universität wien

DISSERTATION | DOCTORAL THESIS

Titel | Title

Ethical Concerns in Economics

verfasst von | submitted by

Philipp Peitler BSc MSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 794 370 140
Dissertationsgebiet lt. Studienblatt | Field of study as it appears on the student record sheet:	(Diss.Gebiet: VWL, DK: Vienna Graduate School of Economics)
Betreut von | Supervisor:	Univ.-Prof. Dipl.-Math. Karl Schlag PhD Univ.-Prof. Dr. Maarten Janssen

Acknowledgments

This thesis would not have been possible without the assistance and support of many people. My deepest gratitude goes to my first supervisor, Karl Schlag, whose mentorship profoundly shaped both my professional journey as an economist and my personal growth. His constant support, encouragement, and friendship have been invaluable to me throughout this journey.

Maarten Janssen, my second supervisor, also deserves special thanks for his unwavering support and invaluable advice that helped steer my path through the PhD program.

I am thankful to the other members of the Economics Department at the University of Vienna for their helpful guidance and valuable feedback. I would like to extend special thanks to Daniel Garcia, Wieland Müller, Philipp Schmidt-Dengler, Justin Leduc, Matan Tsur, Eeva Mauring, Carl Heese, and Juha Tolvanen for their contributions.

My time at the Vienna Graduate School of Economics was greatly enriched by the companionship of my colleagues. I will always cherish the friendly atmosphere and the insightful discussions we shared.

I am also grateful for the financial support provided by the FWF and the OeNB Anniversary Fund (Nr. 18719).

Finally, my heartfelt appreciation goes to my family and friends, whose constant support and encouragement sustained me. I am especially thankful to my parents, Elke and Josef, and to Tiffany, my longtime companion and the love of my life.

Contents

Introduction

At the core of this thesis are two types of statements that guide much of academic inquiry: normative and descriptive. Descriptive statements describe “what is,” whereas normative statements express “what ought to be.” Typically, these two realms are investigated separately. While the sciences strive to uncover descriptive truths about the natural world, it is traditionally the role of ethics, as a branch of philosophy, to deal with normative truths. In economics, however, these two types of truths are often inextricably linked. Endorsing an economic policy requires not only a descriptive theory to predict the policy's effects but also a normative theory to assess whether these effects are desirable.

This interconnection has spurred research on ethical questions among economists and has led to seminal contributions during the latter half of the 20th century. Scholars such as Kenneth J. Arrow, John C. Harsanyi, and Amartya Sen have advanced our understanding by applying rigorous, formal methods to tackle ethical problems. Under the economist's approach to ethics, normative principles are translated into a mathematical framework, which allows for the derivation of precise moral judgments.

Following this tradition, my thesis combines formal inquiries into normative questions with the application of the resulting insights to descriptive models. The thesis comprises three chapters. A recurring theme present in each of these chapters is the conflict between notions of rationality and ethical concerns.

The first chapter, developed in collaboration with Karl Schlag, explores under what conditions a policy is socially desirable. Oftentimes, a policy benefits some at the expense of others. Therefore, assessing the desirability of a policy to society as a whole requires a rule to trade off the gains and losses across individuals. However, according to Kenneth J. Arrow's impossibility theorem (Arrow, 1950, 1963), no such rule exists.

The theorem can be interpreted as a conflict between menu independence - a rationality criterion - and non-dictatorship - an ethical imperative. Menu independence stipulates that the relative desirability of two alternatives should not be influenced by the presence of other alternatives. Non-dictatorship opposes the idea that society should invariably adhere to the preferences of a single individual. We challenge the normative appeal of menu independence and demonstrate that a slight relaxation of this condition allows for a unique rule that assigns equal consideration to all individuals. This rule is then applied to scenarios involving price discrimination, optimal consumption-savings plans, and fair wage determination.

The second chapter considers a group of agents who bargain over the division of a surplus. An arbitrator is appointed to facilitate cooperation, as disagreement would lead to a loss of this surplus. The arbitrator faces a dilemma: she cannot simultaneously act rationally in the face of risk and ensure equal prospects for all agents. Dealing with risk in a rational manner does not permit the arbitrator to rely strictly on randomization. In some situations, however, randomization is required to ensure that individuals are made equal. I resolve this conflict by arguing in favor of a weaker ethical principle. Instead of making agents equal in their expected outcomes, it suffices to give them equal consideration. This modification results in a rational decision rule for the arbitrator. The chapter also describes a bargaining protocol, which allows the arbitrator to achieve an optimal outcome without relying on agents to report their preferences truthfully.

The third chapter adopts a more applied perspective, diverging from the primarily normative inquiries of the earlier chapters. Nevertheless, ethical concerns arise here as well. The chapter models a so-called credence good market, where firms know more about consumer needs than consumers themselves. This informational asymmetry opens the door for unethical practices. An expert firm might be tempted to deceive a layman consumer about their actual condition in order to charge a higher price for the service. The chapter explores the role of price transparency in such markets. Firms can either display a price for a service publicly or choose a price after the consumer has visited the firm and the firm has learned the consumer's condition. I show that opaque pricing can arise in equilibrium without invoking collusion. Furthermore, I find that under certain conditions, consumers can benefit from opaque pricing.

Chapter I

Putting Context into Preference Aggregation

joint with Karl H. Schlag

Abstract

The axioms underlying Arrow's impossibility theorem are very restrictive in terms of what can be used when aggregating preferences. Social preferences may not depend on the menu nor on preferences over alternatives outside the menu. But context matters. So, we weaken these restrictions to allow for context to be included. The context, as we define, describes which alternatives in the menu and which preferences over alternatives outside the menu matter. We obtain unique representations. These are discussed in examples involving markets, the intertemporal well-being of an individual, and bargaining.

I.1 Introduction

A central question in economics is how a social planner should compare alternatives when the preferences of the individuals involved are in conflict. To determine what is best for those individuals involved, the planner needs a rule that aggregates individual preferences into a single social preference. Different strands of the economic literature use very different rules. In the analysis of markets, consumer surplus is used to evaluate the joint welfare of consumers. In the analysis of dynamically inconsistent decision makers, intertemporal welfare is evaluated either by the Pareto criterion or by the long-run utility (O'Donoghue & Rabin, 1999). In bargaining, the Nash bargaining solution (Nash, 1950) and the Kalai-Smorodinsky bargaining solution (Kalai & Smorodinsky, 1975) are prominent. Note here that bargaining problems can be viewed as the search for the best allocation after aggregating preferences. Even within their respective settings, these criteria can only be applied under narrow modeling assumptions. More importantly, each of them fails to conform to a minimal set of desiderata. Consumer surplus violates the neutrality axiom, as it is sensitive to the labels of social alternatives. The Pareto criterion does not generate complete preferences, and long-run utility violates the Pareto principle. Neither the Nash nor the Kalai-Smorodinsky bargaining solution can be interpreted as resulting from the most preferred allocation under a social preference relation that satisfies the von Neumann and Morgenstern (1944) axioms. Some of the above violations are not obvious and will be demonstrated. In each of these settings, we could search separately for other rules. However, we feel that the well-being of individuals should be traded off based on principles that are appealing independently of the application. Therefore, we would like to have a universal rule that can be applied to all settings. In light of Arrow (1950, 1963), it is not clear whether such a rule exists. In the following, we argue that some of Arrow's demands are too stringent and should be relaxed. Before we state our contentions, let us briefly revisit Arrow's theorem.

Arrow (1963) shows that there is no aggregation rule that satisfies completeness, transitivity, the Pareto principle, non-dictatorship, and *independence of irrelevant alternatives* (IIA). If we want the social planner to be rational, benevolent, and impartial, then the first four axioms of Arrow serve as minimal requirements. Consider now the fifth and last axiom, IIA. According to Arrow, IIA demands that when the planner chooses from a menu of alternatives denoted by S, the choice $C(S)$ is not allowed to depend on individual preferences over alternatives outside S. In addition to the above, Arrow assumes that social choice from any menu is made according to a single social preference over all alternatives. Hence, Arrow implicitly assumes one more axiom: *menu independence* (MI). Together, these axioms make it impossible to aggregate preferences. Weakening either IIA or MI is a potential avenue to escape the impossibility. We now argue that one can weaken either of them as neither is as desirable as the first four axioms mentioned above.

First, we assess IIA with the aid of the following example by Pearce (2021). Imagine being tasked with choosing between two social alternatives, x and y, for a group of five kindergarten children. Strict preferences of the children are given in Figure I.1. In isolation, one would be inclined to choose x over y as it is preferred by four of the

1	2	3	4	5
x	x	x	x	y
y	y	y	y	x

Figure I.1: Children's ordinal preferences.

five children. However, we also learn that under x, each of the first four children gets 1001 toys, while under y, each of them only gets 1000 toys. Moreover, we are told that the fifth child has a fatal illness that would be cured under y, whereas under x, the fifth child dies a long and terrifying death. Unquestionably, this additional information would change our preference to y over x. Pearce, therefore, concludes that information besides individual preferences between x and y must be relevant for social choice. But which information exactly? Let us embed the comparison between x and y into a social choice problem in which, for each of the first four children, there exists an alternative z_i where only this child i dies and every other child gets 1001 toys. While the alternatives z_1 to z_4 aren't currently feasible (i.e., they are not in the menu), they are still possible (in the sense of conceivable) and could be in the menu in a different situation. Assume that Figure I.2 depicts the von Neumann-Morgenstern preferences of the children over all these alternatives. The inclusion of alternatives z_1 to z_4 provides a context that

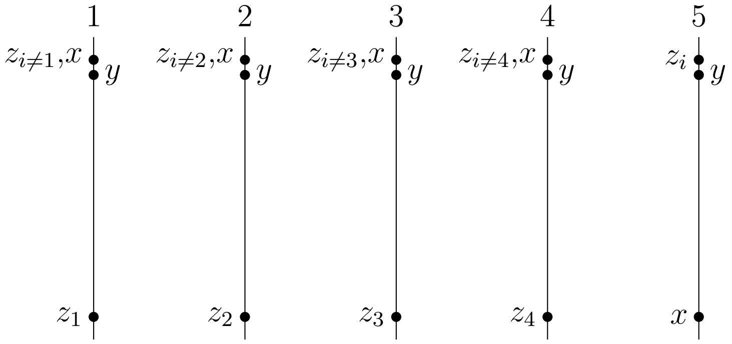


Figure I.2: Children's cardinal utility.

puts the alternatives x and y into perspective. Adding this context, it becomes clear that going from y to x represents a marginal improvement for the first four children at the expense of a substantial loss to the fifth child. IIA demands that we ignore this additional information, which seems unreasonable. In order to include the context provided by the infeasible alternatives, IIA has to be weakened.

Next, we assess MI. To do this, consider the following example by Sen (1993). An individual faces the menu $\{x, y\}$ where y means taking the last remaining apple from the fruit basket at the dinner table, and x means taking nothing instead. Compare this to the situation where there is a second apple in the basket, such that the menu is $\{x, y, z\}$ where z means taking the other apple. One can plausibly prefer x over y under the first menu and y over x under the second. Similarly, our understanding of fairness in a social choice setting might depend on what is currently feasible. When comparing two social alternatives, x and y, information about the feasibility of other alternatives provides a context that helps with the evaluation. In order to make use of that information, MI has to be weakened.

Above we argued that neither IIA nor MI is desirable, as they ignore additional information that can help in the assessment of social alternatives. Weakening either of these conditions opens the possibility for preference aggregation. We are looking for aggregation rules that abide by the von Neumann-Morgenstern axioms, satisfy the strong Pareto condition, and are anonymous. These axioms strengthen Arrow's minimal requirements for rationality, benevolence, and impartiality. We do not want to drop IIA and MI completely, as we wish to limit the information that can be used by the social planner. We, therefore, define two weaker axioms that identify when two social choice problems provide the same relevant information and thus yield the same social preferences. These axioms implicitly define the context as the context captures what is relevant. *Independence of irrelevant comparable alternatives* (IICA) weakens IIA by requiring that preferences over alternatives outside the menu do not influence the ranking if they are comparable for each individual to other alternatives. Specifically, an alternative is comparable to others if each individual is indifferent to some mixture of these other alternatives. *Menu independence of comparable alternatives* (MICA) weakens MI by allowing alternatives to be dropped from the menu if they are comparable to other alternatives in the menu. Each of these axioms, together with our other axioms, uniquely defines an aggregation rule that can be applied to any social choice or aggregation problem. Both of our representations are relative utilitarian, meaning the social welfare of an alternative is identified as the sum of individual utilities. When we weaken MI and maintain IIA, individual utilities are normalized relative to the menu. When we weaken IIA and maintain MI, individual utilities are normalized relative to the set of all possible alternatives within a given setting. Normalizing means setting the utility of the worst alternative in the respective set to 0 and the utility of the best alternative to 1. In a later section, we then use our representations to quantify consumer welfare and the well-being of a dynamically inconsistent decision maker and to derive a fair bargaining solution.

I.1.1 Literature

In the tradition of Arrow (1950, 1963), we axiomatize relative utilitarianism in a *multi-profile* setting, meaning that at least one axiom relates social preferences across different profiles of individual preference (e.g., IIA or IICA). Previous characterizations of relative utilitarianism in this setting are Dhillon and Mertens (1999), Börgers and Choo (2017) and Marchant (2019). Dhillon and Mertens (1999) implicitly assumes menu independence and characterizes an aggregation rule that is the same as the one we obtain when relaxing IIA. They have a bottom-up approach, as their implicit objective is to present the weakest possible axioms that characterize relative utilitarianism. In contrast, we have a top-down approach, as we stay close to Arrow (1963) and wish to present the strongest axioms that don't result in an impossibility. These two different approaches result in distinct axiomatic systems. Dhillon and Mertens (1999) has a higher level of technical sophistication, which is exemplified by their continuity axiom and intricate proof. In contrast, our axioms have a straightforward interpretation, and our proof is much simpler. Note also that Dhillon and Mertens (1999) does not provide a representation when some individuals have either identical or opposing preferences. Börgers and Choo (2017) takes a revealed preference approach and imposes axioms on the planner's revealed marginal rate of substitution between individuals' normalized utilities. Marchant (2019) axiomatizes a relative utilitarian choice correspondence rather than a social preference.

In a *single-profile* setting, meaning without axioms relating different profiles, relative utilitarianism has been axiomatized by Karni (1998), Segal (2000), and Karni and Weymark (2024).

We build on a large body of literature that deals with the aggregation of vNM preferences. The seminal paper is Harsanyi (1955), which shows that vNM rationality of society and a Pareto axiom implies linear aggregation of utilities. Many extensions and variations of this theorem followed (Domotor, 1979; Fishburn, 1984; Border, 1985; Selinger, 1986; Coulhon & Mongin, 1989; Hammond, 1992; Weymark, 1993; Mongin, 1994; Weymark, 1994; De Meyer & Mongin, 1995; Zhou, 1997a; Blackorby et al., 1999; Mandler, 2005). While Harsanyi (1955) shows that social welfare must be a weighted sum of individual utilities, he does not identify the weights. We start from his theorem and impose additional axioms to pin down these weights. Another paper on the aggregation of vNM preferences worthy of mention is Sprumont (2013), which drops the assumption that social preferences are vNM and characterizes the relative egalitarian aggregation rule. Our axiom MICA is quite close to the "independence of inessential expansions" axiom in Sprumont (2013).

Closely related to the aggregation of risk preferences is the aggregation of subjective expected utility (SEU) preferences over uncertain acts. Since beliefs regarding the state can differ across individuals, not only conflicting tastes but also conflicting probability judgments have to be aggregated. This additional complication leads to an impossibility result, even in the single-profile setting. Mongin (1995, 1998) shows that society's preferences cannot simultaneously satisfy the Pareto condition and have a SEU representation. To escape this impossibility, utilitarian aggregation rules either weaken the rationality postulates underlying the SEU representation (Sprumont, 2019) or the Pareto condition (Gilboa et al., 2004; Brandl, 2021; Dietrich, 2021). Most closely related is Sprumont (2019), which allows the social preferences to depend on some reference set of relevant outcomes. This is mathematically similar to our approach, where social preferences can depend on the menu. However, Sprumont's axioms don't tell us what is relevant, as the relevant set is exogenous and part of the social choice problem. Without additional restrictions, this would allow the planner to justify nearly any decision by handpicking the relevant set as needed. One could not simply reinterpret the reference set as the menu, as it is a set of outcomes, not a set of acts.

Related to relative utilitarian aggregation rules is the literature on *range voting* (Smith, 2000; Gaertner & Xu, 2012; Pivato, 2014; Macé, 2018). Under range voting, a voter assigns a score between 0 and 1 to each candidate, and the candidate with the highest sum of scores is selected as the winner.

We also mention a few of the many other papers on axiomatizations of aggregation rules that start similarly to us from Arrow (1950, 1963) and relax axioms therein. Notably, Sen (1993) drops MI and shows that impossibility still follows if IIA is replaced by *binary IIA*.[1] To our knowledge, Sen (1993) is the only previous paper that points out the implicit MI axiom in Arrow (1950, 1963). Saari (1998) and Maskin (2022) relax IIA by allowing the comparison between two alternatives to depend on how many other alternatives lie in between them.

I.2 Axiomatization

There is a society consisting of n individuals, where $n \in \mathbb{N}$. The set of individuals is denoted by $N := \{1, ..., n\}$. Furthermore, there is a set of possible alternatives A, where A is finite.[2] Each individual in society has rational preferences over the

[1]Binary IIA says that the social preference between any two alternatives x and y can only depend on individual preferences between x and y. The literature often uses the notion of IIA and binary IIA interchangeably. Note however that binary IIA is only implied by IIA if MI is assumed as well.

[2]In Appendix I.B we consider the case where A is infinite.

possible alternatives. Formally, let $\triangle A$ denote the set of lotteries over A and let $\mathcal{R}$ denote the set of logically possible von Neumann-Morgenstern (vNM) preferences over $\triangle A$. Individual preferences are then captured by a preference profile $R \in \mathcal{R}^n$. For a given profile R, we sometimes write $\succcurlyeq_i^R$ to denote the i'th element in R. In any given situation, only a subset S of the alternatives in A is feasible and could be implemented by society. We call S the *menu*. For any $S \subseteq A$ and $R \in \mathcal{R}^n$, we call (S, R) the society's *state*, and we denote by Ω the set of all logically possible states. An *aggregation rule* $\succcurlyeq_*$ assigns to each state $(S, R) \in \Omega$ a binary relation $\succcurlyeq_*^{(S,R)}$ over $\triangle S$. For exposition purposes we act as if there was a social planner that employs an aggregation rule to evaluate the alternatives in the menu. Hence, for a given state (S, R) we refer to $\succcurlyeq_*^{(S,R)}$ as the planner's *evaluation*.

We now impose axioms on how the social planner evaluates alternatives. *Rationality* says that the evaluation is rational in the sense of abiding by the vNM axioms.

Axiom RA (Rationality)**.** For each $(S, R) \in \Omega$, $\succcurlyeq_*^{(S,R)}$ satisfies the vNM axioms.

Rationality is normatively desirable and strengthens Arrow's requirement that the planner's evaluation is complete and transitive. Furthermore, we believe that, since individuals are assumed to be rational, an aggregation rule should preserve this characteristic of the individuals.

Our second axiom says that the social planner is benevolent, such that the evaluation respects the individuals' preferences whenever these are not in conflict.

Axiom SP (Strong Pareto)**.** For each $(S, R) \in \Omega$ and $x, y \in \triangle S$, if $x \succcurlyeq_i^R y$ for all $i \in N$, then $x \succcurlyeq_*^{(S,R)} y$. If, in addition, $x \succ_i^R y$ for some $i \in N$, then $x \succ_*^{(S,R)} y$.

SP strengthens Arrow's Pareto condition.

Our third axiom is *anonymity*. Anonymity says that the planner's evaluation must not depend on the individual identities but only on the preferences themselves. Hence, in a counterfactual world, where preferences are interchanged across the individuals, the planner's evaluation must be the same.

Axiom AN (Anonymity)**.** For each $(S, R), (S, R') \in \Omega$, if R' is a permutation of R, then $\succcurlyeq_*^{(S,R)} = \succcurlyeq_*^{(S,R')}$.

AN is an impartiality requirement and strengthens Arrow's non-dictatorship axiom.

Above we have stated our desiderata. Next, we will present the two conditions in Arrow (1950, 1963) that restrict what information can be used in the evaluation. The first condition is *independence of irrelevant alternatives.* We say that two binary relations $\succcurlyeq$ and $\succcurlyeq'$ *agree* on some set of alternatives S if for any $x, y \in \triangle S$, $x \succcurlyeq y$ if and only if $x \succcurlyeq' y$. Furthermore, we say that two preference profiles $R, R' \in \mathcal{R}^n$ agree on S if for each $i \in N$, $\succcurlyeq_i^R$ and $\succcurlyeq_i^{R'}$ agree on S.

Axiom IIA (Independence of Irrelevant Alternatives)**.** Fix $(S, R) \in \Omega$. For any $R' \in \mathcal{R}^n$, if R and R' agree on S, then $\succcurlyeq_*^{(S,R)} = \succcurlyeq_*^{(S,R')}$.

IIA says that the planner's evaluation of the menu cannot depend on individual preferences over alternatives outside the menu. Hence, in a counterfactual world, where individual preferences differ only on alternatives outside the menu, the planner's evaluation must be the same.

Arrow's second condition is that social preferences are menu-independent. Arrow assumes this implicitly, as he writes that the social choice from a menu S is made based on a single preference relation over A, which is independent of the menu. In light of the numerous ways to formalize this notion, we have selected the following axiom, as it best aligns with our subsequent relaxation of the condition.

Axiom MI (Menu Independence)**.** For each $(S, R) \in \Omega$ and $S' \subseteq S$, $\succcurlyeq_*^{(S,R)}$ and $\succcurlyeq_*^{(S',R)}$ agree on S'.

Menu independence says that removing alternatives from the menu does not change the planner's evaluation of the remaining alternatives.

It is well known that Arrow's axioms lead to an impossibility. Unsurprisingly, as our first three axioms strengthen Arrow's rationality, benevolence and impartiality requirements, the above axioms lead to an impossibility as well.[3]

Proposition I.1. There is no aggregation rule $\succcurlyeq_*$ that satisfies RA, SP, AN, IIA and MI.

[3]Note that formally, we weaken Arrow's *universal domain* condition, by assuming that individuals have vNM preferences. However, it has been shown that such a domain restriction is insufficient for escaping the impossibility. See Sen (1970), Kalai and Schmeidler (1977), Hylland (1980), Chichilnisky (1985), and Dhillon and Mertens (1997).

As we have argued in the introduction, we believe that both IIA and MI force the social planner to ignore valuable context and should therefore be reconsidered. We will weaken each of these conditions, using the following notion of comparability. Let $[a]$ denote the lottery that assigns probability 1 to the alternative $a \in A$.

Definition I.1. $a \in A$ is *comparable* relative to $B \subseteq A$ under $R \in \mathcal{R}^n$ if (i) $a \notin B$ and (ii) for every $i \in N$, there exists $x_i \in \triangle B$ such that $x_i \sim_i^R [a]$.

An alternative is comparable to a set if, for each individual, there is a pay-off in the set equal to that of the alternative.

First, we weaken MI.

Axiom MICA (Menu Independence of Comparable Alternatives)**.** For each $(S, R) \in \Omega$ and $S' \subseteq S$ where every $a \in S \setminus S'$ is comparable relative to S', $\succcurlyeq_*^{(S,R)}$ and $\succcurlyeq_*^{(S',R)}$ agree on S'.

MICA says that removing comparable alternatives from the menu does not change the planner's evaluation of the remaining alternatives.

Weakening MI to MICA results in a representation of the planner's evaluation we refer to as *menu-contingent utilitarianism*. For any $R \in \mathcal{R}^n$ and $B \subseteq A$, let $u_i^{B,R}$ denote the representation of $\succcurlyeq_i^R$ where $\max_{a \in B} u_i^{B,R}(a) = 1$ and $\min_{a \in B} u_i^{B,R}(a) = 0$, unless $\succcurlyeq_i^R$ is indifferent on B in which case $u_i^{B,R}(a) = 0$ for all $a \in B$.[4] Furthermore, for any $B \subseteq A$ we denote by $|B|$ the number of elements in B.

Theorem I.1 (Menu-Contingent Utilitarianism)**.** Let $|A| \geq 2n + 4$. An aggregation rule $\succcurlyeq_*$ satisfies RA, SP, AN, IIA and MICA if and only if for each $(S, R) \in \Omega$, $\succcurlyeq_*^{(S,R)}$ is represented by

$$\sum_{i \in N} u_i^{S,R}. \tag{I.1}$$

We sketch the proof of Theorem I.1 in Section I.3. A complete proof can be found in Appendix I.A. Note that no axiom of Theorem I.1 is implied by the other axioms of the theorem. Hence, a subset of the axioms would not suffice for the representation. We prove this in Appendix I.C.

Next, we weaken IIA. For any $B \subset A$ we write B^c to denote $A \setminus B$.

[4]For any binary relation $\succcurlyeq$ on a set X, a utility function $u : X \to \mathbb{R}$ is said to *represent* $\succcurlyeq$ if for all $x, y \in X$, $u(x) \geq u(y)$ if and only if $x \succcurlyeq y$.

Axiom IICA (Independence of Irrelevant Comparable Alternatives)**.** Fix $(S, R) \in \Omega$ and $C \subseteq S^c$ such that every $a \in C$ is comparable relative to C^c under R. For any $R' \in \mathcal{R}^n$, if R and R' agree on C^c and every $a \in C$ is comparable relative to C^c under R', then $\succsim_*^{(S,R')} = \succsim_*^{(S,R)}$.

IICA says that the planner's evaluation of the menu cannot depend on individual preferences over comparable alternatives outside the menu. Hence, in a counterfactual world, where individual preferences over these alternatives are different in a way such that these alternatives are still comparable, the planner's evaluation must be the same.

Weakening IIA to IICA results in a representation we refer to as *setting-contingent utilitarianism.*

Theorem I.2 (Setting-Contingent Utilitarianism)**.** Let $|A| \geq 2n + 4$. An aggregation rule $\succsim_*$ satisfies RA, SP, AN, IICA and MI if and only if for each $(S, R) \in \Omega$, $\succsim_*^{(S,R)}$ is represented by

$$\sum_{i \in N} u_i^{A,R}. \tag{I.2}$$

We prove Theorem I.2 in Appendix I.A. No axiom of Theorem I.2 is implied by the other axioms of the theorem, which we show in Appendix I.C.

In both theorems, we assume that there are at least $2n+4$ possible alternatives. We make this assumption, as it allows us to employ simple and insightful proofs. However, as shown by the following proposition, our proofs require only a few more alternatives than what is necessary for the axioms to be sufficient.

Proposition I.2. Let $|A| < 2n+1$. There exists an aggregation rule $\succsim_*$ which satisfies RA, SP, AN, IIA and MICA and which is not represented by (I.1). Similarly, there exists an aggregation rule $\succsim_*$ which satisfies RA, SP, AN, IICA and MI and which is not represented by (I.2).

We prove Proposition I.2 in Appendix I.A.

Finally, we want to mention *neutrality*, an additional desideratum that is satisfied by both our representations. Neutrality says that the labels of alternatives play no role in the planner's evaluation. Hence, if the labels of a and b were interchanged and a was preferred by the planner before, then b must be preferred afterward. We believe this to be the most important impartiality requirement besides anonymity. If neutrality wasn't implied by our other axioms, we would have imposed it directly. Since we will refer to neutrality in the upcoming sections, a formal definition is in order. We say that $\pi : A \mapsto A$ is a permutation of A if π is bijective and denote by Π the set of permutations of A. We abuse notation and define $\pi(S) := \{\pi(a) : a \in S \subseteq A\}$. We write $\pi(x) \in \triangle A$ to denote the lottery that for every $a \in A$ assigns probability $\mu \in [0, 1]$ to $\pi(a)$ if and only if x assigns probability μ to a. Let $\pi(R) := (\succcurlyeq_i^{\pi(R)})_{i \in N} \in \mathcal{R}^n$ denote the preference profile which has the same preferences on the permuted alternatives as R on the original alternatives. Formally, $\pi(x) \succcurlyeq_i^{\pi(R)} \pi(y)$ if and only if $x \succcurlyeq_i^R y$ for all $i \in N$ and $x, y \in \triangle A$.

Axiom NE (Neutrality)**.** For each $(S, R) \in \Omega$, $x, y \in \triangle S$ and $\pi \in \Pi$, $x \succcurlyeq_*^{(S,R)} y$ if and only if $\pi(x) \succcurlyeq_*^{(\pi(S),\pi(R))} \pi(y)$.

I.2.1 Discussion

Our objective was not to identify the weakest set of axioms leading to either of the above representations. Instead, we wanted to demonstrate that even with minor modifications to Arrow's axioms, preference aggregation becomes possible. To make this point convincingly, it is crucial that our axioms are as strong as they can be. That being said, whether weaker axioms would suffice remains an important question. This holds especially true in view of the existing literature that weakens IIA to *independence of redundant alternatives* (IRA); see Dhillon and Mertens (1999) and Brandl (2021). A redundant alternative is one for which there is a lottery over the remaining alternatives that leaves every individual indifferent between the two. While any redundant alternative is comparable, the converse does not hold. Consequently, IRA is weaker than IICA. We show in Appendix I.D that weakening IICA to IRA would not suffice for our representation. Similarly, weakening MICA to *menu independence of redundant alternatives* (MIRA) wouldn't suffice either. However, in another aspect IICA is indeed stronger than necessary. It would suffice if IICA only applied to cases where all alternatives outside the menu are comparable (i.e., $C = S^c$), akin to the formulation of IRA. We opted for the stronger version of the axiom as it more closely aligns with Arrow's IIA.

I.3 Sketch of Proof

To highlight some insights, this section provides a sketch of the proof. Since the proofs of both theorems are quite similar, we only sketch the proof of Theorem I.1.

Before we prove the theorem, we derive two interim results. Note that SP implies *Pareto indifference* (PI), which says that if every individual in society is indifferent between two lotteries, then so is the social planner. The first interim result states that if both RA and PI are satisfied, then the utility function of the social planner can be expressed as a weighted sum of the individual utility functions. This result is well known and has first been postulated by Harsanyi (1955). However, Harsanyi's original proof contains a mistake, which led to a variety of proofs by the subsequent literature. We, in turn, provide our own proof of Harsanyi's theorem, similar to those by Border (1985), Selinger (1986), and Hammond (1992), albeit ours is self-contained as we do not refer to mathematical theorems. Our proof makes use of the *pay-off matrix*, which indicates for each individual the vNM utility of every alternative in the menu.[5] Figure I.3 shows an example for three individuals and four alternatives. By RA, the planner's

$$\begin{array}{cc} & \begin{array}{cccc} a & b & c & d \end{array} \\ \begin{array}{c} \\ i{=}1 \\ i{=}2 \\ i{=}3 \end{array} & \begin{pmatrix} 1 & 1 & 1 & 1 \\ .5 & .5 & 0 & 1 \\ 1 & 0 & .7 & .3 \\ 0 & 1 & .2 & .8 \end{pmatrix} \end{array}$$

Figure I.3: Pay-off matrix.

evaluation can be represented by a row vector of vNM utilities as well. What needs to be shown is that whenever PI is satisfied, the planner's utility vector is equal to some linear combination of the rows in the pay-off matrix. If individual preferences are sufficiently diverse, such that the rows of the pay-off matrix span the entire vector space, then any logically possible vNM preference over the menu can be expressed by a linear combination of the individual utility functions. If individual preferences are not sufficiently diverse, we show that there exists a *dependent alternative*, whose column

[5]A row of 1's is included in the pay-off matrix to allow for a constant in the representation.

can be expressed by a linear combination of the other columns. For instance, in our example Column d is equal to a plus b minus c. We then sequentially drop dependent alternatives until the columns of the remaining alternatives are linearly independent. Then individual preferences over the remaining alternatives are sufficiently diverse such that any vNM preference over these alternatives can be expressed by a linear combination of rows. For each of the dependent alternatives, we can identify two lotteries, one of them involving the dependent alternative, such that every individual is indifferent between them. In the case of alternative d, these lotteries would be $\frac{1}{2}[a] + \frac{1}{2}[b]$ and $\frac{1}{2}[c] + \frac{1}{2}[d]$. By PI, also the planner must be indifferent between these lotteries, and hence, the planner's utility of the dependent alternative is pinned down by the same linear combination as for the individuals.

For our second interim result, we introduce the concept of *polar* alternatives. An alternative is polar if it is best among the menu for one individual and worst among the menu for everyone else. If the menu consists only of polar alternatives, we call such a state a *polar state*. Our second interim result then says that in a polar state, the planner is indifferent between all alternatives. Consider, for instance, a polar state as described by the pay-off matrix depicted on the left-hand side of Figure I.4. Note

$$
\begin{array}{c c}
 & \begin{array}{ccc} e & f & g \end{array} \\
\begin{array}{c} \\ i{=}1 \\ i{=}2 \\ i{=}3 \end{array} & \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}
\end{array}
\quad \xrightleftharpoons[\text{neutrality}]{\text{anonymity}} \quad
\begin{array}{c c}
 & \begin{array}{ccc} e & f & g \end{array} \\
\begin{array}{c} \\ i{=}1 \\ i{=}2 \\ i{=}3 \end{array} & \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}
\end{array}
$$

Figure I.4: Polar state.

that the pay-off matrix doesn't describe the state completely, as it does not specify individual preferences over infeasible alternatives. However, by IIA, preferences over infeasible alternatives can be ignored. We assume that the planner weakly prefers e over f and show that this implies a weak preference for f over e. This, of course, only leaves indifference. First, notice that the pay-off matrix on the right-hand side of Figure I.4 depicts the state that results from permuting the labels of Individuals 1 and 2 in the left-hand state. If e is weakly preferred to f in the left-hand state, by AN this must be the case in the right-hand state as well. Second, notice that the left-hand state results from permuting the labels of alternatives e and f in the right-hand state. If e is weakly preferred to f in the right-hand state, NE would demand that f must be weakly preferred to e in the left-hand state. With the help of MICA and PI, we show that indeed NE is satisfied in polar states. We leave the details to the formal proof in the appendix.

Finally, we prove the theorem. Consider individual preferences over the menu $\{a, b, c, d\}$ as depicted in Figure I.3 and assume that for each individual, there is a comparable polar alternative outside the menu, denoted by e, f, and g. In the menu $\{e, f, g\}$, the planner must be totally indifferent as we have shown previously. Now we add the original menu, resulting in Figure I.5. Since we have only added comparable

$$\begin{array}{cc}
 & \begin{array}{ccccccc} a & b & c & d & e & f & g \end{array} \\
\begin{array}{c} \\ i{=}1 \\ i{=}2 \\ i{=}3 \end{array} & \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ .5 & .5 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & .7 & .3 & 0 & 1 & 0 \\ 0 & 1 & .2 & .8 & 0 & 0 & 1 \end{pmatrix}
\end{array}$$

Figure I.5: Resulting pay-off matrix.

alternatives, by MICA, the planner must still be indifferent between e, f, and g. This indifference, together with our first interim result, then implies that equal weights on the individual utility functions represent the planner's evaluation. SP ensures that these common weights are positive and can be normalized to 1. Finally, we remove the polar alternatives, and by MICA, the same linear combination must still represent the planner's evaluation. By IIA, this must hold even if there are no comparable polar alternatives outside the initial menu. Note that if there are fewer than n alternatives outside the initial menu, the proof is more involved.

I.4 Applications

We apply our aggregation rules to three classic economic situations: aggregating the welfare of different consumers in a market, identifying the welfare of a dynamically inconsistent decision maker, and finding a fair solution to a bargaining problem. The literature has so far treated each of these problems in isolation. In the previous sections, we provided a unifying framework that aggregates individual preferences consistently across applications. With this methodology we uncover policy recommendations contrary to those by the established welfare criteria.

I.4.1 Consumer Welfare

Consider a society with n individuals and two consumption goods, m and g, where m is the numéraire. A social alternative is an allocation of these goods to the individuals in society, hence an element of $\mathbb{R}_+^{2n}$. Individuals are self-interested and their utility is quasi-linear, formally for each $i \in N$,

$$u_i((m_1, g_1), ..., (m_n, g_n)) = \alpha_i m_i + v_i(g_i) \tag{I.3}$$

for some $v_i : \mathbb{R}_+ \to \mathbb{R}$ and $\alpha_i \in \mathbb{R}_+$. The standard measure of aggregate welfare in this setting is *total consumer surplus*, short TCS, given by

$$\begin{aligned} W_{\text{TCS}}((m_1, g_1), ..., (m_n, g_n)) &= \sum_{i \in N} \frac{1}{\alpha_i} u_i((m_1, g_1), ..., (m_n, g_n)) \\ &= \sum_{i \in N} m_i + \sum_{i \in N} \frac{v_i(g_i)}{\alpha_i}. \end{aligned}$$

Consider two allocations, a and b, such that some individuals strictly prefer a and some strictly prefer b. If $W_{\text{TCS}}(a)$ is higher than $W_{\text{TCS}}(b)$, then there exist monetary transfers after a, resulting in an allocation a', such that a' Pareto dominates b. This makes TCS quite appealing. However, if these transfers aren't implemented, TCS makes a judgment on how the well-being of different individuals should be traded off. We show that the way in which these trade-offs are made violates neutrality. Specifically, TCS is sensitive to which of the two consumption goods is selected as the numéraire. Let $\mathcal{R}^{QL}$ be the subset of $\mathcal{R}^n$ such that individual preferences can be represented by quasi-linear utility functions as in (I.3).

Proposition I.3. There doesn't exist an aggregation rule $\succsim_*$ on the restricted domain $\mathcal{R}^{QL}$ that is represented by W_{TCS} and satisfies NE.

Proof. It is sufficient to show that the axiom is violated in some state, so consider a state $R \in \mathcal{R}^{QL}$ where for each $i \in N$, $v_i(g_i) = \beta_i g_i$ for some $\beta_i > 0$. Next consider the permutation π of alternatives such that $\pi((m_1, g_1), ..., (m_n, g_n)) = ((g_1, m_1), ..., (g_n, m_n))$ and note that $\pi(R) \in \mathcal{R}^{QL}$. For NE to be satisfied, it must hold that

$$W^R_{\text{TCS}}(a) \geq W^R_{\text{TCS}}(b) \text{ if and only if } W^{\pi(R)}_{\text{TCS}}(\pi(a)) \geq W^{\pi(R)}_{\text{TCS}}(\pi(b)). \tag{I.4}$$

Let $a = ((0,1), (0,0), ..., (0,0))$ and $b = ((0,0), (2,0), ..., (0,0))$. Then $W^R_{\text{TCS}}(a) = \frac{\beta_1}{\alpha_1}$, $W^R_{\text{TCS}}(b) = 2$, $W^{\pi(R)}_{\text{TCS}}(\pi(a)) = 1$ and $W^{\pi(R)}_{\text{TCS}}(\pi(b)) = \frac{\alpha_2}{\beta_2}$. This violates (I.4) for instance when $\beta_1 = 3$, $\alpha_1 = 1$, $\alpha_2 = 2$ and $\beta_2 = 1$. □

Now assume that allocations result from the following setting. A monopolist produces the good g at constant marginal cost c. Individuals can be divided into two distinct groups: students s and adults a. Within each group, every individual has the same preferences. The fraction of individuals belonging to Group $J \in \{s, a\}$ is denoted by γ_J. Total demand of Group J is then $D_J(p) := n\gamma_J(v_i')^{-1}(p\alpha_i)$ where $i \in J$. We assume that the monopolist can identify which group an individual belongs to and hence can choose a price pair $(p_s, p_a) \in [c, \infty)^2$, where p_J denotes the price charged to Group J. We denote by $\psi(p_s, p_a) \in \mathbb{R}^{2n}_+$ the allocation resulting from a price pair (p_s, p_a). Let p^*_J denote the monopoly price charged to Group J, let p^* denote the profit-maximizing price if price discrimination was prohibited, and assume

$$p^*_a > p^* > p^*_s.$$

A policymaker can choose whether to allow or prohibit price discrimination, hence $S = \{\psi(p^*_s, p^*_a), \psi(p^*, p^*)\}$. To guide this decision, the policymaker wants to assess whether the prohibition of price discrimination benefits consumers. According to TCS, aggregate welfare is given by

$$W_{\text{TCS}}(\psi(p_s, p_a)) = CS_s(p_s) + CS_a(p_a)$$

where $CS_J(p) := \int_p^\infty D_J(p)dp$. Since total consumer surplus violates neutrality, we propose setting-contingent utilitarianism as an alternative. The setting puts natural bounds on the possible alternatives, namely $A = \{\psi(p_s, p_a) \in \mathbb{R}_+^{2n} : (p_s, p_a) \in [c, \infty)^2\}$. Setting-contingent utilitarianism, short SCU, is then given by

$$W_{\text{SCU}}(\psi(p_s, p_a)) = \frac{n\gamma_s}{CS_s(c)} CS_s(p_s) + \frac{n\gamma_a}{CS_a(c)} CS_a(p_a).$$

We now compare these two welfare measures. To make things simple, we consider the textbook case of linear demand. It is well known that when demand is linear, total consumer surplus is higher under price discrimination if and only if one group isn't served under the uniform price. Our welfare measure, on the other hand, suggests that price discrimination can be socially beneficial, even if both groups are served under uniform pricing. We demonstrate this with the help of the following example. Figure I.6 shows the demand and monopoly prices for $D_s(p) = 2000 - 50p$, $D_a(p) = 7000 - 100p$, and $c = 0$. If the firm is allowed to price discriminate, it will charge

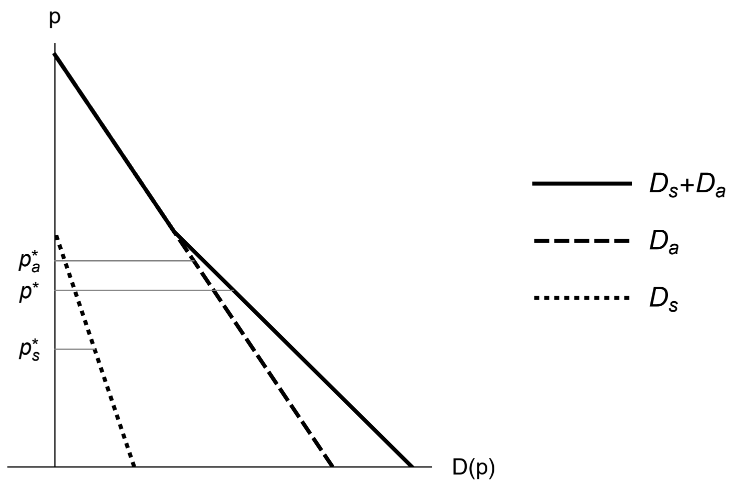


Figure I.6: Linear demand and monopoly prices.

\$20 to students and \$35 to adults, which yields a consumer surplus of $CS_s(\$20) = \$10,000$ and $CS_a(\$20) = \$61,250$. If price discrimination were prohibited, the firm would charge \$30 to both, which yields a consumer surplus of $CS_s(\$30) = \$2,500$ and $CS_a(\$30) = \$80,000$. Note that TCS is higher when price discrimination is prohibited. For both groups to be better off under uniform pricing, adults would have to pay between \$7,500 and \$18,750 to students. However, transfers are typically not implemented, which makes students worse off under uniform pricing. Going from price discrimination to uniform pricing, the normalized utility of students decreases from 0.25 to roughly 0.06, while it increases for adults from 0.25 to roughly 0.33. If both groups consist of an equal number of individuals, then total utility is higher under price discrimination and is therefore favored by SCU. In a sense, SCU penalizes adults through a lower weight on their surplus in order to account for the missing transfers. Note that if transfers would, in fact, be implemented, then SCU would agree with TCS since both satisfy SP. Furthermore, SCU accounts for how many individuals are positively and negatively affected by the policy, whereas TCS ignores the shares of individuals in each group. If, for instance, the share of students was below roughly 0.29, then SCU would favor the prohibition of price discrimination, as sufficiently many adults would be positively affected by the policy. SCU values individuality and has a flavor similar to the majority principle. Finally, note that SCU, unlike TCS, can be meaningfully applied even when individual utilities are not quasi-linear.

I.4.2 Intertemporal Welfare

Consider a finitely lived decision maker (DM) who consumes a good in each period up to period n. As in Laibson (1997), we assume the DM discounts consumption quasi-hyperbolically, such that the utility of a consumption sequence $(c_1, ..., c_n) \in \mathbb{R}^n_+$ experienced in period t is given by

$$u_t(c_t, ..., c_n) = v(c_t) + \beta \sum_{i=t+1}^{n} \delta^{i-t} v(c_i) \tag{I.5}$$

for $\beta, \delta \in [0, 1]$ and some monotone per-period valuation $v : \mathbb{R}_+ \to \mathbb{R}$. If $\beta < 1$, the DM is dynamically inconsistent, meaning there exist sequences $(c_1, ..., c_n)$, $(c_1, ..., c_t, c'_{t+1}, ..., c'_n) \in \mathbb{R}^n_+$ such that the DM in Period t strictly prefers $(c_1, ..., c_n)$ over $(c_1, ..., c_t, c'_{t+1}, ..., c'_n)$, while in Period $t + 1$ her preference is reversed. It is as if the DM consists of different *selves*, Self 1 to Self n, with conflicting interests. If we want to assess which of these two sequences is better for the DM overall, we need a welfare criterion that incorporates the perspective of each self. Two criteria are common in the literature, the Pareto criterion and *long-run utility* (O'Donoghue & Rabin, 1999) given by

$$W_{\text{LRU}}(c_1, ..., c_n) := \sum_{t=1}^{n} \delta^{t-1} v(c_i).$$

Before we discuss these criteria, let us first consider the case where $\beta = 1$, such that the DM is dynamically consistent. It is conventional wisdom among economists that, in this case, there is no conflict between the selves, and the appropriate welfare measure is the utility of Self 1. However, this view ignores the possibility that, even though there is no conflict looking forward, there could be a conflict looking backward. For instance, the DM might regret an earlier decision and prefer they had saved more in the past while simultaneously agreeing with the earlier self on how much to save today. To formalize this idea, let each self t have a vNM preference $\succcurlyeq_t$ over $\triangle(\mathbb{R}^n_+)$. This means that Self t can compare sequences that differ in the consumption levels before Period t. However, note that since one cannot affect one's past, $\succcurlyeq_t$ is only partially revealed. Let $\mathcal{R}^{QH}$ be the subset of $\mathcal{R}^n$ such that the revealed part of the individual preference can be represented by quasi-hyperbolically discounted utility as in (I.5). Formally, for any $(\succcurlyeq_1, ..., \succcurlyeq_n) \in \mathcal{R}^{QH}$, $(c_1, ..., c_n) \in \mathbb{R}^n_+$ and $t \in \{1, ..., n\}$, preferences of Self t over lotteries that assign probability 1 on sequences starting with $(c_1, ..., c_{t-1})$ can be represented by (I.5). Note that the selves are in conflict unless $\succcurlyeq_i = \succcurlyeq_j$ for all $i, j \in \{1, ..., n\}$. Hence, for all selves to agree, the DM would have to value past consumption more than current consumption and value it more the further it lies in the past. Since this doesn't seem very plausible, we would expect backward-looking disagreements to be common. Now that all selves have preferences over the same domain, we can view these conflicting interests from the perspective of preference aggregation. This enables us to demonstrate that taking Self 1's utility as a measure of total welfare whenever $\beta = 1$ violates our Pareto condition.

Proposition I.4. Let $\succcurlyeq_*$ be an aggregation rule on the restricted domain $\mathcal{R}^{QH}$, such that $\succcurlyeq_*^{(S,R)} = \succcurlyeq_1^R$ whenever $\beta = 1$. Then $\succcurlyeq_*$ violates SP.

Proof. Note that it is sufficient to show that the axiom is violated in some state. Consider the state where each self is *past indifferent*, meaning for any $t \in \{2, ..., n\}$, $(c_1, ..., c_n), (c'_1, ..., c'_{t-1}, c_t, ..., c_n) \in \mathbb{R}^n_+$,

$$(c_1, ..., c_n) \sim_t (c'_1, ..., c'_{t-1}, c_t, ..., c_n).$$

Now consider two sequences $s = (c_1, ..., c_n)$ and $s' = (c'_1, c'_2, c_3..., c_n)$ such that $c_1 > c'_1$ and $v(c_1) + \delta v(c_2) = v(c'_1) + \delta v(c'_2)$. Then $s \sim_1 s'$, $s' \succ_2 s$ and $s \sim_t s'$ for all $t \geq 3$ and hence by SP s' should give strictly higher total welfare than s. If, however, u_1 measures total welfare of the DM, then s and s' are equally desirable. $\square$

Let us now return to the aforementioned welfare criteria. The Pareto principle cannot compare every sequence in $\mathbb{R}^n_+$; hence, is incomplete and violates RA. Long-run utility is equal to u_1 for $\beta = 1$, and therefore, as shown by Proposition I.4, violates SP. As an alternative to the established criteria, we propose menu-contingent utilitarianism. We assume that the social planner can distribute an initial endowment of size 1 across periods, hence $S = \{(c_1, ..., c_n) \in [0,1]^n : \sum_1^n c_t \leq 1\}$. Furthermore, in order to apply our criterion, we have to make assumptions on individual preferences over histories. Note that the same is true for the Pareto principle (see Goldman (1979)). As a benchmark, we assume past indifference. Then, preferences over entire sequences are represented by (I.5). Next, we have to normalize each self's utility with respect to the best and worst alternative in S. Without loss of generality, assume that $v(0) = 0$ and let $\bar{u}_t = \max_{s \in S} u_t(s)$ denote the optimal allocation from the perspective of Self t. Then, total welfare of the DM according to menu-contingent utilitarianism is given by

$$W_{\text{MCU}}(c_1, ..., c_n) := \sum_{t=1}^{n} \frac{1}{\bar{u}_t} u_t(c_t, ..., c_n) = \sum_{t=1}^{n} \underbrace{\left(\frac{1}{\bar{u}_t} + \beta \sum_{i=1}^{t-1} \frac{\delta^i}{\bar{u}_{t-i}} \right)}_{\gamma_t} v(c_t).$$

In contrast to long-run utility, the weights γ_t on the per period valuation are increasing in t. Hence, our criterion recommends an increasing consumption profile. This is because future consumption has a positive externality on earlier selves in the form of anticipatory utility, while later selves do not benefit from past consumption. Of course, this is driven by our assumption of past indifference. We leave it to future research to determine whether this assumption is plausible. Finally, note that MCU can be applied even when the DM is not a quasi-hyperbolic discounter.

I.4.3 Bargaining

Consider n individuals who have the possibility to cooperate and create a surplus. In order to generate this surplus, they must agree on an alternative to be implemented. For instance, one could think of a buyer and seller bargaining over the price or a worker and an employer negotiating a wage. If the group cannot reach an agreement, each individual resorts to their respective outside options, which we call the *disagreement* *alternative* and denote by a_0. In the following exposition, we consider bargaining through the eyes of an arbitrator who, to facilitate cooperation, has to choose an alternative for the group. Equivalently, one could phrase the selection as a consequence of fairness postulates.

The literature on axiomatic bargaining suggests that the arbitrator adheres to the following approach. The problem is first reformulated in terms of vectors of utilities, and then a bargaining solution is applied. This reformulation is discussed in more detail below. Here, we recall the bargaining approach. A *bargaining problem* (d, U) specifies a set of utility vectors $U \subset \mathbb{R}^n$, called the *bargaining set*, and a *disagreement point* $d = (d_1, ..., d_n) \in U$. We assume that U is compact and denote by $\mathcal{B}$ the set of all such bargaining problems. A bargaining solution f assigns to each (d, U) a subset of U, hence $f(d, U) \subseteq U$. Two bargaining solutions are prominent in the literature: the *Nash bargaining solution* (Nash, 1950), short Nash, and the *Kalai-Smorodinsky bargaining solution* (Kalai & Smorodinsky, 1975), short KS. Note that both solutions require the bargaining set to be convex. Denote by $\mathcal{B}_{\text{con}}$ the subset of $\mathcal{B}$ for which U is convex. Then Nash is given by

$$f_{\text{Nash}}(d, U) := \underset{(v_1, ..., v_n) \in U}{\arg\max} \prod_{i \in N} (v_i - d_i)$$

for all $(d, U) \in \mathcal{B}_{\text{con}}$. Next, we define KS, which requires some additional notation. For any $(d, U) \in \mathcal{B}$, let $\Lambda(d, U) := \{(v_1, ..., v_n) \in U : v_i \geq d_i \text{ for all } i \in N\}$. Hence, $\Lambda(d, U)$ is the set of utility vectors in U that weakly Pareto-dominate d. Furthermore, let $\alpha_i(d, U)$ denote i's maximal utility among all points in $\Lambda(d, U)$. Note that KS only applies for $n = 2$. So $f_{\text{KS}}(d, U)$ is the intersection of the Pareto frontier of U and the line connecting d and $(\alpha_1(d, U), \alpha_2(d, U))$ for all $(d, U) \in \mathcal{B}_{\text{con}}$.

We now investigate whether these bargaining solutions could be used by an arbitrator who first evaluates the different alternatives and then selects the best according to this evaluation. Let A denote the set of alternatives and note that $a_0 \in A$. Let $\mathcal{U}$ denote the set of possible vNM utility functions over $\triangle A$. A utility profile $u \in \mathcal{U}^n$ specifies for each individual a utility function over the possible alternatives. We say that a bargaining problem $(d, U) \in \mathcal{B}$ is *associated* with a utility profile $u \in \mathcal{U}^n$ (and vice versa) if $d = (u_1(a_0), ..., u_n(a_0))$ and $U = \{(u_1(a), ..., u_n(a)) : a \in A\}$. To ensure that any bargaining problem in $\mathcal{B}$ can be generated by some utility profile in $\mathcal{U}^n$, we require A to be sufficiently rich. In line with our motivating examples, the reader might think of A as the set of prices or wages, i.e., $A = \mathbb{R}_+$. We denote by $\succcurlyeq_+$ the arbitrator's aggregation rule, such that for every $u \in \mathcal{U}^n$, $\succcurlyeq_+^u$ is a binary relation over $\triangle A$. Note that we deviate from the aggregation rule as defined in Section I.2 in two ways. We have indicated this by the use of a different subscript. First, for notational convenience, $\succcurlyeq_+$ is a function of utility profiles rather than preference profiles. Second, $\succcurlyeq_+$ does not depend on the menu because, as we will argue later, in a bargaining setting, the menu is determined endogenously. Neither of these deviations is responsible for our subsequent finding. Furthermore, in line with Section I.2 we require the arbitrator to evaluate lotteries over A. We don't insist that randomization is feasible, but if it were feasible, the arbitrator should be able to evaluate the resulting lotteries. Finally, we connect the aggregation rule to a bargaining solution through the following definition.

Definition I.2. $\succcurlyeq_+$ is *consistent* with f if for any $(d, U) \in \mathcal{B}_{\text{con}}$ and associated $u \in \mathcal{U}^n$, the following holds. For all $x \in \triangle A$,

$$(u_1(x), ..., u_n(x)) \in f(d, U) \text{ if and only if } x \in \{y \in \triangle A : y \succcurlyeq_+^u z \text{ for all } z \in \triangle A\}.$$

The definition captures an implicit assumption of the bargaining approach, namely that alternatives are selected solely based on their utility vectors. Hence, if some lottery's utility vector is chosen by the bargaining solution, this lottery must be among the most preferred. Conversely, if a lottery is among the most preferred, its utility vector must be chosen by the bargaining solution. We have restricted the definition to convex bargaining sets as this is the domain of the prominent bargaining solutions.

We can now investigate whether a given bargain solution is consistent with certain desirable features of our aggregation rule. We find that neither Nash nor KS is consistent with rationality. More generally, any bargaining rule that imposes ex-ante symmetry is inconsistent with the vNM postulates. This is well known; see for instance Diamond (1965), Harsanyi (1975), Broome (1984) or Machina (1989). Nevertheless, we provide the following proposition for the sake of completeness.

Proposition I.5. There doesn't exist an aggregation rule $\succcurlyeq_+$ that satisfies RA and is consistent with either f_{Nash} or f_{KS}.

Proof. We prove the proposition for $n = 2$ as this is required for KS. The proof can be easily extended to a general n. The proof strategy goes as follows. We assume that $\succsim_+$ is consistent with a bargaining solution f that satisfies Nash's *symmetry* and Pareto axiom. Symmetry says that in a symmetric bargaining problem, any element of the bargaining solution must assign equal utility to all individuals. Note that both Nash and KS satisfy symmetry. We then show that $\succsim_+$ must violate RA.

Consider the bargaining problem $((0,0),\mathcal{S})$ where $\mathcal{S}$ is the unit simplex. As $((0,0),\mathcal{S})$ is symmetric, f selects $\{(0.5,0.5)\}$. See Figure I.7 for an illustration. For any

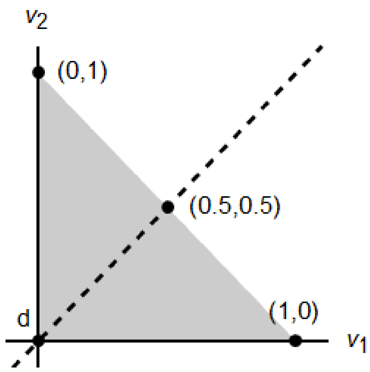


Figure I.7: Symmetric bargaining solution for $((0,0),\mathcal{S})$.

$u \in \mathcal{U}^2$ associated with $((0,0),\mathcal{S})$, there must exist two alternatives $a_1, a_2 \in A$ that generate the extreme points of $\mathcal{S}$, i.e., $(u_1(a_1), u_2(a_1)) = (1,0)$ and $(u_1(a_2), u_2(a_2)) = (0,1)$. Now consider the coin-flip between a_1 and a_2, denoted by $x := \frac{1}{2}[a_1] + \frac{1}{2}[a_2]$. As $(u_1(x), u_2(x)) = (0.5, 0.5) \in f((0,0),\mathcal{S})$, x must be among the most preferred elements according to $\succsim_+^u$. Formally $x \in \{y \in \triangle A : y \succsim_+^u z \text{ for all } z \in \triangle A\}$. But neither a_1 nor a_2 are among the most preferred elements, as neither $(1,0)$ nor $(0,1)$ is selected by f. Therefore,

$$\frac{1}{2}[a_1] + \frac{1}{2}[a_2] \succ_+^u [a_1], \tag{I.6}$$

$$\frac{1}{2}[a_1] + \frac{1}{2}[a_2] \succ_+^u [a_2]. \tag{I.7}$$

By the vNM independence axiom, (I.6) would imply $[a_2] \succ_+^u [a_1]$ and (I.7) would imply $[a_1] \succ_+^u [a_2]$, a contradiction. Hence, the vNM independence axiom and consequently RA must be violated. □

We believe that the arbitrator should make the decision based on a rational evaluation of the alternatives. As this is not reconcilable with the prominent bargaining solutions, we offer menu-contingent utilitarianism as an alternative. The menu arises naturally in this setting. Because an individual would not agree to an alternative worse than a_0, it is the set of alternatives that weakly Pareto-dominate a_0. Formally, given $u \in \mathcal{U}^n$ the menu is $S := \{a \in A : u_i(a) \geq u_i(a_0) \text{ for all } i \in N\}$. Note that a bargaining problem (d, U) contains all relevant information for our rule to evaluate the utility vectors in $\Lambda(d, U)$. Hence, menu-contingent utilitarianism leads to a well-defined bargaining solution, which we abbreviated by MCU and which is given by

$$f_{\text{MCU}}(d, U) := \underset{(v_1,\ldots,v_n)\in\Lambda(d,U)}{\arg\max} \sum_{i\in N} \frac{v_i - d_i}{\alpha_i(d, U) - d_i}$$

for all $(d, U) \in \mathcal{B}$. MCU selects the utility vectors that are most preferred under a menu-contingent utilitarian evaluation. Note that MCU does not require convexity of the bargaining set. The same bargaining solution has been axiomatized before by Pivato (2009) under the assumption of convexity of the bargaining set. A similar bargaining solution has been axiomatized by Cao (1982) and Baris (2018), again assuming convexity of the bargaining set. Their solution differs from ours, as they normalize individual utilities with respect to each individual's highest utility in U, whereas we normalize with respect to each individual's highest utility in $\Lambda(d, U)$.

Like Nash and KS, MCU has a neat geometric interpretation, which we demonstrate with the help of Figure I.8. The disagreement point is the origin and the curve shows

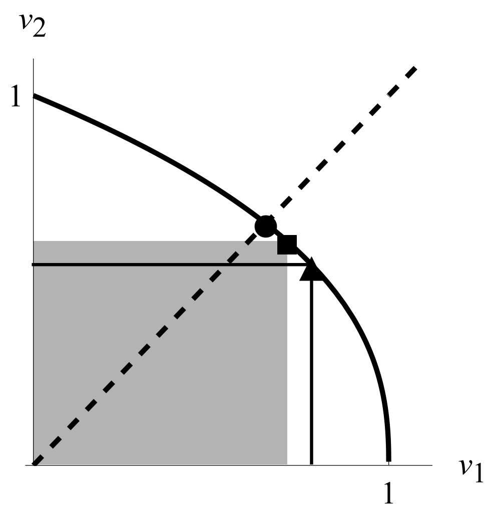


Figure I.8: Nash, KS and MCU.

the Pareto frontier of bargaining set for $u_1(a) = 1 - a$ and $u_2(a) = a^{\frac{2}{5}}$. KS (circle) is the intersection of the Pareto frontier and the dashed 45° line, Nash (square) maximizes the area of the rectangle spanned by d and the Pareto frontier and MCU (triangle) maximizes the circumference of the rectangle spanned by d and the Pareto frontier. If the Pareto frontier is smooth, then the slope of the Pareto frontier at MCU is -1, as it would otherwise be possible to increase the utility of one individual by decreasing the utility of the other individual by a lesser amount. Hence, in this example, MCU favors Individual 1 relative to KS because the slope at KS is flatter than -1 such that Individual 1 can be made significantly better off at the cost of making Individual 2 slightly worse off.

I.5 Conclusion

We initiated this project in pursuit of a rule that is rational, benevolent, and impartial, as these desiderata are not satisfied by the rules that are typically used in applications. We learned from Arrow (1950, 1963) that for a rule to satisfy these desiderata, it must be sensitive to context. We offer two rules, each motivated by adhering closely to Arrow. Each rule results from slightly weakening one of Arrow's two context-restricting axioms, MI and IIA. Each rule makes an explicit normative judgment about what is allowed to matter. We believe that in different situations, different considerations are allowed to matter. Consequently, we find that there is no single unifying rule. This revelation could pave the way for numerous rules derived from various modifications of MI and IIA. However, if we permit a planner to tailor the context to a specific application, almost any social choice can be justified. To maintain discipline, we recommend selecting from our two rules, as each represents a normative judgment on one side of the spectrum.

The menu-contingent utilitarian rule satisfies IIA and thus reflects the judgment that nothing outside the menu is allowed to matter. However, the menu itself does matter. We believe this normative consideration to be relevant for bargaining situations, as bargaining involves finding a compromise that provides each individual with a fair share of the surplus. Our understanding of fairness is relative to what can be achieved and must disregard alternatives that individuals would never agree upon. We also propose this rule for practical convenience in situations where there is no information about individual preferences outside the menu or when the setting does not impose natural bounds on individual utilities.

On the other side of the spectrum is the setting-contingent utilitarian rule. This rule satisfies MI, hence reflects the judgment that the menu itself is not allowed to matter. Individual preferences over all possible alternatives are taken into account, independently of whether an alternative is currently feasible. We believe this normative consideration to be relevant for policymaking, where the policymaker should take a birds-eye view and assess welfare on the basis of the entire setting. In addition, policymakers often face binary decisions sequentially, namely whether to adopt a specific policy or not. Menu independence ensures that these decisions are consistent across time.

Appendix I.A

This section contains the proofs of Theorem I.1, Theorem I.2, Proposition I.2, and some additional propositions and lemmas required for the proofs of the theorems.

First off, note that SP implies *Pareto indifference*.

Axiom PI (Pareto Indifference)**.** For any $(S,R) \in \Omega$ and $x, y \in \triangle S$, if $x \sim_i^R y$ for all $i \in N$, then $x \sim_*^{(S,R)} y$.

We use this axiom for our first interim result.

Proposition I.6. Let RA and PI be satisfied. Then for any $(S,R) \in \Omega$ and for any representations $(u_i)_{i \in N}$ of R and u_* of $\succsim_*^{(S,R)}$, there exists $(\lambda_i)_{i=0}^n \in \mathbb{R}^{n+1}$, such that for all $a \in S$,

$$u_*(a) = \lambda_0 + \sum_{i \in N} \lambda_i u_i(a).$$

Proof. Fix $(S,R) \in \Omega$. We denote the alternatives in S by a_1 to a_m where $m = |S|$. The pay-off vector $\vec{u}_i := (u_i(a_1), ..., u_i(a_m))$ describes individual $i's$ vNM preferences over S. By RA, the representation u_* of $\succsim_*^{(S,R)}$ must be an expected utility representation and is therefore fully described by a pay-off vector as well. We denote this vector by

$\vec{u}_* := (u_*(a_1), ..., u_*(a_m))$. Let

$$M := \begin{bmatrix} (1, ..., 1) \\ \vec{u}_1 \\ \vdots \\ \vec{u}_n \end{bmatrix}$$

be the *pay-off matrix*, which describes individual preferences over S. Note that the proposition states that $\vec{u}_*$ is equal to some linear combination of the rows in M. We distinguish two cases. In the first case, individual preferences are sufficiently diverse, such that there are m linearly independent rows in M, which span the entire m-fold vector space. Hence, a linear combination of rows in M equal to $\vec{u}_*$ exists trivially, even without invoking PI. Next, consider the case where individual preferences are not sufficiently diverse, such that not every vector of length m can be expressed as a linear combination of rows. In this case, the maximal number of linearly independent rows is strictly below m. Hence, the maximal number of linearly independent columns must be strictly below m as well. As M has m columns, linearly dependent columns must exist. Sequentially drop linearly dependent columns from M until all remaining columns are linearly independent. Denote the set of alternatives associated with the remaining columns by B. We call the alternatives in B the *independent alternatives* and the alternatives in $S \setminus B$ the *dependent alternatives*. Note that in the resulting matrix, the rows span the entirety of the reduced vector space. Hence, by dropping the dependent alternatives, preferences are again sufficiently diverse to express any utility vector for the independent alternatives. This means that we can find a linear combination of rows of M that matches $\vec{u}_*$ in the utilities for the independent alternatives. Now consider the dependent alternatives. For each $a \in S \setminus B$, there must exist a linear combination of columns $\gamma^a : B \to \mathbb{R}$ such that

$$\sum_{b \in B} \gamma^a(b) \vec{u}(b) = \vec{u}(a). \tag{I.8}$$

We will now show that each γ^a can be decomposed into two lotteries $\gamma^a_+, \gamma^a_- \in \triangle S$ such that every individual is indifferent between these lotteries. Let $B^a_+ := \{b \in B : \gamma^a(b) \geq 0\}$ and $B^a_- := \{b \in B : \gamma^a(b) < 0\}$. Since by definition the first row of M consists only of 1's, $\sum_{b \in B} \gamma^a(b) = 1$ and furthermore $k := \sum_{b \in B^a_+} \gamma^a(b) = 1 - \sum_{b \in B^a_-} \gamma^a(b)$. Now define two lotteries

$$\gamma_+^a(b) := \begin{cases} \frac{1}{k}\gamma^a(b) & b \in B_+^a \\ 0 & b \notin B_+^a \end{cases} \qquad \text{and} \qquad \gamma_-^a(b) := \begin{cases} -\frac{1}{k}\gamma^a(b) & b \in B_-^a \\ \frac{1}{k} & b = a \\ 0 & b \notin B_-^a \cup \{a\} \end{cases}$$

Note that for each $i \in N$, $\sum_{b\in S} \gamma_+^a(b)u_i(b) = \sum_{b\in S} \gamma_-^a(b)u_i(b)$, meaning each individual is indifferent between the two lotteries. By PI, the planner must be indifferent as well. Therefore, $\sum_{b\in S} \gamma_+^a(b)u_*(b) = \sum_{b\in S} \gamma_-^a(b)u_*(b)$ and furthermore

$$\sum_{b\in B} \gamma^a(b)u_*(b) = u_*(a). \tag{I.9}$$

Compare this to Equation (I.8). The planner's utility for any dependent alternatives is determined from the independent alternatives in the same way as for every individual. Therefore, if a linear combination of rows of M matches the planner's utilities for the independent alternatives, it must also match the utilities for the dependent alternatives. □

We say that a state $(S, R) \in \Omega$ is *polar* if (i) S has exactly n elements, which we denote by p_1 to p_n, (ii) for every $i \in N$, $p_j \sim_i^R p_k$ for all $j, k \in N \setminus \{i\}$ and (iii) either $p_i \succ_i^R p_j$ for all $j \in N \setminus \{i\}$ or $p_i \sim_i^R p_j$ for all $j \in N \setminus \{i\}$. We say that p_i is i's *polar alternative*.

Proposition I.7. Let RA, SP, AN, IIA and MICA be satisfied. Then for any polar state $(S, R) \in \Omega$, if $i, j \in N$ have a strict preference on S, then $p_i \sim_*^{(S,R)} p_j$.

Proof. Assume $(S, R) \in \Omega$ is polar and $i, j \in N$ have strict preferences on S. If $i = j$, the proposition is trivially satisfied, so assume $i \neq j$. We assume $p_i \succcurlyeq_*^{(S,R)} p_j$ and show that $p_j \succcurlyeq_*^{(S,R)} p_i$ is implied, which only leaves $p_i \sim_*^{(S,R)} p_j$. We prove the proposition by going through a sequence of preference profiles and menus. We use Roman numerals as subscripts to keep track of the different preference profiles and menus. Let $R_{\mathrm{I}} \in \mathcal{R}^n$ denote the permutation of R where only preferences of i and j are permuted. By AN, $p_i \succcurlyeq_*^{(S,R_{\mathrm{I}})} p_j$. Let $R_{\mathrm{II}} \in \mathcal{R}^n$ denote a preference profile, which agrees with R_{I} on S and where there is $q_i, q_j \in S^c$ such that $q_i \sim_k^{R_{\mathrm{II}}} p_i$ and $q_j \sim_k^{R_{\mathrm{II}}} p_j$ for all $k \in N$. By our assumption that A has at least $2n + 4$ elements, such alternatives must exist. By IIA, $p_i \succsim_*^{(S,R_{\mathrm{II}})} p_j$. Let $S_{\mathrm{I}} := S \cup \{q_i, q_j\}$. Then $p_i \succsim_*^{(S_{\mathrm{I}},R_{\mathrm{II}})} p_j$ by MICA. Furthermore, $p_i \sim_*^{(S_{\mathrm{I}},R_{\mathrm{II}})} q_i$ and $p_j \sim_*^{(S_{\mathrm{I}},R_{\mathrm{II}})} q_j$ by SP and hence $q_i \succsim_*^{(S_{\mathrm{I}},R_{\mathrm{II}})} q_j$ by RA. Let $S_{\mathrm{II}} := S_{\mathrm{I}} \setminus \{p_i, p_j\}$. Then $q_i \succsim_*^{(S_{\mathrm{II}},R_{\mathrm{II}})} q_j$ by MICA. Let $R_{\mathrm{III}} \in \mathcal{R}^n$ denote a preference profile which agrees with R_{II} on S_{II} and where $q_i \sim_k^{R_{\mathrm{III}}} p_j$ and $q_j \sim_k^{R_{\mathrm{III}}} p_i$ for all $k \in N$. Then $q_i \succsim_*^{(S_{\mathrm{II}},R_{\mathrm{III}})} q_j$ by IIA, $q_i \succsim_*^{(S_{\mathrm{I}},R_{\mathrm{III}})} q_j$ by MICA and $p_j \succsim_*^{(S_{\mathrm{I}},R_{\mathrm{III}})} p_i$ by SP and RA. Furthermore, $p_j \succsim_*^{(S,R_{\mathrm{III}})} p_i$ by MICA. Note that R_{III} and R agree on S. Therefore, $p_j \succsim_*^{(S,R)} p_i$ by IIA, which implies $p_i \sim_*^{(S,R)} p_j$. □

We now show two properties of binary relations that we will need for the proofs of the theorems.

Lemma I.1. Let $\succsim$, $\succsim'$ and $\succsim''$ be binary relations over $\triangle S$. If $\succsim$ and $\succsim'$ agree on $B \subseteq S$ and $\succsim'$ and $\succsim''$ agree on $C \subseteq S$, then $\succsim$ and $\succsim''$ agree on $B \cap C$.

Proof. Consider any $B, C \subseteq S$ s.t. $B \cap C \neq \emptyset$ and any $x, y \in \triangle(B \cap C)$. If $\succsim$ and $\succsim'$ agree on B, then $x \succsim y$ if and only if $x \succsim' y$, and if $\succsim'$ and $\succsim''$ agree on C, then $x \succsim' y$ if and only if $x \succsim'' y$. Hence, for any $x, y \in \triangle(B \cap C)$, $x \succsim y$ if and only if $x \succsim'' y$, meaning $\succsim$ and $\succsim''$ agree on $B \cap C$. □

Lemma I.2. Let $\succsim$ be a binary relation over $\triangle S$ satisfying RA. For any $u : S \to \mathbb{R}$, if for each distinct $b, c, d \in S$ the part of $\succsim$ on $\triangle\{b, c, d\}$ is represented by u, then $\succsim$ is represented by u.

Proof. Let u satisfy the premise above. Select $b, c, d \in S$ such that $b \succ c$. If this is not possible, then $\succsim$ must be totally indifferent on S, in which case the proof is trivial. By assumption, u represents $\succsim$ on $\triangle\{b, c, d\}$. Let $\hat{u} : S \to \mathbb{R}$ denote a representation of $\succsim$ where $\hat{u}(a) = u(a)$ for all $a \in \{b, c, d\}$. Now consider any $e \in S \setminus \{b, c, d\}$. By assumption, both u and $\hat{u}$ represent $\succsim$ on $\triangle\{b, c, e\}$. Since vNM representations are unique up to a positive affine transformation, there must exist $\alpha \in \mathbb{R}_+$ and $\beta \in \mathbb{R}$ such that $u(a) = \alpha\hat{u}(a) + \beta$ for all $a \in \{b, c, e\}$. As $\hat{u}(b) = u(b)$ and $\hat{u}(c) = u(c)$, we find that $(1 - \alpha)u(b) = (1 - \alpha)u(c)$ which further implies $\alpha = 1$ and $\beta = 0$. Hence, $\hat{u}(e) = u(e)$. As this holds true for any $e \in S \setminus \{b, c, d\}$, we find that $\hat{u} = u$ and hence u represents $\succsim$. □

Proof of Theorem I.1

Let RA, SP, AN, IIA and MICA be satisfied and assume that $|A| \geq 2n+4$. We will prove that for each $(S,R) \in \Omega$, $\succsim_*^{(S,R)}$ is represented by

$$\sum_{i \in N} u_i^{S,R}.$$

Fix $(S,R) \in \Omega$ where $|S| \geq 3$. Later, we consider the case $|S| = 2$. We will show that for any distinct $b,c,d \in S$, the part of $\succsim_*^{(S,R)}$ on $\triangle\{b,c,d\}$ is represented by $\sum_{i\in N} u_i^{S,R}$. It then follows from Lemma I.2 that $\succsim_*^{(S,R)}$ is represented by $\sum_{i\in N} u_i^{S,R}$. The proof is done in three steps. First, we define a sequence of states, connecting (S,R) to a polar state. Second, we show that in the final state of that sequence, social preferences on $\triangle\{b,c,d\}$ are represented by $\sum_{i\in N} u_i^{S,R}$. Third, we show that social preferences in the final state agree with the social preferences in the initial state on $\{b,c,d\}$.

We begin by defining the aforementioned sequence of states. We use Roman numerals as subscripts to keep track of the different preference profiles and menus. If $S \neq A$, then define $S_{\mathrm{I}} := S$ and denote an arbitrary alternative in S^c by e. If, on the other hand, $S = A$, we construct S_{I} in the following way. As $|A| \geq 2n+4$, $S \setminus \{b,c,d\}$ must have at least $2n+1$ alternatives. Consequently, at least one alternative $e \in S \setminus \{b,c,d\}$ must be comparable relative to $S \setminus \{e\}$ under R. So let $S_{\mathrm{I}} := S \setminus \{e\}$. Note that either way, there is an alternative e in S_{I}^c. Let $R_{\mathrm{I}} \in \mathcal{R}^n$ denote a preference profile that agrees with R on S_{I} and where for each $i \in N$, $e \sim_i^{R_{\mathrm{I}}} a$ for all $a \in \{a \in S_{\mathrm{I}} : a \succsim_i^{R_{\mathrm{I}}} a' \text{ for all } a' \in S_{\mathrm{I}}\}$. Let $S_{\mathrm{II}} := S_{\mathrm{I}} \cup \{e\}$. Next, identify the smallest subset $S_{\mathrm{III}} \subseteq S_{\mathrm{II}}$ with the property that $\{b,c,d,e\} \subseteq S_{\mathrm{III}}$ and every $a \in S_{\mathrm{II}} \setminus S_{\mathrm{III}}$ is comparable relative to S_{III} under R_{I}. Note that there are at most $n+4$ alternatives in S_{III}, namely b,c,d,e, and n alternatives that are each worst for exactly one individual. Hence, as $|A| \geq 2n+4$, there are at least n alternatives in S_{III}^c. For any state $\omega \in \Omega$, we denote the set of individuals that are not totally indifferent on the menu by $N_\succ^\omega$. Let $R_{\mathrm{II}} \in \mathcal{R}^n$ denote a preference profile that agrees with R_{I} on S_{III} and where there exists $P \subseteq S_{\mathrm{III}}^c$ such that (i) (P, R_{II}) is polar, where $p_i \in P$ denotes i's polar alternative, (ii) for each $i \in N$, $i \in N_\succ^{(P,R_{\mathrm{II}})}$ if and only if $i \in N_\succ^{(S_{\mathrm{III}},R_{\mathrm{II}})}$ and (iii) for each $i \in N$, $p_i \sim_i^{R_{\mathrm{II}}} a$ for all $a \in \{a \in S_{\mathrm{III}} : a \succsim_i^{R_{\mathrm{II}}} a' \text{ for all } a' \in S_{\mathrm{III}}\}$ and $p_j \sim_i^{R_{\mathrm{II}}} a$ for all $a \in \{a \in S_{\mathrm{III}} : a' \succsim_i^{R_{\mathrm{II}}} a \text{ for all } a' \in S_{\mathrm{III}}\}$ and $j \in N \setminus \{i\}$. This concludes the construction of the sequence of states.

Next, we show that the part of $\succsim_*^{(P\cup S_{\mathrm{III}},R_{\mathrm{II}})}$ on $\triangle\{b,c,d\}$ is represented by $\sum_{i\in N} u_i^{S,R}$. By Proposition I.6, there exists weights $(\lambda_i)_{i=0}^n \in \mathbb{R}^{n+1}$ such that $\succsim_*^{(P\cup S_{\mathrm{III}},R_{\mathrm{II}})}$ is represented by

$$\lambda_0 + \sum_{i\in N} \lambda_i u_i^{P\cup S_{\mathrm{III}},R_{\mathrm{II}}}.$$

By Proposition I.7, $p_i \sim_*^{(P,R_{\mathrm{II}})} p_j$ for all $i,j \in N_\succ^{(P,R_{\mathrm{II}})}$. Because every alternative in S_{III} is comparable relative to P under R_{II}, by MICA $p_i \sim_*^{(P\cup S_{\mathrm{III}},R_{\mathrm{II}})} p_j$ for all $i,j \in N_\succ^{(P\cup S_{\mathrm{III}},R_{\mathrm{II}})}$ as well. This implies $\lambda_i = \lambda_j =: \lambda$ for all $i,j \in N_\succ^{(P\cup S_{\mathrm{III}},R_{\mathrm{II}})}$. Note that if $i \notin N_\succ^{(P\cup S_{\mathrm{III}},R_{\mathrm{II}})}$, then $u_i^{P\cup S_{\mathrm{III}},R_{\mathrm{II}}}$ is constant on $P\cup S_{\mathrm{III}}$, and λ_i can be normalized to λ as the planner's vNM representation is unique up to positive affine transformations. Similarly, λ_0 can be normalized to 0. We now show that $\lambda > 0$ and, therefore, λ can be normalized to 1. We distinguish three cases. First, consider the case where $N_\succ^{(P\cup S_{\mathrm{III}},R_{\mathrm{II}})} = N$ and for any alternative $a \in S_{\mathrm{III}}$ there exists a lottery $x_a \in \triangle P$ such that $a \sim_i^{R_{\mathrm{II}}} x_a$ for all $i \in N$. Then $\succsim_*^{(P\cup S_{\mathrm{III}},R_{\mathrm{II}})}$ is totally indifferent on $P\cup S_{\mathrm{III}}$ and any λ represents the same preferences. Second, consider the case where $N_\succ^{(P\cup S_{\mathrm{III}},R_{\mathrm{II}})} = N$ and for some alternative $a \in S$ such a lottery does not exist. Depending on whether $\sum_{i=1}^n u_i^{S_{\mathrm{III}},R_{\mathrm{II}}}(a)$ is greater or smaller than 1, one can construct a lottery over P that either Pareto dominates a or is Pareto dominated by a. Then for SP to be satisfied, λ has to be strictly positive. Third, if $N_\succ^{(P\cup S_{\mathrm{III}},R_{\mathrm{II}})} \neq N$ then for SP to be satisfied, $p_i \succ_*^{(P\cup S_{\mathrm{III}},R_{\mathrm{II}})} p_j$ for any $i \in N_\succ^{(P\cup S_{\mathrm{III}},R_{\mathrm{II}})}$ and $j \notin N_\succ^{(P\cup S_{\mathrm{III}},R_{\mathrm{II}})}$. This requires λ to be strictly positive as well. In any of the three cases, λ can be normalized to 1. Hence, we have shown that $\succsim_*^{(P\cup S_{\mathrm{III}},R_{\mathrm{II}})}$ is represented by

$$\sum_{i\in N} u_i^{P\cup S_{\mathrm{III}},R_{\mathrm{II}}}.$$

Finally, note that we have constructed $(P\cup S_{\mathrm{III}},R_{\mathrm{II}})$ in such a way that for each $i \in N$, $u_i^{P\cup S_{\mathrm{III}},R_{\mathrm{II}}}(a) = u_i^{S,R}(a)$ for all $a \in \{b,c,d\}$. Hence, we have shown that the part of $\succsim_*^{(P\cup S_{\mathrm{III}},R_{\mathrm{II}})}$ on $\triangle\{b,c,d\}$ is indeed represented by $\sum_{i\in N} u_i^{S,R}$.

In the third and final step, we show that $\succsim_*^{(P\cup S_{\mathrm{III}},R_{\mathrm{II}})}$ and $\succsim_*^{(S,R)}$ agree on $\{b,c,d\}$.

$$\succsim_*^{(S,R)} \text{ and } \succsim_*^{(S_{\mathrm{I}},R)} \text{ agree on } \{b,c,d\}. \text{ (MICA)}$$
$$\succsim_*^{(S_{\mathrm{I}},R)} \text{ and } \succsim_*^{(S_{\mathrm{I}},R_{\mathrm{I}})} \text{ agree on } \{b,c,d\}. \text{ (IIA)}$$
$$\succsim_*^{(S_{\mathrm{I}},R_{\mathrm{I}})} \text{ and } \succsim_*^{(S_{\mathrm{II}},R_{\mathrm{I}})} \text{ agree on } \{b,c,d\}. \text{ (MICA)}$$
$$\succsim_*^{(S_{\mathrm{II}},R_{\mathrm{I}})} \text{ and } \succsim_*^{(S_{\mathrm{III}},R_{\mathrm{I}})} \text{ agree on } \{b,c,d\}. \text{ (MICA)}$$
$$\succsim_*^{(S_{\mathrm{III}},R_{\mathrm{I}})} \text{ and } \succsim_*^{(S_{\mathrm{III}},R_{\mathrm{II}})} \text{ agree on } \{b,c,d\}. \text{ (IIA)}$$
$$\succsim_*^{(S_{\mathrm{III}},R_{\mathrm{II}})} \text{ and } \succsim_*^{(P\cup S_{\mathrm{III}},R_{\mathrm{II}})} \text{ agree on } \{b,c,d\}. \text{ (MICA)}$$

Hence, by Lemma I.1 $\succsim_*^{(P\cup S_{\mathrm{III}},R_{\mathrm{II}})}$ and $\succsim_*^{(S,R)}$ agree on $\{b,c,d\}$. This concludes the proof for $|S|\geq 3$.

Now consider $|S|=2$. There must at least be $2n+2$ alternatives in S^c. We consider a different profile R_{I} where there is a set of polar alternatives P outside of S. Then $\succsim_*^{(P\cup S,R_{\mathrm{I}})}$ must be represented by equal weights by the above argument. By MICA, $\succsim_*^{(S,R_{\mathrm{I}})}$ must be represented by equal weights and by IIA $\succsim_*^{(S,R)}$ must be represented by equal weights. This concludes the proof of Theorem I.1.

Next, we derive an interim result required for the proof of Theorem I.2.

Proposition I.8. Let RA, SP, AN, MI, and IICA be satisfied. Then for any polar $(S,R)\in\Omega$, if every $a\in S^c$ is comparable relative to S under R and $i,j\in N$ have a strict preference on S, then $p_i\sim_*^{(S,R)} p_j$.

The proof is nearly identical to the proof of Proposition I.7, with the only difference that IICA is used instead of IIA. This is possible as Proposition I.8 is restricted to polar states where all alternatives outside are comparable.

Proof of Theorem I.2

Let RA, SP, AN, IICA and MI be satisfied and assume that $|A|\geq 2n+4$. We show that for each $(S,R)\in\Omega$, $\succsim_*^{(S,R)}$ is represented by

$$\sum_{i\in N} u_i^{A,R}.$$

Note that it suffices to show that this holds for $S=A$, as by MI the same representation must hold for any $S\subset A$. Furthermore, note that for $S=A$, the representations of both theorems coincide. Hence, we follow the proof of Theorem I.1, with the caveat that R_{II} is selected such that every alternative in $(P \cup S_{\text{III}})^c$ is comparable relative to $(P \cup S_{\text{III}})$ under R_{II}. Then one can simply replace the use of Proposition I.7 with Proposition I.8 and the use of IIA with IICA for the proof to go through. This concludes the proof of Theorem I.2.

Proof of Proposition I.2

We prove the proposition by providing a counterexample. Specifically, we identify an aggregation rule that satisfies the axioms but is not represented by the normalized sum of individual utilities across all states. We begin with a counterexample for the representation of Theorem I.1. Assume that RA, SP, AN, IIA and MICA are satisfied. Fix a state $(A, \hat{R}) \in \Omega$ where no alternative is comparable relative to the remaining alternatives under $\hat{R}$. As $|A| < 2n + 1$, such a state must exist. Let $\pi(\hat{R})$ denote the set containing all permutations of $\hat{R}$ and $\hat{R}$ itself and let $\hat{\Omega} := \{(S, R) \in \Omega : S = A, R \in \pi(\hat{R})\}$. Now consider an aggregation rule where $\succsim_*^{(S,R)}$ is represented by

$$\sum_{i \in N} \left(\sum_{a \in S} u_i^{S,R}(a) \right) u_i^{S,R} \tag{I.10}$$

whenever $(S, R) \in \hat{\Omega}$ and by

$$\sum_{i \in N} u_i^{S,R} \tag{I.11}$$

whenever $(S, R) \notin \hat{\Omega}$. Note that it could be that for all $(S, R) \in \hat{\Omega}$, (I.10) and (I.11) are positive affine transformations of each other. So assume that $(A, \hat{R})$ has been selected such that this is not the case, which is possible by the richness of Ω. If this aggregation rule indeed satisfies our axioms, we have produced a counterexample. Note that AN, MICA and IIA *connect* states, meaning they impose restrictions between states, whereas RA and SP impose restrictions on each state separately. So to prove that no axiom is violated, we will show that (i) no axiom connects a state in $\hat{\Omega}$ to a state outside of $\hat{\Omega}$, (ii) (I.10) satisfies the restrictions imposed between any two states in $\hat{\Omega}$, and (iii) (I.10) satisfies RA and SP for each $(S, R) \in \hat{\Omega}$. First, $\hat{\Omega}$ has been constructed such that AN doesn't connect any state in $\hat{\Omega}$ to a state outside of $\hat{\Omega}$. IIA doesn't connect any state in $\hat{\Omega}$ to another state, as there are no alternatives outside the menu. MICA doesn't connect any state in $\hat{\Omega}$ to another state, as no alternative is comparable relative to the other alternatives in A under any $R \in \pi(\hat{R})$. Second, (I.10) satisfies AN as the weight on each utility function only depends on the utility function itself but not on the index. Third, (I.10) assigns positive weights to all individual utility functions that are not indifferent on A. Hence, RA and SP are satisfied for each $(S, R) \in \hat{\Omega}$. This concludes the proof of Proposition I.2 in case of Theorem I.1.

Next, we provide a counterexample for the representation of Theorem I.2. Let $\hat{\Omega}$ be defined as above. Now consider an aggregation rule where $\succsim_*^{(S,R)}$ is represented by

$$\sum_{i \in N} \left(\sum_{a \in S} u_i^{A,R}(a) \right) u_i^{A,R} \tag{I.12}$$

whenever $(S, R) \in \hat{\Omega}$ and by

$$\sum_{i \in N} u_i^{A,R} \tag{I.13}$$

whenever $(S, R) \notin \hat{\Omega}$. As before, assume that $(A, \hat{R})$ has been selected such that (I.12) and (I.13) are not positive affine transformations of each other. Note that both MI and IICA connect states. As before, it holds that (i) no axiom connects a state in $\hat{\Omega}$ to a state outside of $\hat{\Omega}$, (ii) (I.12) satisfies the restrictions imposed between any two states in $\hat{\Omega}$, and (iii) (I.12) satisfies RA and SP for each $(S, R) \in \hat{\Omega}$. This can be shown, similarly to how it was shown for Theorem I.1 above. Just note that IICA doesn't connect any state in $\hat{\Omega}$ to another state, as no alternative is comparable relative to the other alternatives in A under any $R \in \pi(\hat{R})$. This concludes the proof of Proposition I.2 in the case of Theorem I.2.

Appendix I.B

In this section, we consider the case where A is either countably or uncountable infinite. This requires us to make some adjustments to the framework and the axioms. First, individual utilities might not be bounded, in which case they cannot be normalized as in the representations of Theorems I.1 and I.2. We deal with this by introducing a domain restriction, namely we only impose axioms on states where each individual has a best and worst alternative in A. Formally, we define Ω to be the set of states, such that for each $(S, R) \in \Omega$, both $\{a \in A : a \succsim_i^R b \text{ for all } b \in A\}$ and

$\{a \in A : b \succsim_i^R a \text{ for all } b \in A\}$ are non-empty for all $i \in N$. Second, even if individual preferences have a best and worst alternative in A, they might not have one for $S \subset A$. For example, if $A = [0,1]$ and $u_i^R(a) = a$ for some $R \in \mathcal{R}^n$ and $i \in N$, then $\succsim_i^R$ does not have a best alternative in $S = [0,1)$. The non-existence of a best or worst alternative for an individual in a given menu means that we cannot construct the specific polar state required for the proof of Theorem I.1. Therein, every alternative in the menu must be comparable relative to the set of polar alternatives and every polar alternative must be comparable relative to the menu. For the proof to go through, we introduce a weaker notation of comparability.

Definition I.3. $a \in A$ is *approximately comparable* relative to $B \subseteq A$ under $R \in \mathcal{R}^n$ if $a \notin B$ and for every $i \in N$ and $\varepsilon \in (0,1)$ there exists $x_{i,\varepsilon}, y_{i,\varepsilon} \in \triangle B$ such that

$$(1-\varepsilon)[a] + \varepsilon x_{i,\varepsilon} \sim_i^R y_{i,\varepsilon}.$$

Note that if a is comparable relative to B under R, then a is approximately comparable relative to B under R. If A is finite, the two concepts coincide. Furthermore, if a is approximately comparable, its utility must lie between the supremum and infimum utility in the set for every individual, as shown by the following lemma.

Lemma I.3. For any $R \in \mathcal{R}^n$ and $B \subseteq A$, if $a \notin B$ and for any utility profile $(u_i^R)_{i\in N}$ of R,

$$\sup_{b\in B} u_i^R(b) \geq u_i^R(a) \geq \inf_{b\in B} u_i^R(b)$$

for all $i \in N$ then a is approximately comparable relative to B under R.

Proof. Fix $(u_i^R)_{i\in N}$. Note that $a \in A$ is approximately comparable relative to $B \subseteq A$ under $R \in \mathcal{R}^n$ if and only if for every $i \in N$ and $\varepsilon \in (0,1)$ there exists $x_{i,\varepsilon}, y_{i,\varepsilon} \in \triangle B$ such that

$$(1-\varepsilon)u_i^R(a) + \varepsilon u_i^R(x_{i,\varepsilon}) = u_i^R(y_{i,\varepsilon}). \tag{I.14}$$

If $u_i^R(a)$ is strictly between $\sup_{b\in B} u_i^R(b)$ and $\inf_{b\in B} u_i^R(b)$, one can simply select a $z_i \in \triangle B$ such that $u_i^R(z_i) = u_i^R(a)$ and then set $x_{i,\varepsilon} = y_{i,\varepsilon} = z_i$. Then (I.14) is satisfied for all ε. So assume $u_i^R(a) = \sup_{b\in B} u_i^R(b)$ and fix ε. Choose $x_{i,\varepsilon}$ arbitrarily. The left-hand side of (I.14) is strictly between $\sup_{b\in B} u_i^R(b)$ and $\inf_{b\in B} u_i^R(b)$, and hence there must be a $y_{i,\varepsilon} \in \triangle B$ to satisfy (I.14). The same argument applies when $u_i^R(a) = \inf_{b\in B} u_i^R(b)$. $\square$

In MICA, comparability is then replaced by approximate comparability.

Axiom MICA*. For each $(S,R)\in\Omega$ and $S'\subseteq S$ where every $a\in S\setminus S'$ is approximately comparable relative to S', $\succsim_*^{(S,R)}$ and $\succsim_*^{(S',R)}$ agree on S'.

With these adjustments, we can now state the equivalent of Theorem I.1 when A is infinite. For any $R\in\mathcal{R}^n$ and $B\subseteq A$, let $\hat{u}_i^{B,R}$ denote the representation of $\succsim_i^R$ where $\sup_{a\in B}\hat{u}_i^{B,R}(a)=1$ and $\inf_{a\in B}u_i^{B,R}(a)=0$, unless $\succsim_i^R$ is indifferent on B in which case $\hat{u}_i^{B,R}(a)=0$ for all $a\in B$.

Theorem I.3. Let A be infinite. An aggregation rule $\succsim_*$ satisfies RA, SP, AN, IIA and MICA* if and only if for each $(S,R)\in\Omega$, $\succsim_*^{(S,R)}$ is represented by

$$\sum_{i\in N}\hat{u}_i^{S,R}.$$

Proof. First off, note that the proofs of Proposition I.7, Lemma I.1 and Lemma I.2 go through when there are infinitely many possible alternatives. For a proof of Proposition I.6 under infinite A we refer to Mandler (2005). Now consider the proof of Theorem I.1. When constructing the sequence of states, specifically R_{I} and R_{II}, there might not be a best or worst alternative in the menu for some individuals. We make the following adjustments to the proof. Let $R_{\mathrm{I}}\in\mathcal{R}^n$ denote a preference profile that agrees with R on S_{I} and where there is an $e\in S_{\mathrm{I}}^c$ such that for each $i\in N$ and any $u_i^{R_{\mathrm{I}}}$ representing $\succsim_i^{R_{\mathrm{I}}}$, $u_i^{R_{\mathrm{I}}}(e)=\sup_{a\in S_{\mathrm{I}}}u_i^{R_{\mathrm{I}}}(a)$. Let $R_{\mathrm{II}}\in\mathcal{R}^n$ denote a preference profile that agrees with R_{I} on S_{III} and where there exists $P\subseteq S_{\mathrm{III}}^c$ such that (i) (P,R_{II}) is polar, where $p_i\in P$ denotes i's polar alternative, (ii) for each $i\in N$, $i\in N_{\succ}^{(P,R_{\mathrm{II}})}$ if and only if $i\in N_{\succ}^{(S_{\mathrm{III}},R_{\mathrm{II}})}$ and (iii) for each $i\in N$ and any $u_i^{R_{\mathrm{II}}}$ representing $\succsim_i^{R_{\mathrm{II}}}$, $u_i^{R_{\mathrm{II}}}(p_i)=\sup_{a\in S_{\mathrm{III}}}u_i^{R_{\mathrm{II}}}(a)$ and $u_i^{R_{\mathrm{II}}}(p_j)=\inf_{a\in S_{\mathrm{III}}}u_i^{R_{\mathrm{II}}}(a)$ for all $j\in N\setminus\{i\}$. By Lemma I.3, e is approximately comparable to S_{I} under R_{I} and every alternative in P is approximately comparable relative to S_{III} under R_{II}. Then the proof of Theorem I.1 goes through. □

Theorem I.2 holds without adjustment to the axioms.

Appendix I.C

In this section, we show that none of the axioms can be dispensed with. We do so by separately dropping each axiom and then identifying a representation that satisfies the remaining axioms, different from the representation of the respective theorem.

When SP is dropped in either theorem, total indifference of the planner satisfies the remaining axioms. Formally, for each $(S, R) \in \Omega$ and $x \in \triangle S$, $u_*^{(S,R)}(x) = 0$. When AN is dropped in Theorem I.1, any weighted sum of individual utilities satisfies the remaining axioms as long as the weights are strictly positive. Formally, there exists $(\lambda_i)_{i\in N} \in \mathbb{R}_+^n$ such that for each $(S, R) \in \Omega$, $u_*^{(S,R)}(x) = \sum_{i=1}^n \lambda_i u_i^{S,R}(x)$. Similarly, when AN is dropped in Theorem I.2, $u_*^{(S,R)}(x) = \sum_{i=1}^n \lambda_i u_i^{A,R}(x)$ satisfies the remaining axioms. If RA is dropped in either theorem, then the relative egalitarian rule by Sprumont (2013) satisfies the remaining axioms. The rule ranks alternatives according to the lowest normalized individual utility it generates. If two alternatives generate the same lowest utility, they are further ranked by the second lowest utility, and so on. Depending on the theorem for which we want to provide the counterexample, individual utilities are normalized either with respect to S or A. When IIA is dropped in Theorem I.1, then the representation of Theorem I.2 satisfies the remaining axioms. Similarly, when MI is dropped in Theorem I.2, the then representation of Theorem I.1 satisfies the remaining axioms. For a representation when either MICA is dropped in Theorem I.1 or IICA is dropped in Theorem I.2 we refer to Propositions I.9 and I.10 in Appendix I.D.

Appendix I.D

In this section we consider weaker versions of our context-defining axioms MICA and IICA. The following notion of redundant alternatives has been considered by the literature.

Definition I.4. $a \in A$ is *redundant* relative to $B \subseteq A$ under $R \in \mathcal{R}^n$ if $a \notin B$ and there exists $x \in \triangle B$ such that $x \sim_i^R [a]$ for all $i \in N$.

Given this definition, we can state the central axiom of Dhillon and Mertens (1999). Note that we adapted the axiom to our setting where the menu is explicit.

Axiom IRA (Independence of Redundant Alternatives)**.** Fix $(S, R) \in \Omega$ such that every $a \in S^c$ is redundant relative to S under R. For any $R' \in \mathcal{R}^n$, if R and R' agree on S and every $a \in S^c$ is redundant relative to S under R', then $\succsim_*^{(S,R')} = \succsim_*^{(S,R)}$.

Note that IRA differs from IICA in two ways. First, the independence applies to redundant rather than comparable alternatives. Second, IRA applies only if all alternatives outside the menu are redundant, whereas IICA applies when only some alternatives outside the menu are comparable as well. In order to disentangle these difference, we define two intermediate axioms.

Axiom IICA*. Fix $(S,R) \in \Omega$ such that every $a \in S^c$ is comparable relative to S under R. For any $R' \in \mathcal{R}^n$, if R and R' agree on S and every $a \in S^c$ is comparable relative to S under R', then $\succcurlyeq_*^{(S,R')} = \succcurlyeq_*^{(S,R)}$.

Axiom IIRA (Independence of Irrelevant Redundant Alternatives)**.** Fix $(S,R) \in \Omega$ and $C \subseteq S^c$ such that every $a \in C$ is redundant relative to C^c under R. For any $R' \in \mathcal{R}^n$, if R and R' agree on C^c and every $a \in C$ is redundant relative to C^c under R', then $\succcurlyeq_*^{(S,R')} = \succcurlyeq_*^{(S,R)}$.

It is easy to see that IICA* would suffice in Theorem I.2, since whenever IICA is used in the proof, all alternatives outside the menu are comparable. However, replacing IICA with IIRA would not suffice, as demonstrated by the following proposition. Consequently, replacing IICA with IRA wouldn't suffice either, as IRA is weaker than IIRA.

Proposition I.9. Let $|A| \geq 2n+4$. There exists an aggregation rule $\succcurlyeq_*$ that satisfies RA, SP, AN, IIRA and MI and $\succcurlyeq_*^{(S,R)}$ can not be represented by $\sum_{i\in N} u_i^{A,R}$ for all $(S,R) \in \Omega$.

Proof. The following proof is adapted from Sprumont (2013). We prove the proposition by identifying a representation that satisfies the axioms but is not a positive affine transformation of the normalized sum of individual utility functions. For every $(S,R) \in \Omega$, define

$$\Psi(S,R) := \operatorname*{arg\,max}_{x\in\triangle S} \prod_{i\in N} (u_i^{S,R}(x)+1)$$

and note that $\Psi(S,R)$ is non-empty and for every $x,y \in \Psi(S,R)$, $u_i^{S,R}(x) = u_i^{S,R}(y)$ for all $i \in N$. Hence, $\psi_i^{(S,R)} := u_i^{S,R}(x)+1$ for some $x \in \Psi(S,R)$ is well defined. Now consider the following representation of $\succcurlyeq_*^{(S,R)}$,

$$\sum_{i\in N} \psi_i^{(A,R)} u_i^{A,R}.$$

As $\psi_i^{(A,R)}$ is not necessarily the same for all $i \in N$, the representation is not necessarily a positive affine transformation of $\sum_{i \in N} u_i^{A,R}$. What remains to show is that the representation satisfies the axioms stated in the proposition. RA and SP are satisfied as the representation is a weighted sum of individual utility functions with strictly positive weights. AN is satisfied as any individual's weight does not depend on the individual's index. MI is satisfied as the weights do not depend on the menu. Finally, to see that IIRA is satisfied, assume that for $(S, R) \in \Omega$ there exists a set of redundant alternatives $C \subseteq S^c$. Now consider $R' \in \mathcal{R}^n$ where $R' \neq R$, R and R' agree on C^c and C is still redundant under R'. As any redundant alternative has a lottery that yields the same product, redundant alternatives can be ignored in the maximization problem involved in $\Psi(S, R)$. Hence, $(\psi_i^{(A,R)})_{i \in N} = (\psi_i^{(A,R')})_{i \in N}$. Note that IICA is violated as changing individual preferences over comparable alternatives outside the menu can affect the maximization problem and, consequently, the weights. □

Similarly, we can replace the notion of comparable alternatives in MICA with redundant alternatives, resulting in the following axiom.

Axiom MIRA (Menu Independence of Redundant Alternatives)**.** For each $(S, R) \in \Omega$ and $S' \subseteq S$ where every $a \in S \setminus S'$ is redundant relative to S', $\succsim_*^{(S,R)}$ and $\succsim_*^{(S',R)}$ agree on S'.

Analogously, MIRA does not suffice in Theorem I.1.

Proposition I.10. Let $|A| \geq 2n+4$. There exists an aggregation rule $\succsim_*$ that satisfies RA, SP, AN, IIA and MIRA and $\succsim_*^{(S,R)}$ can not be represented by $\sum_{i \in N} u_i^{S,R}$ for all $(S, R) \in \Omega$.

A representation of $\succsim_*^{(S,R)}$ that satisfies the stated axioms is $\sum_{i \in N} \psi_i^{(S,R)} u_i^{S,R}$. The proof is analogous to that of Proposition I.9.

References

Arrow, K. J. (1950). A difficulty in the concept of social welfare. *Journal of Political Economy*, *58*(4), 328–346.

Arrow, K. J. (1963). *Social choice and individual values* (2nd ed.). Yale University Press.

Baris, O. F. (2018). Timing effect in bargaining and ex ante efficiency of the relative utilitarian solution. *Theory and Decision*, *84*(4), 547–556.

Blackorby, C., Donaldson, D., & Weymark, J. A. (1999). Harsanyi's social aggregation theorem for state-contingent alternatives. *Journal of Mathematical Economics*, *32*(3), 365–387.

Border, K. C. (1985). More on Harsanyi's utilitarian cardinal welfare theorem. *Social Choice and Welfare*, *1*(4), 279–281.

Börgers, T., & Choo, Y.-M. (2017). *Revealed relative utilitarianism* (CESifo Working Paper Series No. 6613).

Brandl, F. (2021). Belief-averaging and relative utilitarianism. *Journal of Economic Theory*, *198*, 105368.

Broome, J. (1984). Uncertainty and fairness. *The Economic Journal*, *94*(375), 624–632.

Cao, X. (1982). Preference functions and bargaining solutions. *1982 21st IEEE Conference on Decision and Control*, 164–171.

Chichilnisky, G. (1985). Von Neumann-Morgenstern utilities and cardinal preferences. *Mathematics of Operations Research*, *10*(4), 633–641.

Coulhon, T., & Mongin, P. (1989). Social choice theory in the case of von Neumann-Morgenstern utilities. *Social Choice and Welfare*, *6*(3), 175–187.

De Meyer, B., & Mongin, P. (1995). A note on affine aggregation. *Economics Letters*, *47*(2), 177–183.

Dhillon, A., & Mertens, J.-F. (1997). An impossibility theorem with von Neumann-Morgenstern preferences. *Economics Letters*, *56*(3), 305–309.

Dhillon, A., & Mertens, J.-F. (1999). Relative utilitarianism. *Econometrica*, *67*(3), 471–498.

Diamond, P. A. (1965). The evaluation of infinite utility streams. *Econometrica*, *33*(1), 170–177.

Dietrich, F. (2021). Fully Bayesian aggregation. *Journal of Economic Theory*, *194*, 105255.

Domotor, Z. (1979). Ordered sum and tensor product of linear utility structures. *Theory and Decision*, *11*(4), 375–399.

Fishburn, P. C. (1984). On Harsanyi's utilitarian cardinal welfare theorem. *Theory and Decision*, *17*(1), 21–28.

Gaertner, W., & Xu, Y. (2012). A general scoring rule. *Mathematical Social Sciences*, *63*(3), 193–196.

Gilboa, I., Samet, D., & Schmeidler, D. (2004). Utilitarian aggregation of beliefs and tastes. *Journal of Political Economy*, *112*(4), 932–938.

Goldman, S. M. (1979). Intertemporally inconsistent preferences and the rate of consumption. *Econometrica*, *47*(3), 621.

Hammond, P. J. (1992). Harsanyi's utilitarian theorem: A simpler proof and some ethical connotations. In R. Selten (Ed.), *Rational Interaction* (pp. 305–319). Springer Berlin Heidelberg.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, *63*(4), 309–321.

Harsanyi, J. C. (1975). Nonlinear social welfare functions. *Theory and Decision*, *6*(3), 311–332.

Hylland, A. (1980). Aggregation procedure for cardinal preferences: A comment. *Econometrica*, *48*(2), 539.

Kalai, E., & Schmeidler, D. (1977). Aggregation procedure for cardinal preferences: A formulation and proof of Samuelson's impossibility conjecture. *Econometrica*, *45*(6), 1431.

Kalai, E., & Smorodinsky, M. (1975). Other solutions to Nash's bargaining problem. *Econometrica*, *43*(3), 513–518.

Karni, E. (1998). Impartiality: Definition and representation. *Econometrica*, *66*(6), 1405.

Karni, E., & Weymark, J. A. (2024). Impartiality and relative utilitarianism. *Social Choice and Welfare*.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, *112*(2), 443–478.

Macé, A. (2018). Voting with evaluations: Characterizations of evaluative voting and range voting. *Journal of Mathematical Economics*, *79*, 10–17.

Machina, M. J. (1989). Dynamic consistency and non-expected utility models of choice under uncertainty. *Journal of Economic Literature*, *27*(4), 1622–1668.

Mandler, M. (2005). Harsanyi's utilitarianism via linear programming. *Economics Letters*, *88*(1), 85–90.

Marchant, T. (2019). Utilitarianism without individual utilities. *Social Choice and Welfare*, *53*(1), 1–19.

Maskin, E. (2022). *Arrow's theorem, may's axioms, and borda's rule* (Working Paper).

Mongin, P. (1994). Harsanyi's Aggregation Theorem: Multi-profile version and unsettled questions. *Social Choice and Welfare*, *11*(4), 331–354.

Mongin, P. (1995). Consistent Bayesian aggregation. *Journal of Economic Theory*, *66*(2), 313–351.

Mongin, P. (1998). The paradox of the Bayesian experts and state-dependent utility theory. *Journal of Mathematical Economics*, *29*(3), 331–361.

Nash, J. (1950). The bargaining problem. *Econometrica*, *18*(2), 155–162.

O'Donoghue, T., & Rabin, M. (1999). Doing it now or later. *American Economic Review*, *89*(1), 103–124.

Pearce, D. (2021). *Individual and social welfare: A Bayesian perspective* (Working Paper).

Pivato, M. (2009). Twofold optimality of the relative utilitarian bargaining solution. *Social Choice and Welfare*, *32*(1), 79–92.

Pivato, M. (2014). Formal utilitarianism and range voting. *Mathematical Social Sciences*, *67*, 50–56.

Saari, D. G. (1998). Connecting and resolving Sen's and Arrow's theorems. *Social Choice and Welfare*, *15*(2), 239–261.

Segal, U. (2000). Let's agree that all dictatorships are equally bad. *Journal of Political Economy*, *108*(3), 569–589.

Selinger, S. (1986). Harsanyi's aggregation theorem without selfish preferences. *Theory and Decision*, *20*(1), 53–62.

Sen, A. (1970). Interpersonal aggregation and partial comparability. *Econometrica*, *38*(3), 393–409.

Sen, A. (1993). Internal consistency of choice. *Econometrica*, *61*(3), 495–521.

Smith, W. D. (2000). *Range voting* (Working Paper).

Sprumont, Y. (2013). On relative egalitarianism. *Social Choice and Welfare*, *40*(4), 1015–1032.

Sprumont, Y. (2019). Relative utilitarianism under uncertainty. *Social Choice and Welfare*, *53*(4), 621–639.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ, USA: Princeton University Press.

Weymark, J. A. (1993). Harsanyi's social aggregation theorem and the weak Pareto principle. *Social Choice and Welfare*, *10*(3), 209–221.

Weymark, J. A. (1994). Harsanyi's social aggregation theorem with alternative Pareto principles. In W. Eichhorn (Ed.), *Models and Measurement of Welfare and Inequality* (pp. 869–887). Springer Berlin Heidelberg.

Zhou, L. (1997a). Harsanyi's utilitarianism theorems: General societies. *Journal of Economic Theory*, *72*(1), 198–207.

Chapter II

Rational Bargaining: Characterization and Implementation

Abstract

The von Neumann-Morgenstern axioms are uncontroversial desiderata for individual decision-making. We say that a bargaining solution is rational if it can be interpreted as the most preferred alternatives under these axioms. Yet, neither the Nash nor the Kalai-Smorodinsky bargaining solution is rational in this sense. We formalize two consequences of rationality, namely that one can neither be strictly better off nor strictly worse off from randomizing over different actions. These two axioms, together with other standard axioms, characterize the relative utilitarian bargaining solution. We then implement this bargaining solution in sub-game perfect equilibrium.

II.1 Introduction

Bargaining is prevalent in many economic settings, for instance, when both sides of a market are concentrated. Typical examples are bargaining over prices between up- and downstream firms, wage bargaining between firms and unions, and bargaining between health care providers and insurers. A common assumption in the applied economic literature is that the outcome of the bargaining process is given by the Nash bargaining solution (Nash, 1950) or a generalization thereof (Kalai, 1977).[1] While multiple bargaining protocols have been identified that support the Nash solution non-cooperatively,[2] the elegance of Nash's axiomatization undoubtedly plays a part in the popularity of the solution.

Another setting in which bargaining is pervasive is the settlement of disputes outside of courts. Oftentimes, in order to facilitate agreement and prevent a costly legal battle, arbitrators or mediators are appointed to make a decision. The international division of the American Arbitration Association (AAA) alone handled over ten thousand cases in the year 2022, with close to 16 billion dollars in total claims.[3] In the context of arbitration, Nash's axioms could be understood as desirable characteristics we would want an arbitrator to display.

Inherent to any economic decision is uncertainty about the final outcome. When up- and downstream firms negotiate prices, they face uncertainty about consumer demand. When firms and unions negotiate wages, they face uncertainty about future inflation rates. Both as a normative desideratum and as a benchmark, economic actors are typically assumed to deal with uncertainty rationally, meaning they act in accordance with the von Neumann and Morgenstern (1944) axioms and maximize expected utility. Similarly, we feel that the collective decision of a group, such as the

[1] Recent applications of the Nash solution for bargaining between up- and downstream firms can be found in Crawford and Yurukoglu (2012), Shang et al. (2016), Crawford et al. (2018), Rogerson (2020) and Grunewald et al. (2023), for wage bargaining in Cahuc et al. (2006), Dobbelaere and Kiyota (2018), de Pinto and Lingens (2019), Piluso et al. (2023) and Terai (2023), and for bargaining between insurers and health care providers in Gaynor et al. (2015), Gowrisankaran et al. (2015), Ho and Lee (2017), Dafny et al. (2019) and Ho and Lee (2019).

[2] See Section II.3 for implementations of the Nash bargaining solution.

[3] See the 2022 AAA-ICDR B2B Case Statistics at https://www.adr.org/research. Accessed on November 5, 2023.

resolution of a bargaining problem, should be rational in this sense. If an arbitrator or mediator decides for the group, they should do so rationally as well.

We define a bargaining solution as rational if it can be interpreted as the most preferred alternatives under some von Neumann-Morgenstern (vNM) preference relation. According to the vNM paradigm, a lottery is evaluated by first considering the value of each of the final outcomes and then taking a convex combination of these values. By reducing a lottery to the final outcomes, it is implicitly ruled out that the decision-maker values the process of randomization. As a consequence, a decision maker would only randomize over different outcomes if she were indifferent between them.[4] The Nash bargaining solution violates this condition. To demonstrate this, consider an arbitrator who has to allocate a single indivisible item to either one of two agents. Let a_i denote the allocation where Agent i receives the item and let the arbitrator's preferences be captured by the utility function u. The Nash solution would prescribe that the arbitrator flips a coin and allocates the item to the winner, i.e., a 50-50 lottery between a_1 and a_2. The Nash solution selects neither a_1 nor a_2, over which the coin randomizes. Hence, it demands that the arbitrator strictly prefers the coin flip over the deterministic allocations. However, by the vNM axioms, the utility of the coin flip is given by $\frac{1}{2}u(a_1) + \frac{1}{2}u(a_2)$, which cannot exceed both $u(a_1)$ and $u(a_2)$. Therefore, the Nash solution is not rational. Note that the same is true for the bargaining solution by Kalai and Smorodinsky (1975).

The incompatibility of rational risk preferences and ex-ante symmetry (i.e., choosing the coin-flip) is well known in the literature on preference aggregation (Diamond, 1967; Harsanyi, 1975; Broome, 1984) and non-expected utility theory (Machina, 1989).[5] Some authors claim that since we have to value ex-ante symmetry as a fairness principle, we need to give up rationality. We argue against this intuition later in this section. We believe that fairness can be understood as the arbitrator giving equal consideration to every agent, which doesn't necessitate that agents are made equal in their expected outcomes. It seems that the AAA shares a similar notion of fairness, as they write that the arbitrator should "achieve a fair, efficient, and economical reso-

[4]Consider for instance mixing in Nash equilibrium, which requires that players are indifferent between all actions over which they randomize.

[5]*Ex-ante* refers to the fact that the coin-flip is only symmetric before the coin has landed.

lution of the dispute", but also makes clear that they "do not split the baby" as they decide "clearly in favor of one party in over 94.5% of the cases".[6]

In this paper, we aim to find a rational bargaining solution. To capture a central aspect of rationality, we propose the *no benefit of randomization* (NBR) axiom. This axiom relates non-convex bargaining sets to their convex hull. A non-convex bargaining set represents a situation where randomization is either not feasible or not permitted. Allowing for randomization means the group or arbitrator can choose from the convex hull of this set. The axiom then says that if an alternative is selected by the bargaining solution in the non-convex set, it must still be selected in the convex hull of the set. Another consequence of the vNM axioms, and the flip-side of NBR, is that the agent is never strictly worse off when randomizing. Hence, when randomization is possible (i.e., the bargaining set is convex) and two alternatives are selected by the bargaining solution, then any mixture (i.e., convex combination) of these two alternatives must be selected as well. We call this axiom *convexity* (CONV). We find that NBR and CONV, together with standard axioms, characterize the *relative utilitarian* (RU) bargaining solution. The RU bargaining solution selects the alternatives with the highest sum of normalized utilities. Utilities are normalized such that the disagreement point has utility 0, and the best alternative has utility 1. The other axioms that underlie our characterization are invariance to the utility-scale, strong Pareto, weak symmetry, and a weaker version of Nash's independence of irrelevant alternatives (IIA) axiom. The proof for two agents is simple and resembles the one of Nash (1950). The proof easily generalizes to any number of agents.

Besides a characterization, we also implement the RU bargaining solution in subgame perfect equilibrium. This means that we identify a bargaining protocol, which, in equilibrium, leads to a RU-optimal alternative. We show that full implementation is not possible and identify a game form that weakly implements our bargaining solution. However, we are quite close to full implementation, as any RU-optimal alternative that is strictly better than the disagreement point for every agent is an equilibrium outcome. In this game, agents simultaneously make a proposal consisting of an alternative and a probability for each agent. In equilibrium, all agents propose the same RU-optimal

[6]See the Commercial Arbitration Rules and Mediation Procedures at https://www.adr.org/Rules, pages 22, 23 and 26, and https://go.adr.org/split-the-baby.html. Accessed on November 5, 2023.

alternative, and this alternative is implemented immediately. If there is disagreement among the agents, the proposal with the highest sum of probabilities has to be sequentially approved by all agents. Agents can choose whether to accept the alternative or receive a utility equal to the probability that the proposal assigns to them. Hence, the probabilities can be understood as a claim regarding how good the alternative is for each agent. If the claim was inflated, the alternative is rejected by at least one agent. If an agent rejects, all other agents receive a utility equal to the disagreement point.

Finally, we consider two other rational solutions. First, we provide the first characterization of the asymmetric relative utilitarian (ARU) solution. Similar to the asymmetric Nash solution by Kalai (1977), the ARU solution generalizes the RU solution by allowing for different weights on the individuals' utilities. This can capture differences in bargaining power, which is important for applications. Second, we characterize the utilitarian solution. This solution maximizes the sum of individual utilities without normalizing them first. It is applicable when utilities have absolute meaning, for instance, in the case where utilities express individuals' willingness to pay to bring about a social alternative. The axiomatization requires only a small tweak to the axioms that characterize RU solution. We simply drop the invariance axiom and impose Nash's IIA in its full strength.

We now come back to the incompatibility of rationality and ex-ante symmetry. Arguments in favor of rational risk preferences are well known, so we do not recapitulate them here. Instead, we will argue that the desirability of the coin flip is, at least in part, due to reasons one is supposed to abstract from. First, in a bargaining situation, individuals might not only receive utility from the final outcome but might also care about the procedure by which this outcome is implemented. An individual might be better off losing a public coin flip compared to the item being allocated to the other agent directly because they care about being treated symmetrically. Furthermore, an individual might be better off winning a public coin flip, compared to the item being allocated to them directly, because they feel better about receiving the item when the other agent had a chance as well. If this is the case, then the procedure of flipping the coin publicly is different from simply mixing over the alternatives, for instance, by flipping the coin in secret. In fact, the former Pareto dominates the latter. A rational arbitrator cannot strictly prefer the secret coin flip to either of the direct allocations, but she could strictly prefer the procedure of flipping the coin publicly if individuals indeed care about procedures. In order to abstract from this possibility, one should think of randomization as being performed in secret. Second, flipping a coin seems desirable because it is the obvious tie-breaking rule when the arbitrator is indifferent between giving the item to either of the agents. However, similar to choice correspondences in the theory of individual decision-making, a set-valued bargaining rule simply does not make a statement about how ties are broken in case of multiplicity. It is perfectly compatible with rationality to impose such a tie-breaking rule as a second-order principle, but only after first-order principles have pinned down the solution. Third, the coin flip would prevent a biased arbitrator from giving the item to her favored agent. However, since we propose a normative theory to resolve the bargaining problem, the arbitrator is unbiased by assumption. To summarize, rationality is compatible with a strict preference for public randomization, with randomization as a matter of tie-breaking and with randomization to limit a biased arbitrator's ability to discriminate. Once we abstract from these aspects, are we still willing to give up rationality in favor of flipping the coin?

II.1.1 Literature

Other axiomatizations of the RU bargaining solution are by Pivato (2009) and Baris (2018). Pivato (2009) considers preferences over bargaining solutions and then imposes axioms on these preferences. This differs from the standard approach, established by Nash (1950), where axioms are imposed on the bargaining solution directly. Baris (2018) adapts the characterization of the utilitarian bargaining solution by Myerson (1981) to a utility-scale invariant setting. Their central axiom can be interpreted as a dynamic consistency condition. When facing uncertainty over what the bargaining set will be, the arbitrator makes a plan, which specifies for each possible bargaining set a utility vector. Then, the expected utility vector must be the solution in the expected bargaining set. Cao (1982) identifies necessary axioms for the RU bargaining solution but does not provide a characterization. Note that Cao (1982), Pivato (2009) and Baris (2018) assume the bargaining set to be convex, whereas we contribute to the literature on bargaining over non-convex sets (Kaneko, 1980; Zhou, 1997b; Mariotti, 1998a, 1998b; Conley & Wilkie, 1996; Denicolò & Mariotti, 2000; Ok & Zhou, 1999; Nagahisa & Tanaka, 2002; Xu & Yoshihara, 2006; Zambrano, 2016).

Related to the RU bargaining solution is a large literature on preference aggregation, which characterizes a relative utilitarian rule (Karni, 1998; Dhillon & Mertens, 1999; Segal, 2000; Börgers & Choo, 2017; Marchant, 2019; Sprumont, 2019; Brandl, 2021; Peitler & Schlag, 2023; Karni & Weymark, 2024). Especially related is Peitler and Schlag (2023). In an application of their aggregation rule, the RU bargaining solution is derived from the most preferred element of a menu-dependent social preference, where the menu consists of the alternatives that are better for every agent than the disagreement point.

Related to rational bargaining is a literature on the rationalizability of bargaining rules (Peters & Wakker, 1991; Bossert, 1994; Sánchez, 2000; Xu & Yoshihara, 2013). A bargaining rule is rationalizable if it can be interpreted as the most preferred alternative under a single preference relation over utility vectors, which applies independently of the bargaining set. In this literature, rationality is understood as satisfying the weak axiom of revealed preference (or similar conditions). We take rationality to mean that the bargaining solution is consistent with the maximization of a vNM preference relation, but we do not insist that it is the same preference relation for every bargaining set.

Implementations of the RU bargaining solution have been provided by Miyagawa (2002) and Hagiwara (2020). They, however, consider the case of only two agents and strictly convex bargaining sets. Note that under strictly convex bargaining sets, there is a unique RU-optimal alternative. Our game, on the other hand, can have multiple equilibrium outcomes, each corresponding to one of the multiple RU-optimal alternatives. Our implementation is similar to the ones by Moulin (1984) and Moore and Repullo (1988). Moulin (1984) implements the Kalai-Smorodinsky solution for convex bargaining sets. As in our implementation, a proposal has to be sequentially approved by all players. Moore and Repullo (1988) provides general results on the implementability of social welfare functions in sub-game perfect equilibrium. The first stage of our game is similar to theirs, however, they require each individual to report the state (i.e., the entire utility function of every player), whereas we ask players to only report the utility for one alternative.

II.2 Axiomatization

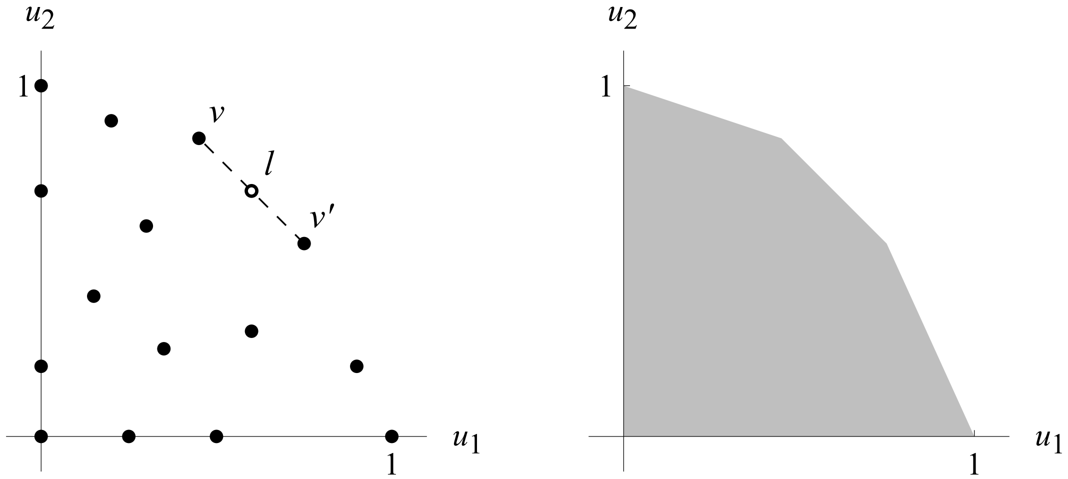
Let $N := \{1, ..., n\}$ be a set of agents where $n \in \mathbb{N}$ and $n \geq 2$. A bargaining problem (S, d) consists of a bargaining set $S \subseteq \mathbb{R}^n$ and a disagreement point $d \in S$. We simplify notation by normalizing the disagreement point to $d = (0, ..., 0)$ and write S instead of (S, d). Let $\mathbb{R}_+$ denote the positive real numbers, including 0. We restrict attention to bargaining sets S where (i) $S \subseteq \mathbb{R}^n_+$ (ii) S is compact, and (iii) for each $i \in N$, there exists a $u \in S$ such that $u_i > 0$. We denote the domain of bargaining sets that satisfy these properties by $\mathcal{S}$. A bargaining solution f is a correspondence that assigns to every $S \in \mathcal{S}$ a non-empty subset of S.

Our central axiom is *no benefit of randomization*. For any $R \subset \mathbb{R}^n$, let $\operatorname{conv} R$ denote the convex hull of R.

Axiom NBR (No Benefit of Randomization)**.** For every $S \in \mathcal{S}$,

$$f(S) \subseteq f(\operatorname{conv} S).$$

We illustrate the axiom with the help of the following example. Consider a finite bargaining set S, arising from the allocation of finitely many indivisible items. Now consider a lottery l that realizes some allocation $v \in S$ with probability $\lambda \in (0, 1)$ and another allocation $v' \in S$ with probability $1 - \lambda$. Since the utilities in v and v' express individuals' vNM preferences, Agent i's utility of l is $\lambda v_i + (1 - \lambda)v'_i$. Hence, if l were feasible, it would be a point in the bargaining set at $\lambda v + (1 - \lambda)v'$. See the left-hand side of Figure II.1 for an illustration when $n = 2$. Consequently, if every lottery over the allocations in S were feasible, the bargaining set would be the convex hull of S. This case is depicted on the right-hand side of Figure II.1. NBR says that if an allocation is optimal in the non-convex bargaining set S, where randomization isn't feasible, then this allocation must still be optimal in the convex bargaining set $\operatorname{conv} S$, where randomization is feasible. Hence, randomization doesn't introduce a lottery, strictly better for the group than the best allocation, as this would contradict that the group acts in accordance with the vNM postulates. Note, however, that randomization



(a) S. **(b)** conv S.

Figure II.1: Convexification of the bargaining set.

can introduce new utility vectors that are equally optimal as an allocation, which is why NBR does not demand $f(S) = f(\operatorname{conv} S)$.

Note that NBR has no bite when the domain is restricted to convex bargaining sets, which is the classic setting of Nash (1950) and Kalai and Smorodinsky (1975). However, these popular bargaining solutions have been extended in various ways to domains that include non-convex sets. We find that these extensions violate NBR. In the following, we demonstrate this for the extensions by Xu and Yoshihara (2006). Let f^{Nash} denote the Nash bargaining solution and f^{KS} denote the Kalai-Smorodinsky bargaining solution as in Xu and Yoshihara (2006).

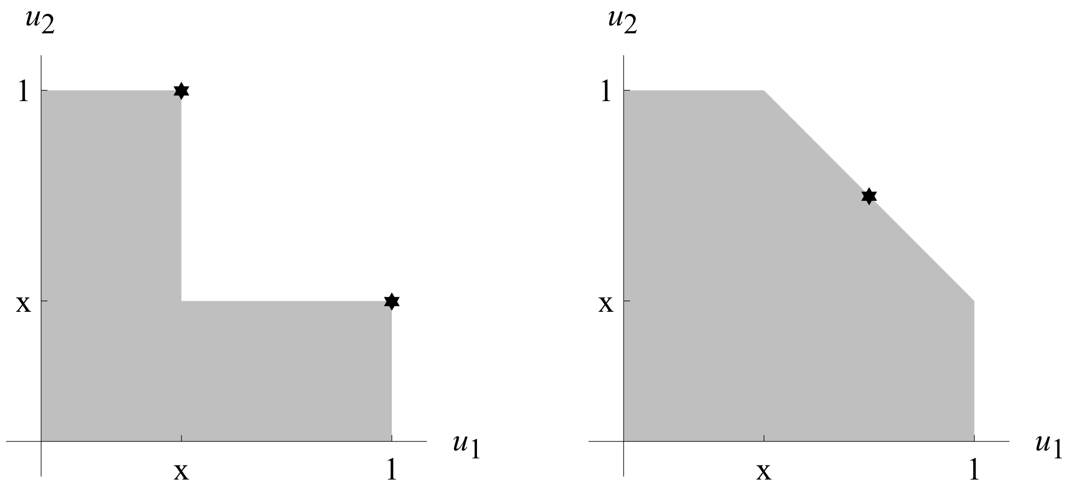
Proposition II.1. Both f^{Nash} and f^{KS} violate NBR.

Proof. Assume $n = 2$ and consider the barging set

$$S = \{(u_1, u_1) \in [0,1]^2 : u_1 \leq x \text{ or } u_2 \leq x\}$$

for some $x \in (0,1)$. Then $f^{\text{Nash}}(S) = \{(1,x),(x,1)\}$ and $f^{\text{KS}}(S) = \{(x,x)\}$. However, $f^{\text{Nash}}(\operatorname{conv} S) = f^{\text{KS}}(\operatorname{conv} S) = \{(\frac{1+x}{2}, \frac{1+x}{2})\}$. Hence, NBR is violated. Figure II.2 illustrates this for the Nash solution. □

Another consequence of the vNM axioms is that the arbitrator cannot be strictly worse off under randomization. If two utility vectors are both optimal from the per-



(a) S. **(b)** $\operatorname{conv} S$.

Figure II.2: Violation of NBR by f^{Nash}. Stars indicate the solution.

spective of the arbitrator, then a lottery over these vectors, assuming it is feasible, must also be optimal. This is captured by the following axiom.

Axiom CONV (Convexity)**.** For every $S \in \mathcal{S}$, $f(S)$ is convex whenever S is convex.

Note that both the Nash and Kalai-Smorodinsky bargaining solutions trivially satisfy this axiom since these solutions are singletons whenever the bargaining set is convex.

Since the prominent bargaining solutions violate NBR and are therefore not rational, we are in need of an alternative solution. Besides the aforementioned axioms, this solution should satisfy agreed-upon desiderata. Both Nash (1950) and Kalai and Smorodinsky (1975) agree that a solution should be Pareto efficient, invariant to the utility-scale, and that it should give equal treatment to symmetric agents. These axioms have been formulated under the assumption that the solution is single-valued. In the following, we translate these axioms to our setting, where the solution can be set-valued.

Axiom PO (Pareto Optimality)**.** For every $S \in \mathcal{S}$, if $u \in f(S)$, then there is no $v \in S$ such that $v \neq u$ and $v_i \geq u_i$ for all $i \in N$.

We say that α is a positive linear transformation if there exists $k_1, ..., k_n > 0$ such that for any $R \subseteq \mathbb{R}^n$, $\alpha(R) = \{u \in \mathbb{R}^n : (k_1 u_1, ..., k_n u_n) \in R\}$.

Axiom INV (Invariance)**.** For every $S \in \mathcal{S}$ and positive linear transformation α,

$$f(\alpha(S)) = \alpha(f(S)).$$

For any $R \subseteq \mathbb{R}^n$, we say that R is symmetric if, for every $u \in R$, any permutation of u is in R as well.

Axiom SYM (Symmetry)**.** For every $S \in \mathcal{S}$, if S is symmetric and $u \in f(S)$, then $u = (x, ..., x)$ for some $x \in \mathbb{R}$.

Note that SYM violates PO in non-convex settings. To see this, consider $n = 2$ and $S = \{(0, 0), (1, 0.9), (0.9, 1)\}$. Then SYM would demand $f(S) = \{(0, 0)\}$, a violation of PO. Therefore, for domains that include non-convex bargaining sets, symmetry needs to be weakened.

Axiom WSYM (Weak Symmetry)**.** For every $S \in \mathcal{S}$, if S is symmetric, then so is $f(S)$.

For the fourth and final axiom, Nash (1950) has independence of irrelevant alternatives (IIA) and Kalai and Smorodinsky (1975) has monotonicity. Note that IIA would not be compatible with the axioms we have imposed so far (NBR, CONV, PO, INV, and WSYM) and there is no obvious extension of monotonicity to set-valued solutions. However, there is a weaker version of IIA that is satisfied by both the Nash and Kalai-Smorodinsky solution, which we call *weak IIA*.[7] Furthermore, this axiom replaces monotonicity in axiomatizations of the Kalai-Smorodinsky solution in non-convex settings (Nagahisa & Tanaka, 2002; Xu & Yoshihara, 2006). Since weak IIA is a common denominator of the Nash and Kalai-Smorodinsky solution, we will impose it as well. Before we introduce this axiom, let us first translate Nash's IIA to a setting with set-valued solutions.

Axiom IIA (Independence of Irrelevant Alternatives)**.** For every $S, S' \in \mathcal{S}$, if $S' \subset S$ and $f(S) \cap S'$ is non-empty, then

$$f(S') = f(S) \cap S'.$$

[7] Weak IIA and variations thereof have already appeared in Yu (1973), Roth (1977), Cao (1982), Imai (1983), Dubra (2001), Nagahisa and Tanaka (2002), Xu and Yoshihara (2006) and Rachmilevitch (2019).

Next, we state weak IIA. For any $S \in \mathcal{S}$ and $i \in N$ let $m_i(S) := \max\{u_i : u \in S\}$. Furthermore, let $m(S) := (m_i(S))_{i \in N}$. Following Yu (1973), we call $m(S)$ the *utopian point* of S.

Axiom WIIA (Weak IIA)**.** For every $S, S' \in \mathcal{S}$, if $S' \subset S$, $m(S') = m(S)$ and $f(S) \cap S'$ is non-empty, then

$$f(S') = f(S) \cap S'.$$

IIA says that removing points from the bargaining set does not change what is optimal from the perspective of the group. WIIA weakens this condition, by imposing the former demand only in cases where the removal of points does not change the utopian point. The utopian point $m(S)$ is the maximal possible utility of each agent under the bargaining set S. While the utopian point typically isn't feasible (i.e., $m(S) \notin S$), it is an anchor point that allows us to relate the utilities of different agents. Since vNM utilities are unique only up to a positive affine transformation, comparisons of absolute utility levels across different individuals are meaningless. However, by INV we can normalize every bargaining set such that $m(S) = (1, ..., 1)$. Then all utilities express how well-off each agent is relative to their best possible outcome. In contrast to absolute utility levels, this measure can be meaningfully compared across individuals. Removing a point $u \in S$ from S such that $m_i(S \setminus \{u\}) < m_i(S)$ for some $i \in N$ would require us to re-normalize the bargaining set, which would change how points other than u are perceived by the group. Unlike IIA, WIIA allows this change of perspective to influence what is collectively optimal.

We show that the above axioms characterize the *relative utilitarian* bargaining solution. Define f^{RU} such that for every $S \in \mathcal{S}$,

$$f^{\text{RU}}(S) = \arg\max_{u \in S} \sum_{i \in N} \frac{u_i}{m_i(S)}.$$

Theorem II.1. f satisfies NBR, CONV, PO, INV, WSYM and WIIA if and only if $f \equiv f^{\text{RU}}$.

For $n = 2$, the proof is short, simple, and resembles the one of Nash (1950). For this reason, we present it here in the main section. A proof for a general number of agents is in Appendix II.A.

Proof. First, consider the bargaining set $S_x = \{(0,0),(1,0),(0,1),(1,x),(x,1)\}$ for some $x \in [0,1]$, as illustrated in Figure II.3. Note that $P_x := \{(1,x),(x,1)\}$ is the set

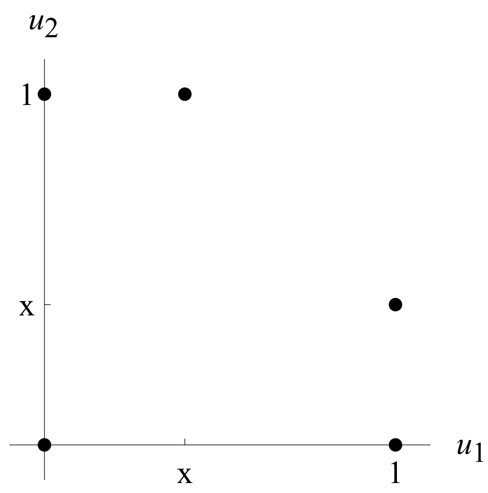


Figure II.3: S_x.

of Pareto optimal points in S_x and by PO, $f(S_x) \subseteq P_x$. Furthermore, S_x is symmetric. So by WSYM,

$$f(S_x) = P_x = \{(1,x),(x,1)\}. \tag{II.1}$$

Second, consider the bargaining set $\operatorname{conv} S_x$. By (II.1) and NBR, $P_x \subseteq f(\operatorname{conv} S_x)$. As $\operatorname{conv} S_x$ is convex, $\operatorname{conv} P_x \subseteq f(\operatorname{conv} S_x)$ by CONV. Note that $\operatorname{conv} P_x = \{u \in [0,1]^2 : u_1 + u_2 = 1 + x\}$ and $\operatorname{conv} S_x = \{u \in [0,1]^2 : u_1 + u_2 \leq 1 + x\}$. Hence, $\operatorname{conv} P_x$ is the set of Pareto-optimal points in $\operatorname{conv} S_x$. Hence, by PO,

$$f(\operatorname{conv} S_x) = \left\{u \in [0,1]^2 : u_1 + u_2 = 1 + x\right\}. \tag{II.2}$$

See Figure II.4 for an illustration. The dashed line indicates the solution.

Third, consider any bargaining set $S \in \mathcal{S}$ where $m(S) = (1,1)$. Let $x^* := \max_{u \in S}(u_1 + u_2) - 1$, which must exist due to our assumption of compactness. Note that $S \subseteq \operatorname{conv} S_{x^*}$ and that $S \cap f(\operatorname{conv} S_{x^*})$ is non-empty. Hence, (II.2) and WIIA imply

$$f(S) = S \cap f(\operatorname{conv} S_{x^*}) = \arg\max_{u \in S}(u_1 + u_2). \tag{II.3}$$

Fourth, note that for any $S \in \mathcal{S}$ there exists a bargaining set $S' \in \mathcal{S}$ with $m(S') = (1,1)$ and a positive linear transformation α such that $\alpha(S) = S'$. Then by (II.3) and INV,

$$f(S) = \arg\max_{u \in S} \left(\frac{u_1}{m_1(S)} + \frac{u_2}{m_2(S)}\right). \tag{II.4}$$

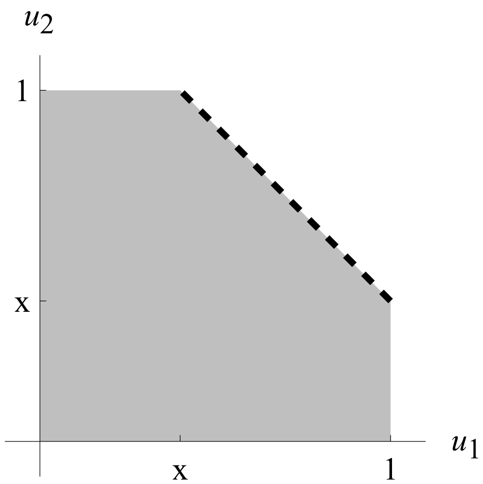


Figure II.4: conv S_x.

□

Above we have identified axioms that characterize the RU bargaining solution. Next, we show that these axioms are independent. We drop each of the axioms in Theorem II.1 and show that there is a solution, other than the RU solution, that satisfies the remaining axioms.

(1) The Nash solution satisfies all axioms but NBR.

(2) Consider the solution that selects all maximal elements of the leximax preorder whenever $m(S) = (1, ..., 1)$. The solution for $m(S) \neq (1, ..., 1)$ is given by INV. This solution satisfies all axioms but CONV.

(3) $f(S) = \{(0, ..., 0)\}$ for all $S \in \mathcal{S}$ satisfies all axioms but PO.

(4) The utilitarian solution (Section II.4.2) satisfies all axioms but INV.

(5) The asymmetric relative utilitarian solution (Section II.4.1) satisfies all axioms but WSYM.

(6) Let $\omega_i(S) = 1 + \max\{u_i : u \in S \text{ and } u_{i+1} = m_{i+1}(S)\} m_i(S)^{-1}$ if $i < n$ and $\omega_n(S) = 1 + \max\{u_n : u \in S \text{ and } u_1 = m_1(S)\} m_n(S)^{-1}$. Then the solution $f(S) = \arg\max_{u \in S} \sum_{i \in N} \omega_i(S) m_i(S)^{-1} u_i$ satisfies all axioms but WIIA.

This proves that the axioms are independent.

II.3 Implementation

In the previous section, we have provided a normative theory on how a group should come to a solution in a bargain problem. However, how a group will arrive at a solution depends on the game form that describes the bargaining process. In the spirit of the Nash program (Nash, 1953), we implement the RU bargaining solution in sub-game perfect equilibrium (SPE), meaning we identify a game form where the RU bargaining solution arises as the SPE outcome. In line with the existing literature on implementation of the Nash bargaining solution (Nash, 1953; Rubinstein, 1982; Binmore et al., 1986; Herrero, 1989; Howard, 1992; Chae & Yang, 1994; Krishna & Serrano, 1996; Trockel, 2000; Miyagawa, 2002; Trockel, 2002; Güth et al., 2004; Gómez, 2006; Britz et al., 2010; Okada, 2010; Anbarci & Sun, 2013; Britz et al., 2014; Abreu & Pearce, 2015; Qin et al., 2019; Hagiwara, 2020; Hu & Rocheteau, 2020; Harstad, 2023) and the Kalai-Smorodinsky bargaining solution (Moulin, 1984; Trockel, 1999; Miyagawa, 2002; Haake, 2009; Anbarci & Boyd, 2011; Hagiwara, 2020), we assume that the players have full information, i.e., know the utility functions of the other players.[8]

The bargaining protocol we propose can be used by an arbitrator who wants to bring about a desirable outcome but does not know the utility functions of the agents. For applicability in actual bargaining situations, it is important that the game is simple and intuitive. For instance, we would not be satisfied with a game where individuals have to report the state of the world, i.e., the entire utility function of every player, as in Moore and Repullo (1988). Note that, as in the previous section, we allow for non-convex bargaining problems, such that multiple alternatives can be optimal. This complicates the game somewhat, as players have to coordinate on one of multiple equilibria.

In the next section, we outline the basic setting.

II.3.1 Preliminaries

Let A be a set of alternatives. We designate one alternative in A as the disagreement alternative and denote it by a_{dis}. Let Θ be a set of states. For every $\theta \in \Theta$ and $i \in N$,

[8] See Serrano (2005, 2021) for a review of the literature on the Nash program.

let $u_i^\theta : A \to [0,1]$ denote Player i's vNM utility function over A, normalized such that $u_i^\theta(a_{\text{dis}}) = 0$ and $\max_{a \in A} u_i^\theta(a) = 1$. For every $\theta \in \Theta$, let S_θ denote the bargaining set associated with θ, i.e., $S_\theta := \{(u_1^\theta(a), ..., u_n^\theta(a)) : a \in A\}$. We assume that for each $\theta \in \Theta$, $S_\theta \in \mathcal{S}$. For any $\theta \in \Theta$, we say that $a \in A$ is *RU-optimal* under θ if $(u_1^\theta(a), ..., u_n^\theta(a)) \in f^{\text{RU}}(S_\theta)$. For a game form g and any $\theta \in \Theta$, we denote by (g, θ) the game with the game form g and players' preferences according to u_1^θ to u_n^θ. We say that a game form g *fully implements* the RU bargaining solution in SPE if, for every $\theta \in \Theta$ and $a \in A$, a is RU-optimal under θ if and only if a is an SPE outcome of (g, θ). Unfortunately, full implementation is not possible, which we discuss in Section II.3.3. Instead, we fully implement the subset of RU-optimal alternatives that are strictly better than a_{dis} for every player. We call these alternatives *strictly RU-optimal.* Note that this *weakly implements* the RU bargaining solution, meaning every SPE outcome is RU-optimal.

A strictly RU-optimal alternative does not always exist. Consider, for example, the case of allocating a single indivisible item without randomization. Giving the item to any of the agents is RU-optimal, but not strictly RU-optimal, as none of the remaining agents improves relative to the disagreement alternative. In order to implement the strictly RU-optimal alternatives, we must ensure that at least one such alternative exists. Hence, we impose the following restriction on Θ.

Assumption A1. For every $\theta \in \Theta$, there exists an $a \in A$ such that a is strictly RU-optimal under θ.

An additional assumption is imposed for convenience. We assume that for each agent there exists an alternative that is among the best for this agent and equal to the disagreement point for the remaining agents.

Assumption A2. For every $\theta \in \Theta$ and $i \in N$, there exists a $b_i^\theta \in A$ such that $u_i^\theta(b_i^\theta) = 1$ and $u_j^\theta(b_i^\theta) = 0$ for all $j \in N \setminus \{i\}$.

A2 holds true for many bargaining situations. For example, when bargaining over a surplus or over the allocation of goods, awarding everything to Player i would correspond to the alternative b_i^θ.

Next, we present the game form.

II.3.2 The Game

We now define the game form that fully implements the set of strictly RU-optimal outcomes. We denote this game form by g^*. The game form has two stages, an initial stage and, depending on the actions in the initial stage, an approval stage.

Initial Stage: Each player simultaneously makes a proposal (a, p), consisting of an alternative $a \in A$ and a list of n strictly positive probabilities $p \in (0, 1]^n$. We distinguish three cases:

(1) Assume all players make the same proposal (a, p). Then a is selected.

(2) Assume there are exactly two distinct proposals (a, p) and (a', p').

  (2a) If $\sum_{i \in N} p_i = \sum_{i \in N} p'_i$, then a_{dis} is selected.

  (2b) If $\sum_{i \in N} p_i \neq \sum_{i \in N} p'_i$, then go to the *approval stage*.

(3) Assume there are three or more distinct proposals. Then the alternative of the proposal with the highest sum of probabilities that lies below n is selected. In case of a tie, one of those proposals is chosen at random.

Approval Stage: For the approval stage to be reached, there must be exactly two distinct proposals (a, p) and (a', p'). Assume without loss of generality that $\sum_{i \in N} p_i > \sum_{i \in N} p'_i$. Assign players into a sequence such that those who proposed (a', p') come before those who proposed (a, p). Any such sequence will do. According to this sequence, let players sequentially decide between *accept* and *reject*. If all players accept, then a is selected. If Player i rejects, then with probability p_i, Player i can choose an alternative and with probability $1 - p_i$, a_{dis} is selected.

Figure II.5 illustrates the approval stage for three players, where Player 1 and 3 proposed (a, p) and Players 2 proposed (a', p').

Now that we have described the game g^*, we can state the following theorem.

Theorem II.2. Assume A1 and A2 are satisfied. Then g^* fully implements the strictly RU-optimal alternatives. Formally, for every $\theta \in \Theta$ and $a \in A$, a is an SPE outcome of (g^*, θ) if and only if a is strictly RU-optimal under θ.

We sketch the proof here and provide a formal proof of the theorem in Appendix II.B. We fix some $\theta \in \Theta$ and omit it from now on. First, consider the approval stage

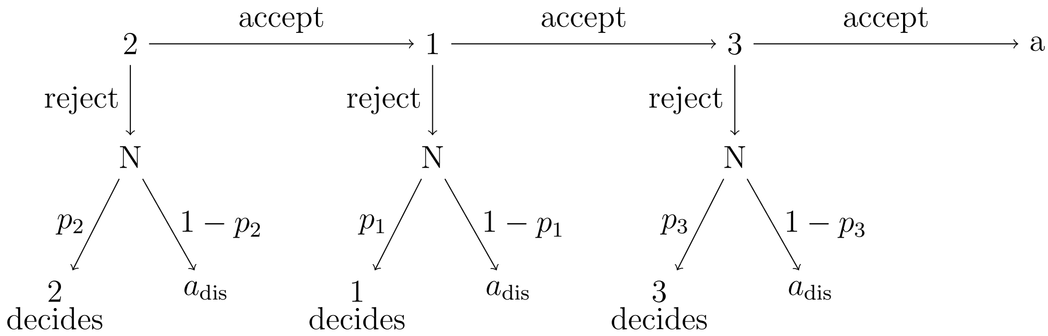


Figure II.5: Example of approval stage.

and let (a, p) denote the proposal with the higher sum of probabilities. Note that Player i's expected utility of rejecting is $p_i \max_{b \in A} u_i(b) + (1 - p_i)u_i(a_{\text{dis}}) = p_i$. Hence, if $p_i > u_i(a)$ for all $i \in N$, then the unique SPE is that all players accept and a is implemented. However, if $p_i < u_i(a)$ for some $i \in N$, then some player will reject. Hence, one can interpret the vector p as a *report* about the utility of a for each player and the approval stage as a test of whether this report was truthful. We say that the report p of a proposal (a, p) is *inflated* if $\sum_{i \in N} p_i > \sum_{i \in N} u_i(a)$ and we say that it is *truthful* if $p = u(a)$.

Next, consider the initial stage and assume that all players propose $(a, u(a))$ for some strictly RU-optimal alternative a. We show that no deviation (a', p') is profitable for Player j. There are three options for Player j, to report a higher sum, an equal sum or a lower sum of probabilities. Reporting an equal sum would lead to a_{dis}. Reporting a higher sum would mean that a' is put to the test in the approval stage and that j will decide last. However, because there is no alternative with a higher sum of utilities than a, p' is necessarily inflated and a' will be rejected by some player before j. Reporting a lower sum would mean that a is put to the test in the approval stage and that j will decide first. However, since the report of the proposal $(a, u(a))$ was truthful, there is an equilibrium where all accept, and a is implemented anyway. In conclusion, there is no profitable deviation for Player j.

Conversely, consider the case where all players propose (a, p), but a isn't strictly RU-optimal. If p is inflated, then there must be one Player j for whom $p_j > u(a)$. This player can then put (a, p) to the test by deviating to a report with a lower sum. Player j is allowed to decide first and rejects. If p is not inflated, then there is some player who would prefer a strictly RU-optimal alternative a' to a. This player can propose a' and report a probability for each player that is slightly below the true utility of a'. Then a' is put to the test in the approval stage and every player accepts.

Above, we have shown that a non-optimal alternative cannot arise in equilibrium through a unanimous proposal, i.e., Case (1). However, we have yet to show that such an alternative cannot arise in equilibrium via Cases (2b) or (3). Note that we have designed Case (3) similar to the integer game in Moore and Repullo (1988), such that it cannot arise in equilibrium. A player can always "outbid" the others by choosing an even higher sum below n. Furthermore, with three or more players, one can deviate from Case (2b) to induce Case (3) and implement their most preferred alternative. Showing that no sub-optimal equilibrium exists for two players is more involved and we leave this to the formal proof in the appendix.

II.3.3 Full Implementation

In this section, we discuss the case of full implementation. The following proposition shows that it is not possible to fully implement the RU bargaining solution.

Proposition II.2. Let $n = 2$ and assume that for every $S \in \mathcal{S}$ there exists a $\theta \in \Theta$ such that $S_\theta = S$. Then there doesn't exist an extensive game form g that fully implements the RU bargaining solution.

A formal proof is in Appendix II.C. The intuition for this result goes as follows. Consider a two-agent bargaining problem $S \in \mathcal{S}$ where both $(1, 0)$ and $(0, 1)$ are in $f^{\mathrm{RU}}(S)$. An example of such a problem is the division of a dollar among risk-neutral agents, where any efficient division is RU-optimal, including allocating the entire dollar to one of the agents. Now consider a state θ such that $S_\theta = S$ and a game form g that implements the solution. Then there must exist two strategy profiles s^+, s^- that are SPE of (g, θ) and where $u^\theta(s^+) := (u_1^\theta(s^+), u_2^\theta(s^+)) = (1, 0)$ and $u^\theta(s^-) = (0, 1)$. Since by assumption, 0 is the minimal utility for both players, $u_1^\theta(s_1, s_2^-) = 0$ for all s_1 and $u_2^\theta(s_1^+, s_2) = 0$ for all s_2. This, in turn, implies that (s_1^+, s_2^-) is a Nash equilibrium with pay-offs $(0, 0)$. In the formal proof, we then use (s_1^+, s_2^-) to construct an SPE with pay-offs $(0, 0)$. Since, $(0, 0)$ is not in $f^{\mathrm{RU}}(S)$, the RU bargaining solution cannot be fully implemented.

We feel that the division of a dollar among two risk neutral agents is a canonical problem that the implementation should be able to address. Hence, we did not want to rule out such problems, for instance by assuming that the bargaining set is strictly convex as in Miyagawa (2002) and Hagiwara (2020). Instead, we decided to exclude solutions that assign a pay-off of 0 to one of the players.

II.4 Other Rational Solutions

II.4.1 Asymmetric Relative Utilitarian Solution

In Section II.2 we have imposed a symmetry axiom as a normative requirement to treat all agents of the group fairly. In applications, however, we might want to take into account that individuals have different bargaining power. This can lead, in otherwise symmetric situations, to asymmetric outcomes. It is, therefore, of interest to generalize a given bargaining solution to a $n-1$ parameter family of solutions, where each parameter is a weight on an individual's utility, representing their bargaining power. Asymmetric generalizations have been provided for the Nash solution (Harsanyi & Selten, 1972; Kalai, 1977) and the Kalai-Smorodinsky solution (Dubra, 2001). In the following, we provide the first generalization of the relative utilitarian solution. We say that f is an *asymmetric relative utilitarian solution* if there exists $(\mu_1, ..., \mu_n) \in (0,1)^n$ with $\sum_{i \in N} \mu_i = 1$ such that for every $S \in \mathcal{S}$,

$$f(S) = \arg\max_{u \in S} \sum_{i \in N} \mu_i \frac{u_i}{m_i(S)}.$$

We denote this solution by f^{ARU}. Unfortunately, for the characterization of f^{ARU} it doesn't suffice to merely drop the symmetry axiom.

Proposition II.3. There exists an f such that $f \neq f^{\text{ARU}}$ and f satisfies NBR, CONV, PO, INV and WIIA.

We prove the proposition by providing two solutions as counterexamples. Each of these solutions can be ruled out by an additional axiom. These two additional axioms, together with NBR, CONV, PO, INV and WIIA then characterize f^{ARU}. For ease of exposition, we provide these counterexamples for $n = 2$.

The first counterexample is the solution f', which selects among the (symmetric) relative utilitarian outcomes the one that most benefits Agent 1, formally

$$f'(S) = \arg\max_{u \in f^{\mathrm{RU}}(S)} u_1.$$

First, we show that the axioms, as stated in Proposition II.3, are satisfied. It is easy to see that PO, INV and WIIA are satisfied. Since the solution is a singleton, CONV is trivially satisfied. While convexification can increase the set of relative utilitarian outcomes, it cannot change the best relative utilitarian outcome of each agent. Hence, NBR is satisfied as well. Next, we identify an axiom that rules out such a solution. Note that the solution is discontinuous, in the sense of vNM continuity. To see this, consider the following example, illustrated by Figure II.6. An arbitrator has to divide a

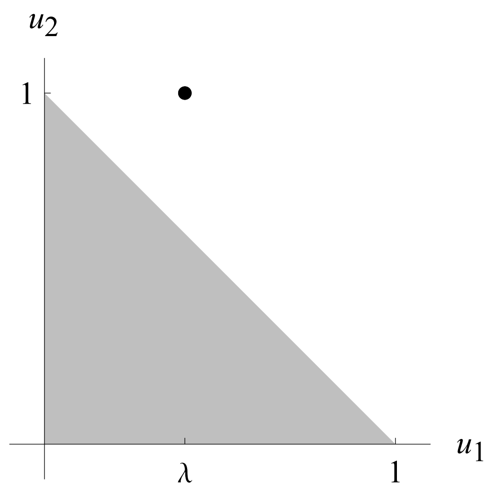


Figure II.6: Division of a dollar, with the risky technology at $(\lambda, 1)$.

dollar among two risk-neutral agents. In addition, the arbitrator could choose to invest the dollar into a risky technology, which gives one dollar to Agent 2 for sure and with probability λ one dollar to Agent 1. If $\lambda = 1$, then the technology is Pareto-dominant and selected by f'. If $\lambda = 0$, the technology is just as good as giving the dollar to Agent 2. Under f', the arbitrator strictly prefers to give the dollar to Agent 1. Continuity would require that for some λ, the arbitrator is indifferent between the technology and giving the dollar to Agent 1, such that both are selected by the solution. However, such a λ does not exist under f', as the solution uniquely selects the technology for every $\lambda > 0$. We impose the following axiom to ensure that a bargaining solution is continuous.

Axiom C (Continuity)**.** For every $S \in \mathcal{S}$ and $u \in S \setminus f(S)$, there exists a $\lambda \in [0,1]$ such that

$$f(S \cup \{\lambda m(S) + (1-\lambda)u\}) = f(S) \cup \{\lambda m(S) + (1-\lambda)u\}.$$

The axiom generalizes our previous example. Consider an arbitrary bargaining set S and some point u in S that is not selected by the solution. Add a convex combination $\lambda m(S) + (1-\lambda)u$ between u and the utopian point $m(S)$ to the original bargaining set. For $\lambda = 1$, any solution satisfying PO must select the convex combination, as it is equal to the utopian point. For $\lambda = 0$, the convex combination is equal to u and is therefore not selected by the solution. Axiom C then states that for some λ between 0 and 1, the convex combination must be in the solution, together with the solution of the original bargaining set S.

The second counterexample is a solution that can be described by linear, but non-parallel indifference curves. See Figure II.7 for an illustration. The solution selects the

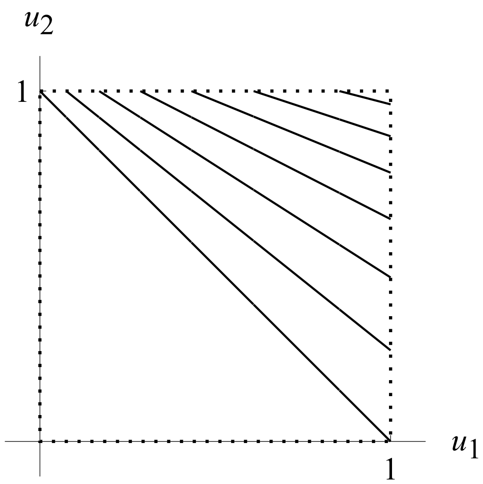


Figure II.7: Non-parallel indifference curves.

utility vectors on the highest indifference curve. These indifference curves are the same across all $S \in \mathcal{S}$ with $m(S) = (1,1)$. Solutions for bargaining sets with $m(S) \neq (1,1)$ are derived from INV. We call this solution f''. INV is then satisfied by assumption. It is easy to see that PO and WIIA are satisfied. Finally, CONV and NBR are satisfied because indifference curves are linear. Note that the fanning-out of indifference curves is reminiscent of weighted expected utility (Chew & MacCrimmon, 1979; Chew, 1983).

Such risk preferences violate the vNM independence axiom. Similarly, we can rule out non-parallel indifference curves through an independence axiom.

Axiom I (Independence)**.** For every $S \in \mathcal{S}$, if $u, v \in f(S)$, then for any $x \in [0, 1]$,

$$\{xm(S)+(1-x)u, xm(S)+(1-x)v\} \subseteq f(S \cup \{xm(S)+(1-x)u, xm(S)+(1-x)v\}).$$

The axiom captures vNM independence in the bargaining setting. Consider any bargaining set, where the solution contains at least two points u and v. It is as if the group was indifferent between these points. Now add two lotteries l_u and l_v that with probability x give the utopian point $m(S)$ and otherwise u, in case of l_u, or v, in case of l_v. The possibility of the irrelevant alternative $m(S)$ in both l_u and l_v does not change the relative desirability between the two. Hence, both l_u and l_v must be in the solution of the new bargaining set.

Imposing axioms C and I in addition, leads to the asymmetric relative utilitarian bargaining solution.

Theorem II.3. f satisfies NBR, CONV, PO, INV, WIIA, I and C if and only if $f \equiv f^{\text{ARU}}$.

We prove the theorem in Appendix II.D.

II.4.2 Utilitarian Solution

In the conventional understanding of the bargaining context, utilities derive from individuals' vNM preferences over the available alternatives. Since a utility representation of a vNM preference relation is unique only up to a positive affine transformation, the invariance axiom ensures that the solution is insensitive to the choice of the representation. However, in alternative scenarios, the utility scale might convey information. For instance, utilities might represent an individual's willingness to pay to bring about a given social alternative. Consider such a setting and think of the two agents who bargain over a single indivisible item. If Agent 1 values the item at \$100 and Agent 2 at \$50, it seems reasonable to award the item to Agent 1. Under the invariance axiom, however, this bargaining problem should be treated identically to one where both value the item at \$50. In order to take the absolute scale of valuations into account, INV must be dropped. Previously we argued that IIA imposes too stringent demands on the solution when utility scales lack significance. However, if utility scales are meaningful, IIA can be assumed in its full strength. Adopting these two modifications leads to the *utilitarian bargaining solution* (Myerson, 1981).

Theorem II.4. f satisfies NBR, CONV, PO, WSYM and IIA if and only if for every $S \in \mathcal{S}$,

$$f(S) = \arg\max_{u \in S} \sum_{i \in N} u_i.$$

Proof. Follow the first three steps of the proof of Theorem II.1 to find that $f(S) = \arg\max_{u \in S} \sum_{i \in N} u_i$ whenever $m(S) = (1, ..., 1)$. Note that an analogous argument can be made when $m(S) = (x, ..., x)$ for any $x > 0$. Then by IIA, $f(S) = \arg\max_{u \in S} \sum_{i \in N} u_i$ even if $m(S) \neq (x, ..., x)$. □

Note that if utilities are indeed valuations, the utilitarian sum is identical to economic surplus, a ubiquitous measure of welfare. Hence, a group that bargains rationally will also bargain welfare-optimally in the typical sense. Conversely, our axiomatization provides justification for the use of economic surplus as a welfare measure.

II.5 Conclusion

This paper considers the implications of rational risk preferences on fair bargaining. While the vNM axioms are formulated in a setting where the primitives are preferences over alternatives, the canonical formulation of the bargaining problem by Nash (1950) considers individuals' utilities directly. It is, therefore, not obvious what it means for a bargaining solution to deal with risk in a rational manner. We propose two axioms, NBR and CONV, which capture the consequences of vNM preferences. Specifically, these axioms describe the role of randomization under the vNM theory. Together with standard axioms from the bargaining literature, these axioms lead to the relative utilitarian bargaining solution.

One might remark that in light of Harsanyi (1955), who shows that a rational group preference must be utilitarian, this result comes as no surprise. Note, however, that NBR and CONV are necessary but not sufficient criteria for rational behavior under risk. In fact, they are quite far from being sufficient, as shown by the solutions that become permissible when WSYM is dropped (Section II.4.1). We, therefore, find it quite remarkable that these two rather weak axioms already pin down the relative utilitarian solution in the presence of standard bargaining axioms.

We believe that the relative utilitarian bargaining solution deserves a place in classrooms and textbooks, next to the solutions by Nash (1950) and Kalai and Smorodinsky (1975). As we have shown, the axioms that characterize the RU solution are straightforward, and the proof is short and simple. Furthermore, like the Nash and Kalai-Smorodinsky solution, the RU solution has a neat geometric interpretation. While the Kalai-Smorodinsky solution is the intersection of the Pareto frontier and the 45° line and the Nash solution maximizes the area of the rectangle spanned by the origin and the Pareto frontier, the RU solution maximizes the circumference of the rectangle spanned by the origin and the Pareto frontier. We illustrate this with an example depicted by Figure II.8. The Pareto frontier is given by $u_2 = (1 - u_1)^{\frac{2}{5}}$. The Kalai-

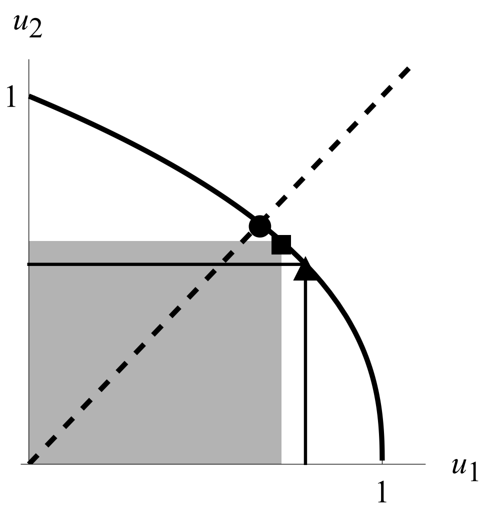


Figure II.8: The Nash, Kalai-Smorodinsky and RU solution.

Smorodinsky solution is indicated by a circle, the Nash solution by a square and the RU solution by a triangle.

We continue our discussion on the normative desirability of randomization, particularly when allocating an indivisible item between two agents. As noted in the introduction, randomization may seem appealing for reasons we are supposed to abstract from: individuals care about procedures, arbitrators might be biased, and randomization can be used to break ties. To disregard these factors, consider the following scenario:

(i) Individuals care only about the final outcomes and are completely indifferent to the procedures by which these outcomes are achieved.

(ii) The arbitrator is entirely unbiased and is unaware of the agents' identities.

(iii) To eliminate the use of randomization merely as a tie-breaking tool, assume randomization is costly. Specifically, a randomization device can be used, which destroys the item with a probability of c and, with a probability of $1-c$, triggers another lottery that assigns the item to either agent with equal probability.

What cost c should we be willing to incur for the sake of randomization? According to both the Nash and Kalai-Smorodinsky solutions, any cost strictly below 1 should be incurred. Therefore, both solutions are willing to sacrifice nearly the entire surplus. According to the RU solution, it is not worth sacrificing any of the surplus. Ultimately, only one of the agents can possess the item, and randomization does not alter this fact.

Appendix II.A

We begin with a definition of S_x and P_x for a general $n \in \mathbb{N}$. For this purpose, the following notation is introduced. For $x \in [1, n]$, let $\lfloor x \rfloor := \max\{k \in \{1, ..., n\} : k \leq x\}$, meaning $\lfloor x \rfloor$ is x rounded down to the closest integer. For any $u \in \mathbb{R}^n$, let $\pi(u) \subset \mathbb{R}^n$ denote the set containing u and all its permutations. For any $k \in \{0, ..., n\}$, let $g(k)$ denote the sequence of length n where $g_i(k) = 1$ if $i \leq k$ and $g_i(k) = 0$ otherwise. Hence, $g(0) = (0, ..., 0)$, $g(1) = (1, 0, ..., 0)$ and so on. For any $x \in [1, n]$, let $h(x)$ denote the sequence of length n where $h_i(x) = 1$ if $i \leq x$, $h_i(x) = x - \lfloor x \rfloor$ if $i = \lfloor x \rfloor + 1$ and $h_i(k) = 0$ otherwise. For any $x \in [1, n]$, let

$$P_x := \pi(h(x)), \qquad S_x := P_x \cup \bigcup_{k=0}^{\lfloor x \rfloor} \pi(g(k)).$$

The following lemma identifies $\operatorname{conv} S_x$ and $\operatorname{conv} P_x$.

Lemma II.1. For any $x \in [1, n]$, $\operatorname{conv} S_x = \{u \in [0, 1]^n : \sum_{i \in N} u_i \leq x\} =: R_x$ and $\operatorname{conv} P_x = \{u \in [0, 1]^n : \sum_{i \in N} u_i = x\} =: Q_x$.

Proof. An element $u \in S \subset \mathbb{R}^n$ is an *extreme point* of S if there doesn't exist $v, w \in S$ and $\lambda \in (0,1)$ such that $v \neq w$ and $u = \lambda v + (1-\lambda)w$. In the following, we show that S_x (resp. P_x) contains all extreme points of R_x (resp. Q_x). Then, by the Krein–Milman theorem, the lemma follows from the fact that $S_x \subseteq R_x$ and $P_x \subseteq Q_x$.

First, we show that an extreme point u of R_x (resp. Q_x) can have at most one coordinate u_j such that $0 < u_j < 1$. Consider $u \in R_x$ (resp. Q_x) with $0 < u_j < 1$ and $0 < u_k < 1$ for some $j \neq k$. Then there exists $v, w \in R_x$ (resp. Q_x) and $\varepsilon > 0$ such that $v_i = w_i = u_i$ whenever $i \notin \{j, k\}$ and

$$\begin{aligned} v_j &= u_j + \varepsilon, & w_j &= u_j - \varepsilon, \\ v_k &= u_k - \varepsilon, & w_k &= u_k + \varepsilon. \end{aligned} \tag{II.5}$$

Since $\frac{1}{2}v + \frac{1}{2}w = u$, u is not an extreme point of R_x (resp. Q_x).

Second, we show that if there is an extreme point of R_x with exactly one coordinate u_j such that $0 < u_j < 1$, then $u \in \pi(h(x))$. Assume $u \notin \pi(h(x))$ and $0 < u_j < 1$. Then there exists $v, w \in R_x$ and $\varepsilon > 0$ such that $v_i = w_i = u_i$ whenever $i \neq j$ and

$$v_j = u_j + \varepsilon, \qquad w_j = u_j - \varepsilon.$$

Since $\frac{1}{2}v + \frac{1}{2}w = u$, u is not an extreme point of R_x.

By the above arguments, the only candidates for extreme points of R_x (resp. Q_x) are the points in S_x (resp. P_x).

□

We now present the proof of Theorem II.1. The proof is nearly identical to the sketch in Section II.2. Nevertheless, we state the proof for the sake of completeness.

Proof of Theorem II.1

First, consider a bargaining set S_x for some $x \in [1, n]$. Note that P_x is the set of Pareto optimal points in S_x. Hence, by PO, $f(S_x) \subseteq P_x$. Furthermore, note that S_x is symmetric. So by WSYM,

$$f(S_x) = P_x. \tag{II.6}$$

Second, consider the bargaining set $\operatorname{conv} S_x$. By (II.6) and NBR, $P_x \subseteq f(\operatorname{conv} S_x)$. As $\operatorname{conv} S_x$ is convex, $\operatorname{conv} P_x \subseteq f(\operatorname{conv} S_x)$ by CONV. From Lemma II.1 we can see that conv P_x is the set of Pareto-optimal points in conv S_x. Then by PO,

$$f(\text{conv } S_x) = \left\{ u \in [0,1]^n : \sum_{i \in N} u_i = x \right\}. \tag{II.7}$$

Third, consider any bargaining set $S \in \mathcal{S}$ where $m(S) = (1, ..., 1)$. Let $x^* := \max_{u \in S} \sum_{i \in N} u_i$, which must exist due to our assumption of compactness. Note that $S \subseteq \text{conv } S_{x^*}$ and that $S \cap f(\text{conv } S_{x^*})$ is non-empty. Hence, (II.7) and WIIA imply

$$f(S) = S \cap f(\text{conv } S_{x^*}) = \arg\max_{u \in S} \sum_{i \in N} u_i. \tag{II.8}$$

Fourth, note that for any $S \in \mathcal{S}$ there exists a bargaining set $S' \in \mathcal{S}$ with $m(S') = (1, ..., 1)$ and a positive linear transformation α such that $\alpha(S) = S'$. Then by (II.8) and INV,

$$f(S) = \arg\max_{u \in S} \left(\sum_{i \in N} \frac{u_i}{m_i(S)} \right). \tag{II.9}$$

This concludes the proof.

Appendix II.B

This section contains the proof of Theorem II.2. We have already shown in the Section II.3.2 that for every strictly RU-optimal alternative, there exists an SPE with this alternative as the outcome. Here, we prove the other direction, namely that every SPE outcome is strictly RU-optimal. We fix some $\theta \in \Theta$ and omit it from now on. Let a denote some alternative that isn't strictly RU-optimal.

First, consider Case (1) of the initial stage, where a is implemented through some unanimous proposal (a, p). If $\sum_{i \in N} p_i > \sum_{i \in N} u_i(a)$ then there exists a Player j for whom $p_j > u_j(a)$. This player can deviate to some proposal (a', p') with $\sum_{i \in N} p'_i < \sum_{i \in N} p_i$ and be the first to reject in the approval stage, which gives an expected utility of p_j. If $\sum_{i \in N} p_i \leq \sum_{i \in N} u_i(a)$, then there is at least one player who strictly prefers some strictly RU-optimal alternative a'. Then this player can deviate to $(a', (u_1(a') - \varepsilon, ..., u_n(a') - \varepsilon))$ for ε sufficiently small such that $\sum_{i \in N} p_i < \sum_{i \in N} (u_i(a') - \varepsilon)$. Then a' is put to the test in the approval stage and the unique SPE outcome is that all players accept and a' is implemented.

Second, consider Case (2a), where there are two distinct proposals (a', p') and (a'', p'') with $\sum_{i \in N} p'_i = \sum_{i \in N} p''_i$, leading to a_{dis}. If $n \geq 3$, then some Player i can deviate in the initial stage to bring about Case (3) and choose b_i. So assume $n = 2$. If $p'_1 + p'_2 = p''_1 + p''_2 > 1$, then a player would be better off proposing a lower sum, be the first to choose in the approval stage and then reject. If $p'_1 + p'_2 = p''_1 + p''_2 \leq 1$, then for some strictly RU-optimal alternative a''' a player can propose $(a''', (u_1(a''') - \varepsilon, u_2(a''') - \varepsilon))$ for ε sufficiently small, leading to a'.

Third, consider Case (2b), where a is implemented through acceptance of the proposal (a, p) by all players in the approval stage. Unless only a single Player j has made the other proposal (a', p'), any player can bring about Case (3) and choose a more preferred alternative. Hence, assume only single Player j has made the other proposal and a is among the best alternatives for all players other than j. There must be some strictly RU-optimal alternative a'' that is preferred to a by j. Since a is unanimously approved, $\sum_{i \in N} p_i \leq \sum_{i \in N} u_i(a)$. Player j can deviate to $(a'', (u_1(a'') - \varepsilon, ..., u_n(a'') - \varepsilon))$ for ε sufficiently small, leading to a''.

Fourth, consider Case (2b), where a is chosen after some player rejects in the approval stage. Let $\Sigma := \sum_{i \in N} u_i(b)$ for any RU-optimal alternative b. First, consider $n = 2$. Let (a', p') be Player 1's proposal and (a'', p'') be Player 2's proposal. Without loss of generality, assume that $p'_1 + p'_2 > p''_1 + p''_2$, such that Player 2 decides first in the approval stage. Player 2 rejects and with probability p'_2 chooses a with $u(a) = (x, 1)$ for some $x \in [0, 1)$. This gives expected utility p'_2 to Player 2 and $p'_2 x$ to Player 1. Consider the case where $p''_1 + p''_2 \geq \Sigma$. Player 1 has an incentive to deviate to a proposal with a lower sum unless

$$p''_1 \leq p'_2 x. \tag{II.10}$$

Furthermore, since $p''_1 + p''_2 \geq \Sigma$,

$$p''_1 + p''_2 \geq x + 1. \tag{II.11}$$

Then (II.10) and (II.11) imply $p''_2 = 1$ and either $x = 0$ or $p'_2 = 1$. But $x = 0$ is not possible since this would imply $p''_1 = 0$ and this is not in the strategy-space of g^*. Hence $p'_2 = 1$ and $p''_1 = x$. Since we consider the case $p''_1 + p''_2 \geq \Sigma$ and have found $p'' = u(a)$, it must be that a is RU-optimal. Furthermore, since $p'_2 = 1$, a is implemented with certainty. This contradicts the assumption that the SPE outcome is sub-optimal. If $p_1'' + p_2'' < \Sigma$, then Player 1 can implement any strictly RU-optimal alternative b by deviating to the proposal $(b, (u_1(b) - \varepsilon, u_2(b) - \varepsilon))$ for ε sufficiently small. The only case in which Player 1 would have no incentive to do so is $p_2' = 1$ and if b is the best strictly RU-optimal alternative for Player 1, which contradicts the assumption that the SPE outcome is sub-optimal. This concludes the case $n = 2$. The argument extends to a general n.

Finally, consider Case (3) of the initial stage, where a is the alternative of the proposal with the highest sum strictly below n. Then at least one Player j prefers b_j to a. This player can deviate to a proposal with an even higher sum below n and the alternative b_j.

This concludes the proof of Theorem II.2.

Appendix II.C

We follow the notation of Moore and Repullo (1988). Consider a two-player extensive game form g. Let T denote the set of nodes. For any $t \in T$, we denote by $g(t)$ the sub-game starting at node t. For any $t \in T$ and $i \in \{1, 2\}$, we denote by $\sigma_i(t)$ the set of actions of Player i. We assume that $|\sigma_i(t)| \geq 1$ for all $t \in T$. If $|\sigma_i(t)| = 1$, then Player i has no decision at t. If both $|\sigma_1(t)| > 1$ and $|\sigma_2(t)| > 1$, then both players move simultaneously at t. Let $\sigma_i := \bigtimes_{t \in T} \sigma_i(t)$ denote the strategy space of Player i and let $\sigma = \sigma_1 \times \sigma_2$ denote the set of strategy profiles. For any $s \in \sigma$ and $t \in T$, we denote by $s|t$ the part of s that specifies the strategy profile for the game $g(t)$. For any $s \in \sigma$ and $t \in T$, we denote by $s(t) \in \sigma_1(t) \times \sigma_2(t)$ the action pair that s prescribes for the node t. Terminal nodes are alternatives in A. For any $s \in \sigma$ and $\theta \in \Theta$, $u^\theta(s) = (u_1^\theta(s), u_2^\theta(s))$ denotes the utility vector of the terminal node that results from s. For any $s \in \sigma$ and $t \in T$, we write $u^\theta(s|t)$ to denote the utility vector of the terminal node that is reached when starting at t and playing according to s.

Proof of Proposition II.2

Fix some $\theta \in \Theta$ such that $f^{\mathrm{RU}}(S_\theta)$ contains both $(1, 0)$ and $(0, 1)$. In the following, we omit θ on the individual utility functions and the game. We have already established in the main section that there must be $s^+, s^- \in \sigma$ such that both are SPE of g with

$u(s^+) = (1,0)$ and $u(s^-) = (0,1)$ and that $(s_1^+, s_2^-) =: s^0$ is a NE with $u(s^0) = (0,0)$. In the following, we construct $s^* \in \sigma$ such that s^* is an SPE with $u(s^*) = (0,0)$.

Let $t_0, t_0, ..., t_k, t_{k+1}$ denote the nodes of the equilibrium path of s^0, where t_0 is the initial node of g and t_{k+1} is a terminal node of g associated with the pay-off $(0,0)$. Consider the second to last node t_k. Let $(x_k, y_k) \in \sigma_1(t_k) \times \sigma_2(t_k)$ denote the action pair that leads to t_{k+1}, formally $t_{k+1} = (t_k, (x_k, y_k))$. If there are only terminal nodes succeeding t_k, then $g(t_k)$ is a one-stage game and $s^0|t_k$ is not only a NE of $g(t_k)$ but also an SPE. Hence, $s^*|t_k = s^0|t_k$ ensures that $s^*|t_k$ is an SPE of $g(t_k)$ with outcome $(0,0)$. If there are non-terminal nodes succeeding t_k, then we construct $s^*|t_k$ as follows.

Choose $s^*(t_k) = (x_k, y_k)$. For any (x_k, y) with $y \in \sigma_2(t_k)$ choose $s^*|(t_k, (x_k, y)) = s^+|(t_k, (x_k, y))$. This ensures that $s^*|(t_k, (x_k, y))$ is an SPE of $g(t_k, (x_k, y))$ for all $y \in \sigma_2(t_k)$. Furthermore, it must be that $u_2(s^*|(t_k, (x_k, y))) = 0$ for all $y \in \sigma_2(t_k)$, because $(t_k, (x_k, y))$ can be reached by a unilateral deviation of Player 2 in the strategy profile s^+. Similarly, choose $s^*|(t_k, (x, y_k)) = s^-|(t_k, (x, y_k))$ for all $x \in \sigma_1(t_k)$. For $s^*|(t_k, (x, y))$ such that neither $x = x_k$ nor $y = y_k$, choose an arbitrary SPE. Note that $s^*|t_k$ has been constructed such that an SPE is played at all nodes succeeding t_k, such that the outcome is $(0,0)$ and such that no Player has an incentive to deviate at t_k. Therefore, $s^*|t_k$ is an SPE of $g(t_k)$ with outcome $(0,0)$.

Finally, consider t_l for any $l \in \{0, ..., k-1\}$. Let $(x_l, y_l) \in \sigma_1(t_l) \times \sigma_2(t_l)$ denote the decision that leads to t_{l+1}, formally $t_{l+1} = (t_l, (x_l, y_l))$. Assume $s^*|t_{l+1}$ is an SPE of $g(t_{l+1})$ with outcome $(0,0)$. Choose $s^*(t_l) = (x_l, y_l)$ and construct $s^*|(t_l, (x, y))$ for $(x, y) \neq (x_l, y_l)$ just as before. Then $s^*|t_l$ is an SPE of $g(t_l)$ with outcome $(0,0)$. By induction, s^* is an SPE of g with outcome $(0,0)$. This concludes the proof.

Appendix II.D

We prove the Theorem II.3 for $n = 2$. The proof for general n follows similarly as that of Theorem II.1. First, consider the bargaining set $S = \{(0,0), (1,0), (0,1)\}$. By PO, there are three possible cases.

Case 1: $f(S) = \{(1,0)\}$

Case 2: $f(S) = \{(0,1)\}$

Case 3: $f(S) = \{(1,0), (0,1)\}$

Assume Case 1 holds true. By NBR, $(1,0) \in f(\operatorname{conv} S)$ and by WIIA, $(0,1) \notin f(\operatorname{conv} S)$. By C, there exists a $\lambda \in [0,1]$ such that $f(\operatorname{conv} S \cup \{(\lambda,1)\}) = f(\operatorname{conv} S) \cup \{(\lambda,1)\}$. Note that $\lambda > 0$, as it would otherwise contradict the assumption that $(0,1) \notin f(\operatorname{conv} S)$. Furthermore, note that $\lambda < 1$ as it would otherwise contradict PO. By NBR, $\{(1,0),(\lambda,1)\} \subseteq f(\operatorname{conv}(S \cup \{(\lambda,1)\}))$ and by CONV,

$$f(\operatorname{conv}(S \cup \{(\lambda,1)\})) = \left\{u \in [0,1]^2 : u_1 + (1-\lambda)u_2 = 1\right\}. \tag{II.12}$$

Second, consider the bargaining set

$$S_x := \operatorname{conv}\left(S \cup \{(\lambda,1)\}\right) \cup \{(1,x),(x+(1-x)\lambda,1)\}$$

for any $x \in [0,1]$. By I and (II.12), $\{(1,x),(x+(1-x)\lambda,1)\} \subseteq f(S_x)$. By NBR, $\{(1,x),(x+(1-x)\lambda,1)\} \subseteq f(\operatorname{conv} S_x)$ and by CONV and PO,

$$f(\operatorname{conv} S_x) = \left\{u \in [0,1]^2 : u_1 + (1-\lambda)u_2 = 1 + (1-\lambda)x\right\}. \tag{II.13}$$

Third, consider any bargaining set $S \in \mathcal{S}$ where $m(S) = (1,1)$. Let $x^* := (1-\lambda)^{-1}\left(\max_{u\in S}(u_1 + (1-\lambda)u_2) - 1\right)$, which must exist due to our assumption of compactness. Note that $S \subseteq \operatorname{conv} S_{x^*}$ and that $S \cap f(\operatorname{conv} S_{x^*})$ is non-empty. Hence, (II.13) and WIIA imply

$$f(S) = S \cap f(\operatorname{conv} S_{x^*}) = \arg\max_{u\in S}(u_1 + (1-\lambda)u_2). \tag{II.14}$$

Fourth, note that for any $S \in \mathcal{S}$ there exists a bargaining set $S' \in \mathcal{S}$ with $m(S') = (1,1)$ and a positive linear transformation α such that $\alpha(S) = S'$. Then by (II.14) and INV,

$$f(S) = \arg\max_{u\in S}\left(\frac{u_1}{m_1(S)} + (1-\lambda)\frac{u_2}{m_2(S)}\right). \tag{II.15}$$

This concludes Case 1.

Note that by an analogous argument, we can find an analogous solution for Cases 2 and 3. We can summarize the results for the different cases as follows. There exists $(\mu_1,\mu_2) \in (0,1)^2$ where $\mu_1 + \mu_2 = 1$ such that for any $S \in \mathcal{S}$,

$$f(S) = \arg\max_{u\in S}\left(\mu_1\frac{u_1}{m_1(S)} + \mu_2\frac{u_2}{m_2(S)}\right). \tag{II.16}$$

This concludes the proof.

References

Abreu, D., & Pearce, D. (2015). A dynamic reinterpretation of Nash bargaining with endogenous threats. *Econometrica*, *83*(4), 1641–1655.

Anbarci, N., & Boyd, J. H. (2011). Nash demand game and the Kalai–Smorodinsky solution. *Games and Economic Behavior*, *71*(1), 14–22.

Anbarci, N., & Sun, C.-j. (2013). Asymmetric Nash bargaining solutions: A simple Nash program. *Economics Letters*, *120*(2), 211–214.

Baris, O. F. (2018). Timing effect in bargaining and ex ante efficiency of the relative utilitarian solution. *Theory and Decision*, *84*(4), 547–556.

Binmore, K., Rubinstein, A., & Wolinsky, A. (1986). The Nash bargaining solution in economic modelling. *The RAND Journal of Economics*, *17*(2), 176–188.

Börgers, T., & Choo, Y.-M. (2017). *Revealed relative utilitarianism* (CESifo Working Paper Series No. 6613).

Bossert, W. (1994). Rational choice and two-person bargaining solutions. *Journal of Mathematical Economics*, *23*(6), 549–563.

Brandl, F. (2021). Belief-averaging and relative utilitarianism. *Journal of Economic Theory*, *198*, 105368.

Britz, V., Herings, P. J.-J., & Predtetchinski, A. (2010). Non-cooperative support for the asymmetric Nash bargaining solution. *Journal of Economic Theory*, *145*(5), 1951–1967.

Britz, V., Herings, P. J.-J., & Predtetchinski, A. (2014). On the convergence to the Nash bargaining solution for action-dependent bargaining protocols. *Games and Economic Behavior*, *86*, 178–183.

Broome, J. (1984). Uncertainty and fairness. *The Economic Journal*, *94*(375), 624–632.

Cahuc, P., Postel-Vinay, F., & Robin, J.-M. (2006). Wage bargaining with on-the-job search: Theory and evidence. *Econometrica*, *74*(2), 323–364.

Cao, X. (1982). Preference functions and bargaining solutions. *1982 21st IEEE Conference on Decision and Control*, 164–171.

Chae, S., & Yang, J.-A. (1994). An n-person pure bargaining game. *Journal of Economic Theory*, *62*(1), 86–102.

Chew, S. H. (1983). A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the Allais paradox. *Econometrica*, *51*(4), 1065–1092.

Chew, S. H., & MacCrimmon, K. R. (1979). *Alpha-nu choice theory: A generalization of expected utility* (University of British Columbia Working Paper).

Conley, J. P., & Wilkie, S. (1996). An extension of the Nash bargaining solution to nonconvex problems. *Games and Economic Behavior*, *13*(1), 26–38.

Crawford, G. S., Lee, R. S., Whinston, M. D., & Yurukoglu, A. (2018). The welfare effects of vertical integration in multichannel television markets. *Econometrica*, *86*(3), 891–954.

Crawford, G. S., & Yurukoglu, A. (2012). The welfare effects of bundling in multichannel television markets. *American Economic Review*, *102*(2), 643–685.

Dafny, L., Ho, K., & Lee, R. S. (2019). The price effects of cross-market mergers: Theory and evidence from the hospital industry. *The RAND Journal of Economics*, *50*(2), 286–325.

Denicolò, V., & Mariotti, M. (2000). Nash bargaining theory, nonconvex problems and social welfare orderings. *Theory and Decision*, *48*(4), 351–358.

de Pinto, M., & Lingens, J. (2019). The impact of unionization costs when firm-selection matters. *Labour Economics*, *60*, 50–63.

Dhillon, A., & Mertens, J.-F. (1999). Relative utilitarianism. *Econometrica*, *67*(3), 471–498.

Diamond, P. A. (1967). Cardinal welfare, individualistic ethics, and interpersonal comparison of utility: Comment. *Journal of Political Economy*, *75*(5), 765–766.

Dobbelaere, S., & Kiyota, K. (2018). Labor market imperfections, markups and productivity in multinationals and exporters. *Labour Economics*, *53*, 198–212.

Dubra, J. (2001). An asymmetric Kalai–Smorodinsky solution. *Economics Letters*.

Gaynor, M., Ho, K., & Town, R. J. (2015). The industrial organization of health-care markets. *Journal of Economic Literature*, *53*(2), 235–284.

Gómez, J. C. (2006). Achieving efficiency with manipulative bargainers. *Games and Economic Behavior*, *57*(2), 254–263.

Gowrisankaran, G., Nevo, A., & Town, R. (2015). Mergers when prices are negotiated: Evidence from the hospital industry. *American Economic Review*, *105*(1), 172–203.

Grunewald, A., Lanning, J. A., Low, D. C., & Salz, T. (2023). *Auto dealer loan intermediation: Consumer behavior and competitive effects* (Working Paper).

Güth, W., Ritzberger, K., & van Damme, E. (2004). On the Nash bargaining solution with noise. *European Economic Review*, *48*(3), 697–713.

Haake, C.-J. (2009). Two support results for the Kalai–Smorodinsky solution in small object division markets. *Mathematical Social Sciences*, *57*(2), 177–187.

Hagiwara, M. (2020). Subgame-perfect implementation of bargaining solutions: Comment. *Games and Economic Behavior*, *122*, 476–480.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, *63*(4), 309–321.

Harsanyi, J. C. (1975). Nonlinear social welfare functions. *Theory and Decision*, *6*(3), 311–332.

Harsanyi, J. C., & Selten, R. (1972). A generalized Nash solution for two-person bargaining games with incomplete information. *Management Science*, *18*, 80–106.

Harstad, B. (2023). Pledge-and-review bargaining. *Journal of Economic Theory*, *207*, 105574.

Herrero, M. J. (1989). The Nash program: Non-convex bargaining problems. *Journal of Economic Theory*, *49*(2), 266–277.

Ho, K., & Lee, R. S. (2017). Insurer competition in health care markets. *Econometrica*, *85*(2), 379–417.

Ho, K., & Lee, R. S. (2019). Equilibrium provider networks: Bargaining and exclusion in health care markets. *American Economic Review*, *109*(2), 473–522.

Howard, J. (1992). A social choice rule and its implementation in perfect equilibrium. *Journal of Economic Theory*, *56*(1), 142–159.

Hu, T.-W., & Rocheteau, G. (2020). Bargaining under liquidity constraints: Unified strategic foundations of the Nash and Kalai solutions. *Journal of Economic Theory*, *189*, 105098.

Imai, H. (1983). Individual monotonicity and lexicographic maxmin solution. *Econometrica*, *51*(2), 389.

Kalai, E. (1977). Nonsymmetric Nash solutions and replications of 2-person bargaining. *International Journal of Game Theory*, *6*(3), 129–133.

Kalai, E., & Smorodinsky, M. (1975). Other solutions to Nash's bargaining problem. *Econometrica*, *43*(3), 513–518.

Kaneko, M. (1980). An extension of the Nash bargaining problem and the Nash social welfare function. *Theory and Decision*, *12*(2), 135–148.

Karni, E. (1998). Impartiality: Definition and representation. *Econometrica*, *66*(6), 1405.

Karni, E., & Weymark, J. A. (2024). Impartiality and relative utilitarianism. *Social Choice and Welfare*.

Krishna, V., & Serrano, R. (1996). Multilateral bargaining. *The Review of Economic Studies*, *63*(1), 61–80.

Machina, M. J. (1989). Dynamic consistency and non-expected utility models of choice under uncertainty. *Journal of Economic Literature*, *27*(4), 1622–1668.

Marchant, T. (2019). Utilitarianism without individual utilities. *Social Choice and Welfare*, *53*(1), 1–19.

Mariotti, M. (1998a). Extending Nash's axioms to nonconvex problems. *Games and Economic Behavior*, *22*(2), 377–383.

Mariotti, M. (1998b). Nash bargaining theory when the number of alternatives can be finite. *Social Choice and Welfare*, *15*(3), 413–421.

Miyagawa, E. (2002). Subgame-perfect implementation of bargaining solutions. *Games and Economic Behavior*.

Moore, J., & Repullo, R. (1988). Subgame perfect implementation. *Econometrica*, *56*(5), 1191.

Moulin, H. (1984). Implementing the Kalai-Smorodinsky bargaining solution. *Journal of Economic Theory*, *33*(1), 32–45.

Myerson, R. B. (1981). Utilitarianism, egalitarianism, and the timing effect in social choice problems. *Econometrica*, *49*(4), 883.

Nagahisa, R.-i., & Tanaka, M. (2002). An axiomatization of the Kalai-Smorodinsky solution when the feasible sets can be finite. *Social Choice and Welfare*, *19*(4), 751–761.

Nash, J. (1950). The bargaining problem. *Econometrica*, *18*(2), 155–162.

Nash, J. (1953). Two-person cooperative games. *Econometrica*, *21*(1), 128–140.

Ok, E. A., & Zhou, L. (1999). Revealed group preferences on non-convex choice problems. *Economic Theory*, *13*(3), 671–687.

Okada, A. (2010). The Nash bargaining solution in general n-person cooperative games. *Journal of Economic Theory*, *145*(6), 2356–2379.

Peitler, P., & Schlag, K. H. (2023). *Putting context into preference aggregation* (Working Paper).

Peters, H., & Wakker, P. (1991). Independence of irrelevant alternatives and revealed group preferences. *Econometrica*, *59*(6), 1787–1801.

Piluso, N., Colletis, G., Blanchet, A., & Badra, M. (2023). Mobility and income distribution. *Journal of Quantitative Economics*.

Pivato, M. (2009). Twofold optimality of the relative utilitarian bargaining solution. *Social Choice and Welfare*, *32*(1), 79–92.

Qin, C.-Z., Tan, G., & Wong, A. C. L. (2019). Implementation of Nash bargaining solutions with non-convexity. *Economics Letters*, *178*, 46–49.

Rachmilevitch, S. (2019). Egalitarianism, utilitarianism, and the Nash bargaining solution. *Social Choice and Welfare*, *52*(4), 741–751.

Rogerson, W. P. (2020). Modelling and predicting the competitive effects of vertical mergers: The bargaining leverage over rivals effect. *Canadian Journal of Economics/Revue canadienne d'économique*, *53*(2), 407–436.

Roth, A. E. (1977). Independence of irrelevant alternatives, and solutions to Nash's bargaining problem. *Journal of Economic Theory*, *16*(2), 247–251.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica*, *50*(1), 97–109.

Sánchez, M. (2000). Rationality of bargaining solutions. *Journal of Mathematical Economics*, *33*(4), 389–399.

Segal, U. (2000). Let's agree that all dictatorships are equally bad. *Journal of Political Economy*, *108*(3), 569–589.

Serrano, R. (2005). Fifty years of the Nash program, 1953-2003. *Investigaciones Económicas*, (29), 219–258.

Serrano, R. (2021). Sixty-seven years of the Nash program: Time for retirement? *SERIEs*, *12*(1), 35–48.

Shang, W., Ha, A. Y., & Tong, S. (2016). Information sharing in a supply chain with a common retailer. *Management Science*, *62*(1), 245–263.

Sprumont, Y. (2019). Relative utilitarianism under uncertainty. *Social Choice and Welfare*, *53*(4), 621–639.

Terai, K. (2023). Investment in human capital by a shareholder. *Applied Economics*.

Trockel, W. (1999). Unique Nash implementation for a class of bargaining solutions. *International Game Theory Review*, *01*, 267–272.

Trockel, W. (2000). Implementations of the Nash solution based on its Walrasian characterization. *Economic Theory*, *16*(2), 277–294.

Trockel, W. (2002). A universal meta bargaining implementation of the Nash solution. *Social Choice and Welfare*, *19*(3), 581–586.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ, USA: Princeton University Press.

Xu, Y., & Yoshihara, N. (2006). Alternative characterizations of three bargaining solutions for nonconvex problems. *Games and Economic Behavior*, *57*(1), 86–92.

Xu, Y., & Yoshihara, N. (2013). Rationality and solutions to nonconvex bargaining problems: Rationalizability and Nash solutions. *Mathematical Social Sciences*, *66*(1), 66–70.

Yu, P. L. (1973). A class of solutions for group decision problems. *Management Science*, *19*(8), 936–946.

Zambrano, E. (2016). 'Vintage' Nash bargaining without convexity. *Economics Letters*, *141*, 32–34.

Zhou, L. (1997b). The Nash bargaining theory with non-convex problems. *Econometrica*, *65*(3), 681.

Chapter III

Price Transparency and Consumer Search in Service Markets

Abstract

In many service markets, a firm's marginal cost of providing service depends on consumer-specific characteristics. Additionally, expert firms can often better judge consumers' relevant characteristics than the consumers themselves. This paper explores the consequences of this informational asymmetry on price transparency. Firms can choose whether to display a single price for all consumers or to make individualized offers after learning the consumer's characteristics. We find that an equilibrium can exist in which no firm displays a price, resulting in monopoly prices and profits. Remarkably, this equilibrium can be better for consumers compared to pricing at the expected marginal cost.

III.1 Introduction

An extensive literature on consumer search suggests that improving consumers' access to pricing information increases competition among firms, leading to reduced prices and improved welfare. Therefore, the topic of price transparency holds significant relevance as a policy concern. The advent of online stores and price comparison tools might have improved price transparency across various markets in recent years. Yet, it appears that service markets are lagging behind. For example, a report by the United Kingdom's Competition & Markets Authority revealed that a mere 17% of legal service providers published prices online.[1] Another example is health services, where price transparency is a frequent subject of regulatory efforts.[2]

The persistence of opaque pricing schemes in the digital era, where price information could be easily disseminated by firms, is somewhat puzzling. One would expect that in markets lacking price transparency, firms would be motivated to disclose their prices in order to divert consumers from their competitors. A readily available answer to this puzzle is that firms, instead of directly colluding on prices, collude on hiding prices, which raises search costs and, in turn, increases profits. However, this explanation applies to services and tangible goods alike. We are interested in exploring alternative mechanisms, taking into account the peculiarities of service markets.

One potential explanation for service providers' reluctance to share pricing information lies in the variability of cost based on consumer-specific characteristics. Providing the same service to different consumers may require providers to employ different *procedures*, which may vary in cost. A provider could publish a menu stating a price for each distinct procedure. However, in many service markets, consumers lack the technical expertise to discern the appropriateness of a procedure for their condition or to verify the procedure's execution post-service.

A typical example of the information asymmetry described above is found in health services. Patients typically do not know which medical procedure is appropriate for their specific condition. Hence, even if medical service providers were to disclose pric-

[1] https://www.gov.uk/cma-cases/legal-services-market-study (Accessed April 4, 2024).

[2] For recent examples of regulation in the U.S. health sector see the Hospital Price Transparency rule (https://www.cms.gov/priorities/key-initiatives/hospital-price-transparency, Accessed April 4, 2024) and the No Surprise Act (https://www.cms.gov/medical-bill-rights, Accessed April 4, 2024).

ing information for all possible procedures, patients could neither accurately predict the final price of treatment nor evaluate whether the most cost-effective procedure was utilized. Note that healthcare expenditures constitute a significant part of the GDP, accounting for 17.3% of the U.S. GDP in 2022.[3] Furthermore, the information asymmetry we explore is present in many other service markets, such as legal services, repair services, IT services, cosmetic services, and home improvement.

Goods and services that exhibit such an information asymmetry are referred to as *credence goods* (Dulleck & Kerschbamer, 2006). The existing literature on credence goods assumes either that prices are exogenous or that firms have to display prices publicly. Hence, it cannot account for a lack of price transparency in these markets.

To understand firms' incentives to display prices in a credence goods market, we propose the following model. Consumers differ in their *cost type*, a characteristic of which they are unaware and which determines a firm's cost of service. Firms can choose whether to publicly display a single price for the service or not.[4] Displaying a price has two effects. First, it informs consumers about the price of that firm at no cost to the consumer. In contrast, to learn the price of a firm that doesn't display a price, the consumer has to pay a search cost. Second, it commits the firm to charging this price to consumers irrespective of the consumer's cost type. In contrast, if a firm chooses not to display a price, it can charge a different price to each type of consumer.

Under these conditions, displaying a price to attract consumers presents a trade-off to the firm: serve a larger consumer base at a uniform price or serve a smaller consumer base with cost-based differential pricing. Furthermore, the impact of price transparency on welfare becomes unclear. Mandating firms to display a single price for their services might indeed stimulate competition. However, it cannot decrease prices to marginal costs for each distinct consumer.

We find that an equilibrium can exist where no firm displays a price. If no firm displays a price, then, because of search frictions, firms charge the respective monopoly price to each type of consumer. This allows firms to share monopoly profits among

[3] https://www.cms.gov/data-research/statistics-trends-and-reports/national-health-expenditure-data/historical (Accessed April 4, 2024).

[4] Under the information asymmetry we consider, firms cannot credibly commit to a menu. We elaborate on this point later in this section.

them. We call this the *monopolistic equilibrium*. Note the close similarities to the model of consumer search by Diamond (1971). Now consider a deviation of a firm to displaying a price, with the intention of attracting consumers away from its competitors. As this deviation is observed by the other firms, they can adjust their pricing strategy and will price at most the search cost above the displayed price of the deviating firm. If there was just a single cost type, such as in Diamond (1971), then displaying a price just slightly below the monopoly price would attract all consumers, and the deviating firm would receive close to monopoly profits. However, as there are multiple cost types, there are multiple monopoly prices. If the displayed price is above some of these monopoly prices, it can still be beneficial for a consumer to visit a flexible (i.e., non-committed) firm over the committed firm. We find that there is a threshold between the lowest and highest monopoly price, such that a deviating firm attracts consumers if and only if its price is below this threshold. Depending on the parameters of the model, it is possible that pricing below this threshold yields lower profits than the shared monopoly profits. In that case, a firm has no incentive to deviate to displaying a price, explaining the lack of price transparency in service markets.

The monopolistic equilibrium does not always exist. Conditions favorable for existence are a smaller number of firms and more dispersed costs. Furthermore, we find an interesting relationship between existence and search cost. While a strictly positive search cost is necessary for firms to enjoy any profits, the monopolistic equilibrium exists for a smaller parameter space as search costs increase. The reason is as follows. As search costs decrease, visiting flexible firms in the presence of committed firms becomes more attractive, as flexible firms price at most the search cost above the lowest displayed price. Hence, to attract consumers, a deviating firm would have to display an even lower price, making a deviation less attractive.

Besides the monopolistic equilibrium, there always exists a *competitive equilibrium*, where all firms display a price equal to the expected marginal cost. Note that pricing at expected marginal cost would also result from a policy that requires firms to display a single price for the service. Therefore, for our analysis of welfare, we consider pricing at expected marginal cost as the benchmark, against which we compare the monopolistic equilibrium. We find that in the region of the parameter space, where both equilibria co-exist, the monopolistic equilibrium can yield not only higher total surplus but also higher consumer surplus. To see this, assume that for some consumers the cost of service is so large that the expected marginal cost exceeds the monopoly price for consumers with low cost of service. Then, low-cost consumers are better off under the monopolistic equilibrium than the competitive equilibrium. In aggregate, the benefit to low-cost consumers can outweigh the detriment of high-cost consumers. Whether this is indeed the case depends on the level of cost dispersion. For uniformly distributed valuations, we identify a simple statistic, namely the normalized standard deviation of the cost distribution. This statistic completely determines which of the two equilibria has higher consumer and total surplus. For the monopolistic equilibrium to yield higher welfare, the statistic needs to be sufficiently large. Hence, the conditions that make the existence of the monopolistic equilibrium more likely make the monopolistic equilibrium more desirable as well.

We now come back to the case of firms displaying a menu. If consumers lack the knowledge of which procedure is appropriate for their condition, then a menu would not allow consumers to predict the price of service accurately. Furthermore, if consumers are not able to verify the procedure's execution post-service, then firms could deceive consumers by claiming that a more expensive procedure was performed. Under these conditions, firms cannot credibly commit to a menu. Instead, displaying a menu would amount to displaying a set of prices. We consider this case as an extension to the basic model in Section III.5.1. When committing to a set of prices, a firm can choose any price from the set after the consumer has visited the firm and the firm has learned the consumer's cost type. We show that under some conditions, consumers can rationally believe that a committed firm will always charge the highest price from its set. This ensures that the results from the basic model go through, at least qualitatively.

Another extension is explored in Section III.5.2. In the basic model, we assume that a consumer's valuation and their cost type are realized independently and that all consumers have the same outside option (normalized to 0). While we consider this to be a natural base-line for many of the markets we have in mind (e.g. IT services, cosmetic services, home improvement), it is less plausible in markets for medical services, where the cost of treatment is oftentimes positively correlated with the seriousness of the patient's medical condition. In the second extension, we take this into account by assuming that the consumer's outside option is negatively proportional to the cost of the procedure that is appropriate for their type. This means that forgoing treatment is worse for high-cost types with a more serious medical condition than it is for low-cost types with a less serious medical condition. We show that the results of the basic model go through as long as the factor of proportionality is sufficiently small.

III.1.1 Literature

This paper connects to multiple strands of literature. First, the market we describe is a specific kind of *credence good* market, according to the classification by Dulleck and Kerschbamer (2006). The defining feature of a credence good is the presence of an information asymmetry, where the firm knows more about the relevant characteristics of consumers than the consumers themselves.[5] In contrast to our paper, the existing credence good literature either assumes that prices are exogenous (Pitchik & Schotter, 1987; Sülzle & Wambach, 2005) or that firms have to publicly announce prices for procedures (Wolinsky, 1993; Taylor, 1995; Glazer & McGuire, 1996; Fong, 2005; Dulleck & Kerschbamer, 2006). Hence, none of the existing credence good literature can account for service markets, where firms do not disclose any information on prices before the consumer's visit.

Second, this paper is closely related to the literature on consumer search, where consumers are uncertain about firms' marginal cost due to industry-wide cost shocks (Benabou & Gertner, 1993; Dana, 1994; Tappata, 2009; Janssen et al., 2011, 2017). Note that a model of industry-wide cost shocks, where marginal costs are the same across firms in every state of the world, is formally identical to our framework. In both cases, consumers are uncertain about a firm's marginal cost, and in the absence

[5] Some confusion about the term 'credence good' might arise from a competing definition by Darby and Karni (1973), which distinguishes between search, experience and credence goods. While the consumer learns the value of an experience good after purchase, the value of a credence good is "expensive to judge even after purchase" (Darby & Karni, 1973, p. 69). Note that in our model, consumers learn the value before purchase. Hence, there is little relation to experience goods and credence goods in this sense.

of commitment, firms make a price offer conditional on the cost realization. Since firms maximize expected profits, they behave the same in a market where the realized marginal cost is identical across consumers and where the marginal cost is independently realized across consumers. In contrast to our paper, this literature assumes that a fraction of consumers have no search cost, so-called shoppers, in order to prevent market collapse. This leads to a mixed pricing strategy of firms and a complicated search strategy for consumers, who learn about the cost incrementally as they search different firms and receive different price offers. This literature, however, does not consider the incentives of firms to commit to a single price across states of the world in order to attract consumers. Hence, under a different reading of our paper, we contribute to the literature on consumer search under industry-wide cost shocks by extending the model to allow firms to commit to prices.

Third, there is a small literature that considers the welfare implications of cost-based differential pricing (Chen & Schwartz, 2015; Chen et al., 2021). This literature compares uniform pricing to differential pricing, conditional on a given level of competition. However, the equilibria of our model require us to compare competitive uniform pricing to monopolistic differential pricing. Nevertheless, we are able to build on some of the insights in Chen and Schwartz (2015).

Finally, this paper is related to the literature on consumer search with price advertisement (Robert & Stahl, 1993; Anderson & Renault, 2006; Janssen & Non, 2008) and consumer search with price commitment (Obradovits, 2014; Myatt & Ronayne, 2019). In both strands of literature, marginal costs are identical across consumers and common knowledge. Advertisement, unlike commitment, is costly for firms and only reaches a fraction of consumers. Unlike in our model, where firms could inform consumers about prices at no cost but choose not to, firms want to inform consumers about prices in order to prevent market collapse. Both Obradovits (2014) and Myatt and Ronayne (2019) consider a two-stage Varian (1980) model of search, where all consumers are either shoppers or captive to a given firm. Firms commit to an upper bound in the first stage and, after observing the commitment decisions of competing firms, choose a price below their upper bound in the second stage. Note that commitment does not serve to inform consumers about prices and does not influence second-period demand per se since shoppers observe all prices at no cost and captives only observe the price of their firm, independent of whether firms commit or not. Instead, commitment only restricts the choice set of the firm in the second stage and thereby signals other firms the second stage equilibrium the firm intends to play.

III.2 Model

Consider a market with $n \geq 2$ identical firms and a unit mass of consumers. Each firm offers a single, homogeneous service to consumers, for which consumers have unit demand. Consumer i has a cost type τ_i and a valuation for the service v_i. The cost type determines a firm's cost of serving this consumer. For ease of exposition, we assume that there are two cost types: l (for *low*) and h (for *high*).[6] We denote by c_τ a firm's cost of serving a consumer of cost type $\tau \in \{l, h\}$ and assume $0 \leq c_l < c_h$. Note that the cost of serving a given consumer is identical across firms. The fraction of type l consumers is denoted by λ, where $\lambda \in (0, 1)$. Consumers' valuations are drawn independently from a distribution F with support $[0, \bar{v})$, where $\bar{v}$ could be infinite. If $\bar{v}$ is finite, we assume $\bar{v} > c_l$. Note that a consumer's cost type and valuation are assumed to be independent. We denote the density of F by f, assume that f is continuously differentiable, and assume that $(1 - F)^{-1}$ is convex.[7] Furthermore, we normalize each consumer's outside option to 0.

The market plays out in two stages: a *commitment stage* and a *search stage*. In the commitment stage, each firm simultaneously chooses an action from the set $\mathbb{R}_+ \cup \{\emptyset\}$. Let a_j denote the action of Firm j. If $a_j \in \mathbb{R}_+$, then Firm j publicly displays the price a_j and thereby commits to charging a_j to all consumers in the subsequent stage, irrespective of their cost type. We call such firms *committed firms*. If $a_j = \emptyset$, then Firm j displays no price and makes no such commitment. We call such firms *flexible firms*. The decisions of the commitment stage $(a_1, ..., a_n)$ are observed by all firms and consumers, making the subsequent stage a subgame.

At the beginning of the search stage, flexible firms simultaneously choose a price

[6] The extension to more than two cost types is straight forward. See Appendix III.B.

[7] Convexity of $(1 - F)^{-1}$ ensures that monopoly profits are quasi-concave and is weaker than the assumption of log-concavity of $1 - F$ (Caplin & Nalebuff, 1991).

vector consisting of a price for each cost type. Let $(p_{j,l}, p_{j,h})$ denote the price vector chosen by Firm j. This decision is neither observed by consumers nor by any other firm. Then, consumers search firms sequentially with search cost $s > 0$ and costless recall. Consumers can only buy from firms which they have visited and consumers incur s when visiting committed firms as well.[8] Consumers can identify firms by their decision in the commitment stage and can choose the order in which they visit these firms. Firms with identical decisions in the commitment stage are indistinguishable to consumers and are visited with equal probability. We assume that a consumer's valuation is realized on their first visit at any firm, hence only after they've incurred the search cost for the first time.[9] Consumers do not observe their own cost type and do not observe the price that a flexible firm would have offered if they were a different cost type. Instead, consumers update their belief about their cost type, given the offers they receive from flexible firms and the expectations about these firms' pricing strategies. To summarize, a visit of Consumer i at Firm j unfolds as follows.

1. Consumer i incurs the search cost s.

2. If Firm j is the first firm visited by Consumer i, then Consumer i learns v_i.

3. Firm j offers the service at a price p, where $p = p_{j,\tau_i}$ if j is flexible and $p = a_j$ if j is committed.

4. Given the offer p, the consumer updates their belief about their cost type.

5. The consumer decides whether to buy from Firm j at price p, search another firm, buy from any previously visited firm, or leave the market without purchase.

III.3 Equilibria

The firms' decisions in the commitment stage can create three distinct market environments in the search stage. When all firms display a price, we speak of a *committed*

[8]This is because a firm can only serve a consumer after it has learned the consumer's type, which requires a visit from the consumer.

[9]Alternatively, we could assume that consumers know their valuation from the start and that the first search is free. We show in Appendix III.C that the equilibria we identify persist under this alternative model.

market. A committed market resembles Bertrand competition, as consumers will simply buy from the cheapest firm. When no firm displays a price, we speak of a *flexible market*. A flexible market is reminiscent of Diamond (1971), where consumers sequentially search identical firms for the lowest price. However, the presence of different cost types in our model complicates the analysis, as consumers have to form and update beliefs about their cost type. Finally, we are in a *partially committed market*, when there are both committed and flexible firms present. In comparison to a flexible market, the presence of committed firms creates an additional outside option for consumers. This effectively creates a price ceiling, as flexible firms cannot profitably charge more than the search cost above the lowest displayed price.

Whether or not a firm finds it beneficial to display a price depends on the firm's profits in the ensuing market environments. Since the search stage is a subgame, we solve for an equilibrium by backwards induction. First, we determine the possible equilibria for each of the three environments of the search stage. Then, taking the outcome of the search stage as given, we determine the equilibria of the commitment stage.

We restrict attention to perfect Bayesian equilibria in pure and symmetric strategies, both for the search and the commitment stage. This restriction leaves two kind of equilibrium candidates, one where all firms display a price and one where no firm displays a price. Even though a partially committed market will not arise on any equilibrium path, it is crucial for understanding firms incentive to deviate in the commitment stage.

Besides focusing on pure and symmetric equilibria, we impose two additional equilibrium refinements. First, we assume that consumers have *passive beliefs* about the prices of flexible firms. This means that when a consumer observes a deviation by a flexible firm in the search stage, they do not change their expectations about the pricing strategy of other flexible firms. Second, we assume that consumers' beliefs about their cost type satisfy the *intuitive criterion* (Cho & Kreps, 1987). Specifically, when a consumer observes a deviation by a flexible firm in the search stage, which could not possibly be profitable under some cost type, the consumer's belief assigns zero probability to this cost type.[10]

[10]The intuitive criterion is applicable, as the interaction between the consumer and the firm resem-

III.3.1 Search Stage

Committed Market

We begin with committed markets, where all firms display a price. We denote by $\underline{p}$ the lowest price displayed among all firms. Consumers will either search once or not at all. Since consumers are initially uncertain about their valuation, they have to compare the expected surplus from buying at $\underline{p}$ with s. Let $CS(p) := \int_p^\infty (v-p)dF(v)$, such that $CS(p)$ denotes the expected surplus of visiting a firm charging p. If

$$CS(\underline{p}) \geq s \tag{III.1}$$

then there exists an equilibrium where each consumer i randomly visits a firm displaying $\underline{p}$, buys if $v_i \geq \underline{p}$ and leaves the market without purchase otherwise. Let $c_e := \lambda c_l + (1-\lambda)c_h$ denote the expected marginal cost. Each firm committed to $\underline{p}$ receives an equal share of the industry profits

$$\left(1 - F\left(\underline{p}\right)\right)\left(\underline{p} - c_e\right).$$

If $CS(\underline{p}) \leq s$, then there exists an equilibrium where consumers don't make an initial search and firms make no profits. These are the only possible equilibria.

Flexible Market

In the flexible market, each firm initially chooses a price vector. Since we restrict attention to equilibria in pure and symmetric strategies, we consider equilibrium candidates where all firms choose an identical vector (p_l, p_h). We then characterize the search behavior of consumers who expect firms to price according to (p_l, p_h) and consider a firm's incentive to deviate. If a firm has no incentive to deviate from its pricing strategy, we have identified an equilibrium.

Consumers initially expect all firms to charge the same price, hence expect to search at most once. Let

$$\Delta ES(p_l, p_h) := \lambda CS(p_l) + (1-\lambda)CS(p_h) - s$$

bles the original sender-receiver game. The firm is the sender, having private information about the state of the world, which is the consumer's cost type. The state is pay-off relevant for the consumer, as it determines what other firms will charge.

denote the expected surplus net of search cost. Consumers make an initial search if $\Delta ES(p_l, p_h) \geq 0$. Assume this holds. As firms are indistinguishable in the flexible market, each consumer i will randomly select one of the firms for an initial visit. Upon visit at the first firm, the consumer learns v_i, receives an offer, and updates their belief about their cost type using Bayes rule whenever possible. We denote by $\mu_i \in [0, 1]$ the consumer's belief that they are of the low-cost type. If the consumer receives an offer of either p_l or p_h, they believe with certainty that they are of the corresponding cost type (assuming $p_l \neq p_h$). They buy if v_i is above the offer and leave the market without purchase otherwise, as they expect to receive the same offer from any of the other firms.

If the firm has deviated to a price p^d that is neither p_l nor p_h, then we are out of equilibrium and the consumer's beliefs, both about their cost type and the pricing strategy of other firms, are not determined by Bayes rule. With passive beliefs about other firms' prices and for some belief $\mu_i \in [0, 1]$ about their cost type, the consumer will search an additional firm if and only if the expected gain in surplus is greater than the search cost.[11] Formally,

$$\mu_i \max\{v_i - p_l, v_i - p^d, 0\} + (1 - \mu_i) \max\{v_i - p_h, v_i - p^d, 0\} - \max\{v_i - p^d, 0\} \geq s.$$

Independent of v_i and μ_i, an additional search is never beneficial if $p^d < \min\{p_l, p_h\} + s$. This means that a firm can deviate away from the lower price by up to the search cost, without losing the consumer to a competing firm. Furthermore, since consumers do not observe deviations of firms they do not visit, a firm cannot attract additional consumers for a visit through a deviation. Hence, for the fraction of $1/n$ consumers that initially visit the firm and for prices locally around the lower price, the firm acts like a monopolist and receives expected profits of

$$\Pi_\tau(p) := (1 - F(p))(p - c_\tau)$$

per consumer of type $\tau \in \{l, h\}$. Let $p^m_\tau := \arg\max_p \Pi_\tau(p)$ denote the monopoly price for type τ consumers.[12] Our assumptions on F imply that $\Pi_\tau(p)$ is single-peaked,

[11] Without passive beliefs, nearly any pricing strategy can be supported in equilibrium, by the belief that all other firms have deviated to charging low prices.

[12] If $c_h > \bar{v}$ then let p^m_h denote c_h.

hence increases as p moves closer to p_τ^m, and that $p_l^m < p_h^m$. Therefore, unless $p_l = p_l^m$, a firm has a profitable deviation.

Lemma III.1. In any equilibrium of the flexible market where consumers make an initial search, $p_l = p_l^m$ and $p_h > p_l$.

For a detailed proof, see Appendix III.A.

The mechanism that pushes the equilibrium price for low-cost consumers towards the monopoly price is the same as in Diamond (1971). When facing a slight deviation, the consumer does not find it profitable to incur the search cost for visiting another firm, and hence, the price must maximize monopoly profits for firms to not have an incentive to deviate. Note that this argument does not apply to the price charged to high-cost consumers. Assume that $p_h^m > p_l^m + s$ and consider $p_h \in [p_l^m + s, p_h^m)$. If a firm deviates slightly upward from p_h, then whether an additional search is beneficial depends on the consumer's belief μ_i. One could punish a deviating firm with the belief that the consumer is certainly a low-cost type ($\mu_i = 1$). However, the intuitive criterion rules out such beliefs. As a firm could not possibly increase its profits from low-cost consumers when $p_l = p_l^m$, the intuitive criterion implies that $\mu_i = 0$ after the consumer observes a deviation that could be profitable if they were a high-cost type. Therefore, unless $p_h = p_h^m$, a firm has a profitable deviation. Note that in addition to the equilibrium where firms charge monopoly prices, there always exists an equilibrium where consumers do not search. The following proposition states all equilibria of the flexible market.

Proposition III.1. In the flexible market there exists at most two equilibria.

(i) There exists an equilibrium where firms price sufficiently high such that consumers do not search, i.e., $\Delta ES(p_l, p_h) \leq 0$, and the market collapses.

(ii) If $\Delta ES(p_l^m, p_h^m) \geq 0$, then there exists an equilibrium where firms choose monopoly prices, i.e., $p_\tau = p_\tau^m$ for $\tau \in \{l, h\}$, and consumers search exactly once.

For a detailed proof, see Appendix III.A.

Finally, we consider firms' profits in the equilibrium where consumers search. Let $\Pi_\tau^m := \Pi_\tau(p_\tau^m)$ denote monopoly profits from consumers of type τ. Each firm earns an equal share of the monopoly profits, which is

$$\frac{1}{n}\left(\lambda\Pi_l^m + (1-\lambda)\Pi_h^m\right).$$

This concludes the analysis of the flexible market.

Partially Committed Market

Finally, we come to the analysis of partially committed markets, where both flexible and committed firms are present. As in the previous sections, $\underline{p}$ refers to the lowest displayed price, and (p_l, p_h) refers to the pure and symmetric pricing strategy of flexible firms. Consumers will never visit firms displaying a price above $\underline{p}$ and we, therefore, ignore these firms from now on. Taking $\underline{p}$ as given, we describe the search behavior of consumers who expect all flexible firms to charge (p_l, p_h) and then consider flexible firms' incentives to deviate from this pricing strategy.

We begin with the trivial case where both p_l and p_h are above $\underline{p}$. Then consumers have no reason to visit flexible firms, as they never make a better offer than committed firms. A flexible firm cannot attract consumers by lowering prices in the search stage, because a deviation from the pricing strategy is unobserved. We call this a *pessimistic equilibrium*, as it is supported by consumers' pessimistic expectations about the prices of flexible firms. Either expected consumer surplus is above the search cost, in which case committed firms are visited by consumers, or it is not, in which case the market collapses.

Next, we consider equilibria where flexible firms are visited by at least some consumers on the equilibrium path. We call these *optimistic equilibria*, as they require consumers to be sufficiently optimistic about the prices of flexible firms. Note that consumers are identical before the first search, as they have the same prior beliefs about their type and their valuation. Therefore, our restriction to pure and symmetric equilibria implies that all consumers take the same initial action. This leaves two candidates for optimistic equilibria. Either (i) all consumers initially visit flexible firms or (ii) all consumers initially visit committed firms and then some consumers make an additional visit at flexible firms. The following lemma rules out equilibria of the second kind.

Lemma III.2. Fix $\underline{p}$, p_l, and p_h. Assume there exists a $v_i \in [0, \bar{v})$, such that Consumer i would benefit from visiting a flexible firm, after already having visited a committed firm. Then under the optimal search strategy, consumers would either visit a flexible firm first or not participate in the market at all.

We prove Lemma III.2 in Appendix III.A.

Above we have shown that in any optimistic equilibrium, flexible firms are visited first. Next, we consider the pricing strategy of flexible firms in such an equilibrium. We build on the insights of Section III.3.1. Under passive beliefs, firms have an incentive to deviate unless $p_l = p_l^m$. Given $p_l = p_l^m$ and the intuitive criterion, firms can slightly deviate away from p_h without losing the consumer to a competing firm, and hence, p_h must maximize monopoly profits. However, in contrast to Section III.3.1 there is no demand at prices above $\underline{p} + s$, as consumers would be better off paying s to visit a committed firm and buying at $\underline{p}$. Hence, p_h must maximize $D_{\underline{p}}(p)(p - c_h)$ where

$$D_{\underline{p}}(p) := \begin{cases} 1 - F(p) & \text{for } p \leq \underline{p} + s \\ 0 & \text{otherwise} \end{cases}.$$

Finally, if $\underline{p} + s < c_h$ then the firm wants to deter consumers from buying. While any price above $\underline{p} + s$ would have the desired effect, we simplify notation and assume that, in this case, firms charge p_h^m. Note that committed firms are only visited if the price of a flexible firm is strictly above $\underline{p} + s$. Since this occurs only if $\underline{p} + s < c_h$, committed firms never make a profit and might even make a loss. The following lemma summarizes the above conditions on the optimistic equilibrium.

Lemma III.3. If there exists an optimistic equilibrium, then in this equilibrium $p_l = p_l^m$ and

$$p_h = \begin{cases} \underline{p} + s & \text{if } p_h^m \geq \underline{p} + s \geq c_h \\ p_h^m & \text{otherwise} \end{cases}.$$

In this equilibrium, committed firms make no profits.

Next, we identify the conditions under which an optimistic equilibrium exists. We begin by stating a necessary condition, namely, that consumers are better off first searching a flexible firm compared to only searching a committed firm. Note that the expected surplus of high-cost consumers facing a price strictly above $\underline{p}+s$ is $CS(\underline{p}+s)$, as consumers can pay s to visit a committed firm and buy at $\underline{p}$. Hence, the condition is given by

$$\lambda CS(p_l^m) + (1-\lambda)CS(\min\{\underline{p}+s, p_h^m\}) \geq CS(\underline{p}). \tag{III.2}$$

Note that (III.2) is trivially satisfied if $\underline{p} \geq p_h^m$, as the consumer will never receive a better offer from the committed firm. Similarly, (III.2) is trivially violated if $\underline{p} \leq p_l^m$, as the consumer will never receive a better offer from the flexible firm. Therefore, interesting cases are $\underline{p} \in (p_l^m, p_h^m)$. We find that (III.2) has a single crossing property, meaning there exists a threshold $t \in (p_1^m, p_h^m)$ such that (III.2) is satisfied if and only if $\underline{p} \geq t$. We prove this in Appendix III.A. The threshold t is the unique solution to the equation

$$\lambda CS(p_l^m) + (1-\lambda)CS(\min\{t+s, p_h^m\}) = CS(t). \tag{III.3}$$

We analyze t further down in this section.

Note that by Lemma III.2, it can never be optimal to first search a committed firm and then sometimes, depending on the valuation, make an additional visit at a flexible firm. Hence, besides the two search strategies compared by (III.2), the only other candidate for the optimal search strategy is to not search at all. Therefore, an optimistic equilibrium exists if and only if both $\underline{p} \geq t$ and $\Delta ES(p_l^m, \min\{\underline{p}+s, p_h^m\}) \geq 0$. We summarize the possible equilibria of partially committed markets in the following proposition.

Proposition III.2. In any partially committed market, there exists at most three equilibria.

(i) If $CS(\underline{p}) \geq s$, then there exists a pessimistic equilibrium where each consumer visits a firm committed to $\underline{p}$, buys if $v_i \geq \underline{p}$, and leaves the market without purchase otherwise. Flexible firms price sufficiently high to never be visited.

(ii) If $CS(\underline{p}) \leq s$, then there exists a pessimistic equilibrium where consumers don't make an initial search, and the market collapses. Flexible firms price sufficiently high to never be visited.

(iii) If $\underline{p} \geq t$ and $\Delta ES(p_l^m, \min\{\underline{p}+s, p_h^m\}) \geq 0$, then there exists an optimistic equilibrium where each consumer initially visits a flexible firm. Committed firms do not make a profit. Flexible firms price as described in Lemma III.3.

We now come back to the threshold t. We find that t is weakly monotonically increasing in s. To see this, consider (III.3). Increasing s, while holding the other parameters fixed, weakly decreases the left-hand side of (III.3) while the right-hand side remains the same. Due to the single crossing property of (III.2), the right-hand side decreases faster in t than the left-hand side, so in order to achieve equality, t must increase. Denote the bounds of t in relation to s by $\underline{t}$ and $\bar{t}$. It is easy to see that for $s \to 0$, $t \to p_l^m$ and hence $\underline{t} = p_l^m$. On the other hand, if s is sufficiently large, then $\min\{t+s, p_h^m\} = p_h^m$ and hence $\bar{t}$ solves

$$\lambda CS(p_l^m) + (1-\lambda)CS(p_h^m) = CS(\bar{t}). \tag{III.4}$$

Additionally, we know that $\bar{t} \leq \lambda p_l^m + (1-\lambda)p_h^m$ by the fact that $CS(p)$ is convex in p.[13]

This concludes the analysis of the search stage. In the following section, we identify the equilibria of the commitment stage.

III.3.2 Commitment Stage

Now that we have determined the outcomes of the different markets, we come back to the commitment stage to understand firms' incentives to display a price. Note that our restriction to pure and symmetric equilibria greatly restricts the space of equilibrium candidates. Either all firms display the same price in the commitment stage or no firm displays a price.

First, consider the equilibrium candidate where every firm displays the same price $\underline{p}$. If $\underline{p}$ is above the expected marginal cost c_e, then a firm can increase its profit by slightly decreasing its displayed price and thereby attracting all consumers. Hence, the only candidate is $\underline{p} = c_e$. Deviating to not displaying a price would result in a partially committed market. If consumers are pessimistic, which they could rationally

[13] $\frac{\partial CS(p)}{\partial p} = -D^m(p)$ and $D^m(p)$ is downward sloping.

be in every partially committed market, this deviation will not result in any profit for the deviating firm. Therefore, every firm displaying a price of c_e is an equilibrium. We call this the *competitive equilibrium*, as all firms price at (expected) marginal cost and make no profits.

Proposition III.3 (Competitive Equilibrium)**.** There exists an equilibrium where all firms display a price of c_e. In this equilibrium, firms make no profits.

Next, consider the equilibrium candidate where no firm displays a price in the commitment stage. Assume $\Delta ES(p_l^m, p_h^m) \geq 0$, such that on the equilibrium path, this results in a flexible market where each firm earns an equal share of the monopoly profits. Whether a firm finds it beneficial to deviate to displaying a price depends on its profits in the resulting partially committed market. In an effort to deter a deviation, we specify that the optimistic equilibrium is played in all partially committed markets where it exists. In a partially committed market where the optimistic equilibrium doesn't exist, the pessimistic equilibrium is played. Let Π^* denote the highest profits of a firm among all partially committed markets that can be reached through a unilateral deviation of that firm. If these profits are still lower than the shared monopoly profits, then no firm displaying a price is an equilibrium. We call this the *monopolistic equilibrium*, as all firms charge the respective monopoly price to each type of consumer and share monopoly profits.

Proposition III.4 (Monopolistic Equilibrium)**.** If $\Delta ES(p_l^m, p_h^m) \geq 0$ and

$$\Delta\Pi := \frac{1}{n}\left(\lambda\Pi_l^m + (1-\lambda)\Pi_h^m\right) - \Pi^* \geq 0$$

then there exists an equilibrium where no firm displays a price. In this equilibrium, each firm earns an equal share of the monopoly profits, i.e., $\frac{1}{n}\left(\lambda\Pi_l^m + (1-\lambda)\Pi_h^m\right)$.

In order to understand the conditions for which the monopolistic equilibrium exists, we need to determine Π^*. For the deviating firm to make a profit, it must display a price $\underline{p}$ below t, leading to a market where the optimistic equilibrium doesn't exist. Note that from $\Delta ES(p_l^m, p_h^m) \geq 0$ and (III.4) it follows that $CS(\underline{p}) \geq s$ for all $\underline{p} \leq t$. Hence, the only equilibrium in such a partially committed market is the pessimistic one, where all consumers visit the deviating firm and where this firm makes a profit of

$$\lambda\Pi_l(\underline{p}) + (1-\lambda)\Pi_h(\underline{p}) \equiv \left(1 - F\left(\underline{p}\right)\right)\left(\underline{p} - c_e\right).$$

Consider the case $t > c_e$ such that $\Pi^* > 0$. Then there are two candidates for optimal deviations leading to Π^*. Let $p^m(c) := \arg\max_p (1 - F(p))(p - c)$.[14] If $p^m(c_e) < t$, then the optimal deviation is to display the price $p^m(c_e)$ in the commitment stage. If $p^m(c_e) \geq t$, then the optimal deviation is to display a price just below t in the commitment stage, as our assumptions of F ensure that profits are single-peaked. We find that if $\partial p^m(c)/\partial c$, the pass-through rate from marginal cost to the monopoly price, is non-increasing in c, then $p^m(c_e) \geq t$.

Lemma III.4. If $p^m(c)$ is concave then $p^m(c_e) \geq t$ and

$$\Pi^* = \max\{(1 - F(t))(t - c_e), 0\}.$$

Proof. If $p^m(c)$ is concave, then $\lambda p^m(c_l) + (1-\lambda)p^m(c_h) \leq p^m(\lambda c_l + (1-\lambda)c_h)$. Since $t \leq \lambda p^m(c_l) + (1-\lambda)p^m(c_h)$, this implies $t \leq p^m(c_e)$. □

Note that $p^m(c)$ is concave for many common demand functions, for instance, all demand functions with a constant cost pass-through rate (see Bulow and Pfleiderer (1983)).

In the remainder of this section, we discuss how $\Delta\Pi$ depends on the primitives of the model. We assume throughout that $p^m(c)$ is indeed concave such that

$$\Delta\Pi = \frac{1}{n}\left(\lambda\Pi_l^m + (1-\lambda)\Pi_h^m\right) - \max\{(1 - F(t))(t - c_e), 0\}.$$

First, it is easy to see that $\Delta\Pi$ is decreasing in the number of firms n. In a flexible market, monopoly profits are shared among firms, whereas the best deviation will attract all consumers and hence deviation profits are independent of the number of competitors.

Second, consider the relation between $\Delta\Pi$ and s. As shown in Section III.3.1, t is weakly increasing in s. The intuition is that, as s increases, flexible firms can charge more on top of $\underline{p}$ without losing the consumer to a committed firm. This, in turn, makes it less beneficial to visit flexible firms in the first place, increasing t. Since t is assumed to be below $p^m(c_e)$, increasing t increases deviation profits $(1 - F(t))(t - c_e)$. Hence, $\Delta\Pi$ is weakly decreasing in s.

[14] If $c \geq \bar{v}$ then let $p^m(c)$ denote c.

Third, consider the extreme case where λ is close to 0 (resp. 1). Since consumers expect to be of the high-cost type (resp. low-cost type) with near certainty, a deviating firm can attract consumers by pricing closely below p_h^m (resp. p_l^m). See (III.3) to confirm that $t \to p_h^m$ (reps. $t \to p_l^m$) as $\lambda \to 0$ (reps. $\lambda \to 1$). Since the deviating firm serves nearly only high-cost (resp. low-cost) consumers, it earns close to monopoly profits. In the flexible market, on the other hand, firms would have to share these same monopoly profits. Hence, as $\lambda \to 0$ (reps. $\lambda \to 1$), $\Delta\Pi \to -\frac{n-1}{n}\Pi_h^m$ (reps. $\Delta\Pi \to -\frac{n-1}{n}\Pi_l^m$).

The following proposition summarizes these results.

Proposition III.5. Let $p^m(c)$ be concave. Then $\Delta\Pi$ strictly decreases in n and weakly decreases in s. Furthermore, for λ sufficiently close to 1, $\Delta\Pi < 0$ and if $\Pi_h^m > 0$, then for λ sufficiently close to 0, $\Delta\Pi < 0$ as well.

Finally, we consider $\Delta\Pi$ for s close to 0. This simplifies the analysis in two ways. First, it ensures $\Delta ES(p_l^m, p_h^m) \geq 0$, and hence, the monopolistic equilibrium exists if and only if $\Delta\Pi \geq 0$. Second, it pins down t, as we have shown in Section III.3.1 that $t \to p_l^m$ as $s \to 0$. We find that for both λ and c_h, the existence of the monopolistic equilibrium can be approximated by a threshold condition with arbitrary precision.

Proposition III.6. Let $p^m(c)$ be concave. Then for every $\varepsilon > 0$ there exists an $s > 0$ such that the following holds.

(i) There exists $\bar{c}_h$ such that $\Delta\Pi > 0$ if $c_h \geq \bar{c}_h + \varepsilon$ and $\Delta\Pi < 0$ if $c_h \leq \bar{c}_h - \varepsilon$.

(ii) There exists $\bar{\lambda}$ such that $\Delta\Pi < 0$ if $\lambda \geq \bar{\lambda} + \varepsilon$ and $\Delta\Pi > 0$ if $\lambda \in [\varepsilon, \bar{\lambda} - \varepsilon]$.

We prove Proposition III.6 in Appendix III.A. We illustrate the proposition with the following example. Assume $c_l = 0$, $n = 2$, $s = 10^{-4}$ and that valuations are uniformly distributed on $[0, 1]$. Then the free parameters are c_h and λ. The grey area in Figure III.1 indicates the parameter space under which the monopolistic equilibrium exists. Indeed, for each λ the monopolistic equilibrium exists if and only if c_h is above some threshold. For large λ, this threshold is above 1, i.e., the highest possible valuation in this example. Furthermore, for each c_h the monopolistic equilibrium exists if λ is below some threshold, but not too close to 0. Note that for small c_h, $\bar{\lambda}$ is below

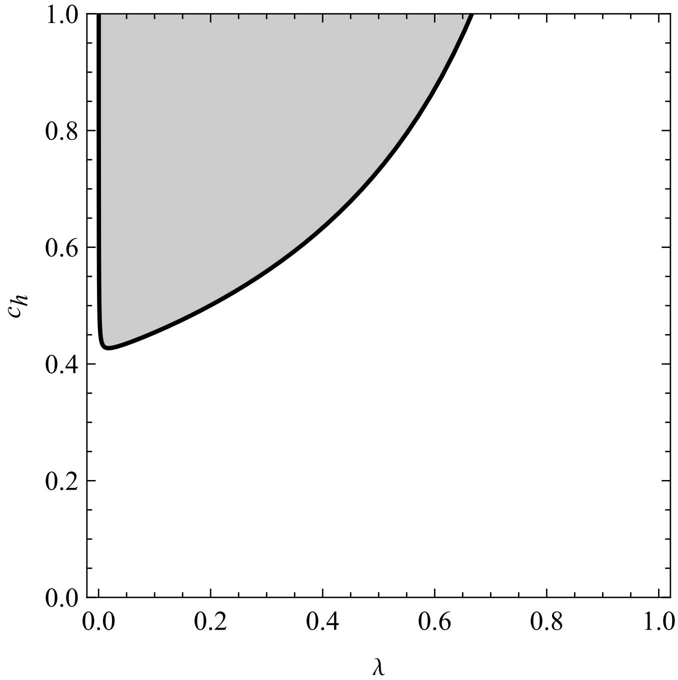


Figure III.1: Existence of the monopolisitc equilibrium.

2ε such that the monopolistic equilibrium doesn't exist for any $\lambda \in (0, 1)$. To gain some intuition for this result, note that the deviating firm serves consumers with an average cost c_e at the monopoly price for c_l. Hence, a deviation becomes unattractive if c_e is sufficiently far above c_l, which is the case if either c_h is large or if λ is low.

III.4 Welfare

In this section, we compare consumer surplus and total surplus between the monopolistic and competitive equilibrium. As a first step, we compare our setting to the existing literature on cost-based differential pricing. Chen and Schwartz (2015) find that when monopoly consumer surplus is convex in marginal cost, both consumer surplus and total surplus are larger under monopolistic cost-based differential pricing compared to monopolistic uniform pricing across cost types. In a perfectly competitive market, cost-based differential pricing always yields higher consumer and total surplus than uniformly pricing at expected cost. This is due to the fact that consumer surplus is convex in price and that under both pricing schemes firms make no profits. The equilibria of our setting require us to compare surplus under monopolistic differential pricing (i.e., the monopolistic equilibrium) to surplus under uniform pricing in a perfectly competitive market (i.e., the competitive equilibrium). To our knowledge, this has not been done by the literature.

Assume for now that the monopolistic equilibrium exists. Let $CS_\tau^m := CS(p_\tau^m)$ such that CS_τ^m is consumer surplus for type τ consumers under the monopoly price. Consumer surplus is larger under the monopolistic equilibrium than the competitive equilibrium if and only if

$$\Delta CS := \lambda CS_l^m + (1-\lambda) CS_h^m - CS(c_e) \geq 0 \tag{III.5}$$

and total surplus is larger if and only if

$$\Delta TS := \lambda \left(CS_l^m + \Pi_l^m\right) + (1-\lambda)\left(CS_h^m + \Pi_h^m\right) - CS(c_e) \geq 0. \tag{III.6}$$

First, note that neither ΔCS nor ΔTS depends on the number of firms, as equilibrium prices are independent of the number of firms (conditional on the monopolistic equilibrium existing). Second, note that neither ΔCS nor ΔTS depends on the search cost, as every consumer searches in either equilibrium exactly once. Therefore, conditional on the monopolistic equilibrium existing, ΔCS and ΔTS depends only on the distribution of valuations F and the distribution of costs (c_l, c_h, λ). We denote by $\mathrm{Var}[c] := \lambda c_l^2 + (1-\lambda)c_h^2 - c_e^2$ the variance and by $\mathrm{SD}[c] := \sqrt{\mathrm{Var}[c]}$ the standard deviation of the cost distribution. Note that a mean preserving spread of the cost distribution neither changes consumer nor total surplus of the competitive equilibrium. How it affects surplus in the monopolistic equilibrium depends on the curvature of the monopoly price with respect to marginal cost, $p^m(c)$.

Proposition III.7. Assume $CS(p^m(c))$ is convex in c. A mean preserving spread of the cost distribution increases both ΔCS and ΔTS.

We prove Proposition III.7 in Appendix III.A. Chen and Schwartz (2015) shows that $CS(p^m(c))$ is convex in c for many common demand functions. Note that $CS(p^m(c))$ is always convex in c if $p^m(c)$ is concave.

Proposition III.7 informs us that the relative social desirability of the monopolistic equilibrium (compared to the competitive equilibrium) increases as costs become more dispersed. To be able to quantify the exact amount of cost dispersion necessary for the monopolistic equilibrium to yield higher consumer or total surplus, we have to make assumptions on the demand. We find that if valuations are distributed uniformly on $[0, \bar{v}]$, then there exists a measure of cost dispersion

$$\sigma := \frac{\mathrm{SD}[c]}{\bar{v} - c_e}$$

that uniquely determines the relative social desirability of the two equilibria.

Proposition III.8. Assume valuations are uniformly distributed on $[0, \bar{v}]$. Then $\Delta CS \geq 0$ if and only if $\sigma \geq \sqrt{3}$ and $\Delta TS \geq 0$ if and only if $\sigma \geq \frac{1}{\sqrt{3}}$.

We prove Proposition III.8 in Appendix III.A. The proof easily generalizes to more than two cost types. Hence, when valuations are uniformly distributed, the conditions hold true for any distribution of cost types.

Until now we have compared consumer and total surplus under the assumption that the monopolistic equilibrium exists. However, if, for instance, the monopolistic equilibrium would only exist whenever it was socially desirable, then there wouldn't be a need for a policy intervention. Conversely, if the monopolistic equilibrium would only exist in a parameter range where the competitive equilibrium yields higher surplus, then welfare could be improved by indiscriminately mandating price commitments in markets where firms currently don't post prices. We show in the following that neither of these cases obtains. Even when restricting attention to uniformly distributed valuations, either equilibrium can yield higher consumer or total surplus in the range of parameters where both equilibria co-exist.

We continue with the example given in Section III.3, where $c_l = 0$, $n = 2$, $s = 10^{-4}$ and where valuations are uniformly distributed on $[0, 1]$. Figure III.2 partitions the parameter space for λ and c_h based on existence of the monopolistic equilibrium and social desirability of the monopolistic equilibrium relative to the competitive equilibrium. As before, the monopolistic equilibrium exists in the area above the solid curve. The dashed curve represents the condition $\sigma = \frac{1}{\sqrt{3}}$, such that above the dashed curve, total surplus is larger under the monopolistic equilibrium. The dotted curve represents $\sigma = \sqrt{3}$, such that above the dotted curve, consumer surplus is larger under the monopolistic equilibrium. The example confirms that, indeed, either equilibrium can yield higher consumer or total surplus in the range of parameters where both equilibria co-exist.

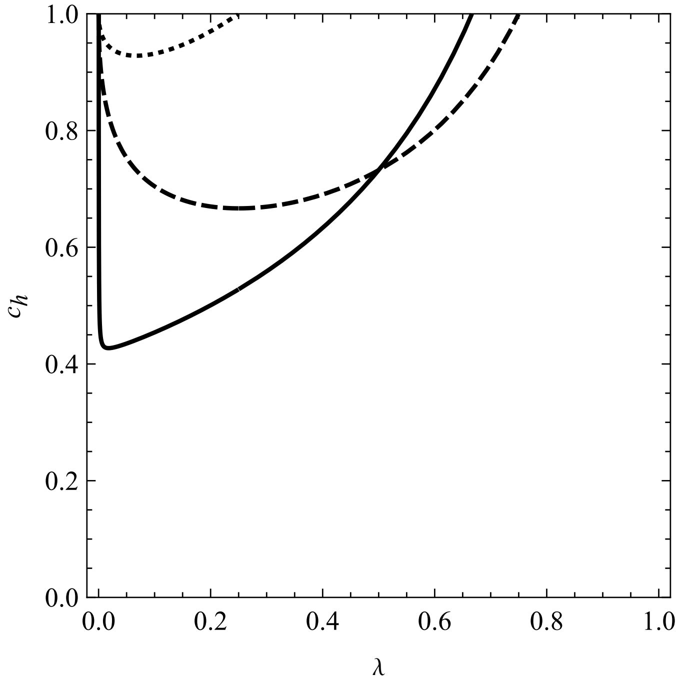


Figure III.2: Existence and surplus of equilibria.

III.5 Robustness

III.5.1 Commitment to Sets

In this section, we consider the case where firms, instead of committing to a single price, can commit to a set of prices, from which they can freely choose in the search stage. Note that commitment to a menu, in the sense of a pricing strategy $p(\tau)$, cannot be enforced. Consumers do not learn their type and hence could not confirm whether they were charged the price corresponding to their type or corresponding to any other type.

We now formally present the model. In the commitment stage, each Firm j simultaneously chooses an action A_j, where A_j is a compact subset of $\mathbb{R}_+$. If $A_j \neq \emptyset$, then Firm j publicly displays a set of prices A_j and thereby commits to charging a price from the set in the subsequent stage. Note that if A_j has two or more elements, the firm could choose different prices for different cost types. We still refer to these firms as committed firms, as they made some commitment rather than none. If $A_j = \emptyset$, then Firm j displays no price and makes no such commitment. At the beginning of the search stage, both committed and flexible firms choose a pricing strategy $(p_{j,l}, p_{j,h})$, where $(p_{j,l}, p_{j,h}) \in \mathbb{R}_+^2$ if Firm j is flexible and $(p_{j,l}, p_{j,h}) \in A_j^2$ if Firm j is committed. We show that both the competitive and the monopolistic equilibrium survive in this framework.

We begin with the competitive equilibrium. Assume that on the equilibrium path, all firms commit to the singleton set $\{c_e\}$. Consider a deviation of Firm j in the commitment stage to A_j with $\max A_j > c_e$. If the maximal element of A_j was below c_e, the firm could never make a profit from this deviation. The deviation can be deterred by consumer expectations that Firm j will charge $\max A_j$ to both types in the search stage. With these expectations, consumers never visit Firm j, and hence, the deviating firm makes no profits.

Next, consider the monopolistic equilibrium and a deviation of Firm j to A_j in the commitment stage. If $\max A_j \geq t$, then an equilibrium analogous to the optimistic equilibrium of Section III.3.1 exists. The flexible firms choose prices as described in Lemma III.3. Firm j chooses the pricing strategy $(\max A_j, \max A_j)$ and consumers never visit this firm unless $c_h > \max A_j + s$. If $c_h > \max A_j + s$, the deviating firm would make a loss in any case. Hence, assume $c_h \leq \max A_j + s$. Because the committed firm is not visited on the equilibrium path, a deviation to a different pricing strategy in the search stage would be unobserved by consumers. If $\max A_j < t$, then only the pessimistic equilibrium as described in Section III.3.1 exists and the deviating firm could make a profit. Note that maximal deviation profits are higher compared to the baseline model, as the committed firm can charge $p_l^m < t$ to low-cost consumers. Hence, the monopolistic equilibrium exists for a smaller region of the parameter space.

III.5.2 Correlated Outside Options

For some goods and services, the outside option of consumers is correlated with consumers' cost types. A typical example is a medical condition, which can vary in its severity. If the condition is more severe, then it is more costly for the doctor to treat the patient, and it is more costly for the patient to leave the condition untreated. To incorporate this into our model, assume that the outside option of a consumer with type τ is $-\phi c_\tau$ where $\phi \geq 0$. In the absence of competition, the consumer would buy if $v - p \geq -\phi c_\tau$. Note that for $\phi = 0$, this reduces to the baseline model.

Since consumers are assumed to not know their type, we assume that they don't know their outside option either. Instead, they infer both their type and outside option from the price they are being offered by a flexible firm. In a flexible market, this opens up incentives for firms to charge the equilibrium price of high-cost consumers to low-cost consumers in order to deceive consumers into believing that they are of the high-cost type, effectively raising the consumer's valuation. Let $p_\tau^m := \arg\max_p (1 - F(p - \phi c_\tau))(p - c_\tau)$. If ϕ is sufficiently small, then

$$(1 - F(p_l^m - \phi c_l))(p_l^m - c_l) \geq (1 - F(p_h^m - \phi c_h))(p_h^m - c_l)$$

and firms are still better off charging the corresponding monopoly price to each cost type. Similarly, if ϕ is sufficiently small, then the optimistic equilibrium of a partially committed market persists. Hence, the results of the baseline model go through as long as ϕ isn't too large.

III.6 Conclusion

In this paper, we propose a model of service markets to explain why, in some markets, firms do not publicly display prices. Our explanation does not rely on collusion in an infinitely repeated setting. Instead, it hinges on an information asymmetry between firms and consumers. Consumers differ in their cost type, which determines the firms' marginal cost of serving a consumer. Firms know a consumer's cost type, whereas consumers do not. Displaying a price comes with the consequence of serving all consumers at that price. This situation enables an equilibrium where no firm displays a price. Because firms can react to displayed prices, they charge at most the search cost above the displayed price. To attract consumers, the deviating firm must display a price so low that it would be better off sharing monopoly profits. The existence of this equilibrium requires that costs are sufficiently dispersed. The social desirability of this equilibrium, relative to pricing at the expected marginal cost, depends on the level of cost dispersion as well. In fact, if costs are sufficiently dispersed, even consumers benefit from firms not displaying prices compared to a price-transparent market.

The information asymmetry plays a crucial role in our model. To see this, assume consumers would know their own type. Then, each type would constitute a separate market. The model would then reduce to Diamond (1971), with the possibility of committing to a price. If no firm displays a price, then a firm could attract all consumers by displaying a price slightly below the monopoly price. The unique equilibrium outcome of such a market would be that of Bertrand competition, i.e., at least two firms announce a price equal to the marginal cost.

In our model, firms cannot credibly commit to a menu, i.e., a price for each type. This is because consumers do not learn their type even after the service has been provided. A firm could deceive a consumer by claiming that the consumer was of the cost type associated with the highest price on the menu. Therefore, displaying a menu reduces to displaying a set of prices, which does not qualitatively change the results, as we have shown. If firms could credibly commit to a menu, then there could not be an equilibrium where no firm displays prices. A firm could attract all consumers by displaying a price slightly below the monopoly price for each type.

Our explanation relies on the fact that firms make profits in the absence of displayed prices. This necessitates some assumptions to prevent a hold-up problem, such as in Diamond (1971). If consumers knew their valuation before the first search, then the only equilibrium of a flexible market would be market collapse. Firms would then display a price to prevent market collapse, and hence, the model could not explain a lack of price transparency. Therefore, we assume that consumers learn their valuation after the first search, meaning the service is a search (or inspection) good. We believe this to be a reasonable assumption in the context of credence goods. Only after a consultation with the service provider can the consumer fully understand the personal value of the service. The model is similar to Wolinsky (1986), where firms are horizontally differentiated and consumers learn a firm-specific match value after inspecting the product of a given firm. We abstract from horizontal differentiation and assume that the same value is realized for all firms. A model similar to ours can be found in Preuss (2023). To show that this assumption doesn't drive our results, we have analyzed an alternative model, where consumers know their valuation before the first search, and the first search is free. We haven't chosen this as the baseline model as more equilibria arise in partially committed markets, which complicates the analysis.

Next, we discuss the role of search cost in our model. For a firm to provide the service, it first has to learn the consumer's type. This typically requires the consumer to provide the firm with information, for example, by physically visiting a store or by filling out a questionnaire. We interpret the search cost s as the consumer's ordeal of providing this information. We have assumed that consumers can observe displayed prices at no cost. The rationale is that the cost of learning a displayed price, for instance, through a visit to a website, is negligible compared to the ordeal of providing information. However, an equilibrium where no firm displays a price exists, even if there was a separate search cost for learning displayed prices. If no firm displays a price, consumers do not search for displayed prices. Hence, a deviation to displaying a price would go unobserved and not attract any consumers.

Finally, we consider the policy implications of our model. Our analysis suggests that indiscriminately mandating firms to display a single price for the service may backfire. In markets with high levels of cost dispersion, consumers might benefit from a lack of price transparency. Alternatively, consider a more nuanced policy that mandates that firms display a price for each procedure. A recent example of such a policy would be the U.S. Hospital Price Transparency regulation. In the context of our model, this would have the same effect as mandating firms to display a single price for the service. Competitive pressures would drive the price of each procedure to the expected marginal cost across procedures. Decreasing the price of a cheaper procedure below this expected marginal cost would not attract additional consumers, as they would rationally expect the firm to charge expected marginal cost in any case. This result hinges on firms' ability to deceive consumers about their actual condition and the procedures performed. Note that consumer deception in markets for credence goods is a persistent issue. For example, research by Gottschalk et al. (2020) indicates that 28% of Zurich dentists recommended unnecessary treatments in their study. Therefore, to ensure the effectiveness of price transparency regulations, it is crucial to implement safeguards that prevent deceptive practices by service providers.

Appendix III.A

Proof of Lemma III.1

First, consider $p_h > p_l$. If $p_l \neq p_l^m$, then the firm can deviate $\varepsilon < s$ closer to p_l^m

when facing a low-cost consumer without losing the consumer to a competitor. As this would present a profitable deviation, it must be that $p_l = p_l^m$.

Second, we rule out $p_h \leq p_l$. If $p_h \leq p_l$ and $p_h \neq p_h^m$, then again, the firm can profitably deviate $\varepsilon < s$ closer to p_h^m when facing a high-cost consumer. So assume $p_h = p_h^m$ and note that $p_l^m < p_h^m$. Hence $p_l^m < p_h \leq p_l$ and the firm can profitably deviate to p_l^m when facing a low-cost consumer. Therefore $p_h > p_l$, which concludes the proof.

Proof of Proposition III.1

By Lemma III.1, $p_l = p_l^m$ in any equilibrium. So assume $p_h \neq p_h^m$ and consider a deviation $\varepsilon < s$ closer to p_h^m when facing a high-cost consumer. This deviation could be profitable when facing a high-cost consumer (e.g., if they would believe they are a high-cost type). Hence, the intuitive criterion implies $\mu = 0$, making the deviation indeed profitable. For (p_l^m, p_h^m), there is no profitable deviation, as firms make maximal profits. This concludes the proof.

Proof of Lemma III.2

Upon visit at a committed firm, the consumer learns her valuation but learns nothing about her type. The net benefit of visiting a flexible firm in addition is given by

$$B(v_i) := \lambda \max\{v_i - p_l, v_i - \underline{p}, 0\} + (1-\lambda)\max\{v_i - p_h, v_i - \underline{p}, 0\} - \max\{v_i - \underline{p}, 0\} - s.$$

Note that $B(v_i)$ is increasing in v_i up to $\underline{p}$ and is constant from then on. Hence, for any consumer to additionally search a flexible firm it must be that $B(\underline{p}) \geq 0$. This implies that every consumer who does not make an additional visit at a flexible firm must leave the market without purchase, as their valuation is below $\underline{p}$, such that they wouldn't buy from the committed firm. Hence, if there exists a consumer who would benefit from visiting a flexible firm after already having visited a committed firm, then no consumer would buy at their initial visit at a committed firm. Then, it would be an ex-ante better strategy for consumers to first visit a flexible firm over first visiting a committed firm, as they would sometimes save on search costs. This concludes the proof.

Proof of single crossing property

We show that (III.2) has a single crossing property. Define

$$g(\underline{p}) := \lambda CS(p_l^m) + (1-\lambda)CS(\min\{\underline{p}+s, p_h^m\}) - CS(\underline{p}). \tag{III.7}$$

Note that $g(p_l^m) < 0$ and $g(p_h^m) > 0$. If g is monotonically increasing, then (III.2) has a single crossing property. Taking the first derivative we find

$$g'(\underline{p}) = -(1-\lambda)\left(1 - F(\underline{p}+s)\right) + \left(1 - F(\underline{p})\right) \tag{III.8}$$

for $\underline{p} < p_h^m - s$ and $g'(\underline{p}) = 1 - F(\underline{p})$ for $\underline{p} > p_h^m - s$. Since $1 - F(\underline{p}) \geq 0$ and $\left(1 - F(\underline{p})\right) \geq \left(1 - F(\underline{p}+s)\right)$, $g'(\underline{p}) \geq 0$ for all $\underline{p}$. This concludes the proof.

Proof of Proposition III.6

We begin with the proof of Part i). We prove this part of the proposition by showing that shared monopoly profits and deviation profits have a single crossing property in c_h when s is small. Shared monopoly profits are given by

$$\Pi_{\text{shared}} := \frac{1}{n}\left(\lambda \Pi_l^m + (1-\lambda)\Pi_h^m\right).$$

For every $\varepsilon > 0$ there exists a sufficiently small s such that for $c_h \geq \varepsilon$, deviation profits are at most ε away from

$$(1 - F(p_l^m))(p_l^m - c_e).$$

Hence, deviations profits are approximately linear in c_h, decreasing and 0 for $c_h \geq \frac{p_l^m - \lambda c_l}{1-\lambda}$. Shared monopoly profits are decreasing in c_h to a lower bound of $\frac{1}{n}\lambda \Pi_l^m > 0$. Finally, we show that shared monopoly profits are convex in c_h, which then implies that they intersect the approximation of deviation profits exactly once. Note that by definition of $p^m(c)$,

$$f\left(p^m(c)\right)\left(p^m(c) - c\right) = \left(1 - F(p^m(c)\right).$$

Substituting this expression in $\partial \Pi_{\text{shared}}/\partial c_h$ gives

$$\partial \Pi_{\text{shared}}/\partial c_h = -\frac{1}{n}(1-\lambda)(1 - F(p^m(c_h))).$$

The second derivative is then $\frac{1-\lambda}{n} f(p^m(c_h))\frac{\partial p^m(c_h)}{\partial c_h} \geq 0$. This concludes the proof of Proposition III.6 i).

Next, we prove Part ii). As pointed out before, for every $\varepsilon > 0$ there exists a sufficiently small s such that for $c_h \geq \varepsilon$, deviation profits are at most ε away from $(1 - F(p_l^m))(p_l^m - c_e)$. Then $\Delta\Pi$ is approximately linear in λ with the exception of $\lambda < \varepsilon$ and intersects the horizontal axis at most once. This concludes the proof.

Proof of Proposition III.7

First, we consider the effect of a mean preserving spread of the cost distribution on ΔCS. $CS(c_e)$ is unaffected by a mean preserving spread. If $CS(p^m(c))$ is convex in c, then $\lambda CS_l^m + (1-\lambda)CS_h^m$ increases with a mean preserving spread. The argument is the same as in the case of risk preferences and second-order stochastic dominance.[15] This implies that ΔCS must increase. Chen and Schwartz (2015) (Conditions A1a and A1b) show that if monopoly consumer surplus is convex in cost, then total surplus is convex in cost as well. Hence the same argument goes through for ΔTS. This concludes the proof.

Proof of Proposition III.8

First, we show that (III.5) collapses to $\sigma \geq \sqrt{3}$ when valuations are uniformly distributed on $[0, \bar{v}]$. For uniformly distributed valuations, $CS(p) = \frac{1}{2\bar{v}}(\bar{v} - p)^2$ and $p^m(c) = \frac{\bar{v}+c}{2}$. Substituting in (III.5) gives

$$\begin{aligned}
\lambda \frac{1}{2\bar{v}} \left(\bar{v} - p_l^m\right)^2 + (1-\lambda)\frac{1}{2\bar{v}}\left(\bar{v} - p_h^m\right)^2 &\geq \frac{1}{2\bar{v}}\left(\bar{v} - c_e\right)^2 \\
\lambda \frac{1}{4}\left(\bar{v} - c_l\right)^2 + (1-\lambda)\frac{1}{4}\left(\bar{v} - c_h\right)^2 &\geq \left(\bar{v} - c_e\right)^2 \\
\mathbb{E}\left[(\bar{v} - c)^2\right] &\geq 4\mathbb{E}[\bar{v} - c]^2 \\
\text{Var}\left[\bar{v} - c\right] &\geq 3\mathbb{E}[\bar{v} - c]^2 \\
\frac{\text{SD}\left[c\right]}{\bar{v} - c_e} &\geq \sqrt{3}.
\end{aligned}$$

Next, we show that (III.6) collapses to $\sigma \geq \frac{1}{\sqrt{3}}$. Let $\lambda_l := \lambda$ and $\lambda_h := 1 - \lambda$.

[15] See for instance Proposition 6.D.2. in Mas-Colell et al. (1995).

$$\sum_{\tau\in\{l,h\}} \lambda_\tau \left(\frac{1}{2\bar{v}}\left(\bar{v}-p_\tau^m\right)^2 + \left(1-\frac{p_\tau^m}{\bar{v}}\right)\left(p_\tau^m - c_\tau\right)\right) \geq \frac{1}{2\bar{v}}\left(\bar{v}-c_e\right)^2$$

$$\sum_{\tau\in\{l,h\}} \lambda_\tau \left(\frac{1}{8\bar{v}}\left(\bar{v}-c_\tau\right)^2 + \frac{1}{4\bar{v}}\left(\bar{v}-c_\tau\right)^2\right) \geq \frac{1}{2\bar{v}}\left(\bar{v}-c_e\right)^2$$

$$\frac{3}{4}\sum_{\tau\in\{l,h\}} \lambda_\tau(\bar{v}-c_\tau)^2 \geq \left(\bar{v}-c_e\right)^2$$

$$\mathbb{E}\left[(\bar{v}-c)^2\right] \geq \frac{4}{3}\mathbb{E}[\bar{v}-c]^2$$

$$\mathrm{Var}\left[\bar{v}-c\right] \geq \frac{1}{3}\mathbb{E}[\bar{v}-c]^2$$

$$\frac{\mathrm{SD}\left[c\right]}{\bar{v}-c_e} \geq \frac{1}{\sqrt{3}}.$$

This concludes the proof.

Appendix III.B

In this section, we consider the case of more than two cost types. To save on notation, we equate a consumer's cost type τ with the firm's cost of serving this consumer c_τ. Hence, a consumer's cost type is an element of $\mathbb{R}_+$. Cost types, or simply costs, are distributed with c.d.f. G, such that $G(c)$ is the fraction of consumers with an associated cost of weakly less than c. Let $C \subseteq \mathbb{R}_+$ denote the support of G and assume that C is closed. We define $\underline{c} = \min C$. Note that this notation encompasses the case of finitely many cost types as well.

Committed Market

The equilibrium, where all consumers visit firms displaying the lowest displayed price $\underline{p}$, still exists. Expected costs are now given by $c_e := \int_0^\infty c dG(c)$.

Flexible Market

Let $p : C \to \mathbb{R}_+$ denote the pure and symmetric pricing strategy of flexible firms. Let $p^m(c) := \arg\max_p \left(1 - F\left(p\right)\right)\left(p - c\right)$ denote the monopoly price for cost type c.

Note that a firm can deviate to any price in $\{p(c) : c \in C\}$ without losing the consumer to a competing firm. So assume a firm deviates to a price that is not in

$\{p(c) : c \in C\}$. Let $C' := \{c \in C : p(c) \neq p^m(c)\}$. Under passive beliefs and the intuitive criterion, the strongest possible punishment is that consumers believe that the next firm will charge $\arg\inf_{c \in C'} p(c)$. First, assume $\arg\min_{c \in C'} p(c)$ exists and denote it by c'. Then a firm can deviate by up to s away when facing a consumer of type c', without losing the consumer to a competing firm. Since $p(c') \neq p^m(c')$, the firm has an incentive to do so. Second, assume $\arg\min_{c \in C'} p(c)$ doesn't exists. Then there must exist a $c' \in C'$, such that $p(c') < \arg\inf_{c \in C'} p(c) + s$. In this case, a firm can deviate by up to $\arg\inf_{c \in C'} p(c) + s - p(c')$ away when facing a consumer of type c', without losing the consumer to a competing firm. Since $p(c') \neq p^m(c')$, the firm has an incentive to do so. This shows that the only pricing strategy for which firms have no incentive to deviate is $p(c) \equiv p^m(c)$.

Partially Committed Market

Clearly, the pessimistic equilibrium still exists. Furthermore, Lemma III.2 holds true without caveats. Hence, the only other equilibrium candidate is the optimistic equilibrium where all consumers initially visit flexible firms. The arguments from Section III.3.1 apply to show that in this equilibrium the pricing strategy of flexible firms must be

$$p(c) = \begin{cases} \underline{p} + s & \text{if } p^m(c) \geq \underline{p} + s \geq c \\ p^m(c) & \text{otherwise} \end{cases}.$$

The condition for consumers to indeed prefer visiting a flexible firm first over visiting a committed firm first is given by

$$\int_0^{c^*(\underline{p})} CS(p^m(c))dG(c) + \int_{c^*(\underline{p})}^{\infty} CS(\underline{p} + s)dG(c) \geq CS(\underline{p}), \tag{III.9}$$

where $c^*(\underline{p}) = p^{m-1}(\underline{p} + s)$. If the left-hand side is differentiable w.r.t. $\underline{p}$ at $\underline{p}$, then the derivative is given by $-(1 - F(\underline{p} + s))(1 - G(c^*(\underline{p})))$, hence is clearly larger than the derivative of the right-hand side, which is $-(1 - F(\underline{p}))$. Since the left-hand side is monotonically decreasing without jumps, this proves the single crossing property of III.9. The bounds of the threshold t w.r.t. s are given by $\underline{t} = p^m(\underline{c})$ and $\bar{t}$ solves

$$\int_0^{\infty} CS(p^m(c))dG(c) = CS(\bar{t}).$$

Appendix III.C

In this section, we consider an alternative model, where consumers know their valuation ex-ante and the first search is free. We show that the equilibria of Section III.3.2 persist under this alternative assumption. Note that after the first search, both models are identical. Hence, the only difference lies in the search strategy of consumers before the first search. Under the model described in Section III.2, consumers are ex-ante identical and base their initial search on the expected surplus of visiting a given firm. In order to participate in a market and make an initial search, s has to be sufficiently small. Under the alternative model described in this section, consumers differ even before the first search and might choose different search strategies based on their valuation. However, as the first search is free, they might participate in the market for any s.

Committed Market

The equilibrium, where all consumers visit firms displaying the lowest displayed price $\underline{p}$, still exists. However, in contrast to Section III.3.1, it exists for any s.

Flexible Market

It is easy to see that the equilibrium described in Section III.3.1 still exists. Firms charge the respective monopoly price to each cost type. Consumers make an initial search at a random firm, buy if their valuation is above the monopoly price and leave the market without purchase otherwise. Note that even though consumers with valuations below p_l^m do not expect to make a purchase, they visit a firm anyway, as the first search is free. In contrast to Section III.3.1, this equilibrium exists for any s.

Partially Committed Market

The pessimistic equilibrium still exists and, in contrast to Section III.3.1, it exists for any s.

For the optimistic equilibrium, the analysis departs from Section III.3.1. Consider the case where consumers expect flexible firms to sometimes make better offers than $\underline{p}$. Assume $p_l < \underline{p} < p_h$, such that low-cost consumers would be better off visiting flexible firms and high-cost consumers would be better off visiting committed firms.

We find that a consumer has only two relevant search strategies. One strategy is to initially visit a flexible firm and potentially make another visit at a committed firm after receiving an offer of p_h. The consumer's expected utility of this strategy is

$$u_{\text{flex}}(v_i) := \lambda \max\{v_i - p_l, 0\} + (1 - \lambda) \max\{v_i - p_h, v_i - (\underline{p} + s), 0\}.$$

Alternatively, the consumer could initially visit a committed firm. If, after the initial visit, it was optimal to go on to a flexible firm, it would be even better to visit the flexible firm first. Hence, the only other potentially optimal strategy is to initially visit a committed firm and never search on. The consumer's expected utility of this strategy is

$$u_{\text{com}}(v_i) := \max\{v_i - \underline{p}, 0\}.$$

Let $\Delta u(v) := u_{\text{flex}}(v_i) - u_{\text{com}}(v_i)$ denote the net benefit of initially visiting a flexible firm over only visiting a committed firm. Figure III.3 shows $\Delta u(v)$ for two values of λ. Up to p_l, $\Delta u(v)$ is constant at 0, since consumers with valuations below p_l don't expect

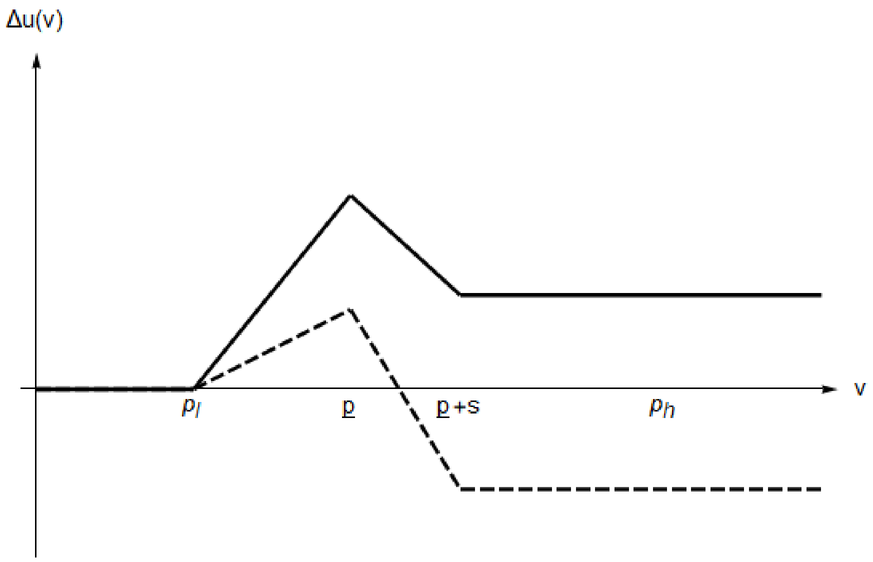


Figure III.3: Net utility of visiting a flexible firm depending on v.

an acceptable offer of either firm. Then $\Delta u(v)$ increases at a slope of λ because at these valuations consumers would only receive an acceptable offer from flexible firms and if they are of the low-cost type. From $\underline{p}$ to $\min\{p_h, \underline{p} + s\}$, $\Delta u(v)$ decreases at a slope of $\lambda - 1$. These consumers would always buy from a committed firm, but only from a flexible firm if they are a low-cost type. Finally, from $\min\{p_h, \underline{p} + s\}$ onward,

$\Delta u(v)$ is constant as these consumers would buy at any of the three prices, and their decision only depends on which firm offers the lower price in expectation. We can distinguish two cases.

In the first case, $\Delta u(v)$ intersects the horizontal axis at

$$\tilde{v}(p_l) := \frac{\underline{p} - \lambda p_l}{1-\lambda},$$

such that consumers with valuations above $\tilde{v}(p_l)$ initially visit a committed firm, and furthermore $\tilde{v}(p_l) < \bar{v}$, such that consumers with such valuations actually exist. Consumers with valuations below $\tilde{v}(p_l)$ are weakly better off visiting a flexible firm. This is true when

$$\tilde{v}(p_l) < \min\{p_h, \underline{p} + s, \bar{v}\} \tag{III.10}$$

and is depicted by the dashed line.

In the second case, (III.10) is violated such that either $\Delta u(v)$ does not intersect the horizontal axis or only intersects it after $\bar{v}$. Either way, all consumers are weakly better off visiting a flexible firm. This is depicted by the solid line. As we want to reproduce the results of Section III.3.1, we will focus on the second case. Note however, that under the alternative model, more equilibrium can arise in a partially committed market, compared to the baseline model.

Assume that (III.10) is violated and consider the incentives of flexible firms to deviate from (p_l, p_h). Since a flexible firm is visited by consumers of all valuations, incentives are similar to a flexible market. For low-cost consumers, the search frictions make marginal deviations towards the monopoly prices p_l^m profitable and hence in equilibrium $p_l = p_l^m$. The price for high-cost consumers is also pushed up towards p_h^m, but it is bounded by $\underline{p} + s$, as otherwise consumers would buy from a committed firm. We call such an equilibrium a *pooling optimistic equilibrium* because consumers rationally expect flexible firms to make good offers, and all consumers take the same initial action, independent of their valuation. Solving (III.10) for $\underline{p}$ provides a threshold t for the existence of this equilibrium.

Proposition III.9 (Pooling Optimistic Equilibrium)**.** In partially committed markets with

$$\underline{p} \geq t := \min\{p_l^m + \frac{1-\lambda}{\lambda} s, \lambda p_l^m + (1-\lambda)p_h^m, \lambda p_l^m + (1-\lambda)\bar{v}\}$$

there exists an equilibrium where flexible firms choose $p_l = p_l^m$ and

$$p_h = \begin{cases} \underline{p} + s & c_h \leq \underline{p} + s \leq p_h^m \\ p_h^m & \text{otherwise} \end{cases}$$

and all consumers initially visit flexible firms. In partially committed markets with $\underline{p} < t$, this equilibrium does not exist.

The pooling optimistic equilibrium resembles the optimistic equilibrium of Section III.3.1, with the difference that the threshold t has a closed-form solution even without specifying F. The results on t partially carry over to this setting. As in Section III.3.1, t is increasing in s and $\underline{t} = p_l^m$. Furthermore, $t \to p_l^m$ as $\lambda \to 1$ and $t \to p_h^m$ as $\lambda \to 0$. What differs is that now $\bar{t} = \min\{\lambda p_l^m + (1-\lambda)p_h^m, \lambda p_l^m + (1-\lambda)\bar{v}\}$.

Commitment Stage

All results of Section III.3.2 go through, with the only difference that t might be different whenever s is not small, and hence, the monopolistic equilibrium might exist in a different region of the parameter space.

References

Anderson, S. P., & Renault, R. (2006). Advertising content. *American Economic Review*, *96*(1), 93–113.

Benabou, R., & Gertner, R. (1993). Search with learning from prices: Does increased inflationary uncertainty lead to higher markups. *The Review of Economic Studies*, *60*(1), 69.

Bulow, J. I., & Pfleiderer, P. (1983). A note on the effect of cost changes on prices. *Journal of Political Economy*, *91*(1), 182–185.

Caplin, A., & Nalebuff, B. (1991). Aggregation and imperfect competition: On the existence of equilibrium. *Econometrica*, *59*(1), 25.

Chen, Y., Li, J., & Schwartz, M. (2021). Competitive differential pricing. *The RAND Journal of Economics*, *52*(1), 100–124.

Chen, Y., & Schwartz, M. (2015). Differential pricing when costs differ: A welfare analysis. *The RAND Journal of Economics*, *46*(2), 442–460.

Cho, I.-K., & Kreps, D. M. (1987). Signaling games and stable equilibria. *The Quarterly Journal of Economics*, *102*(2), 179.

Dana, J. D. (1994). Learning in an equilibrium search model. *International Economic Review*, *35*(3), 745.

Darby, M. R., & Karni, E. (1973). Free competition and the optimal amount of fraud. *The Journal of Law & Economics*, *16*(1), 67–88.

Diamond, P. A. (1971). A model of price adjustment. *Journal of Economic Theory*, *3*(2), 156–168.

Dulleck, U., & Kerschbamer, R. (2006). On doctors, mechanics, and computer specialists: The economics of credence goods. *Journal of Economic Literature*, *44*(1), 5–42.

Fong, Y.-f. (2005). When do experts cheat and whom do they target? *The RAND Journal of Economics*, *36*(1), 113–130.

Glazer, J., & McGuire, T. G. (1996). *Price contracts and referrals in markets for services* (Working Paper No. No. 10/96). The Israel Institute of Business Research.

Gottschalk, F., Mimra, W., & Waibel, C. (2020). Health Services as Credence Goods: A Field Experiment. *The Economic Journal*, *130*(629), 1346–1383.

Janssen, M. C., & Non, M. C. (2008). Advertising and consumer search in a duopoly model. *International Journal of Industrial Organization*, *26*(1), 354–371.

Janssen, M. C., Parakhonyak, A., & Parakhonyak, A. (2017). Non-reservation price equilibria and consumer search. *Journal of Economic Theory*, *172*, 120–162.

Janssen, M. C., Pichler, P., & Weidenholzer, S. (2011). Oligopolistic markets with sequential search and production cost uncertainty. *The RAND Journal of Economics*, *42*(3), 444–470.

Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic theory*. Oxford University Press.

Myatt, D. P., & Ronayne, D. (2019). *A theory of stable price dispersion* (Economics Series Working Papers No. No. 873). University of Oxford, Department of Economics.

Obradovits, M. (2014). Austrian-style gasoline price regulation: How it may backfire. *International Journal of Industrial Organization*, *32*, 33–45.

Pitchik, C., & Schotter, A. (1987). Honesty in a model of strategic information transmission. *American Economic Review*, *77*(5), 1032–1036.

Preuss, M. (2023). Search, learning, and tracking. *The RAND Journal of Economics*, *54*(1), 54–82.

Robert, J., & Stahl, D. O. (1993). Informative price advertising in a sequential search model. *Econometrica*, *61*(3), 657.

Sülzle, K., & Wambach, A. (2005). Insurance in a market for credence goods. *Journal of Risk and Insurance*, *72*(1), 159–176.

Tappata, M. (2009). Rockets and feathers: Understanding asymmetric pricing. *The RAND Journal of Economics*, *40*(4), 673–687.

Taylor, C. R. (1995). The economics of breakdowns, checkups, and cures. *Journal of Political Economy*, *103*(1), 53–74.

Varian, H. R. (1980). A Model of Sales. *American Economic Review*, *70*(4), 651–659.

Wolinsky, A. (1986). True monopolistic competition as a result of imperfect information. *The Quarterly Journal of Economics*, *101*(3), 493.

Wolinsky, A. (1993). Competition in a market for informed experts' services. *The RAND Journal of Economics*, *24*(3), 380.

Abstract

While the sciences strive to uncover descriptive truths about nature, it is typically the role of ethics, as a branch of philosophy, to investigate normative truths. However, in economics, descriptive and normative truths are often inextricably linked. For instance, supporting an economic policy requires not only a descriptive theory to predict the effects of the policy but also a normative theory to evaluate the desirability of these effects. In this vein, my dissertation combines inquiries into ethical questions with the application of the resulting insights to descriptive models. The normative issues I consider come in the form of dilemmas, where it is impossible for a decision-maker to simultaneously abide by competing criteria. This dissertation comprises three separate chapters, each revolving around a conflict between rationality and fairness concerns.

The first chapter, in collaboration with Karl Schlag, addresses the problem of aggregating individual preferences into a single societal preference. We overcome Arrow's impossibility theorem by slightly weakening an implicit rationality requirement, namely menu independence. This adjustment leads to a unique rule that satisfies strong ethical criteria. We apply this rule to assess price discrimination, optimal consumption-savings plans, and fair wage determination.

The second chapter considers the dilemma faced by an arbitrator who cannot both have rational risk preferences and a desire for egalitarian outcomes. I argue in favor of a weaker fairness requirement and show that this is compatible with a rational solution. I then implement this solution in a subgame-perfect equilibrium.

The third chapter models a credence good market, where firms know more about consumer needs than consumers themselves. This informational asymmetry opens the door for unethical practices, such as deceiving consumers about their condition. The

chapter explores the role of price transparency in such markets and finds that under certain conditions, consumers might benefit from opaque pricing.

Zusammenfassung

Während die Naturwissenschaften danach streben, deskriptive Wahrheiten über die Natur zu enthüllen, ist es typischerweise die Aufgabe der Ethik, als Zweig der Philosophie, normative Wahrheiten zu untersuchen. In der Ökonomie sind jedoch deskriptive und normative Wahrheiten oft untrennbar miteinander verbunden. Zum Beispiel erfordert die Unterstützung einer bestimmten Wirtschaftspolitik nicht nur eine deskriptive Theorie, um die Auswirkungen der Politik vorherzusagen, sondern auch eine normative Theorie, um die Wünschbarkeit dieser Effekte zu bewerten. In diesem Sinne kombiniert meine Dissertation Untersuchungen zu ethischen Fragen mit der Anwendung der daraus gewonnenen Erkenntnisse auf deskriptive Modelle. Die normativen Probleme, die ich betrachte, nehmen die Form von Dilemmata an, bei denen es für einen Entscheidungsträger unmöglich ist, gleichzeitig konkurrierenden Kriterien gerecht zu werden. Diese Dissertation besteht aus drei separaten Kapiteln, in denen jeweils der Konflikt zwischen Rationalität und Gerechtigkeitsbedenken ein zentrales Thema darstellt.

Das erste Kapitel, in Zusammenarbeit mit Karl Schlag, befasst sich mit dem Problem, individuelle Präferenzen in eine einzige gesellschaftliche Präferenz zu aggregieren. Wir überwinden Arrows Unmöglichkeitstheorem, indem wir eine implizite Rationalitätsanforderung, nämlich die Menüunabhängigkeit, leicht abschwächen. Diese Anpassung führt zu einer Regel, die starke ethische Kriterien erfüllt. Wir wenden diese Regel an, um Preisdiskriminierung, optimale Konsum-Spar-Pläne und faire Lohnbestimmung zu analysieren.

Das zweite Kapitel untersucht das Dilemma eines Schlichters, der nicht gleichzeitig rationale Risikopräferenzen und den Wunsch nach egalitären Ergebnissen haben kann. Ich trete für eine schwächere Gerechtigkeitsanforderung ein und zeige, dass diese mit

einer rationalen Lösung vereinbar ist. Anschließend implementiere ich diese Lösung in einem teilspielperfekten Gleichgewicht.

Das dritte Kapitel modelliert einen Markt für Vertrauensgüter, in dem Unternehmen mehr über die Bedürfnisse der Verbraucher wissen als die Verbraucher selbst. Diese Informationsasymmetrie öffnet die Tür für unethische Praktiken, wie das Täuschen von Verbrauchern über ihren wahren Zustand. Das Kapitel erforscht die Rolle der Preistransparenz auf solchen Märkten und zeigt, dass Verbraucher von undurchsichtiger Preisgestaltung profitieren könnten.