \setminus \{i\} : i \in \xi^*_{\tilde{\imath}}\}) \; R(i) \; (\xi_i \cap (\{\tilde{\imath} \in I \setminus J : i \in \xi^*_{\tilde{\imath}}\} \cup \{\tilde{\imath} \in J \setminus \{i\} : i \in \xi_{\tilde{\imath}}\})) \, .$$

The set of all the PNE of $\Gamma(P)$ is denoted as $PNE(P)$.

We close this section with the following result.

Proposition 5. *Let $P \in \mathscr{P}$ be a weakly separable preference profile. Then, the set of PNE of $\Gamma(P)$ is nonempty. Furthermore, a matching $\mu \in \mathscr{M}$ is the outcome of a PNE of $\Gamma(P)$ if, and only if, $\mu \in S(P)$.*

6. Concluding Remarks

We have studied an ordinal theory of network formation with a matching approach. We provided sufficient conditions for the nonemptiness of several solution sets including the setwise stable set and the bargaining set, two core-like concepts, and proposed a non-cooperative game in strategic form that implements the pairwise stable set. We showed that almost all the inclusion results for the solution sets proposed in the two-sided many-to-many matching model of Echenique and Oviedo (2006) carry over to the one-sided many-to-many matching model under the assumption that preferences are (strongly) substitutable. However, in our framework, we have shown that the nonemptiness of the solution sets may not hold in the strongly substitutable preference domain. In order for the solution sets proposed in Echenique and Oviedo (2006) to be nonempty in our setup, the preferences should, in addition to strong substitutability, satisfy a property called desirability, where the two together are equivalent to a weak separability condition, which is weaker than separability, a well-known property in the social choice and matching literature. Furthermore, we provided some properties of the set of stable matchings.

In addition to the results we provided, we note that any restriction on the set of agents that a player may match with does not change our results. In particular, any arbitrary partitioning that defines the feasible matchings between the agents has no effect on our results since, for any agent, those from the same partition would easily be deemed as unacceptable. Therefore, our results also hold for general many-to-many matching problems.[17]

[17] By general many-to-many matching problems, we mean matching problems which are with bilateral consent. So, this precludes partitioning problems such as Drèze and Greenberg's (1980) hedonic games, where forming a group may require the consent of more than two agents, or Alkan's (1988) three-sided husband-wife-child problem, where forming a threesome apparently requires the consent of three agents.

We summarize these observations below.

Observation 4. *Let* $(I, (I_i, P(i))_{i \in I})$ *be a general many-to-many matching problem, where* I *is the set of agents and, for each* $i \in I$*,* I_i *is the list of feasible partners and* $P(i)$ *is the preference ordering over the subsets of* I_i*. If* $(P(i))_{i \in I}$ *is strongly substitutable, then,*

i. $SW(P) = B(P) = S(P) = S^*(P) = \mathscr{E}(P) \subseteq IRC(P)$.

Furthermore, if the preferences additionally satisfy desirability, then

ii. $S(P) = IRC(P)$*, and is a nonempty, singleton set,*

iii. *the set of PNE of* $\Gamma(P)$ *as described in Section 5.5 is nonempty,*

iv. *a matching* $\mu \in \mathscr{M}$ *is the outcome of a PNE of* $\Gamma(P)$ *if, and only if,* $\mu \in S(P)$*, and*

v. $\mathcal{S}^s$*, the stable rule (restricted to the weakly separable domain), is strategy-proof.*

A further research question would be whether there is a larger preference domain than the weakly separable one where nonemptiness of the solution sets continues to hold. Also, it would be interesting to investigate whether we would obtain an implementation result with an equilibrium concept such as coalition-proof Nash equilibrium, which is similar in nature to strong Nash equilibrium, yet, is based on the credibility and stability of coalitional deviations. We leave these questions as future research.

Appendix

A1. Proof of Theorem 1

The proof of Theorem 1 follows from Lemmata 1–7 below.

Lemma 1. *For all* $P \in \mathscr{P}$*, if* $\nu \in \mathscr{E}(P)$*, then* ν *is a matching and is individually rational.*

Proof. Let $\nu \in \mathscr{E}(P)$. Pick any $i \in I$ with $\nu(i) \neq \varnothing$ (otherwise there is nothing to prove), and any $j \in \nu(i)$.

First, we show that ν is individually rational. The condition $\nu \in \mathscr{E}(P)$ implies that

$$j \in \nu(i) = (T\nu)(i) = Ch(U(i, \nu), P(i)). \tag{1}$$

The definition of $U(i, \nu)$ implies

$$i \in Ch(\nu(j) \cup \{i\}, P(j)) \, R(j) \, \nu(j). \tag{2}$$

Apply $Ch(\cdot, P(i))$ to both sides of (1) and use Observation 1 in order to obtain

$$Ch(\nu(i), P(i)) = Ch(Ch(U(i, \nu), P(i)), P(i)) = Ch(U(i, \nu), P(i)). \tag{3}$$

Combining (1) and (3), we obtain

$$\nu(i) = Ch(v(i), P(i)), \tag{4}$$

i.e., ν is individually rational.

Next, we show that ν is a matching; i.e., for all $i, j \in I$ with $i \neq j$, $i \in \nu(j)$ if, and only if, $j \in \nu(i)$. We shall show the 'if' part. The 'only if' part holds by symmetry.

Observe that $j \in \nu(i)$ and (4) imply

$$Ch(\nu(i), P(i)) = Ch(\nu(i) \cup \{j\}, P(i)). \tag{5}$$

Hence, (4) and (5) together with $j \in \nu(i)$ imply that

$$i \in U(j, \nu). \tag{6}$$

Since $\nu \in \mathscr{E}(P)$,

$$\nu(j) = (T\nu)(j) = Ch(U(j, \nu), P(j)),$$

so that

$$\nu(j) \subseteq U(j, \nu). \tag{7}$$

(6) and (7) yield

$$Ch\,(\nu(j) \cup \{i\}, P(j)) \subseteq \nu(j) \cup \{i\} \subseteq U(j, \nu). \tag{8}$$

Using (8) and Observation 1, we write

$$\nu(j) = Ch(U(j, \nu), P(j))\, R(j)\, Ch(\nu(j) \cup \{i\}, P(j)). \tag{9}$$

(2), (9), and the antisymmetry of $P(j)$ imply that $i \in \nu(j)$. $\square$

Lemma 2. *For all $P \in \mathscr{P}$, $S^*(P) \subseteq S(P)$.*

Proof. If $\mu \in S^*(P)$, it is individually rational and there is no block* of μ. By definition, a pairwise block is a block*, so there is no pairwise block of μ. Therefore, $\mu \in S(P)$. $\square$

Lemma 3. *For all $P \in \mathscr{P}$, $\mathscr{E}(P) \subseteq S^*(P)$.*

Proof. Let $\mu \in \mathscr{E}(P)$. By Lemma 1, μ is an individually rational matching. We shall show that there is no block* of μ. Suppose for a contradiction that $(D, i) \in 2^{I \setminus \{i\}} \times I$ blocks* μ. In particular, $D \neq \varnothing$,

$$D \cap \mu(i) = \varnothing \text{ and } D \subseteq Ch(\mu(i) \cup D, P(i)). \tag{10}$$

Also, for all $j \in D$, $i \in Ch(\mu(j) \cup \{i\}, P(j))$; i.e., $j \in U(i, \mu)$. So, $D \subseteq U(i, \mu)$. As $\mu \in \mathscr{E}(P)$, $\mu(i) = Ch(U(i, \mu), P(i))$, so that $\mu(i) \subseteq U(i, \mu)$. Hence, $\mu(i) \cup D \subseteq U(i, \mu)$. By Observation 1,

$$\mu(i) = Ch(U(i, \mu), P(i))\, R(i)\, Ch(\mu(i) \cup D, P(i)).$$

As $\mu(i) \subseteq \mu(i) \cup D$ is preferred to the best subset of $\mu(i) \cup D$, $Ch(\mu(i) \cup D, P(i))$, it follows that $\mu(i) = Ch(\mu(i) \cup D, P(i))$. This contradicts (10). $\square$

Lemma 4. *For all $P \in \mathscr{P}$, $S^*(P) \subseteq \mathscr{E}(P)$.*

Proof. Let $\mu \in S^*(P)$ and suppose that $\mu \notin \mathscr{E}(P)$. So, $\mu \neq T\mu$ and, hence, there is $i \in I$ such that

$$\mu(i) \neq (T\mu)(i) = Ch\,(U(i, \mu), P(i)) \subseteq U(i, \mu). \tag{11}$$

Since $\mu \in S^*(P)$, μ is individually rational.

Let $D = Ch\,(U(i, \mu), P(i)) \setminus \mu(i)$. We shall show that (D, i) blocks* μ, yielding a contradiction. By definition of D, $D \cap \mu(i) = \varnothing$. As $D \subseteq U(i, \mu)$, for all $j \in D$, $i \in Ch(\mu(j) \cup \{i\}, P(j))$. To show that $D \neq \varnothing$, it suffices, by (11), that $\mu(i) \subseteq U(i, \mu)$. Thus, let $j \in \mu(i)$. As μ is a matching, $i \in \mu(j)$. By individual rationality of μ, it follows that

$$i \in \mu(j) = Ch(\mu(j), P(j)) = Ch(\mu(j) \cup \{i\}, P(j)),$$

i.e., $j \in U(i, \mu)$.

So, it remains to show that $D \subseteq Ch(\mu(i) \cup D, P(i))$. Since $Ch\,(U(i, \mu), P(i)) \subseteq \mu(i) \cup D \subseteq U(i, \mu)$, by Observation 1,

$$Ch(\mu(i) \cup D, P(i)) = Ch(U(i, \mu), P(i)) \supseteq D.$$

$\square$

Lemma 5. *For all $P \in \mathscr{P}$, $SW(P) \subseteq \mathscr{E}(P)$.*

Proof. Let $\mu \in \mathscr{M}$ be a matching such that $\mu \notin \mathscr{E}(P)$. Then, there is $i \in I$ such that $\mu(i) \neq Ch(U(i, \mu), P(i))$. We shall prove that $\mu \notin SW(P)$. If μ is not individually rational, there is nothing to prove. Suppose then that μ is an individually rational matching.

As μ is individually rational, $\mu(i) \subseteq U(i, \mu)$ since, for all $j \in \mu(i)$, $i \in \mu(j)$ and, hence, $i \in Ch(\mu(j) \cup \{i\}, P(j))$.

Let $S' = (Ch(U(i, \mu), P(i)) \setminus \mu(i)) \cup \{i\}$. Observe that $S' \setminus \{i\}$ is nonempty for $\mu(i) \neq Ch(U(i, \mu), P(i))$ and $\mu(i) \subseteq U(i, \mu)$. To see this, suppose by contradiction that

$$Ch(U(i, \mu), P(i)) \subseteq \mu(i) \subseteq U(i, \mu).$$

As μ is individually rational, by Observation 1, $\mu(i) = Ch(U(i,\mu), P(i))$, a contradiction.

Now, we shall construct $\mu' \in \mathscr{M}$ such that $\langle S', \mu' \rangle$ is a setwise block of μ. Let $\mu'(i) = Ch(U(i,\mu), P(i))$ and, for all $j \in S' \setminus \{i\} = \mu'(i) \setminus \mu(i)$, $\mu'(j) = Ch(\mu(j) \cup \{i\}, P(j))$. Define, for all $j \in I \setminus S'$, $\rho(j) := \{j' \in S' : j \in \mu(j'),\ j \notin \mu'(j')\}$ and let $\mu'(j) = \mu(j) \setminus \rho(j)$.

By construction, $\mu'(i) \setminus \mu(i) \subseteq S'$. Note that, for all $j \in S' \setminus \{i\}$, $i \in Ch(\mu(j) \cup \{i\}, P(j))$ as $j \in U(i,\mu)$, so that $i \in \mu'(j)$ and $j \in \mu'(i)$. It follows that $\mu'(i)\, P(i)\, \mu(i)$ and, for all $j \in S' \setminus \{i\}$, $\mu'(j) \setminus \mu(j) \subseteq S'$ and $\mu'(j)\, P(j)\, \mu(j)$. By construction, μ' is a matching and, for all $j \in S'$, it is individually rational. Therefore, $\langle S', \mu' \rangle$ forms a setwise block to μ. $\square$

Lemma 6. *For all $P \in \mathscr{P}$, $SW(P) \subseteq B(P)$.*

Proof. Let $\mu \in \mathscr{M}$ be such that $\mu \notin B(P)$. We shall prove that $\mu \notin SW(P)$. If μ is not individually rational, there is nothing to prove. Suppose then that μ is an individually rational matching.

As $\mu \notin B(P)$, there is $\langle S', \mu' \rangle$ where S' is nonempty and μ' is a matching such that, for all $\tilde{\imath} \in S'$, $\mu'(\tilde{\imath}) \setminus \mu(\tilde{\imath}) \subseteq S'$ and $\mu'(\tilde{\imath})\, P(\tilde{\imath})\, \mu(\tilde{\imath})$, and there is no counterobjection to μ (i.e., μ' has no objection within S'). Since there is no counterobjection to μ, for all $\tilde{\imath} \in S'$, μ' is also individually rational. To see this, let $i \in S'$ be an agent with $\mu'(i) \neq Ch(\mu'(i), P(i))$. Then, set $S'' = \{i\} \subseteq S'$ and consider the pair $\langle S'', \mu'' \rangle$ with $\mu''(i) = Ch(\mu'(i), P(i))$, $\mu''(\tilde{\imath}) = \mu'(\tilde{\imath}) \setminus \{i\}$ for all $\tilde{\imath} \in \mu'(i) \setminus Ch(\mu'(i), P(i))$, and $\mu''(\tilde{\imath}) = \mu'(\tilde{\imath})$ for all $\tilde{\imath} \in I \setminus ((\mu'(i) \setminus Ch(\mu'(i), P(i))) \cup \{i\})$. Clearly, μ'' is a matching, $\mu''(i)\, P(i)\, \mu'(i)$, and $\varnothing = \mu''(i) \setminus \mu'(i) \subseteq S'' \neq \varnothing$. So, $\langle S'', \mu'' \rangle$ is a counterobjection to μ, a contradiction.

Accordingly, $\langle S', \mu' \rangle$ is a setwise block to μ. $\square$

Lemma 7. *For all $P \in \mathscr{P}$, $B(P) \subseteq S(P)$.*

Proof. Let $\mu \in \mathscr{M}$ be such that $\mu \notin S(P)$. We shall prove that $\mu \notin B(P)$ by showing that there exists an objection to μ without any counterobjection. If μ is not individually rational, there is nothing to prove. Suppose then that μ is an individually rational matching.

As $\mu \notin S(P)$, there exist $i, j \in I$ with $i \neq j$ and $i \notin \mu(j)$ such that $i \in Ch(\mu(j) \cup \{i\}, P(j))$ and $j \in Ch(\mu(i) \cup \{j\}, P(i))$. Define $S' := \{i, j\}$, $\mu'(i) := Ch(\mu(i) \cup \{j\}, P(i))$, $\mu'(j) := Ch(\mu(j) \cup \{i\}, P(j))$ and, for all $\tilde{\imath} \in I \setminus S'$,

$$\mu'(\tilde{\imath}) := \begin{cases} \mu(\tilde{\imath}) \setminus \{i,j\} & \text{if } \tilde{\imath} \in (\mu(i) \setminus \mu'(i)) \cap (\mu(j) \setminus \mu'(j)), \\ \mu(\tilde{\imath}) \setminus \{i\} & \text{if } \tilde{\imath} \in (\mu(i) \setminus \mu'(i)) \setminus (\mu(j) \setminus \mu'(j)), \\ \mu(\tilde{\imath}) \setminus \{j\} & \text{if } \tilde{\imath} \in (\mu(j) \setminus \mu'(j)) \setminus (\mu(i) \setminus \mu'(i)), \\ \mu(\tilde{\imath}) & \text{otherwise.} \end{cases}$$

Clearly, μ' is a matching and, for all $\tilde{\imath} \in S'$, $\mu'(\tilde{\imath})\, P(\tilde{\imath})\, \mu(\tilde{\imath})$. Hence, $\langle S', \mu' \rangle$ is an objection to μ.

Next, we shall prove that there is no objection within $\langle S', \mu' \rangle$. Since, $j \in \mu'(i)$, $i \in \mu'(j)$ and,for any counterobjection $\langle S'', \mu'' \rangle$, $\mu''(\tilde{\imath}) \setminus \mu'(\tilde{\imath}) \subseteq S''$ for all $\tilde{\imath} \in S''$, it must be that $\mu''(\tilde{\imath}) \subseteq \mu'(\tilde{\imath})$ for all $\tilde{\imath} \in S''$. Notice that μ' is an individually rational matching for i and j. It follows from Observation 1 that $\mu'(\tilde{\imath})\, R(\tilde{\imath})\, Ch(\mu''(\tilde{\imath}), P(\tilde{\imath}))\, R(\tilde{\imath})\, \mu''(\tilde{\imath})$ for all $\tilde{\imath} \in S''$. Thus, $\langle S'', \mu'' \rangle$ is not a counterobjection to μ. $\square$

$\square$

A2. Proof of Theorem 2

The proof of Theorem 2 follows from Theorem 1 and Lemma 8 below.

Lemma 8. *If P is substitutable, then $S(P) \subseteq S^*(P)$.*

Proof. Let $\mu \notin S^*(P)$. We shall prove that $\mu \notin S(P)$. If μ is not individually rational there is nothing to prove. Suppose then that μ is individually rational.

Since $\mu \notin S^*(P)$, there exists $(D, i) \in 2^I \times I$ with $i \notin D \neq \varnothing$ that blocks* μ. This implies for all $j \in D$,

$$i \in Ch(\mu(j) \cup \{i\}, P(j)),$$

and

$$D \subseteq Ch(\mu(i) \cup D, P(i)).$$

But $P(i)$ is substitutable, so there is $j' \in D$ with

$$j' \in Ch(\mu(i) \cup \{j'\}, P(i)).$$

Thus, $\mu \notin S(P)$. $\square$

$\square$

A3. Proof of Proposition 1

Assume that P satisfies strong substitutability. Let $i \in I$, $S, S' \in 2^{I \setminus \{i\}}$ with $S \subseteq S'$, and $j \in I \setminus \{i\}$. Suppose that $j \in Ch(S' \cup \{j\}, P(i))$. We shall show that $j \in Ch(S \cup \{j\}, P(i))$.

By Observation 1, $S \cup \{j\} \subseteq S' \cup \{j\}$ implies $Ch(S' \cup \{j\}, P(i))\, R(i)\, (S \cup \{j\})$. By definition of $R(i)$, there are two cases. First, let $Ch(S' \cup \{j\}, P(i)) = S \cup \{j\}$. Then, by Observation 1,

$$j \in Ch(S' \cup \{j\}, P(i)) = Ch(S \cup \{j\}, P(i)).$$

Second, let $Ch(S' \cup \{j\}, P(i))\, P(i)\, (S \cup \{j\})$. By Observation 1, $j \in Ch(S' \cup \{j\}, P(i)) = Ch(Ch(S' \cup \{j\}, P(i)), P(i))$. It follows from strong substitutability that $j \in Ch(S \cup \{j\}, P(i))$. $\square$

A4. Proof of Theorem 3

The proof of Theorem 3 follows from Proposition 1, Theorem 2 and Lemmata 9 and 10 below.

Lemma 9. *If P is strongly substitutable, then $\mathscr{E}(P) \subseteq SW(P)$.*

Proof. Let $\mu \in \mathscr{E}(P)$. By Lemma 1, μ is an individually rational matching. Suppose for a contradiction that $\mu \notin SW(P)$. Let $\langle S', \mu' \rangle$ be a setwise block of μ.

Pick any $i \in S'$, so that $\mu'(i)\, P(i)\, \mu(i)$. The matching μ is individually rational, so $\mu'(i)\, P(i)\, \mu(i)$ implies

$$Ch(\mu(i) \cup \mu'(i), P(i)) \not\subseteq \mu(i),$$

and there is $j \in \mu'(i) \setminus \mu(i)$ such that $j \in Ch(\mu(i) \cup \mu'(i), P(i))$. As strong substitutability implies substitutability by Proposition 1, $j \in Ch(\mu(i) \cup \{j\}, P(i))$. So,

$$i \in U(j, \mu). \tag{12}$$

By definition of setwise block, $j \in \mu'(i) \setminus \mu(i)$ implies that $j \in S'$ and, hence,

$$\mu'(j)\, P(j)\, \mu(j). \tag{13}$$

As $\langle S', \mu' \rangle$ is a setwise block of μ, $\mu'(\tilde{\imath}) = Ch(\mu'(\tilde{\imath}), P(\tilde{\imath}))$ for all $\tilde{\imath} \in S'$; in particular, $\mu'(j) = Ch(\mu'(j), P(j))$. Moreover, μ' is a matching, hence $j \in \mu'(i) \setminus \mu(i)$ implies $i \in \mu'(j) \setminus \mu(j)$. Therefore, $i \in Ch(\mu'(j) \cup \{i\}, P(j))$. So, (13) implies, by strong substitutability, that

$$i \in Ch(\mu(j) \cup \{i\}, P(j)). \tag{14}$$

Furthermore, as $\mu \in \mathscr{E}(P)$, we have $\mu(j) = Ch(U(j, \mu), P(j))$. So, by (12), $\mu(j) \cup \{i\} \subseteq U(j, \mu)$. It follows from Observation 1 and (14) that $i \in Ch(\mu(j) \cup \{i\}, P(j)) = Ch(U(j, \mu), P(j)) = \mu(j)$, which is a contradiction for $i \notin \mu(j)$. This completes the proof. $\square$

Lemma 10. *If P is strongly substitutable, then $\mathscr{E}(P) \subseteq IRC(P)$.*

Proof. Let $\mu \in \mathscr{E}(P)$. By Lemma 1, μ is an individually rational matching. Suppose for a contradiction that $\mu \notin IRC(P)$. Let $\langle S', \mu' \rangle$ be an individually rational block to μ.

Pick $i \in S'$ such that $\mu'(i)\, P(i)\, \mu(i)$. The matching μ is individually rational, so $Ch(\mu(i) \cup \mu'(i), P(i)) \not\subseteq \mu(i)$. Therefore, there is $j \in Ch(\mu(i) \cup \mu'(i), P(i))$ such that

$j \in \mu'(i) \setminus \mu(i)$. By Proposition 1, strong substitutability implies substitutability. So, $j \in Ch(\mu(i) \cup \mu'(i), P(i))$ implies $j \in Ch(\mu(i) \cup \{j\}, P(i))$ and, hence, $i \in U(j, \mu)$. Moreover, $j \in \mu'(i) \setminus \mu(i)$ implies $i \in \mu'(j) \setminus \mu(j)$ and $j \in S'$, hence,

$$\mu'(j)\, P(j)\, \mu(j). \tag{15}$$

By individual rationality of μ', we have

$$i \in \mu'(j) = Ch(\mu'(j), P(j)) = Ch(\mu'(j) \cup \{i\}, P(j)). \tag{16}$$

As P satisfies strong substitutability, (15) and (16) imply

$$i \in Ch(\mu(j) \cup \{i\}, P(j)). \tag{17}$$

Moreover, $\mu \in \mathscr{E}(P)$ implies that

$$\mu(j) = Ch(U(j, \mu), P(j)). \tag{18}$$

Now, as $i \notin \mu(j)$, (18) implies that $i \notin Ch(U(j, \mu), P(j))$. But $i \in U(j, \mu)$, so

$$Ch(U(j, \mu), P(j))\, P(j)\, U(j, \mu).$$

It follows from (17) and (18) that $i \in Ch(Ch(U(j, \mu), P(j)) \cup \{i\}, P(j))$. By strong substitutability, $i \in Ch(U(j, \mu) \cup \{i\}, P(j))$. But then, $i \in Ch(U(j, \mu), P(j)) = \mu(j)$ for $i \in U(j, \mu)$, a contradiction for $i \notin \mu(j)$. So, we conclude that there are no individually rational blocks of μ and, hence, $\mu \in IRC(P)$. $\square$

$\square$

A5. Proof of Proposition 2

Suppose P satisfies strong substitutability and desirability. Pick any $i, j \in I$ with $i \neq j$ and any $S \subseteq I \setminus \{i\}$. We shall show

$$j \in Ch(S \cup \{j\}, P(i)) \text{ if, and only if, } \{j\}\, P(i)\, \varnothing, \tag{19}$$

where the 'if' part is immediate from desirability. We now show the 'only if' part of (19). Suppose that $j \in Ch(S \cup \{j\}, P(i))$. By Observation 1, $j \in Ch(Ch(S \cup \{j\}, P(i)) \cup \{j\}, P(i)) = Ch(S \cup \{j\}, P(i))$. By assumption, $Ch(S \cup \{j\}, P(i)) \neq \varnothing$. Then, by the definition of the choice correspondence, $Ch(S \cup \{j\}, P(i))\, P(i)\, \varnothing$. It follows from strong substitutability that $j \in Ch(\varnothing \cup \{j\}, P(i)) = Ch(\{j\}, P(i))$. This implies, by the definition of the choice correspondence, that $\{j\}\, P(i)\, \varnothing$, as we seek.

Now, suppose P satisfies weak separability. Desirability immediately follows from weak separability. We shall show that weak separability implies strong substitutability. Pick any $i, j \in I$ with $i \neq j$ and any $S, S' \subseteq I \setminus \{i\}$ such that $S\, P(i)\, S'$ and $j \in Ch(S \cup \{j\}, P(i))$. Weak separability implies, in turn, that $\{j\}\, P(i)\, \varnothing$ and $j \in Ch(S' \cup \{j\}, P(i))$, as we seek. $\square$

A6. Proof of Theorem 4

We start with two lemmas that will be employed in the proof of Theorem 4, and then we shall provide the proof of Theorem 4.

Lemma 11. *If $P \in \mathscr{P}$ satisfies desirability and strong substitutability, then, for all $\mu, \mu' \in S(P)$ and all $i \in I$, $\mu(i) = \mu'(i)$.*

Proof. Let $P \in \mathscr{P}$ satisfy desirability and strong substitutability. Suppose by contradiction that there exists $i \in I$ and $\mu, \mu' \in S(P)$ such that $\mu(i) \neq \mu'(i)$. Assume, without loss of generality, that $\mu(i) \setminus \mu'(i) \neq \varnothing$. Pick any $j \in \mu(i) \setminus \mu'(i)$, so that $i \in \mu(j)$. By individual rationality of μ, $\mu(j)\, P(j)\, \varnothing$ and $\mu(i)\, P(i)\, \varnothing$. So, by strong substitutability, $\{i\}\, P(j)\, \varnothing$ and $\{j\}\, P(i)\, \varnothing$. But then, desirability implies that $i \in Ch(\mu'(j) \cup \{i\}, P(j))$ and $j \in Ch(\mu'(i) \cup \{j\}, P(i))$. It follows that $\{i, j\}$ forms a pairwise block to μ', a contradiction to $\mu' \in S(P)$. Therefore, $\mu = \mu'$. $\square$

Lemma 12. *If P satisfies strong substitutability and desirability, then $IRC(P) \subseteq \mathscr{E}(P)$.*

Proof. Let $P \in \mathscr{P}$ satisfy desirability and strong substitutability. Suppose that $\mu \notin \mathscr{E}(P)$. We shall show that $\mu \notin IRC(P)$. If μ is not individually rational, there is nothing to prove. So, assume that μ is individually rational.

Since $\mu \notin \mathscr{E}(P)$, there exists at least one $i \in I$ such that $\mu(i) \neq Ch(U(i, \mu), P(i))$. Define $\tilde{I} := \{\tilde{\imath} \in I : \mu(\tilde{\imath}) \neq Ch(U(\tilde{\imath}, \mu), P(\tilde{\imath}))\}$. Fix any $i' \in \tilde{I}$. Since μ is individually rational, for all $j \in \mu(i')$,

$$i' \in \mu(j) = Ch(\mu(j) \cup \{i'\}, P(j)).$$

Therefore, for any $i' \in I$, $\mu(i') \subseteq U(i', \mu)$.

Now, we shall construct a pair $\langle S', \mu' \rangle$ with $\mu' \in \mathscr{M}$ that is an individually rational block to μ. Let $S'_i = Ch(U(i, \mu), P(i)) \setminus \mu(i)$ for all $i \in \tilde{I}$. We need the following result.

Claim 1. For all $i \in \tilde{I}$, $S'_i \subseteq \tilde{I}$.

Proof. Pick any $i \in \tilde{I}$ and any $j \in S'_i$. We shall show that $j \in \tilde{I}$. Note that $j \notin \mu(i)$ and, hence, $i \notin \mu(j) = Ch(\mu(j), P(j))$. Suppose, for a contradiction, that $Ch(\mu(j), P(j)) = Ch(U(j, \mu), P(j))$. As $j \in Ch(U(i, \mu), P(i))$, $j \in Ch(U(i, \mu) \cup \{j\}, P(i))$. By Proposition

1, strong substitutability implies substitutability. So, $j \in Ch(U(i,\mu) \cup \{j\}, P(i))$ and $\mu(i) \subseteq U(i,\mu)$ imply that $j \in Ch(\mu(i) \cup \{j\})$. Therefore, $i \in U(j,\mu)$. Moreover, $j \in Ch(U(i,\mu), P(i))$ implies

$$i \in Ch(\mu(j) \cup \{i\}, P(j)). \tag{20}$$

As $i \notin \mu(j)$ and μ is individually rational, it follows that

$$Ch(\mu(j) \cup \{i\}, P(j))\, P(j)\, Ch(\mu(j), P(j)) = Ch(U(j,\mu), P(j)) = Ch(U(j,\mu) \cup \{i\}, P(j)). \tag{21}$$

Since P is substitutable, $Ch(Ch(\mu(j) \cup \{i\}, P(j)) \cup \{i\}, P(j)) = Ch(\mu(j) \cup \{i\}, P(j))$ by Observation 3. So, from (20), we get

$$i \in Ch(Ch(\mu(j) \cup \{i\}, P(j)) \cup \{i\}, P(j)). \tag{22}$$

By strong substitutability, (21) and (22) imply $i \in Ch(Ch(U(j,\mu) \cup \{i\}, P(j)) \cup \{i\}, P(j))$. By Observation 3, $Ch(Ch(U(j,\mu) \cup \{i\}, P(j)) \cup \{i\}, P(j)) = Ch(U(j,\mu) \cup \{i\}, P(j))$. But then, as $i \in U(j,\mu)$, $i \in Ch(U(j,\mu), P(j)) = Ch(\mu(j), P(j))$, a contradiction for $i \notin \mu(j)$. So, we conclude that $Ch(\mu(j), P(j)) \neq Ch(U(j,\mu), P(j))$, hence $j \in \tilde{I}$. □

Now, set $S' = I$ and, for all $i \in \tilde{I}$, $\mu'(i) = Ch(U(i,\mu), P(i))$ and, for all $j \in S' \setminus \tilde{I}$, $\mu'(j) = \mu(j) = Ch(U(j,\mu), P(j))$. Recall that, for any $i \in \tilde{I}$, $\mu(i) \subseteq U(i,\mu)$ by individual rationality of μ. So, it follows from strong substitutability and desirability that, for any $i \in \tilde{I}$, $\mu(i) \subseteq Ch(U(i,\mu), P(i))$. This, together with Claim 1 ($S'_i \subseteq \tilde{I}$ for all $i \in \tilde{I}$), imply that all the previous links at μ are kept at μ' and the new links are established only between the agents from $\tilde{I}$. Therefore, μ' is a matching. By construction, $\mu'(i) \subseteq S'$ for all $i \in S'$. Note that, for all $j \in S' \setminus \tilde{I}$, $\mu'(j)\, R(j)\, \mu(j)$. For any $i \in \tilde{I}$, since $\mu(i) \neq Ch(U(i,\mu), P(i))$ and $\mu(i) \subseteq U(i,\mu)$, we have $\mu'(i)\, P(i)\, \mu(i)$. (By assumption, $\tilde{I}$ is nonempty.) Finally, by Observation 1, μ' is individually rational. Hence, $\langle S', \mu' \rangle$ is an individually rational block to μ, as we seek. This completes the proof. □

Now, we proceed with the proof of Theorem 4.

Set the pre-matching ν as $\nu(i) = I \setminus \{i\}$ and the matching μ as $\mu(i) = Ch(U(i,\nu), P(i))$ for all $i \in I$. We first show that there exists such a matching μ. Pick any $i, j \in I$ with $i \neq j$ such that $i \in \mu(j)$. We are to show that $j \in \mu(i)$. By definition, μ is individually rational. Then, $i \in \mu(j)$ implies $i \in Ch(\mu(j) \cup \{i\}, P(j))$. By strong substitutability, $\{i\}\, P(j)\, \varnothing$. By desirability, we have $i \in Ch(\nu(j) \cup \{i\}, P(j)) = Ch(\nu(j), P(j))$; so, $j \in U(i,\nu)$. Note that $i \in U(j,\nu)$ for $i \in \mu(j) = Ch(U(j,\nu), P(j))$. Therefore, $j \in Ch(v(i) \cup \{j\}, P(i)) = Ch(v(i), P(i))$. By strong substitutability, $\{j\}\, P(i)\, \varnothing$. By desirability, $\{j\}\, P(i)\, \varnothing$ and $j \in U(i,\nu)$ imply that $j \in Ch(U(i,\nu), P(i)) = \mu(i)$. So, μ is a matching with

$$\forall i, j \in I: \quad j \in \mu(i) \Leftrightarrow \{i\}\, P(j)\, \varnothing,\ \{j\}\, P(i)\, \varnothing,$$

which is exactly the one defined in Theorem 4.

Next, we show that μ is pairwise stable. Consider any pair of agents $i, j \in I$ with $i \neq j$ such that $j \notin \mu(i)$. Suppose for a contradiction that $i \in Ch(\mu(j) \cup \{i\}, P(j))$ and $j \in Ch(\mu(i) \cup \{j\}, P(i))$. Then, $j \in U(i, \nu)$ and $i \in U(j, \nu)$. But, strong substitutability and desirability imply that $j \in Ch(U(i, \nu), P(i)) = \mu(i)$ and $i \in Ch(U(j, \nu), P(j)) = \mu(j)$, a contradiction. So, no pair of agents (pairwise) blocks μ.

Strong substitutability implies, by Theorem 3, that the pairwise stable set is equivalent to the setwise stable set, the bargaining set, the fixpoint set, and the stable* set, and is included in the individually rational core. By Lemma 12, we also established the equivalence of the fixpoint set and the individually rational core. By the foregoing argument, these sets are also nonempty. Finally, that μ is the unique pairwise stable matching follows from Lemma 11. $\square$

A7. Proof of Proposition 3

Let P be a weakly separable preference profile that reflects the truthful preferences of the agents. Suppose $i \in I$ reports her preferences untruthfully as $P'(i)$ instead of $P(i)$, where $P'(i) \neq P(i)$, and the remaining ones report truthfully. With a slight abuse of notation, set $\mu = \mathcal{S}^s(P)$ and $\mu' = \mathcal{S}^s(P(I \setminus \{i\}), P'(i))$. Suppose for a contradiction that $\mu'(i)\, P(i)\, \mu(i)$. As μ is individually rational, there is $j \in Ch(\mu'(i), P(i)) \setminus \mu(i)$ and, hence, $i \in \mu'(j) \setminus \mu(j)$. So, $j \in Ch(\mu'(i) \cup \{j\}, P(i))$ and, by weak separability, $j \in Ch(\mu(i) \cup \{j\}, P(i))$. Since j reports her preferences truthfully, $\mu'(j)$ is individually rational. So, $i \in \mu'(j) = Ch(\mu'(j) \cup \{i\}, P(j))$. By weak separability, $i \in Ch(\mu(j) \cup \{i\}, P(j))$. But then, i and j forms a pairwise block to μ, a contradiction to $\mu \in S(P)$. $\square$

A8. Proof of Proposition 4

Let P be strongly substitutable. Fix any $\mu \in \mathcal{S}(P)$. Suppose by contradiction that there is an individually rational matching $\mu' \in \mathscr{M}$ such that

$$\mu'(i)\, P(i)\, \mu(i), \tag{23}$$

for some $i \in I$. We shall show that there is $\tilde{j} \in I \setminus \{i\}$ such that $\mu(\tilde{j})\, P(\tilde{j})\, \mu'(\tilde{j})$.

As the stable rule is by definition individually rational, μ is individually rational. So, (23) implies $\mu'(i) \setminus \mu(i) \neq \varnothing$. Fix $j \in \mu'(i) \setminus \mu(i)$ and note that $i \in \mu'(j) \setminus \mu(j)$. So, either $\mu(j)\, P(j)\, \mu'(j)$ or $\mu'(j)\, P(j)\, \mu(j)$ holds. If the former holds, there is nothing to prove. So, suppose that

$$\mu'(j)\, P(j)\, \mu(j). \tag{24}$$

As μ and μ' are individually rational, (23) and (24) imply, by strong substitutability, that $j \in Ch(\mu(i) \cup \{j\}, P(i))$ and $i \in Ch(\mu(j) \cup \{i\}, P(j))$, a contradiction to $\mu \in \mathcal{S}(P)$. Hence, $\mu(j)\, P(j)\, \mu'(j)$. $\square$

A9. Proof of Proposition 5

We start with the proof of the 'only if' part of the second claim. Suppose that ξ^* is a PNE of $\Gamma(P)$ and $\mu \in \mathscr{M}$ is the outcome of ξ^*. So, for all $j \in I$, $\mu(j) = \xi^*_j \cap \{\tilde{\imath} \in I \setminus \{j\} : j \in \xi^*_{\tilde{\imath}}\}$. It is immediate that $\tilde{\xi} := \mu$ is also a PNE, since, given the equilibrium strategy profile ξ^*, each $j \in I$ is indifferent between proposing ξ^*_j and $\tilde{\xi}_j$ as they both result in the same set of partners.

Now, we prove that $\mu \in S(P)$. Notice that, for all $j \in I$ and all $\xi_j \in \Xi_j$, $\xi_j \cap \{\tilde{\imath} \in I \setminus \{j\} : j \in \tilde{\xi}_{\tilde{\imath}}\} \subseteq \mu(j)$ and, as $\tilde{\xi} = \mu$ is also a PNE, $\mu(j)\, R(j)\, (\xi_j \cap \{\tilde{\imath} \in I \setminus \{j\} : j \in \tilde{\xi}_{\tilde{\imath}}\})$. In particular, for any $j \in I$ and any $S \subseteq \mu(j)$, $\mu(j)\, R(j)\, S$. Therefore, μ is individually rational. Next, we show that μ is immune to any pairwise block. Suppose for a contradiction that there are $i, j \in I$ such that $i \notin \mu(j)$, $i \in Ch(\mu(j) \cup \{i\}, P(j))$ and $j \in Ch(\mu(i) \cup \{j\}, P(i))$. Put $\xi_j := Ch(\mu(j) \cup \{i\}, P(j))$ and $\xi_i := Ch(\mu(i) \cup \{j\}, P(i))$. It follows that $\xi_i\, P(i)\, \tilde{\xi}_i$ and $\xi_j\, P(j)\, \tilde{\xi}_j$. Notice that $\xi_i = (\xi_i \cap (\{\tilde{\imath} \in I \setminus \{i,j\} : i \in \tilde{\xi}_{\tilde{\imath}}\} \cup \{j\})$ and $\xi_j = (\xi_j \cap (\{\tilde{\imath} \in I \setminus \{i,j\} : j \in \tilde{\xi}_{\tilde{\imath}}\} \cup \{i\}))$. Also, by definition of $\tilde{\xi}$, for any $i \in I$, $\tilde{\xi}_i = \tilde{\xi}_i \cap \{\tilde{\imath} \in I \setminus \{i\} : i \in \tilde{\xi}_{\tilde{\imath}}\}$. But then,

$$(\xi_i \cap (\{\tilde{\imath} \in I \setminus \{i,j\} : i \in \tilde{\xi}_{\tilde{\imath}}\} \cup \{j\}))\, P(i)\, (\tilde{\xi}_i \cap \{\tilde{\imath} \in I \setminus \{i\} : i \in \tilde{\xi}_{\tilde{\imath}}\})$$

and

$$(\xi_j \cap (\{\tilde{\imath} \in I \setminus \{i,j\} : j \in \tilde{\xi}_{\tilde{\imath}}\} \cup \{i\}))\, P(j)\, (\tilde{\xi}_j \cap \{\tilde{\imath} \in I \setminus \{j\} : j \in \tilde{\xi}_{\tilde{\imath}}\}),$$

a contradiction for $\tilde{\xi}$ is a PNE.

Next, we prove the 'if' part. Let $\mu \in S(P)$ and ξ^* be the strategy profile defined as $\xi^*_j := \mu(j)$ for all $j \in I$. As μ is a matching, for any $i \in I$, $\xi^*_i = \{\tilde{\imath} \in I \setminus \{i\} : i \in \xi^*_{\tilde{\imath}}\}$. Notice that, for any $i \in I$ and any $\xi_i \in \Xi_i$, $\xi_i \cap \{\tilde{\imath} \in I \setminus \{i\} : i \in \xi^*_{\tilde{\imath}}\} \subseteq \mu(i)$. As μ is individually rational, it follows that, for any $i \in I$ and any $\xi_i \in \Xi_i$,

$$(\xi^*_i \cap \{\tilde{\imath} \in I \setminus \{i\} : i \in \xi^*_{\tilde{\imath}}\})\ R(i)\ (\xi_i \cap \{\tilde{\imath} \in I \setminus \{i\} : i \in \xi^*_{\tilde{\imath}}\}).$$

Therefore, ξ^* is immune to individual deviations. Now, suppose for a contradiction that there exist $i, j \in I$ and $\xi_i \in \Xi_i$, $\xi_j \in \Xi_j$ such that $i \notin \xi_j^*$,

$$(\xi_i \cap (\{\tilde{\imath} \in I \setminus \{i,j\} : i \in \xi_{\tilde{\imath}}^*\} \cup \{j\}))\ P(i)\ (\xi_i^* \cap \{\tilde{\imath} \in I \setminus \{i\} : i \in \xi_{\tilde{\imath}}^*\}) \tag{25}$$

and

$$\left(\xi_j \cap (\{\tilde{\imath} \in I \setminus \{i,j\} : j \in \xi_{\tilde{\imath}}^*\} \cup \{i\})\right)\ P(j)\ \left(\xi_j^* \cap \{\tilde{\imath} \in I \setminus \{j\} : j \in \xi_{\tilde{\imath}}^*\}\right). \tag{26}$$

Notice that $\xi_i \cap (\{\tilde{\imath} \in I \setminus \{i,j\} : i \in \xi_{\tilde{\imath}}^*\} \cup \{j\}) \subseteq \mu(i) \cup \{j\}$ and $\xi_j \cap (\{\tilde{\imath} \in I \setminus \{i,j\} : j \in \xi_{\tilde{\imath}}^*\} \cup \{i\}) \subseteq \mu(j) \cup \{i\}$. So, these along with (25) and (26) imply that $Ch(\mu(i) \cup \{j\}, P(i))\, P(i)\, \mu(i)$ and $Ch(\mu(j) \cup \{i\}, P(j))\, P(j)\, \mu(j)$. As μ is individually rational, it follows that $i \in Ch(\mu(j) \cup \{i\}, P(j))$ and $j \in Ch(\mu(i) \cup \{j\}, P(i))$ but $i \notin \mu(j)$, a contradiction for $\mu \in S(P)$. So, ξ^* is immune to pairwise deviations. Therefore, ξ^* is a PNE of $\Gamma(P)$.

Finally, that $PNE(P) \neq \varnothing$ follows from the foregoing result and Theorem 4. □

References

H. Adachi, On a characterization of stable matchings, *Economics Letters* **68** (2000), 43–49.

J. Alcalde, S. Barberà, Top dominance and the possibility of strategy-proof stable solutions to matching problems, *Economic Theory* **4** (1994), 417–435.

A. Alkan, Non-existence of stable threesome matchings, *Mathematical Social Sciences* **16** (1988), 207–209.

A. Alkan, On the properties of stable many-to-many matchings under responsive preferences, in A. Alkan, C.D. Aliprantis and N.C. Yannelis (Eds.), *Current Trends in Economics: Theory and Applications*, Berlin, Springer-Verlag, 1999.

R.J. Aumann, M. Maschler, The bargaining set for cooperative games, in M. Dresher, L.S. Shapley and A.W. Tucker (Eds.), *Advances in Game Theory*, Princeton, Princeton University Press, 1964.

S. Banerjee, H. Konishi, T. Sönmez, Core in a simple coalition formation game, *Social Choice and Welfare* **18** (2001), 135–153.

C. Blair, The lattice structure of the set of stable matchings with multiple partners, *Mathematics of Operations Research* **13** (1988), 619–628.

A. Bogomolnaia, M. Jackson, The stability of hedonic coalition structures, *Games and Economic Behavior* **38** (2002), 201–230.

A. Calvó-Armengol, R. İlkılıç, Pairwise-stability and Nash equilibria in network formation, *International Journal of Game Theory* **38** (2009), 51–79.

J.H. Drèze, J. Greenberg, Hedonic coalitions: optimality and stability, *Econometrica* **48** (1980), 987–1003.

L. Dubins, D. Freedman, Machiavelli and the Gale-Shapley algorithm, *American Mathematical Monthly* **88** (1981), 485–494.

F. Echenique, J. Oviedo, Core many-to-one matchings by fixed point methods, *Journal of Economic Theory* **115** (2004), 358–376.

F. Echenique, J. Oviedo, A theory of stability in many-to-many matching markets, *Theoretical Economics* **1** (2006), 233–273.

D. Gale, L. Shapley, College admissions and the stability of marriage, *The American Mathematical Monthly* **69** (1962), 9–15.

M.O. Jackson, B.W. Rogers, The economics of small worlds, *Journal of the European Economic Association 3* (2005), 617–627.

M.O. Jackson, A. Wolinsky, A strategic model of social and economic networks, *Journal of Economic Theory*, **71** (1996), 44–74.

C. Johnson, R.P. Gilles, Spatial social networks, *Review of Economic Design*, **5** (2000), 273–299.

A.S. Kelso, V.P. Crawford, Job matching, coalition formation, and gross substitutes, *Econometrica* **50** (1982), 1483–1504.

B. Klaus, M. Walzl, Stable many-to-many matchings with contracts, *Journal of Mathematical Economics* **45** (2009), 422–434.

F. Klijn, J. Massó, Weak stability and a bargaining set for the marriage model, *Games and Economic Behavior* **42** (2003), 91–100.

H. Konishi, M.U. Ünver, Credible group stability in many-to-many matching problems, *Journal of Economic Theory* **129** (2006), 57–80.

S. Lahiri, The core of directed network problems with quotas, *Economics Bulletin* **3** (2004), 1–8.

R. Martínez, J. Massó, A. Neme, J. Oviedo, Single agents and the set of many-to-one stable matchings, *Journal of Economic Theory,* **91** (2000), 91–105.

R. Martínez, J. Massó, A. Neme, J. Oviedo, On group strategy-proof mechanisms for a many-to-one matching model, *International Journal of Game Theory* **33** (2004), 115–128.

D. Persitz, Power in the heterogeneous connections model: the emergence of core-periphery networks, Tel Aviv University (2008), mimeo.

A.E. Roth, The economics of matching: stability and incentives, *Mathematics of Operations Research* **7** (1982), 617–628.

A.E. Roth, Stability and polarization of interests in job matching, *Econometrica* **52** (1984a), 47–57.

A.E. Roth, The evolution of the labor market for medical interns and residents: a case study in game theory, *Journal of Political Economy* **92** (1984b), 991–1016.

A.E. Roth, The college admissions problem is not equivalent to the marriage problem, *Journal of Economic Theory* **36** (1985), 277–288.

M. Sotomayor, Three remarks on the many-to-many stable matching problem, *Mathematical Social Sciences* **38** (1999), 55–70.

L. Zhou, A new bargaining set of an n-person game and endogenous coalition formation, *Games and Economic Behavior* **6** (1994), 512–526.

The Stockholm School of Economics

A complete publication list can be found at *www.hhs.se/research/publications*.
Books and dissertations are published in the language indicated by the title and can be ordered via e-mail: efi.publications@hhs.se.

A selection of recent publications

Books

Barinaga, Ester. *Powerful dichotomies.*
Ericsson, Daniel. *Den odöda musiken.*
Ericsson, Daniel. *Scripting Creativity.*
Melén, Sara. *Globala från start. Småföretag med världen som marknad.* Forskning i Fickformat.
Mårtensson, Pär, Mähring, Magnus. *Mönster som ger avtryck: Perspektiv på verksamhetsutveckling.*
Sjöström, Emma. *Ansiktslösa men ansvarsfulla.* Forskning i fickformat.
Törn, Fredrik. *Utmana varumärket.* Forskning i fickformat.
Wijkström, Filip. *Civilsamhällets många ansikten.*

Dissertations

Alexandersson, Gunnar. *The Accidental Deregulation. Essays on Reforms in the Swedish Bus and Railway Industries.*
Bohman, Claes. *Attraction.*
Ejenäs, Markus. *Ledning av kunskapsintegration – förutsättningar och hinder: en studie av en fusion mellan IT- och managementkonsulter.*
Engvall, Anders. *Poverty and Conflict in Southeast Asia.*
Glassér, Charlotte. *The fountainhead of Innovation health.*
Juks, Reimo. *Corporate Governance and the Firm´s Behavior towards Stakeholders.*
Lychnell, Lars-Olof . *IT- relaterad verksamhetsförändring. Processer som formar växelspelat mellan utveckling och användning.*
Magnusson Bernard, Kristin. *Remittances, Regions and Risk Sharing.*
Mohlin, Erik. *Essays on Belief Formation and Pro-Sociality.*
Nakatani, Tomoaki. *Four essays on building Conditional Correlation GARCH models.*
Siming, Linus. *Private Equity and Advisors in Mergers and Acquisitions.*
Sjöqvist Rafiqui, Pernilla. *Evolving Economic Landscapes: Institutions and localized economies in time and space.*
Sunesson, Daniel. *School networks and active investors.*
Tolstoy, Daniel. *International entrepreneurship in networks.*
Öhman, Niclas. *Considering intentions.*