Essays on Individual and Collective Decision Making

Zur Erlangung des akademischen Grades
eines **Doktors der Wirtschaftswissenschaften**

(Dr. rer. pol.)

von der KIT-Fakultät für Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

Dissertation

von

M.Sc. Claudio Kretz

Tag der mündlichen Prüfung:	06. Juni 2024
Referent:	Prof. Dr. Clemens Puppe
Korreferent:	Prof. Dr. Klaus Nehring, University of California, Davis
Karlsruhe	(2024)

To Ella & Viki

In loving memory of Fini

Contents

List of Figures

List of Tables

Acknowledgements

This dissertation wouldn't have come together without the help and support of some wonderful people, and I want to take a moment to thank them.

First and foremost, I am profoundly grateful to my advisor, Clemens Puppe, for his unwavering support, insightful feedback, and consistent encouragement. You have been a constant source of inspiration, guidance and patience throughout the entire process, both academically and personally. To some extent, completing this thesis feels like coming full circle. After all, it was your lectures that helped convince me that economics is an intriguing intellectual endeavor to undertake when I was an undergraduate student at KIT. I would also like to thank the members of my dissertation committee, Klaus Nehring, Martin Ruckes and Stefan Nickel, for their thoughtful comments, challenging questions, and constructive criticism. Your collective expertise has deepened my understanding of the topics covered in this dissertation, its strengths and soft spots.

To my (former) colleagues and friends at KIT thank you for creating a stimulating and cooperative work and research environment. Special shoutout to Matthias Ammann, Jessica Jachimowicz, Michael Müller, Christoph Neumann, Danisz Okulicz, Paula Peters and Jana Rollmann for making this journey a lot more fun and much less solitary. I owe special thanks to Philipp Reiss for his continued presence during my seminar talks at KIT. Your feedback and constructive criticism has really helped me on this journey. To my fellow students at the University of Mannheim, Tomasz Olma and Daniil Larionov, thank you for keeping my spirits up when my excitement seemed to slip amidst the deluge of problems sets and coursework during our first year of PhD studies. To Andrey Alexandrov and Fabian Greimel I am deeply grateful for the countless insightful and enjoyable discussions, and the convivial moments we shared during our second year of Phd studies at Yale. This was real fun.

Over the years many people have contributed to the different parts of this Thesis through discussions or comments. Although I wish to thank them all, the following list is likely to be incomplete. I thank Pierre Bardier, Mira Frick, Hans Gersbach, Ryota Iijima, Kevin Leportier, Yusufcan Masatlioglu, Jawwad Noor, Luca Pegorari, Hannes Rau, Hendrik Rommeswinkel, Arkadii Slinko and Thomas Tröger.

I also want to extend a heartfelt thanks to Karl-Martin Ehrhart and John Geanakoplos. You both played a big role in sparking my interest in economics and pushing me to think critically about the world around me. Your passion for the subject was infectious, and I am deeply grateful for the inspiration you provided. Your lectures have been a highlight of my studies.

I am indebted to my family, especially to my parents Sabine and Jörg, and to my grandparents Fini and Martin, for their love and patience. You have really made all this possible. To Benedikt Nell, thank you for always being there to listen and cheer me up over a drink or two. And to you and Moritz Deinhard, thank you for many of the most enjoyable nights of my life. You have made Karlsruhe a truly great time and place to live in. To Danny Schulz, Florian Müller and Raphael Mack, thank you for reminding me of what it really is all about and for having always been there to give me a nudge. This really feels like being part of a pack of wolves.

And last but very much not least to Viki, thank you for your endless patience and support. You have been part of this journey from the very start and you have never given up on me. You have always been there to pick me up when things got tough. This means a lot. I am so lucky to have you by my side.

To everyone who has been part of this journey, thank you from the bottom of my heart.

Part I

Introduction

Introduction to This Thesis

> "So economics is a choice between alternatives all the time."[1]
>
> *Paul Samuelson*

One of the organizing principles of modern economic theory is to understand human behavior in terms of choice. That is, it is interpreted against the background of what alternatives are available. For example, by writing up this thesis, the author chooses to forego other ways of passing time; after all, he could read a book, study to get a fishing license or apply for a private-sector job instead. In this sense, economics holds, to behave is to choose. This paradigm is epitomized by the opportunity cost principle that is taught to every undergraduate student in economics. According to it, every potential choice comes at that cost which is incurred by foregoing the best available alternative.[2] As no alternative is therefore truly costless, choice involves decision making.

From a methodological point of view, economics is built on the idea that an economy – or any single part of it, such as the market for labor, capital, consumer goods etc. – is best understood as a system of decision-making individuals that interact (through markets or otherwise), rather than a holistic entity governed by its own invariant laws. Thus, the behavior and interaction of aggregate economic variables (e.g., wages, interest rates, commodity prices) is derived from assumptions on individual behavior.[3] If relying on econometric estimates based on historic data for these relationships instead, economic analysis is inadequate to study changes in the economic environment (e.g., an economic policy) because it does not account for how individuals change their decisions in response to it. Popularized by Lucas (1976), this critique gained momentous influence within economics in the seventies and was accordingly named after him. It has lent support to the methodological stance that economic phenomena at the macro level must be explained

[1] Source: `https://www.pbs.org/newshour/economy/the-basics-of-economics-with-p` accessed on 16 April, 2024.

[2] Note that this cost need not be pecuniary. However, in perfect and frictionless markets, opportunity costs equal prices.

[3] By 'individual' we do not necessarily mean that of an individuum. Instead, we use the term to refer to a single decision-making entity. What is (and what is not) a single such entity, may depend on the context. For example, form a market perspective, a firm may be treated as one decision maker. At the same time, the decisions made by the firm can result from incorporating those of several decision makers within the firm.

in terms of behavior at the micro level. The search for such 'microfoundations' (see also Janssen, 2016) puts individual decision making center-stage. More generally speaking, if we grant that what matters to a group (or society) as a whole is the interests of its individual members, then collective decisions should be based in a systematic and transparent way on individual ones.[4]

The present thesis compiles some novel contributions to the theoretical research on individual and collective decision making. To facilitate putting them into the context of their wider research agenda, we start by giving a brief introduction to choice theory first. Every attempt is made to keep it non-technical. However, mathematical notation is used where it helps to understand a concept or avoids ambiguity. As we feel is common practice within economics, no attempt is being made at drawing a clear distinction between 'choice' and 'decision making'. Indeed, we will treat them as synonyms. Note that we will use 'choice' to refer to both the situation of having to choose (i.e. a decision problem) and the alternative being chosen. Our presentation will focus on decisions under certainty and abstract from strategic aspects, that is, from situations where individual decisions may depend on the actions of others.

'Rational' Choice Theory

The principal object of investigation in choice theory is that of a *choice function*. Suppose X is some universe of alternatives. Every non-empty set of alternatives $A \subseteq X$ is called a *decision problem*. For instance, a restaurant goer is faced with the problem of deciding between dishes on the menu. (It is due to this analogy that decision problems are also referred to as *menus*.) Or consider, somewhat more abstractly speaking, a consumer who, endowed with a certain income and at given prices, faces a budget set of possible consumption choices (over a certain period of time). A choice function c is an abstract representation of how a decision maker chooses from all conceivable decision problems. That is, c specifies, for every decision problem A, which alternative(s) $c(A) \subseteq A$ are chosen from it. In general, this allows for the case that the decision maker is unable to make a choice ($c(A) = \emptyset$) or her choice is indeterminate ($c(A)$ contains more than one alternative). However, we assume that choice sets $c(A)$ are non-empty at least for decision problems A with finitely many alternatives.[5] If choice sets contain only a single element, we say that they are singleton-valued.[6]

A particularly simple and prominent class of choice functions arises when the decision maker has stable preferences $\succsim$ over all alternatives in X and chooses the best or maximal

[4] This approach to collective decision making is also referred to as 'methodological individualism' (see Janssen, 2016, and the references given there).
[5] If X is finite, this implies that all $c(A)$ are non-empty. In this case, the assumption of non-empty choice sets is often built into the definition of a choice function (cf. Part II).
[6] It is not uncommon to consider this to be a defining property of a choice function (we do so in Part II) and refer to the more general concept as a *choice correspondence*.

ones available to her.[7] In this case, the preference relation $\succsim$ is said to *rationalize* the choice function c. This concept is so fundamental that preferences are often taken as primitives of an economic model. Indeed, we also do so in various parts of this thesis. If preferences are complete and transitive, that is, allow to rank (or order) alternatives from best to worst (allowing for ties or indifferences), they are sometimes referred to as 'rational', corresponding choice behavior as 'rational choice'.[8] 'Rational' preferences are behaviorally equivalent to the maximization of a utility function.[9]

This use of the label 'rational' is somewhat unfortunate for it almost certainly paints too narrow a picture of rationality. For example, consider an individual who likes bigger cars better than smaller ones. However, if two cars do not differ by much in terms of size, she is indifferent. As many small size differences can add up to a large one, there may exist a sequence of cars such that the decision maker is always indifferent between subsequent ones yet strictly prefers the first car to the last. This makes for intransitive preferences (to be precise, intransitive indifferences). At the same time, there seems to be nothing evidently irrational about them and the choices they induce.[10] So, more generally, we might wonder what types of preference can rationalize choice behavior at all, given our assumption that some alternative must be chosen from every finite decision problem?[11] As it turns out, the answer is that there must not be any cycles of strict preference. In more precise terms, that is: if and only if a preference relation is acyclic, it rationalizes some choice function. Note that this fact is borne out by our example above. Indeed, if some car x is sizably bigger than another car y and car y is sizably bigger than yet another car z, then car x must be sizably bigger than car z. So *strict* preferences are in fact transitive in our example. As matter of fact, such preferences (with a transitive strict part) are not only acyclic but referred to as quasi-transitive.[12]

At a closer look, the sense of rationality that is transported by preference-induced choice is that an *invariable* relation between alternatives explains choice behavior *across all* decision problems. In turn, this imparts a stringent notion of *consistency* across decision

[7] An alternative x is best in A if it is preferred to all other alternatives in A ($\forall y \in A : x \succsim y$); it is maximal in A if there does not exist some other alternative that is *strictly* preferred to it ($\nexists y \in A : y \succ x$). If $\succsim$ is complete, the two notions coincide.

[8] Preferences $\succsim$ are said to be complete if they allow to compare every pair of alternatives. They are said to be transitive if, for all alternatives $x, y, z \in X$, $x \succsim y$ and $y \succsim z$ imply $x \succsim z$.

[9] Technically speaking, if X is uncountable, preferences additionally need to satisfy a continuity assumption in order to be represented by some utility function. Intuitively speaking, this means that there must not be a sudden change of preference when altering alternatives slightly.

[10] Admittedly, one might point out to her: "You like bigger cars better than smaller ones. So, even if the difference in size is marginal, you should like the marginally bigger car better. Hence the indifferences you display are irrational." She might reply: "That's true. However, while I can tell straight away when two cars are only marginally different in size, I cannot say without considerable further (cognitive) effort which one is the (marginally) bigger one. Since this does not matter much to me anyway, it is perfectly reasonable for me to be indifferent."

[11] Note that, when decision problems are infinite, a best alternative may not exist even when preferences order alternatives from best to worst. For example, an individual preferring more money to less cannot choose a best (real-valued) alternative from $[\$0, \$10)$.

[12] Evidently, quasi-transitive preferences are acyclic. For formal details, see also Part III.

problems. To see this, note that preferences relate pairs of alternatives. Thus, in effect, preference-induced choice requires that all choice is consistent with binary choice (in the sense that the decision maker chooses from decision problem A exactly those alternatives that are chosen in binary comparisons with all other alternatives in A). In general, however, whether one alternative is chosen over another may depend on the menu from which this choice is made. As we explain in Part II, this is natural for decision makers that need to exert self-control since the cost of self-control may depend on what other alternatives are available. In any case, this discussion suggests to study choice behavior in terms of (internal) consistency requirements directly. The probably most well-known one such condition is the Weak Axiom of Revealed Preference (WARP). It requires that if alternative x is chosen from some decision problem in which alternative y is present, then y may be chosen from any other decision problem containing x only if the latter is chosen as well. (Sometimes this condition is also simply referred to as Choice Consistency.) It is well known that a choice function c satisfies WARP if and only if it is rationalized by some complete and transitive preference relation. As was first shown by Amartya Sen (1969), WARP can be decomposed into conditions that require consistency under contraction and expansion respectively. The former, known as (Sen's) condition α demands that if x is chosen from decision problem A, then so it must be from every $B \subseteq A$ that contains it. The latter, known as (Sen's) condition β, requires that if x is chosen from $B \subseteq A$ and y was available at B, then y is chosen from A only if x is. The following example may serve to illustrate the logic behind the two conditions: if economics is among your favorite sciences, then it must be among your favorite social sciences (α). Moreover, if economics is among your favorite social sciences and sociology – another social science – is among your favorite sciences overall, then so must be economics (β).

The equivalence results described above capture 'rational' choice in terms of internal consistency conditions (WARP or, respectively, conditions α and β). If we are willing to grant that this is not the *only* kind of behavior that qualifies as rational, however, where exactly do we draw the line between rational and irrational behavior? In somewhat of a modern positivist take on it, Itzhak Gilboa (2012) suggests to throw the question back to the decision maker herself. Behavior may count as rational as long as the decision maker has no desire to *change* it even after it has been analyzed and laid out for her (including all of its potential flaws and inconsistencies).[13] On the other hand, if the decision maker wishes to revise her decision upon further analysis, it was irrational in the first place. The allure of this approach to rationality is that it allows to broaden the scope of rational analysis while not rendering the term 'irrational' void of any empirical content. For example, many known decision biases would likely count as irrational under this view. Indeed, it is hard to imagine that decision makers would not be embarrassed by decisions that are explained to them to have been biased by anchors, win-loss frames or confounded

[13] Note that our defense of intransitive preference over cars of different sizes above very much is in line with this notion of rationality (cf. Footnote 10).

conditional probabilities. In effect, this notion captures rationality in terms of stability of behavior. When behavior is stable, it can be meaningfully analyzed in terms of the consistency conditions that underlie it and the optimality rationales these correspond to. This idea informs much of the analysis carried out in this Thesis (in particular, Parts II and III).

Contributions of this Thesis

Part II – Planning to Self-Control

As we outlined above, modern economic analysis ultimately rests on representations of individual behavior. For this reason, it seems crucial to develop appropriate models of it. In intertemporal decision making – that is, for decision problems composed of alternatives that differ along a time dimension (for example, dated streams of consumption) – exponential discounted utility continues to be the benchmark model. At the same time, it is irreconcilable with two robust empirical findings: present bias and the magnitude effect. *Present bias* refers to the fact that individuals tend to choose a smaller, sooner reward over a larger, later one when the former is immediate but reverse their choice when both are delayed. This is commonly seen as evidence of preferences that are dynamically inconsistent, that is, change over time.[14] Individuals exhibit the *magnitude effect* if they choose smaller, sooner rewards when payoffs are small but reverse to larger, later ones when the stakes are increased.[15] In particular, the magnitude effect can make present bias disappear at high stakes. Part II gives a theoretical explanation of magnitude effects (in intertemporal choice tasks and more generally) in terms of costly self-control. It provides a behavioral characterization of decision makers who optimally (plan to) bring self-control to bear on choices that are considered sub-optimal at the planning stage. When self-control costs are fixed, the simplest specification consistent with our model, decision makers may optimally forego self-control when little is at stake but exert it when the stakes are large. For example, if decision makers are present biased in the absence of self-control, this may explain why they choose immediate rewards when the stakes are low but choose to wait for larger rewards as the stakes rise. In the context of an infinite-horizon consumption-savings problem such decision makers save sufficiently only when their current wealth level is high enough (because much is at stake) but else over-consume. This creates a poverty trap at the individual level. That is, unless they are endowed with sufficient initial wealth, decision makers continually run down their wealth as time progresses.

[14] This assumes that individuals would choose the immediate reward again when given the chance to reconsider their choice in future (when the delay has passed). However, this delayed choice task is usually not part of the experiment (cf. Halevy, 2015).

[15] Standard exponential discounted utility is inconsistent with present bias. While it is not inconsistent with the magnitude effect per se, the amount of utility curvature needed to accommodate it produces unrealistic predictions (individuals would need to turn down opportunities to save even at astronomical rates of interest).

Part III – Aggregation of Choice Functions

Putting individual choices at the center of economic analysis entails a natural paradigm for how groups or whole societies ought to make decisions. Indeed, if they are to respect their constituting members' interests and values, collective decisions should be based in a systematic and predictable way on individual ones. This idea lies at the heart of modern social choice theory as pioneered by the late Kenneth Arrow. In his path-breaking work on 'Social Choice and Individual Values' (Arrow, 1951/1963), he envisions social choice as a process of forming a collective preference order over alternatives based on individual rankings of these.[16] His famous (im-)possibility result states that if one insists that: (1) whether society prefers one alternative over another can only depend on how individuals rank *these two* alternatives (known as the INDEPENDENCE condition), (2) society cannot revoke its preference for one alternative over another as support for the former increases (POSITIVE RESPONSIVENESS), and (3) all profiles of individual preference orders are admitted (UNIVERSAL DOMAIN); then the aggregation procedure must be either imposed (the social preference over at least one pair of alternatives is pre-determined) or dictatorial (the social ranking is that of one individual).[17] As we discussed above, the assumption of 'rational' choice (that is, choice rationalized by complete and transitive preferences) can be hard to justify in general at the individual level. At the collective level, it seems even more debatable. After all, the methodological stance underlying social choice theory, that collective choices be based on individual ones, makes no claim to the effect that collective behavior resemble that of a single individual. Indeed, when weakening collective consistency requirements slightly to the extent of being rationalizable by some quasi-transitive preference, possibilities emerge; for example, by choosing, for every decision problem, all alternatives that are not Pareto-dominated in it (Sen, 1969). More generally, every aggregation procedure that yields quasi-transitive social preferences and satisfies INDEPENDENCE, UNIVERSAL DOMAIN and ranks one state strictly over another if all individuals do (WEAK PARETO), must be oligarchic (Gibbard, 1969/2014). That is, there exists a (privileged) group of individuals, the *oligarchs*, such that society strictly prefers one social state over the other if all oligarchs do, and otherwise, is indifferent.

[16] Arrow proposed an all-encompassing notion of an alternative as a social state specifying everything that is relevant to society (allocation of goods etc.). More generally speaking, alternatives should account for everything that is relevant to the decision(s) under consideration. For example, a university admissions committee may only need to determine a ranking of applicants.

[17] In fact, Arrow's original formulation of the INDEPENDENCE condition demands that what society chooses from a given menu can only depend on how individuals rank alternatives *within* that menu. Given Arrow's implicit assumption that collective choices are rationalized by some (collective) preference order, this is equivalent to (1) (as all choices can be reduced to binary choices). In an updated version (1963) of the original text (1951), Arrow gives an alternative account of his (im-)possibility result which states that INDEPENDENCE, UNIVERSAL DOMAIN and the condition that society strictly prefers one alternative over another if all individuals do (WEAK PARETO; in particular, requiring that the aggregation procedure is not imposed) imply dictatorship. As it is the stronger result (note that INDEPENDENCE, POSITIVE RESPONSIVENESS and non-imposition imply WEAK PARETO), this version has gained more prominence. However, we present the original version here as we study choice functions that satisfy MONOTONE INDEPENDENCE, a combined version of POSITIVE RESPONSIVENESS and INDEPENDENCE, in Part III.

Building on joint work with Clemens Puppe, Part III presents a first attempt at reformulating the Arrow problem in the general context of choice functions. Here, a stronger but very natural notion of independence would demand that whether some alternative x is collectively chosen from decision problem A can only depend on whether individuals choose x from A. We study the (im-)possibility of positive responsive and independent aggregation when imposing the same consistency requirements on choice at the individual and collective level. Interestingly, even on the domain of choice functions that can be rationalized by some acyclic relation such aggregation is necessarily dictatorial. This is in sharp contrast to the possibilities described above for Arrow's notion of independence. In our framework, oligarchic rules emerge as the only consistent independent procedures for choice functions that satisfy a condition of *path-independence*, which requires choice to be invariant to how decision problems are split up by choosing from a sub-problem first and pitting the chosen against all remaining alternatives.[18] Yet these rules are not the same as those described by Gibbard. Indeed, the latter violate both our strong notion of independence as well as a natural weakening to a condition of 'independence across menus' that allows for collective choice of alternative x from menu A to depend on individual choice behavior for other alternatives $y \in A$ (but not on choice behavior from other menus). When restricting attention to individual and collective choice functions that are rationalized by some preference order, this notion is equivalent to Arrow's independence condition. More possibilities emerge under independence across menu; for example, in the form of approval voting. However, collective choices under approval voting fail to satisfy the contraction property α and cannot be rationalized by any binary relation.

Part IV – Consistent Rights

Arrow's independence axiom is central to his approach to social choice theory, not only from a technical perspective (for obtaining his impossibility result) but also from a methodological point of view. It suggests to study collective decisions in terms of how they may (or may not) be allowed to vary when individual decisions change. This has proven to be of great importance for guaranteeing that individual strategic motives do not interfere with the aggregation procedure; a problem wo do not study in this Thesis. For example, if a procedure for aggregating preferences is not independent, individuals may want to manipulate the social preference over a given pair of alternatives by misrepresenting their preference over *another* pair. Yet, for all its merits, the classical Arrow approach neglects other aspects of collective decision making that seem of no less fundamental importance.

First, in many groups (and most societies) subgroups of individuals (or single individuals) are granted *rights* to determine certain components of the collective decision. That is to say that, if a subgroup agrees on an issue it has a right to, the group as a whole (society) must follow their lead on that issue. Much like the Arrow axioms, this seems innocuous a

[18] Such choice functions have an interesting interpretation in terms of multiple 'selves'.

demand to make. Yet Amartya Sen (1970) has shown that a minimal amount of such rights can conflict with even the most basic welfarist principle. To be precise, he has shown that no procedure of aggregating individual preference orders into an acyclic social preference that satisfies UNIVERSAL DOMAIN and WEAK PARETO can let two individuals be decisive for the social preference over at least one pair of alternatives each. This fact has become known as the (Sen) liberal paradox.

In Part IV, we analyze the problem of rights in the more general framework of judgment aggregation. This emerging field is concerned with the analysis of aggregating sets of individual judgments into a collective one, subject to restrictions on what combination of judgments are consistent. For example, judgments may correspond to strict preference statements between pairs of alternatives and the consistent sets of judgments be given by all strict rankings of alternatives. Or judgments may refer to whether some alternative x is chosen from decision problem A. As this shows, judgment aggregation includes the aggregation of (rational) choice functions as a special case but is not limited to it. We model judgment aggregation on *property spaces* (Nehring and Puppe, 2007, 2010). Here, every alternative (consistent set of judgments) corresponds to a unique combination of properties (judgments). Going beyond impossibility results, we fully characterize when a system of rights to such properties is consistent to the effect that there exists some aggregation rule that grants them. Indeed, we show that consistency is equivalent to the following well known 'Intersection Property' of Nehring and Puppe (2007, 2010): whenever rights are granted to a minimally inconsistent (critical) family of properties, the corresponding rights holding (sub-)groups must intersect to at least one common member. In the Arrowian spirit, we also study when rights are consistent with monotone independent (that is, positively responsive and independent) aggregation. As it turns out, the characterizing condition of non-empty intersection remains in tact when extended to hold for a more general concept of criticality. We derive important impossibility results from the literature as corollaries of our more general characterizations. Moreover, we employ our results to relate properties of consistent rights to classes of underlying property spaces. For example, on spaces for which there are conditional logical entailments between all properties (totally blocked spaces), only trivial rights (all rights holding groups have at least one common member) are consistent with monotone independent aggregation. On the other hand, when all logical entailments are unconditional (median spaces), the consistent rights are exactly the independent rights (no rights to logically dependent properties for disjoint groups). This part of the present Thesis has appeared in the *Journal of Economic Theory* (Kretz, 2021) under the title "Consistent Rights on Property Spaces".[19]

[19] The published article can be accessed at `https://doi.org/10.1016/j.jet.2021.105323`.

Second, traditionally, classical social choice theory is concerned with how collective decisions can be made in ways that are justifiable based on, or fair to, individual views on matters of societal concern, such as how to distribute goods or whom to elect into office. Such questions are a matter of (individual) norms or tastes and, as such, defy an objective framework for what is right or wrong.[20] On the other hand, in many situations, groups may need to judge a measurable quantity (the weight of an object, the sales potential of a new product etc.) or assess the truth content of a verifiable proposition. In these circumstances, all that may matter about the collective decision is that it be *correct*. For example, suppose we are interested in ascertaining whether a certain measure is effective in mitigating climate change. Hypothetically speaking, if we new that everyone in society incorrectly judged the measure to be ineffective, then we should come to the conclusion that it *is* effective, even though this overrules unanimous consent.[21] In similar fashion, when subgroups of individuals have expert knowledge, they may be given greater weight in collective decisions to improve the likelihood of these decisions being correct. In contrast to the classical concern with fairness, here the focus is on the *epistemic* properties of the aggregation procedure; that is, on how likely it is to uncover the truth. This line of analysis goes back to work by the Marquis de Condorcet (1785/2014) on majority decisions over a simple binary issue. He is credited with being the first to formulate the result, now formally known as the *Condorcet Jury Theorem*, that the probability of a correct majority decision increases in the number of voters and approaches one as the number of voters grows infinitely large.

Yet in most real-world scenarios, the issue to be decided upon is analytical in nature in the sense that it is the logical conclusion of a number of premises. For example, society may want to implement a certain measure against climate change if it is feasible (the first premise) and effective (the second premise). In such a case, two particularly simple and intuitive ways to implement the majoritarian idea are to arrive at a collective judgment on the conclusion (1) through direct majority voting on it (*Conclusion Based Procedure*, CBP) or (2) by logically deriving it from majority decisions on the premises (the so called *Premise Based Procedure*, PBP). In analogy to Condorcet's classical version, the literature that studies Jury Theorems in this more general judgment aggregation setting has so far mostly assumed (the correctness of) voters' judgments to be independent of each other. Yet as voters usually have access to common sources of information that raise or lower their conditional competence for correct judgments, this seems unrealistic.

[20] As the saying goes, there is no accounting for tastes. For example, if I prefer yellow buses and you prefer red ones, there is no meaningful sense in which one of us is 'right'.

[21] This assumes that there is a shared interest in identifying correct judgments (making this a purely factual matter). However, when individual stakes in the decision are high, the distinction between taste and fact may not be easy to draw. For example, we can imagine a person who *prefers* not to learn the truth about a matter, possibly because it could call for action she dislikes (in our example: that the measure is enacted).

In Part V we model correlated votes at the premise level by means of binary latent random variables and study how reliable the PBP and CBP are as epistemic procedures. Our main finding is that, for most agendas, the asymptotic performance (that is, as the number of voters tends to infinity) of the CBP crucially depends on the objectively true state of the world (truth scenario). As this state is unknown, the reliability of the CBP is at times highly uncertain. On the other hand, the (asymptotic) properties of the PBP are independent of the agenda and the truth scenario. In the benchmark case of independent voters, the PBP yields the correct conclusion with asymptotic certainty. The CBP performs well in truth scenarios that are 'truth-conducive' to the effect that competent signals on the premises translate into a competent conclusion signal. For example, if the conclusion is the logical conjunction of two premises, this is the case for the first three out of the following four possible truth scenarios: no premise is true, only the first premise is true, only the second premise is true, or both premises are true.[22] In the first three scenarios, the CBP is asymptotically infallible and grows more reliable when increasing the number of voters in finite samples. In the last truth scenario, the exact opposite is the case. This uncertainty in behavior is often disguised, however, when considering *weighted* behavior. For example, when assigning equal prior probability to all four possible states of the world, the CBP is superior in finite samples of moderately competent voters (cf. Bovens and Rabinowicz, 2006; Hartmann et al., 2010). While the picture that emerges for correlated voters is somewhat more complex and the PBP loses its status as asymptotically immaculate procedure, the general intuition, that the CBP is (more) sensitive to the agenda and the truth scenario, remains intact. However, interestingly, the presence of correlation is somewhat of a moderating force on the variance of the CBP's reliability (across truth scenarios). Moreover, there exist agendas for which the CBP is more reliable in all states of the world in finite samples of moderately competent voters when correlation is present (and for which this was not true in the absence of correlation).

[22] Note that, in the last truth scenario, voters judge the conclusion correctly only if judging both premises correctly. Even if they do so with probability $p > 0.5$, they reach the correct conclusion only with probability p^2: Thus, for $p \in (0.5, 1/\sqrt{2})$, voters are competent on both premises but not on the conclusion.

Part II

Planning to Self-Control

1 Introduction

We make plans on a daily basis. Indeed, planning is predominant in thoughts about the future.[23] Recent research in psychology and human decision processes suggests that planning acts as a form of 'pragmatic prospection'. Individuals engage in it "so as to guide actions to bring about desirable outcomes" (Baumeister et al., 2016, p. 3). That is, plans play an active part in steering future own choices. For example, implementation intentions in the form of simple plans have been found to help overcome self-regulatory problems and lead to better self-control (Gollwitzer, 1999; Gollwitzer and Sheeran, 2006; Gollwitzer et al., 2010; Gollwitzer and Oettingen, 2011).

To illustrate, consider an individual contemplating potential after-work activities at lunch-time. Suppose that she may either watch a movie or go for a run. At lunch-time, when she is still energized, she prefers to exercise. At the same time, she anticipates her preference to change come the evening. This creates an obvious incentive to commit to the after-work run (for example, by arranging to meet with a friend). If external commitment is out of reach (her friend might be busy), however, she is left to her own devices. By making a *plan*, she may induce herself to go for the run. That is, planning may serve as an *internal* commitment device allowing to exert *self-control*.

Yet planning has received little explicit treatment in economics. Generally speaking, in intertemporal decision problems, plans capture what decision makers desire or predict about own future choices (e.g., how much to save and consume in the future) but do not actually influence these choices. For example, a present-biased saver may plan to start saving more tomorrow. However, this plan does not actually affect her savings decision tomorrow. Thus, if her future self is present-bias again, she will keep under-saving.[24]

In contrast, we present a model in which plans serve as internal commitment devices that enable the decision maker to exert self-control in subsequent choices.[25]

We characterize *Planning To Self-Control* in terms of three simple and intuitive Axioms on choice between and from decision problems. Our model generalizes temptation-based models of self-control pioneered by Gul and Pesendorfer (2001) to allow for magnitude effects: self-control increases when upping the stakes of a decision problem. For example,

[23] Baumeister et al. (2020) report that close to three quarters of thoughts about the future involve plans and that thinking about the future is common in individuals.

[24] In such cases, when subsequent choices do not conform to the plan made beforehand, the decision maker is said to be *dynamically inconsistent*.

[25] The fact that anticipating such a change of preference creates an incentive to commit has been well understood since the seminal work by Strotz (1955). However, as already pointed out by Thaler and Shefrin (1981), even in the absence of external commitment devices, decision makers typically have at their disposal strategies short of external commitments: self-control.

in intertemporal choice tasks, individuals are consistently found to switch from sooner, smaller rewards to larger, later ones when rewards are scaled up.[26] We show that – given a particularly simple specification, our model produces such behavior for agents that are present-biased unless using self-control to overcome it.[27] When applied to a simple infinite-horizon consumption-savings model, it produces a poverty trap at the individual level: decision makers forego self-control and over-consume (running down their wealth) unless they are endowed with sufficient initial wealth.

Preview of Main Results

Consider an individual faced with a decision problem A. At the planning stage she evaluates alternatives in A according to some utility function u but anticipates to choose from A according to v at the choice stage. By making a plan, the decision maker is able to restrict subsequent choices. That is, plan $P \subseteq A$ induces choice $c(A) = \max(P, v)$.[28] For instance, a restaurant goer may plan to 'choose a vegetarian dish'. If, say, the restaurant offers steak, salad and a veggie lasagna, this restricts her to choose between the salad and lasagna. At the same time, plan $P \subseteq A$ comes at a cost $\kappa(P, A)$. When evaluating decision problem A, the decision maker makes an optimal plan by trading off its cost against the benefit of (increased) self-control:

$$\begin{aligned} U(A) = \ & \max_{P \subseteq A} u(x_P) - \kappa(P, A) \\ & \text{s.t. } x_P = \max(P, v). \end{aligned} \tag{$\star$}$$

We show that *Planning to Self-Control* is equivalent to three simple and intuitive Axioms on preferences over decision problems and subsequent choices from them: decision makers (1) always weakly prefer to commit, (2) strictly prefer to commit only if this rids them of a self-control problem, and (3) are made strictly better off when adding an alternative to a decision problem A that is preferred to their choice from A both at the planning and the choice stage.

Decision makers who evaluate decision problems according to ($\star$) act as if being unconstrained in their ability to induce self-control: every alternative $x \in A$ is choosable given appropriate planning (e.g., consider $P = \{x\}$). The minimal planning cost to induce $x \in A$ may be interpreted as its self-control cost $C(x, A)$. We show that Planning To Self-Control is behaviorally equivalent to a decision maker who, at every decision problem A, acts as if choosing the optimal level of self-control:

$$U(A) = \max_{x \in A} u(x) - C(x, A). \tag{$\star\star$}$$

[26] We discuss the empirical evidence in Chapter 4 below.
[27] Recent research in neuroscience links intertemporal decisions in general, and the magnitude effect in particular, to self-control (Figner et al., 2010; Ballard et al., 2017, 2018).
[28] For every $f : X \to \mathbb{R}$ and all $A \in \mathcal{A}$, we define $\max(A, f) = \{x \in A | \nexists y \in A : f(y) > f(x)\}$.

The restrictions our model puts on $C(x, A)$ allow for an interpretation in terms of self-control costs but are weak enough to nest other axiomatic models of self-control. For example, in Gul and Pesendorfer (2001), $C(x, A) = \max_A v - v(x)$. Here, v is interpreted as a measure of how much other alternatives tempt the decision maker at the choice stage. Subsequent work (Takeoka, 2008; Noor and Takeoka, 2010, 2015) has considered more general functional forms for the cost function but has retained convexity (in foregone temptation utility) implying that decision makers use less self-control when more tempting alternatives are added to A.[29]

This seems at odds with the intuition that self-control may become more attractive as the *stakes* of a decision increase because the costs of self-control decrease relative to its benefit. The simplest possible specification in our model, a fixed cost $C(x, A) = k > 0$ if $v(x) < \max_A v$ and $C(x, A) = 0$ else, implies such a *magnitude effect* for self-control.[30] To see this, note that if, say $x = \max(A, u)$ and $y = \max(A, v)$, then $(\star\star)$ reduces to $U(A) = \max\{u(x) - k, u(y)\}$. Thus, self-control is worthwhile if (and only if) the utility stakes $u(x) - u(y)$ exceed k.[31]

For example, reconsider our individual making plans for after-work activities at lunch-time. To put a twist on the story, suppose now that, initially, she is not very concerned about missing a run (r) when only faced with the alternative to watch a movie (m) and ends up in front of the screen. As a third option, to go out for drinks (d), becomes available to her, however, the stakes of the problem increase. While she expects to prefer the social occasion come the evening, she is profoundly concerned about its alcohol-related health risks in advance, so that $v(d) > v(m) > v(r)$ but $u(d) \ll u(m) < u(r)$. Consequently, she may use self-control in this situation seeing that its benefits have increased substantially.[32] Indeed, a fixed cost of self-control would imply such behavior given that $u(r) - u(d) > k > u(r) - u(m)$.

The possibly most well-known and empirically best documented case of increasing self-control is that of magnitude effects in intertemporal choice tasks: individuals tend to choose a smaller immediate reward (s) when the stakes are low but self-control to a larger later

[29] For details, see our discussion of temptation-driven self-control models in Appendix A.1.

[30] While simple, we believe that a fixed cost of self-control is also compelling on intuitive grounds as it may reflect the cost of engaging a 'self-control system' in the human brain. Findings that exertion of self-control is linked to heightened activity in certain areas of the human pre-frontal cortex (Figner et al., 2010; Ballard et al., 2017, 2018) provide tentative evidence in support of the existence of such a system.

[31] In the finite choice setting we use for our axiomatization below, utility is identified only up to positive monotone transformations hence may carry no *cardinal* information. Still $u(x) - u(y)$, as an (ordinal) measure of the stakes at menu A, is guaranteed to increase as more options z are added to A that present self-control problems (in the sense that if $v(z) > v(y)$, then $u(z) < u(y)$).

[32] Note that this constitutes a violation of the Weak Axiom of Revealed Preference (WARP) caused by increasing self-control: $c(\{m, r\}) = m$ but $c(\{d, m, r\}) = r$ where $u(r) > u(m)$. How is this possible? Note that, from a self-control perspective, watching the movie is a fundamentally different alternative in the two choice problems. This is due to the fact that it requires no self-control when the only alternative is to go for the run, while it *does* require self-control when being presented with the option to go out for drinks. At the same time, going for the run requires self-control in both instances. Thus, the *marginal* cost of self-controlling from m to r may decrease as another option d becomes available. Given a fixed cost of self-control, for example, the marginal cost is zero when d is present.

one (l) when the stakes are high. For example, if, say $s = \$30$ and $l = \$50$, you might prefer receiving s immediately over receiving l in a year. At the same time, if $s = \$150$ and $l = \$250$, you prefer to wait a year to receive the additional \$100. While this behavior is hard to square with discounted utility maximization, it may be perfectly reasonable if you are struggling with self-control problems related to immediate rewards.[33] When little is at stake, choosing the immediate reward is of little consequence. Yet, as the stakes grow, self-control becomes a worthwhile exercise. Indeed, we show in Chapter 4 below that this is natural for decision makers who exhibit present-bias unless engaging in self-control at a fixed cost to overcome it.

More generally, in a simple infinite-horizon consumption-savings problem, such decision makers over-consume (due to present-bias) unless they are sufficiently wealthy to make self-control attractive. This produces a poverty trap (at the individual level).

Relation to the Literature

Psychology and Human Decision Processes. Recent work in psychology and human decision processes has identified several strategies and processes engaged by individuals to facilitate self-control (Duckworth et al., 2014, 2016; Hennecke and Bürgler, 2020).[34] This 'process model' of self-control distinguishes between preventive strategies (e.g., commitment) seeking to avoid situations involving conflicts, and interventive, intra-psychic ones employed to deal with self-control problems if they occur (cf. Inzlicht et al., 2021).[35] While, traditionally, goal setting has been considered central among the latter, goal achievement has been found to improve substantially when forming implementation intentions through making plans (Gollwitzer, 1999; Gollwitzer and Sheeran, 2006; Gollwitzer et al., 2010; Gollwitzer and Oettingen, 2011).[36] Masicampo and Baumeister (2011) argue that this might be due to the fact that plans can effectively turn control of goal pursuit over to automatic unconscious processes that can be called upon when the need arises. In support of the effectiveness of plans, Sjåstad and Baumeister (2018) show that willingness to plan is as-

[33] While not inconsistent with discounted utility maximization per se, such choices would imply unreasonable curvature of the utility function (see Noor, 2011).

[34] Earlier findings that individuals subjected to cognitively laborious tasks (requiring self-control to stay focused) exhibited lower levels of self-control in subsequent experiments initially led the literature to theorize about the existence of a limited stock of cognitive (energy) resource that gets depleted ('ego-depletion') when exercising self-control (Baumeister et al., 1998, 2007). For a recent axiomatic treatment of choice behavior given a limited stock of 'willpower', see Masatlioglu et al. (2020). However, it should be noted that the term is not always used consistently. For example, Bermúdez et al. (2023) seem to refer to any kind of internal psychological mechanisms for resolving self-control problems as 'willpower' (as opposed to more externally-rooted devices such as commitment or extrinsic incentives).

[35] In this sense, our model is primarily concerned with an intra-psychic/interventive strategy for self-control: planning. Findings by Bermúdez et al. (2023) suggest that this internal psychological aspect is also more representative of every-day notions of self-control. However, the situational/preventive aspect features indirectly in our model to the effect that menu preferences reflect the decision maker's desire to pre-commit (in order to rid herself of self-control problems).

[36] In a related vein, Taylor et al. (1998) report that (mental) process simulations improved exam performance by psychology students substantially better than outcome simulations.

sociated with good self-control. Other measures of propensity to plan have been shown to be predictive of goal achievement (Ludwig et al., 2018) and good credit scores (Lynch et al., 2010). In an early study of self-control in children, Mischel and Patterson (1976) find that four-year old children are better at sticking with a boring task (and resisting a distraction) when given a simple if-then plan.

Economics. In economics, the study of conflicting self-interests as exemplified by dynamically inconsistent tastes and the resulting desire to pre-commit to a course of action goes back at least to Strotz (1955). Pointing out that focusing on pre-commitment only provides part of the story, Thaler and Shefrin (1981) develop an early theory of self-control. Gul and Pesendorfer (2001) pioneer an axiomatic treatment based on the idea that decision makers take self-control into account when ranking decision problems (menus). In the strand of literature their work has inspired (see, e.g., Gul and Pesendorfer, 2004, 2006; Noor and Takeoka, 2010, 2015) self-control is costly since decision makers need to resist more tempting alternatives. While this literature considers costs that are convex (in foregone temptation utility) and lead to a loss of self-control when adding more tempting options to a menu, our model allows for more general specifications consistent with increasing self-control. Moreover, this literature takes as behavioral primitive preferences over menus of *lotteries* and invokes an Independence assumption. While we build on the same general framework, our axiomatic treatment is set in a (risk-free) finite-choice environment. *Prima facie*, the notions of self-control and risk seem unrelated to us. Thus, we believe it is of interest to develop an axiomatic foundation in a deterministic setting. While, in our model, self-control is enacted by making plans, Nehring (2006b) considers a more general approach in which decision makers optimize over preferences according to which they choose subsequently. He characterizes this model in terms of the 'positive' component of the the 'Set Betweenness' Axiom of Gul and Pesendorfer (2001): that the union of two menus can never be strictly preferred to both of them. While our model fails the fully-fledged 'Set Betweenness' Axiom, it always satisfies the weakening considered by Nehring (2006b). Thus, in terms of generality, our model ranks in between the two.

A natural interpretation of representation ($\star$) is that of a planning-stage self gaming a choice-stage self. The planning self is faced with selecting a self-control action $P \subseteq A$ (the plan) to which the choice-stage self reacts optimally (choosing $x_P = \max(P, v)$).This connects our work to games-of-multiple-selves models in the literature. For example, Fudenberg and Levine (2006, 2012) capture self-control in terms of equilibria of a game played by a long-run self (the "planner") and multiple sequential short-run selves (the "doers"). The "planner" can steer future selves through some costly self-control action entering their utility functions. Thus, in their model, self-control actions affect choice-stage *preferences*. Hsiaw (2013) and Koch and Nafziger (2011) study the role of goals. They show that strategically setting goals can help attenuate present bias assuming that they provide reference points for future utility. In contrast, we capture self-control actions in terms of *partial* (internal) *commitment* devices: plans restrict the choice-stage self to choose from a sub-menu according to *stable* preferences. In general, while game-of-multiple-selves models derive their results from assumptions on utility functions and associated costs of self-control, our axiomatic treatment takes observable choice behavior as primitives of the model and derives the representation from testable axioms.

Like we do here, Benhabib and Bisin (2005) argue that self-control can be enacted through internal commitment mechanisms, although they do not model them explicitly. In analogy to our analysis of a fixed cost of self-control, they assume that decisions otherwise made through automatic responses can be overridden through activation of controlled processing (that is, internal commitment acts through cognitive control) at some given cost. Interestingly, when applied to a consumption-savings problem where automatic decisions are driven by (stochastic) temptation of immediate consumption, the optimal self-control behavior is determined by a cut-off rule: the decision maker makes use of self-control when enough is at stake; i.e. when the temptation is sufficiently large. In contrast, in the simple consumption-savings model we present below, stakes are determined by the decision maker's wealth. That is, self-control is exerted only by the sufficiently wealthy.[37]

The rest of Part II is structured as follows. Chapter 2 presents the framework and discusses our Axioms. Chapter 3 contains our two Representation Theorems (Theorem 1 for ($\star$) and Theorem 2 for ($\star\star$)). Chapter 4 applies the assumption of a fixed cost of self-control to intertemporal consumption(-savings) choices. Here, we show that the fact that self-control increases in the stakes of a decision translates into a magnitude effect and leads to poverty traps in infinite-horizon problems. Lastly, Chapter 5 concludes. All proofs are contained in the Appendix.

[37] Unlike us, Benhabib and Bisin assume cognitive control costs to be proportional to the wealth stakes of the problem, thereby focusing on varying temptation.

2 The Model

2.1 Preliminaries

We consider a simple dynamic setting with two stages: a planning stage and a choice stage. Let X be the set of alternatives at the choice stage; we assume that $|X| < \infty$. Define $\mathcal{A} = 2^X \backslash \{\emptyset\}$ to be the collection of all non-empty decision problems (henceforth also: *menus*) over X. We generically denote alternatives from X by x, y, z and menus from $\mathcal{A}$ by A, B. At the planning stage, the decision maker ranks menus according to some weak order $\succsim \subseteq \mathcal{A} \times \mathcal{A}$. When restricted to singletons, $\succsim$ reveals the decision maker's planning-stage (commitment) preference over alternatives. For simplicity, we assume that it does not display any indifferences. For ease of notation, we also denote singleton sets $\{x\}$ simply by x and write $x \succsim y$ instead of $\{x\} \succsim \{y\}$.

Axiom 0.1. WEAK ORDER. *$\succsim$ is complete and transitive. Its restriction to singletons is anti-symmetric.*

At the choice stage, the decision maker chooses from menus. We model this by means of a (non-empty and singleton-valued) choice function $c : \mathcal{A} \to X$, $A \mapsto c(A) \in A$. Dynamic choice behavior is captured by the tuple $(\succsim, c(\cdot))$.

We denote by $\geq$ the anticipated default choice-stage preference over alternatives in the absence of self-control. Note, however, that since the decision maker might optimally bring self-control to bear on observed choice behavior, $\geq$ is not revealed by observing $c(\cdot)$ alone. That is, in general, $\geq$ does not rationalize $c(\cdot)$.[38] Indeed, $c(\cdot)$ may not be rationalized by *any* preference relation as optimal planning can lead to plausible violations of the Weak Axiom of Revealed Preference (WARP). We elaborate further below. At the same time, $\geq$ is revealed by observing $(\succsim, c(\cdot))$ jointly. Consider any $x \succ y$. First, note that choosing x from the menu $\{x, y\}$ may require costly planning; thus $x \succsim \{x, y\}$. Second, the default choice (in the absence of costly planning) from $\{x, y\}$ is guaranteed to be no worse than y; thus $\{x, y\} \succsim y$. Consequently, $x \succsim \{x, y\} \succsim y$.

If $x \sim \{x, y\}$, the decision maker anticipates to choose x from $\{x, y\}$ without the need for self-control (i.e. by default: $x > y$). In this case, planning-stage and choice-stage preferences agree; we say that x *dominates* y:

$$x \gg y :\iff x \sim \{x, y\} \succ y.$$

[38] We say that a preference relation R rationalizes choice function $c(\cdot)$ if, for all menus A, $c(A) = \{x \in A | \forall y \in A : xRy\}$.

On the other hand, if $x \succ \{x, y\}$, the decision maker anticipates that x is choosable only under costly planning because y is preferred at the choice-stage: $y > x$. We denote such anticipated preference reversals by:

$$x \gtrless y :\iff x \succ \{x, y\} \succsim y.$$

If these relations are to reveal a consistent choice-stage (default) *ranking* of alternatives, the decision maker must have a transitive perception of both dominance and reversals.

Axiom 0.2. Transitive Reversals and Dominance. $[x \gtrless y$ *and* $y \gtrless z] \implies x \gtrless z$. $[x \gg y$ *and* $y \gg z] \implies x \gg z$.

In turn, this ensures that $\geq = (> \cup -)$, is a linear order; where, for all $x \neq y$,

$$x > y :\iff [x \gg y \text{ or } y \gtrless x]$$

and, for all $x \in X$: $x - x$.[39] Indeed, vindicating the intuitions presented above, $x \gtrless y \iff [x \succ y$ and $x < y]$ and $x \gg y \iff [x \succ y$ and $x > y]$.

2.2 Main Axioms

Anticipating a preference reversal $x \gtrless y$ presents the decision maker with a self-control problem. (Henceforth, we will also simply refer to $x \gtrless y$ as a self-control problem.) Planning allows to exert self-control but is costly. Thus, in making an optimal plan, the decision makers weighs said costs against the benefits of (increased) self-control. Observed choices as captured by the choice function $c(\cdot)$ reflect optimal resolution of this trade-off. The intuition driving the main Axioms of our model is that the costliness of a plan is linked to the self-control problems it helps overcome. That is, planning is the costlier the more self-control it allows to exert. For example, when considering binary menus, this is implicit in our definition of a self-control problem: $x \gtrless y \iff x \succ \{x, y\} \succsim y$. Indeed, plan $P = \{x, y\}$ (which we may also think of as making no plan) does not induce self-control ($y = \max(\{x, y\}, >)$) but is free of cost.[40] Consequently, $\{x, y\} \succsim y$. On the other hand, plan $P = \{x\}$ allows to exert self-control ($x = \max(\{x\}, >)$) but is costly; thus $\{x, y\} \prec x$.

Generally speaking, if planning costs are tied to the self-control problems it help overcome, the decision maker cannot be left worse off by committing to a sub-menu given it contains what is chosen from the original one. To see this, suppose we observe the decision maker choosing alternative $x \in A$ from menu $A \cup B$. Then, in a revealed sense, the decision maker prefers to plan so as to induce x in menu $A \cup B$.[41] Potentially, this involves self-controlling away from alternatives $y \in B$. The need to do so is removed, however,

[39] See Lemma 3 in the Appendix.
[40] For every $P \subseteq X \times X$ and all $A \in \mathcal{A}$, we define $\max(A, P) = \{x \in A | \nexists y \in A : yPx\}$.
[41] We say that $P \subseteq A$ induces x (in menu A) if $x = \max(P, >)$.

when committing to the sub-menu A. Thus, planning to induce x in A comes at a lesser cost. Therefore, A should be no worse to the decision maker than $A \cup B$. This is Axiom 1.

Axiom 1. Weak Preference for Commitment. $c(A \cup B) \in A \implies A \succsim A \cup B.$

Moreover, if preference for commitment is strict, this must be due to the fact that it rids the decision maker of some self-control problem that need not be overcome by planning when committing. In other words, it is self-control problems that make planning costly. This is Axiom 2.

Axiom 2. Costly Planning/Self-Control. $A \succ A \cup B \implies \exists y \in B : c(A) \gtrless y.$

Equivalently, the counter-positive of Axiom 2 states that if, for all $y \in B$, $c(A) > y$ or $y \succ c(A)$, then $A \cup B \succsim A$. This reflects the intuition that additional options can never hurt as long as the decision maker either does not need to plan to self-control away from them or has no incentive to do so.

In contrast, when $x \gg y$, there is no self-control problem. The decision maker strictly prefers x both at the planning stage and at the choice stage. Axiom 3 states that adding x to some menu from which y is chosen must make the decision makers better off. Roughly speaking, this reflects the intuition that there exists some plan for menu $x \cup A$ inducing x that is no more costly than the optimal plan inducing y in menu A.[42] As x is strictly preferred at the planning stage, the decision maker values the addition of x to A.

Axiom 3. Unequivocally Better Choice. $x \gg c(A) \implies x \cup A \succ A.$

A fortiori, when $\{x, y\} \subseteq A$ and $x \gg y$, then $y \neq c(A)$. That is, alternatives that are dominated by some other available alternative are never chosen.

As we show below our axioms are not only necessary but also sufficient for a decision maker to be Planning to Self-Control. Before we turn to our representation, we discuss our Axioms further.

2.3 Discussion of Main Axioms

Axiom 1 implies a weakened version of the 'Set-Betweenness' Axiom that is central in the characterization of temptation-self-control preferences in Gul and Pesendorfer (2001). 'Set-Betweenness' requires that the union of any two menus $A \succsim B$ rank between them: $A \succsim A \cup B \succsim B$. Indeed, Axiom 1 implies its 'positive' part: $A \succsim A \cup B$. Nehring (2006b) shows that it is necessary and sufficient for $\succsim$ to be rationalized in terms of a

[42] To be precise, we cannot make this inference from the fact that the decision maker strictly prefers adding x to A. The latter only implies that, if planning to induce x in $A \cup \{x\}$ is more costly than the optimal plan for A (inducing y), then this is dominated by the corresponding increase in planning-stage utility that x offers over y. This fact is reflected by Property 3.b) of a planning/self-control cost function in our model (cf. Definition 1/2).

general model of 'second-order preference'.[43] However, the corresponding 'negative' part fails. To see why $A \succsim B \succ A \cup B$ is reasonable within our model, suppose that the decision maker self-controls to x in menus A and B (that is $x = c(A) = c(B)$, which requires that $x \in A \cap B$, of course). Planning to induce x in the menu $A \cup B$ is more costly, however, seeing that the decision maker now needs to self-control away from alternatives in A *and* B at the same time. In contrast, this is not possible if $A \cup B$ contains only two alternatives. Indeed for all $x \succsim y$, Axiom 2 implies that $\{x, y\} \succsim y$. Axiom 1 implies that $x \succsim \{x, y\}$. In other words, 'Set-Betweenness' holds for binary menus.

Axiom 1 implies that if $A \succ y$ for some $y \in A$, then $y \neq c(A)$. For example, if $x \succsim \{x, y\} \succ y$, then we must have $x = c(\{x, y\})$. On the other hand, if $\{x, y\} \sim y$, our Axioms do not put any restrictions on choice from $\{x, y\}$. Note that if $\{x, y\} \sim y$ and $x = c(\{x, y\})$, the decision maker plans to self-control to x at a cost that exactly equals the benefit of self-control to her. Hence she is indifferent between self-control and no self-control ($y = c(\{x, y\})$).

Other than the ones discussed above, our Axioms imply only weak (consistency) restrictions on choice from menus. In particular, $c(\cdot)$ may fail to satisfy WARP known to characterize choice behavior that is rationalized by some preference relation. In the presence of self-control problems it is natural to allow for violations of WARP. To see this, reconsider our introductory example from above. While our decision maker self-controls to go for the run (r) in the presence of an alternative she considers seriously detrimental to her health (to go for drinks: d), she optimally decides to forego self-control when only ending up watching a movie (m). This choice pattern, $r = c(\{d, m, r\})$ and $m = c(\{m, r\})$, constitutes a violation of WARP. Seeing that $r \gneqq m$, it is an example of *increasing* self-control: the decision maker has more self-control at the menu $\{d, m, r\}$ than at the sub-menu $\{m, r\}$. To give another example of where such behavior can be plausible, consider an individual struggling with being tempted by desserts. It may be easier to resist having a dessert altogether when there are 10 of them available (including one that is particularly unhealthy but tempting) as compared to when there is only one (moderately unhealthy but tempting one).

At the same time, our Axioms are equally consistent with violations of WARP resulting from decreasing self-control. Such choice patters are generally consistent with generalizations of Gul and Pesendorfer (2001) (see also Appendix A for more details). For instance, suppose some product is available at different prices (e.g., because it is available from different brands). There is a high-price (h), medium-price (m) and low-price (l) variant. At the planning stage, a consumer regards all variants as perfect substitutes (i.e., there are no perceived differences in terms of quality) so that she ranks them according to price:

[43] Nehring considers a second condition ('singleton monotonicity') on menu preferences. When adding to some menu an alternative that is preferred to all alternatives in the menu, this never leaves the decision makers worse off. Formally, if for all $y \in A$, $x \succsim y$, then $x \cup A \succsim A$. Nehring calls this subclass of preferences 'self-command' preferences. Our Axiom 2 implies 'singleton monotonicity'. Thus, our model is a special case of 'self-command' preference.

$l \succ m \succ h$. However, she expects her ranking to reverse at the time of choice: $h \succ m \succ l$ (e.g., driven by product packaging and presentation or impulsive inferences about quality). Suppose further that a corner store only offers l and m, a bigger supermarket sells all three variants. While our consumer may optimally (plan to) self-control to the low-cost option l in the corner store ($l = c(\{l, m\})$), doing so in the supermarket ($\{l, m, h\}$) might require a higher planning effort so as to exclude both m and h at the choice stage. Thus, she may only choose to self-control to the medium-price option ($m = c(\{l, m, h\})$) but optimally forego a higher level of self-control (to l). As $l \gtrless m$, this choice pattern constitutes a loss of self-control when adding the high-price option h to the menu $\{l, m\}$.

3 Representation

3.1 Planning to Self-Control

Our representation builds on the idea that plans allow the decision maker to restrict subsequent choices. We model this by *identifying* plans with the choice restrictions (commitment) they entail. That is, for every decision problem A and every non-empty $P \subseteq A$, we call P a *plan* for A. If plan P is made, the decision maker chooses the best available alternative that is consistent with P. Given our Axioms above, (default) preferences at the choice stage are represented by some utility function v. Thus, plan $P \subseteq A$ *induces* $x_P := \max(P, v)$ at the choice stage. At the same time, it comes at a cost $\kappa(P, A)$. Our Axioms imply restrictions on the planning-cost function $\kappa(\cdot, \cdot)$ (cf. Definition 1); we discuss them in detail below. Intuitively speaking, planning incurs a cost (if and) only if it allows to exert self-control and is the costlier the more it does so. As preferences over decision problems $\succsim$ are a weak order, they are representable by some utility function U. We let u be the restriction of U to singleton menus; that is, for all $x \in X$, $u(x) := U(\{x\})$. Thus, u represents the decision maker's planning-stage (commitment) preference over alternatives. Our first main result shows that the tuple $(\succsim, c(\cdot))$ satisfies our Axioms if and only if it is *Planning to Self-Control* (PTSC). That is, $U(A)$ is the indirect utility from maximizing planning-stage utility $u(x_P)$ net of planning costs $\kappa(P, A)$ and optimal planning rationalizes subsequent choice behavior: $c(A) = x_{P^\star}$ for some optimal $P^\star$.

Theorem 1. *If and only if $(\succsim, c(\cdot))$ satisfies Axioms 0.1,0.2,1-3, there exist strictly increasing*[44] *utility functions $u, v : X \to \mathbb{R}$ and planning-cost function κ such that:*

1. *$\succsim$ is represented by*

$$U(A) = \max_{P \subseteq A} u(x_P) - \kappa(P, A) \quad \text{s.t. } x_P = \max(P, v). \tag{$\star$}$$

2. *$c(A) = x_{P^\star}$ for some $P^\star$ that solves ($\star$).*

Definition 1. *We say that $\kappa : \mathcal{A} \times \mathcal{A} \ni (P \subseteq A, A) \mapsto \kappa(P, A) \geq 0$ is a planning-cost function if*

1. *$\kappa(P, A) > 0 \iff \exists y \in A \backslash P : v(y) > v(x_P)$*

[44] That is, $u, v : X \to \mathbb{R}$ allow only for trivial indifferences: $u(x) = u(y) \implies x = y$ and $v(x) = v(y) \implies x = y$.

2. $x_P \in A \implies \kappa(P \cap A, A) \leq \kappa(P, A \cup B)$

3. *a)* $u(x_{P \cup B}) \leq u(x_P) \implies \kappa(P \cup B, A \cup B) \leq \kappa(P, A)$

b) $u(x_{P \cup B}) > u(x_P) \implies \kappa(P \cup B, A \cup B) - \kappa(P, A) < u(x_{P \cup B}) - u(x_P)$

The first property requires that P incurs a cost at menu A if and only if the alternative x_P it induces is not the default choice; that is, if there exists some $y \in A \backslash P$ that is chosen over x_P unless it is excluded from P (as $v(y) > v(x_P)$). Second, plan $P \subseteq A \cup B$ is no more costly when projected onto the sub-decision problem A as long as it induces the same alternative: note that if $x_P \in A$, then $x_{P \cap A} = x_P$. In particular, this implies that some fixed P is (weakly) less costly at smaller menus as compared to larger ones. That is, if $P \subseteq A$, then $\kappa(P, A) \leq \kappa(P, A \cup B)$ for every menu B. Equivalently put, the planning cost associated with restricting choices to some given P (weakly) increases as more alternatives are available at the choice stage. This reflects the intuition that P may need to exclude additional alternatives when choosing from $A \cup B$ as compared to A (namely those in B; unless $P \subseteq B$). On the other hand, when modifying plan $P \subseteq A$ so as to be consistent with all alternatives that are added to menu A, that is when considering $P \cup B \subseteq A \cup B$, the associated planning cost should be no higher. The third property establishes this for the case that $x_P = x_{P \cup B}$ or $x_P \gtrless x_{P \cup B}$ (part a)). However, if $x_{P \cup B} \gg x_P$ (i.e. if B contains an alternative that dominates x_P), then our Axioms only allow us to ascertain that the cost differential $\kappa(P \cup B, A \cup B) - \kappa(P, A)$ is bounded above by the planning-stage utility differential $u(x_{P \cup B}) - u(x_P)$ (part b)).[45] Note that an immediate consequence of Property 3a) is that if $P' \subseteq P \subseteq A$ induce the same alternative, then $\kappa(P', A) \geq \kappa(P, A)$ (let $B \subseteq A \backslash P'$).

Planning is costly. For this reason, rational decision makers use it only to the extent necessary to induce self-control. We say that plan $P \subseteq A$ is *efficient* if for all alternatives $y \in A \backslash P$ excluded by P it holds that self-control is both necessary ($v(y) > v(x_P)$) and beneficial ($u(y) < u(x_P)$), that is, $x_P \gtrless y$.[46] Without loss of generality, we can restrict a PTSC decision maker to choose among all efficient plans in representation ($\star$). To see this, note that for any $P \subseteq A$, we can construct an efficient plan $P' \supseteq P$ such that $A \backslash P' = \{y \in A : x_P \gtrless y\} = \{y \in A : u(x_P) > u(y) \textit{ and } v(x_P) < v(y)\}$. If $x_{P'} = x_P$, the cost of plan P' is (weakly) less than that of P (let $B = P' \backslash P \subseteq A$ and use property 3a) for planning-cost functions) but induces the same alternative. Hence P can be optimal only if P' is. Else if $x_{P'} \neq x_P$, then $x_{P'}$ must dominate x_P ($u(x_{P'}) > u(x_P)$ and $v(x_{P'}) > v(x_P)$). By property 3b) for planning-cost functions, the cost of plan P' exceeds that of P (if at all) by less than the planning-stage utility benefit of $x_{P'}$ with respect to x_P: $u(x_{P'}) - \kappa(P', A) < u(x_P) - \kappa(P, A)$. Here, P cannot be optimal.

[45] Note that, by definition, $v(x_{P \cup B}) \geq v(x_P)$, with equality only if $v(x_{P \cup B}) = v(x_P)$. If $u(x_{P \cup B}) > u(x_P)$, then $x_{P \cup B} \neq x_P$; hence $x_{P \cup B} \gg x_P$.

[46] In other words, all $y \in A \backslash P$ present the decision maker with a self-control problem: $x_P \gtrless y$.

In our model, plan $P \subseteq A$ induces $x_P = \max(P, v)$ at the choice stage. Prima facie, this seems to assume that decision makers (rationally expect to) stick to the plan they made beforehand. This seems implausible. However, it is not the only possible interpretation. Alternatively, we can identify with $P \subseteq A$ the plan for which P is the collection of alternatives that are acceptable (deviations) given said plan. After all, what is important about a plan in our model is to what extent it restricts subsequent choices but not how it is represented in the decision maker's mind.[47] Given this interpretation, however, it is important to note that our concept of optimality (for plans) implicitly assumes that every non-empty choice restriction $P \subseteq A$ is induced by *some* plan the decision makers can make. In effect, this ensures that self-control is *potentially unlimited* (but possibly too costly). Fudenberg and Levine make an analogous assumption for their 'planner-doer' model: every action by a future self can be elicited by the long-run self through some appropriate self-control action (2006, Assumption 2).

As can be expected in our finite choice context, the additive form in ($\star$) is not identified. That is, while Theorem 7 shows that a representation of this from always exists, it does not ensure that U is of the prescribed form whenever U represents $\succsim$. In general, there are multiple solutions to ($\star$). However, the optimal *efficient* plan $P^\star$ that solves ($\star$) and rationalizes observed choice behavior (i.e. $c(A) = x_{P^\star}$) is unique.[48]

3.2 Equivalent Self-Control Cost Model

Plans are instrumental in inducing self-control. That is, a PTSC decision maker plans in order to induce alternatives that would otherwise not be chosen. Consider decision problem A. For every alternative $x \in A$, let $P_x := \{y \in A : v(y) \leq v(x)\}$. Note that if $P \subseteq A$ is some plan inducing x, we need to have $P \subseteq P_x$; hence $\kappa(P, A) \geq \kappa(P_x, A)$.[49] Thus, P_x is the cost-minimal plan inducing x. From an abstract point of view, we may also think of the decision maker as optimizing directly over eventual choices $x \in A$ incurring a self-control cost $C(x, A) = \kappa(P_x, A)$. Then, she evaluates menus according to

$$U(A) = \max_{x \in A} u(x) - C(x, A). \tag{$\star\star$}$$

Indeed, Theorem 2 shows that this is an equivalent representation of planning-stage menu preferences. Moreover, if $x^\star$ solves ($\star\star$), then there exists some (efficient) plan $P^\star$ solving ($\star$) such that $x^\star = x_{P^\star}$ (hence $x^\star = c(A)$). Vice versa, if $P^\star$ solves ($\star$), then $x_{P^\star}$ solves ($\star\star$).

Theorem 2. *Let κ be a planning-cost function, then there exists a self-control-cost function C such that:*

[47] That is, we adopt an extensional definition of plans.
[48] If we do not restrict ourselves to efficient plans in ($\star$), there may be several equally costly plans that induce the choice $c(A)$. However, as noted above, restriction to efficient plans is without loss of generality.
[49] To see this, let $B = P_x \setminus P \subseteq A \setminus P$ and use Property 3a) for $\kappa(\cdot, \cdot)$.

1. $\max_{P \subseteq A} u(x_P) - \kappa(P, A) = \max_{x \in A} u(x) - C(x, A)$

2. a) $P^\star$ *solves* $\max_{P \subseteq A} u(x_P) - \kappa(P, A) \implies x_{P^\star}$ *solves* $\max_{x \in A} u(x) - C(x, A)$

   b) $x^\star$ *solves* $\max_{x \in A} u(x) - C(x, A) \implies \exists P^\star \subseteq A : x_{P^*} = x^\star$ *and* $P^\star$ *solves* $\max_{P \subseteq A} u(x_P) - \kappa(P, A)$

Vice versa, if C *is a self-control-cost function, then there exists a planning-cost function* κ *such that 1. and 2. hold.*

Definition 2. *We say that* $C : X \times \mathcal{A} \ni (x \in A, A) \mapsto C(x, A) \geq 0$ *is a self-control-cost function if*

1. $C(x, A) > 0 \iff \exists y \in A : v(y) > v(x)$

2. $C(x, A) \leq C(x, A \cup B)$ *and* $C(x, A) < C(x, A \cup B) \implies \exists y \in B : v(y) > v(x)$

3. *if* $v(x) > v(y)$*, then:*

   a) $u(x) < u(y) \implies C(x, x \cup A) \leq C(y, A)$

   b) $u(x) > u(y) \implies C(x, x \cup A) - C(y, A) < u(x) - u(y)$

In analogy to planning-cost functions, self-control costs increase in the number of self-control problems that need to be overcome. First, the cost of self-controlling to $x \in A$ is strictly positive if and only if at least one alternative $y \in A$ is preferred at the choice stage. Second, self-control costs are greater at larger menus; and strictly so only if the larger menu contains an alternative that is preferred at the choice stage. Lastly, suppose that x is choice-stage preferred to $y \in A$. Then x incurs a smaller self-control cost when added to menu A than y does in menu A given that x presents the decision makers with a self-control problem vis-Ã -vis y: $y \gtrless x$. If x dominates y, the self-control cost for x in $x \cup A$ exceeds that for y in A by strictly less than the additional planning-stage utility offered by x.

Theorem 2 is helpful as it facilitates comparison of the PTSC model with other models of self-control by capturing those in terms of additional assumptions on the self-control-cost function in ($\star\star$). Theorem 2 characterizes the constrained optimization problem ($\star$) in terms of an unconstrained one. This is of particular interest in applications. To simplify both the verbal and technical exposition, we mostly suppress the planning aspect for the rest of Part II and formulate our assumptions in terms of (generic) self-control costs. While our definition of a self-control-cost function involves minimal criteria to be consistent with our Axioms, it allows for a large variety of functional forms. For example, one interesting class of functions arises when considering self-control costs driven by the need to resist temptation. In Appendix A.1 we show how some prominent temptation-driven models of self-control correspond to specific assumptions on self-control costs. Another particularly simple case arises when costs are fixed; that is, do not vary by 'how much' self-control need be exerted. We turn to this case next and show that it generates well-known stylized facts about intertemporal decision making.

4 Fixed Costs Imply Increasing Self-Control: Magnitude Effect and Poverty Traps in Intertemporal Choice

In general, Representation ($\star\star$) allows for self-control costs to increase in the 'amount' of self-control needed to follow through with a certain choice. That is, self-control may become more costly when including alternatives that pose additional self-control problems (cf. Definition 2, in particular, Property 2).[50] In this chapter we consider the particular case of a fixed cost of self-control; that is, when all self-control is equally costly. Intuitively speaking, this could reflect the existence of a self-control system in the human brain that can be activated at some given cognitive cost.[51] Once engaged, this system takes over all decision making (at no additional variable cost).

Formally speaking, we say that the cost of self-control is fixed if there exists some $k > 0$ such that for all $A \in \mathcal{A}$ and $x \in A$:

$$C(x, A) = \begin{cases} 0 & \text{if } x = \max(A, v) \\ k & \text{else} \end{cases}.$$

A fixed cost greatly simplifies the self-control decision faced at some decision problem A. As all self-control is equally costly, the decision maker never exhibits intermediate levels of it. The self-control decision reduces to a binary choice between full self-control at cost k (choosing the planning-stage optimum $\max(A, u)$), and no self-control (choosing $\max(A, v)$). Let $W(A) := \max_{x \in A} u(x)$ be the indirect utility given (full) self-control but net of cost k and $V(A) := u(x_A) = u(\max(A, v))$ denote the indirect utility given no self-control. Self-control is optimal at menu A if and only if

$$W(A) - V(A) \geq k. \tag{4.1}$$

Importantly, note that, while v determines choices given no self-control, these are still evaluated according u when considering whether to self-control (or not). The overall indirect utility at menu A – reflecting self-control behavior – can be conveniently expressed as

$$U(A) = \max\{W(A) - k, V(A)\}.$$

[50] Equivalently, planning is more costly when more self-control problems are present (cf. Definition 1, in particular, Property 2).

[51] Presumably, the latter might best be understood as an opportunity cost for not engaging those parts of the brain involved in self-control in other cognitive activities (Boureau et al., 2015).

Consequently, the decision maker gains self-control as the (utility) stakes of decision problem A, $W(A) - V(A)$ increase.

Although a simplification, the fixed-cost model is general enough to account for a variety of stylized facts about intertemporal choice behavior. In what follows we consider applications to typical intertemporal choice tasks in experimental settings and to consumption-savings decisions.

4.1 The Magnitude Effect

Consider an individual faced with choosing between pairs of dated money rewards. Let the tuple $(m, t) \in \mathbb{R} \times \mathbb{N}$ denote an alternative that pays $\$m$ at time t. In a typical experimental setting, individuals are asked to choose between a smaller sooner (s, t) and a larger later $(l, t+1)$ reward. Individuals exhibit *present bias* when preferring the sooner reward when it is immediate but choose the larger reward when all payoffs are delayed by the same amount of time $\tau \geq 1$. If the decision is made at time t, that is:

$$(s, t) \succ_t (l, t+1) \textit{ and } (s, t+\tau) \prec_t (l, t+\tau+1) \qquad (\textit{present bias})$$

where $0 < s < l$.

For example, an individual may prefer receiving \$30 immediately to receiving \$50 in a week but prefer \$50 in a year and a week to \$30 in a year. For an overview of the early literature finding present bias, see, for example, Frederick et al. (2002). In more recent studies, the status of present bias for *money* rewards has been somewhat contested (see, e.g., Andreoni and Sprenger, 2012; Sutter et al., 2013; Meyer, 2015; Sun and Potters, 2022; however, cf. Meier and Sprenger, 2010; Benhabib et al., 2010; Andersen et al., 2013; Augenblick et al., 2015). Yet present bias is consistently found in studies involving *real* rewards (see, e.g., McClure et al., 2007; Augenblick and Rabin, 2019).

The possibly most well documented finding about intertemporal choice behavior, however, is that preferences between smaller, sooner and larger, later rewards change systematically when increasing the stakes of a decision. For example, an individual preferring \$30 immediately to \$50 in a week may switch to the larger, later one when the rewards are \$150 and \$250 respectively (i.e. scaled up by factor 5). Such *magnitude effects* have been found both for money and real rewards across a variety of settings (see, e.g., Thaler, 1981; Green et al., 1997; Kirby, 1997; Benhabib et al., 2010; Andersen et al., 2013; Sutter et al., 2013; Meyer, 2015; Sun and Potters, 2022).[52]

Formally, individuals exhibit a magnitude effect if

$$(s, t) \succ_t (l, t+1) \textit{ and } \exists \lambda > 1 : (\lambda s, t) \prec_t (\lambda l, t+1). \qquad (\textit{magnitude effect})$$

[52] Studies that elicit money discount rates through indifference statements which ask subjects to specify an amount x such that $(s, t) \sim_t (x, t+1)$, refer to the magnitude effect as an increasing money discount factor $\frac{s}{x}$ (decreasing money discount rate $\frac{x}{s} - 1$).

While, in principle, magnitude effects can be accounted for in the standard (exponentially discounted utility) model by curvature of the utility function, the extreme curvature this would require creates implausible predictions (Noor, 2011). At the same time, magnitude effects are natural in a model with self-control problems that can be overcome at a fixed cost. For simplicity, consider a discounted value maximizer who, given self-control, is an exponential discounter and, given no self-control, discounts quasi-hyperbolically. That is, she evaluates dated rewards according to $u_t(m, t+\tau) = D_\delta(\tau) \cdot m$ and $v_t(m, t+\tau) = D_{\beta,\delta}(\tau) \cdot m$ where

$$D_\delta(\tau) = \delta^\tau \text{ and } D_{\beta,\delta}(\tau) = \begin{cases} \beta\delta^\tau & \tau \geq 1 \\ 1 & \tau = 0 \end{cases}$$

for some $0 \leq \beta < 1$ and $0 < \delta \leq 1$.

Suppose that

$$0 \leq \beta\delta l < s < \delta l \tag{4.2}$$

Then, the decision maker would prefer the later reward $(l, t+1)$ given self-control but choose the immediate reward (s, t) under no self-control. If both rewards are delayed, however, the later reward is preferred no matter what the self-control decision. Thus, the decision maker exhibits present bias if choosing to forego self-control in the former case. This is case if the benefit of self-control, $\delta l - s$, falls short of the self-control cost k.[53]

Now suppose that rewards are scaled up by some factor $\lambda > 1$. Note that this leaves preferences regarding $(\lambda s, t)$ and $(\lambda l, t+1)$ given self-control and no self-control unaffected (and still governed by Equation (4.2)) while making self-control worthwhile if

$$\lambda(\delta l - s) \geq k. \tag{4.3}$$

Thus, as the stakes rise, the decision maker will eventually find it optimal to exert self-control (no matter what the cost of self-control k). For instance, suppose $\beta = 0.5$, $\delta = 1$, $k = 80$ and reconsider our example from above. As $0.5 \cdot 50 < 30 < 50$, Equation (4.2) is satisfied. Moreover, since $50 - 30 < k = 80 < 5 \cdot (50 - 30) = 250 - 150$, the decision maker foregoes self-control in the low-stakes ($\lambda = 1$: $(\$30, t) \succ_t (50\$, t+1)$) but gains self-control in the high-stakes ($\lambda = 5$: $(\$150, t) \prec_t (250\$, t+1)$) condition (cf. Equation (4.3)).

More generally, when choosing at time t from a menu of dated rewards $\mathcal{M}_t = \{(m^{(i)}, t + \tau^{(i)}), i \in \mathcal{I}\}$ (where $\mathcal{I}$ is some index set) such that an immediate reward is preferred under no self-control: $(m^{(i^*)}, t) = \max(\mathcal{M}_t, D_{\beta,\delta}(\tau) \cdot m)$ for some $i^* \in \mathcal{I}$. Then, the choice from $\mathcal{M}$ (given optimal self-control) maximizes the discounted value $D(\tau, m) \cdot m$ according to

[53] Here, linear utility implies that the benefit of self-control is the discounted value it produces in excess of the no-self-control option. Vice versa, we may interpret the cost of self-control k as the premium the decision makers demands to make self-control worthwhile.

the (magnitude-dependent) discount function

$$D(\tau, m) = \begin{cases} \delta^\tau - \frac{k}{m} & \tau \geq 1 \\ 1 & \tau = 0 \end{cases}. \tag{4.4}$$

Benhabib et al. (2010) show that such a fixed component is a better fit to their experimental data than both hyperbolic and quasi-hyperbolic discounting.[54] Discounting according to (4.4) produces a magnitude effect as the discount factor is increasing in the size of the reward. That is, larger rewards are discounted at lower rates.

4.2 Optimal Self-Control in Consumption-Savings Decisions

In this section, we consider a decision maker with Epstein-Zin preferences and a fixed cost of self-control $k > 0$ facing a simple consumption-savings problem. In each period t, the decision maker decides how much of her wealth w (carried over from the previous period) to consume and how much to save (at gross interest rate $R \geq 1$) for the future. As for the previous section, we assume that preferences given self-control and no self-control differ only in terms of how they discount the future. While the decision maker discounts exponentially at rate $0 < \delta \leq 1$ given self-control, no self-control behavior is based on quasi-hyperbolic β-δ discounting where $0 \leq \beta < 1$.

To build intuition, we start with the simple case of two periods. In the second period, the decision maker simply consumes her savings carried over from the first period. Every initial wealth level $w > 0$ presents the decision maker with a menu of first-period consumption levels $c \in [0, w]$. In the second period, she consumes her savings carried over from the first period: $R(w - c)$. Thus, the decision maker solves a single self-control problem in the first period. By slight abuse of notation, we write $U(w) = U([0, w])$ to denote the value function of this problem. Given a fixed cost of self-control, we have

$$U(w) = \max\{W(w) - k, V(w)\} \tag{4.5}$$

where

$$W(w) = \max_{c \in [0,w]} [c^\sigma + \delta(R(w - c))^\sigma]^{\frac{1}{\sigma}} \tag{4.6}$$

and

$$V(w) = [c_{NSC}^\sigma + \delta(R(w - c_{NSC}))^\sigma]^{\frac{1}{\sigma}} \tag{4.7}$$

such that c_{NSC} solves

$$\max_{c \in [0,w]} [c^\sigma + \beta\delta(R(w - c))^\sigma]^{\frac{1}{\sigma}}. \tag{4.8}$$

where $\delta R \leq 1$.

[54] The hypothesis of exponential discounting can be rejected as it does not allow for present bias which they find in their data.

The parameter $\sigma < 1$ captures how readily optimal growth c_{t+1}/c_t responds to changes in the interest rate R: $\sigma = 1 - 1/\gamma$ where $\gamma = d\ln\left(\frac{c_{t+1}}{c_t}\right)/d\ln R > 0$ is the Elasticity of Intertemporal Substitution (EIS). When σ decreases, the decision maker becomes less inclined to readjust to a changing interest rate. In the limit, as σ tends to $-\infty$ (i.e. γ tends to 0) there is extreme consumption smoothing to the effect that $c_{t+1} \approx c_t$ in the optimum (irrespective of R). In the opposite case, as σ tends to 1 (i.e. γ tends to $+\infty$) consumption is perfectly substitutable across time.[55] When $\beta = 1$, the model reduces to the standard consumption-savings (pie-eating) problem (hence we assume $\beta < 1$).

As preferences are homothetic, the optimal solutions c_{SC} to (4.6) and c_{NSC} to (4.8) are to consume a constant fraction (depending on R, σ, β, δ) out of wealth. That is, there exist $0 < \mu_{SC}, \mu_{NSC} < 1$ such that $c_{SC} = \mu_{SC} \cdot w$ and $c_{NSC} = \mu_{NSC} \cdot w$. As the decision maker is less patient in the absence of self-control ($\beta < 1$), she over-consumes: $\mu_{NSC} > \mu_{SC}$.[56] Since utility is homogeneous of degree one, indirect utilities $W(w), V(w)$ are affine in wealth. That is, there exist $b_{SC}, b_{NSC} > 0$ such that $W(w) = b_{SC} \cdot w$ and $V(w) = b_{NSC} \cdot w$. As b_{NSC} reflects sub-optimal resolution of the intertemporal problem (based on $\beta\delta < \delta$), we have $b_{NSC} < b_{SC}$. Consequently, the (net) benefit of self-control

$$W(w) - V(w) = (b_{SC} - b_{NSC}) \cdot w$$

is an increasing affine function in wealth.

As self-control is worthwhile only if its benefits exceed its cost $k > 0$, optimal self-control behavior is given by a simple cut-off rule: self-control is exerted if and only if the wealth stakes exceed the critical value $\bar{w} = \frac{k}{b_{SC} - b_{NSC}}$:

$$c^*(w) = \begin{cases} \mu_{NSC} \cdot w & w \leq \bar{w} \\ \mu_{SC} \cdot w & w > \bar{w} \end{cases}.$$

Thus, poor decision makers over-consume (relative to the self-control benchmark enacted by richer decision makers). As the marginal utility of an additional unit of wealth is higher when using self-control than when not ($b_{,SC} > b_{NSC}$), the value function is kinked upwards at $\bar{w}$:

$$V(w) = \begin{cases} b_{NSC} \cdot w & w \leq \bar{w} \\ b_{SC} \cdot w - k & w > \bar{w} \end{cases}.$$

[55] In this limit case, decisions are made based maximizing discounted values (of consumption streams). Thus, our discussion of the magnitude effect in the last section is a limiting case of the model presented here.

[56] As may be easily verified by computing the First Order Conditions (FOCs): $\mu_{SC} = \frac{\delta^{-\gamma} R^{1-\gamma}}{1+\delta^{-\gamma} R^{1-\gamma}} < \frac{(\beta\delta)^{-\gamma} R^{1-\gamma}}{1+(\beta\delta)^{-\gamma} R^{1-\gamma}} = \mu_{NSC}$.

Infinite Horizon: Poverty Traps

Suppose now that the problem is infinitely lived. Depending on her wealth stock w, the decision maker decides how much to consume in the current period and how much to save. Her savings $R(w-c)$ become the wealth endowment of the continuation problem faced in the next period. As the decision maker faces the same problem in every period, we have

$$U(w) = \max\{W(w) - k, V(w)\} \tag{4.9}$$

where W, V obey the Bellman Equations

$$W(w) = \max_{c\in[0,w]} [(1-\delta)c^\sigma + \delta U(R(w-c))^\sigma]^{\frac{1}{\sigma}} \tag{4.10}$$

and

$$V(w) = [(1-\delta)c_{NSC}^\sigma + \delta U(R(w-c_{NSC}))^\sigma]^{\frac{1}{\sigma}} \tag{4.11}$$

such that c_{NSC} solves

$$\max_{c\in[0,w]} [(1-\delta)c^\sigma + \beta\delta U(R(w-c))^\sigma]^{\frac{1}{\sigma}}.^{57} \tag{4.12}$$

Note that multiplying utility from current consumption by $(1-\delta)$ in Equations (4.10) and (4.11) is a convenient normalization in the infinite-horizon context as it ensures that self-control/commitment utility is measured in units of stationary consumption: $U(\bar{c}, \bar{c}, \dots) = \bar{c}$.[58] The term $(1-\delta)$ appears in Equation (4.12) as well to ensure that $\beta\delta$ is the Marginal Rate of Intertemporal Substitution (hence may be interpreted as the rate of time preference) in the no-self-control problem.

Under conditions we identify in the Appendix, the solution to the infinite-horizon problem is simple and mirrors the two-period case discussed above: when the decision maker is relatively poor, little is at stake financially and she optimally decides to forego self-control. This leads to over-consumption and under-saving, thus running down her wealth stock more. In turn, this makes self-control even less attractive in the future and the pattern repeats. As a consequence, wealth decreases and diminishes asymptotically. The decision maker is stuck in a poverty trap: Insufficient wealth makes self-control unattractive and lack of self-control leads to lower wealth (due to over-consumption). On the other hand, when decision makers are sufficiently wealthy, the stakes are high and self-control is op-

[57] That is, in every period t, preferences over consumption c_t and continuation problems A_{t+1} given self-control (u_t) and given no self-control (v_t) are represented by: $u_t(c_t, A_{t+1}) = [(1-\delta)c_t^\sigma + \delta U_{t+1}(A_{t+1})^\sigma]^{\frac{1}{\sigma}}$ and $v_t(c_t, A_{t+1}) = [(1-\delta)c_t^\sigma + \beta\delta U_{t+1}(A_{t+1})^\sigma]^{\frac{1}{\sigma}}$.

[58] By slight abuse of notation, we write $U(\bar{c}, \bar{c}, \dots)$ to denote the utility derived from being committed to the stationary consumption stream $(\bar{c}, \bar{c}, \dots)$. More formally, let $A_{\bar{c}}$ denote the (degenerate) menu that commits to current consumption $\bar{c}$ and continuation problem $A_{\bar{c}}$. Then $U(A_{\bar{c}}) = [(1-\delta)c^\sigma + \delta U(A_{\bar{c}})^\sigma]^{\frac{1}{\sigma}}$; thus, $U(A_{\bar{c}}) = \bar{c} = U(\bar{c}, \bar{c}, \dots)$. Since there are no self-control decisions to be made, we have $U(\bar{c}, \bar{c}, \dots) = V(\bar{c}, \bar{c}, \dots) = W(\bar{c}, \bar{c}, \dots)$.

timal. In turn, this induces them to save enough to make self-control worthwhile in the future. As a marginal dollar of wealth is used sub-optimally when not self-controlling, the value function is kinked upwards at a critical wealth level. That is, there exist some $a > 0$ and $b_{SC} > b_{NSC} > 0$ such that

$$U(w) = \begin{cases} b_{NSC} \cdot w & \text{if } w < \bar{w} \\ -a + b_{SC} \cdot w & \text{if } w \geq \bar{w} \end{cases} \tag{4.13}$$

for $\bar{w} = \frac{a}{b_{SC} - b_{NSC}}$.

Here, a captures the cumulative cost of exerting self-control in the current and all future periods ($a = \frac{k}{1-R^{-1}}$, cf. Equation (A.3) in the Appendix). For $w > \bar{w}$, the marginal utility of wealth, b_{SC}, is equal to that of a standard decision maker without self-control problems.[59] The fact that the marginal utility of wealth is lower for poor decision makers ($b_{NSC} < b_{SC}$) reflects the fact that they over-consume (under-save) due to a lack of self-control.

While the kink introduces a non-differentiability at $w = \bar{w}$, U is differentiable everywhere else. Moreover, at $w = \bar{w}$ the right and left derivatives exist and are given by b_{SC} and b_{NSC} respectively. As $b_{SC} > b_{NSC}$, it is clear that $R(w - c) = \bar{w}$ for no solution c to Problems (4.10) and (4.12). Intuitively, for every c such that $R(w - c) = \bar{w}$, the fact that $b_{SC} > b_{NSC}$ creates a wedge between the marginal utility of a dollar saved, $\left[-(1-\delta)c^{\sigma-1} + \delta U(\bar{w})^{\sigma-1} b_{SC} R\right]((1-\delta)c^{\sigma} + \delta U(\bar{w})^{\sigma})$, and that of a dollar consumed, $\left[(1-\delta)c^{\sigma-1} - \delta U(\bar{w})^{\sigma-1} b_{NSC} R\right]((1-\delta)c^{\sigma} + \delta U(\bar{w})^{\sigma})$. As at least one of them must be positive, the decision maker is better off by consuming slightly more or less. In particular, this implies that any optimal solution c^* to Problems (4.10) and (4.12) satisfies a First-Order Condition (FOC). However, potentially, we may need to consider two candidate FOCs; one for $R(w - c) > \bar{w}$ and one for $R(w - c) < \bar{w}$. Thus, for current wealth levels close to $\bar{w}$, there are two candidate solutions; one for which the DM saves enough to reach wealth levels that allow for self-control in the future and one where she saves less and ends up with future wealth that entails optimally foregoing self-control. On the other hand, when w is small ($w \gtrapprox 0$) or very large ($w \gg \bar{w}$), one of the FOCs does not have a solution. Intuitively speaking, very poor decision makers never save enough to pass $\bar{w}$ tomorrow while very rich decision makers never consume so much so as to fall below $\bar{w}$.

To analyze this in more detail, note that the general FOC to Problems (4.10) and (4.12) is given by

$$(1-\beta)c^{\sigma-1} = \xi U(R(w-c))^{\sigma-1} U'(R(w-c))R$$

where $\xi = \delta$ (for Equation (4.10)) or $\xi = \beta\delta$ (for Equation (4.12)). Given Equation (4.13) for U, this results in the following two candidate FOCs

$$c^{\sigma-1} = \frac{\xi}{1-\delta}(b_{NSC} R(w-c))^{\sigma-1} b_{NSC} R \text{ where } R(w-c) < \bar{w} \tag{4.14}$$

[59] That is, the solution to the benchmark problem where $k = 0$ is given by $V(w) = b_{SC} \cdot w$.

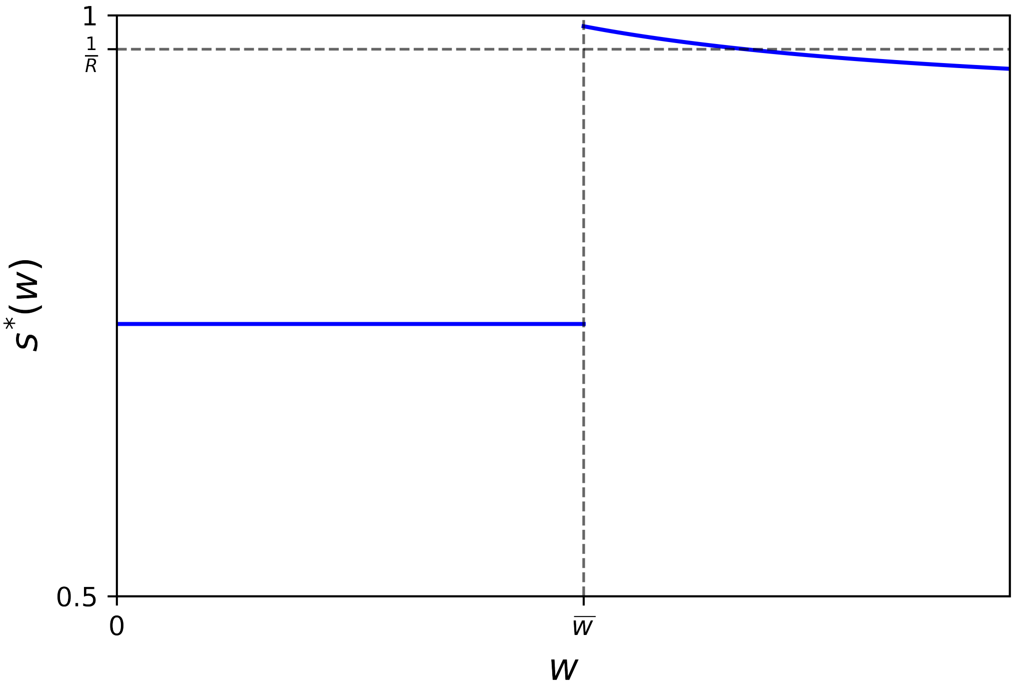

Figure 4.1: Optimal savings rate for $\beta = 0.1$, $\delta = 0.9$, $\gamma = 0.8$, $k = 0.1$ and $R = 1.03$

and

$$c^{\sigma-1} = \frac{\xi}{1-\delta}(-a + b_{SC}R(w-c))^{\sigma-1} b_{SC}R \text{ where } R(w-c) > \bar{w}. \tag{4.15}$$

As the solution to (4.14) entails a future level of wealth in the No-Self-Control region of U, we denote its solution by using the subscript NSC (and, likewise, the subscript SC for the solution to (4.15)). Note, however that this refers to the *future* exertion of self-control. The solutions depend on the current exertion of self-control only through their dependence on the discount parameter ($\xi = \delta$ given self-control and $\xi = \beta\delta$ given no self-control). To make this explicit, we denote the candidate solutions to (4.10) and (4.12) by $c_{\xi,NSC}$ and $c_{\xi,SC}$.

Solving Equations (4.14) and (4.15) for c and noting that $1/(\sigma - 1) = -\gamma$, we obtain

$$c_{\xi,NSC} = \mu(\xi, b_{NSC}) \cdot w \text{ and } c_{\xi,SC} = \mu(\xi, b_{SC}) \cdot \left[-\frac{a}{b_{SC}R} + w.\right] \tag{4.16}$$

where, for all $0 < \xi \leq 1$ and all $b > 0$,

$$\mu(\xi, b) := \frac{(\frac{\xi}{1-\delta})^{-\gamma}(bR)^{1-\gamma}}{1 + (\frac{\xi}{1-\delta})^{-\gamma}(bR)^{1-\gamma}} \tag{4.17}$$

denotes the Marginal Propensity to Consume (MPC) depending on discount factor ξ and marginal utility of (future) wealth b.

As we show in the Appendix, the optimal consumption policy $c^*(w)$ is such that (i) the

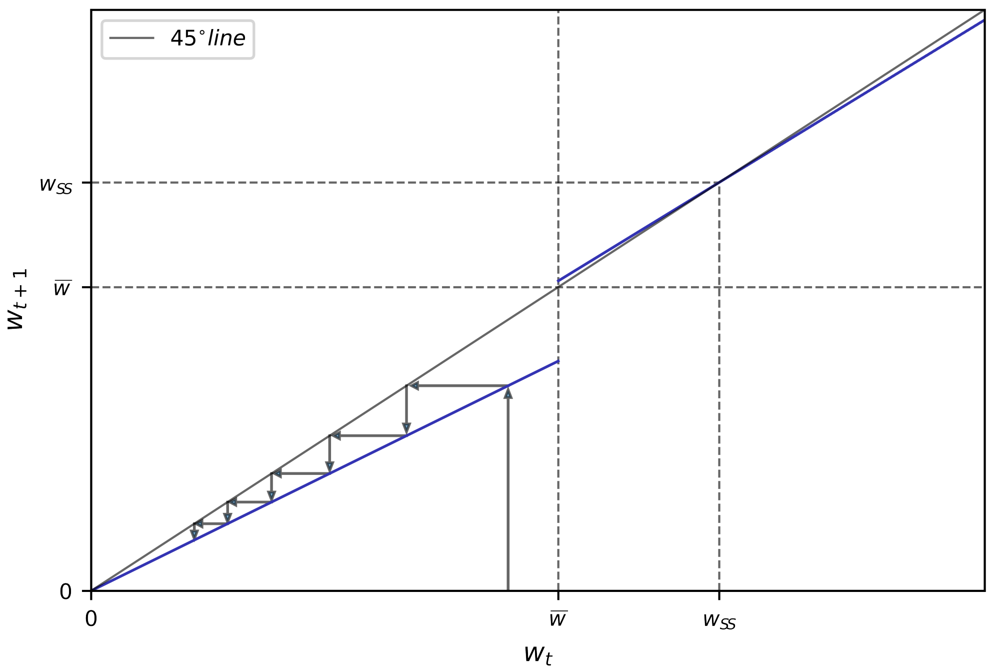

Figure 4.2: Wealth dynamics for $\beta = 0.1$, $\delta = 0.9$, $\gamma = 0.8$, $k = 0.1$ and $R = 1.03$

decision maker exerts self-control if and only if $w \geq \bar{w}$ (ii) under self-control, the solution is given by $c_{\delta,SC}$ (i.e. such that $R(w - c) > \bar{w}$) (iii) under no self-control, the solution is given by $c_{\beta\delta,NSC}$ (i.e. such that $R(w - c) < \bar{w}$). That is,

$$c^*(w) = \begin{cases} \mu(\beta\delta, b_{NSC}) \cdot w & \textit{if } w < \bar{w} \\ \mu(\delta, b_{SC}) \cdot \left(-\frac{a}{b_{SC}R} + w\right) & \textit{if } w \geq \bar{w} \end{cases}.$$

The optimal savings rate $s^*(w) = 1 - \frac{c^*(w)}{w}$ is given by

$$s^*(w) = \begin{cases} 1 - \mu(\beta\delta, b_{NSC}) & \text{if } w < \bar{w} \\ \underbrace{1 - \mu(\delta, b_{SC})}_{\substack{\text{savings rate} \\ \text{w/o SC problem}}} + \underbrace{\mu(\delta, b_{SC}) \frac{a}{b_{SC}R} \frac{1}{w}}_{\text{excess savings}} & \text{if } w \geq \bar{w} \end{cases}.$$

Note that the savings rate in the benchmark problem ($k = 0$) is $\frac{1}{R}(\delta R)^{\gamma} = 1 - \mu(\delta, b_{SC})$.[60] So the need to self-control creates excess savings. As the amount of additional savings is fixed, the savings rate is decreasing for $w \geq \bar{w}$ and asymptotes to the benchmark level as $w \to +\infty$. Figure 4.1 depicts the savings rate for our exemplary parameter combination.

[60] As we show in the Appendix, we have $b_{SC} = (1 - \delta)^{-\frac{\gamma}{1-\gamma}} \left[1 - R^{-1}(R\delta)^{\gamma}\right]^{\frac{1}{1-\gamma}}$. Consequently, $1 - \mu(\delta, b_{SC}) = \frac{1}{R}(\delta R)^{\gamma}$.

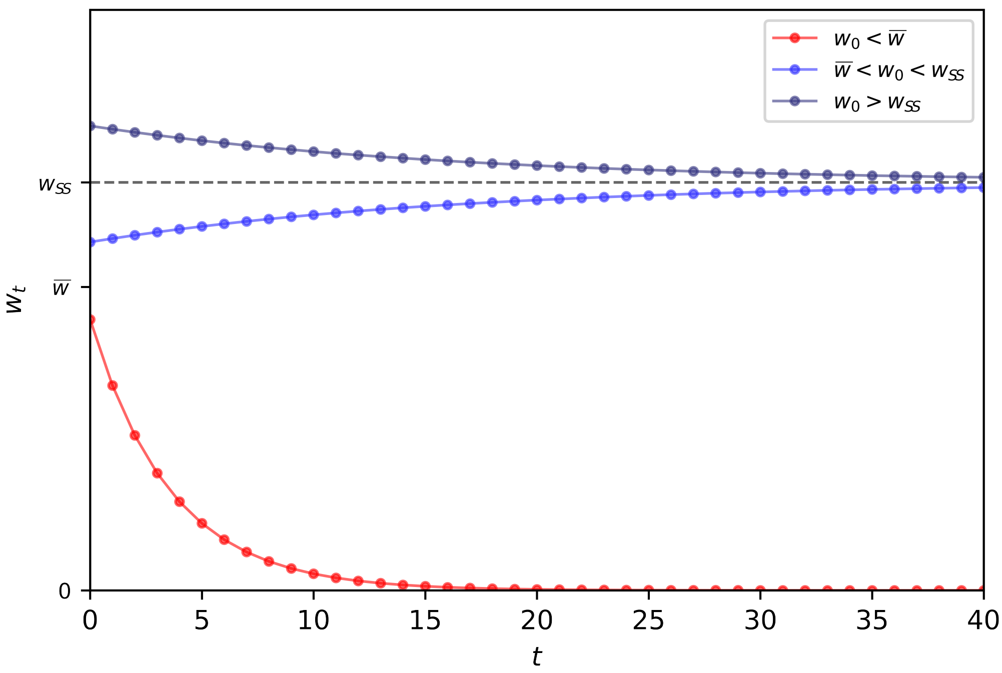

Figure 4.3: Wealth paths for $\beta = 0.1$, $\delta = 0.9$, $\gamma = 0.8$, $k = 0.1$, $R = 1.03$ and three initial wealth levels w_0

As $1 - \mu(\delta, b_{NSC}) < 1 - \mu(\beta, b_{SC}) \leq \frac{1}{R}$ for $\delta R \leq 1$,[61] the model implies poverty traps for the poor. Decision makers whose initial wealth is low ($w < \bar{w}$) run down their wealth exponentially at rate $R(1 - \mu(\delta, b_{NSC})) < 1$. On the other hand, affluent decision makers ($w \geq \bar{w}$) are not be prone to asymptotically diminishing wealth. If $\delta R < 1$, their wealth converges to the steady state level

$$w_{SS} = \frac{R - (\delta R)^\gamma}{1 - (\delta R)^\gamma} \frac{a}{b_{SC} R}.$$

For our exemplary parameter combination Figure 4.2 illustrates the optimal wealth dynamics $w^*_{t+1}(w_t)$; Figure 4.3 depicts the optimal wealth paths for initial wealth levels below $\bar{w}$ (blue), above $\bar{w}$ but below w_{SS} (orange) and above w_{SS} (green).

In steady state, we have

$$c_{SS} = c^*(w_{SS}) = \frac{k}{b_{SC} R} \frac{R - (\delta R)^\gamma}{1 - (\delta R)^\gamma}.^{62}$$

If $\delta R = 1$, affluent decision makers keep on accumulating wealth without converging to a

[61] See the derivation following Equation (A.8) in the Appendix.

[62] Thus, steady state consumption is increasing in the self-control cost k. However, note that, as $\bar{w} = \frac{k}{1 - \frac{1}{R}} \frac{1}{b_{SC} - b_{NSC}}$ (and b_{SC}, b_{NSC} are independent of k), the initial wealth level required to converge to this steady state increases as well.

finite steady state level. However, asymptotically, their wealth grows at a diminishing rate (i.e. the gross growth rate approaches 1 as $w \to +\infty$).

5 Conclusion

We have shown three simple and intuitive Axioms on dynamic choice behavior in a simple two-period setting to be equivalent to a decision maker who is strategically Planning to Self-Control. Planning is optimal to the effect that, for every decision problem (menu), it maximizes planning-stage utility of the choice it induces, net of a planning cost. Planning to Self-Control is behaviorally equivalent to model with self-control costs that allows for more general specifications of the cost function that have not been consider in other axiomatic models of self-control. Under a particularly simple and intuitive such cost specification, a fixed cost, self-control is increasing in the stakes of the problem. When applied to intertemporal problems faced by a present-biased decision maker, this produces a magnitude effect: the empirical finding that reverse their preferences for smaller sooner rewards vs. larger later ones when both rewards are scaled up. In a simple consumption-savings problem, increasing self-control means that self-control is exerted only by decision makers with sufficiently high wealth. The poor forego self-control and over-consume, thus running down their wealth even more. This results in a poverty trap.

Part III

Aggregation of Choice Functions

6 Introduction

At the heart of social choice theory lies Arrow's preference aggregation problem. Indeed, it is fair to say that Arrow's (1951/63) formulation of the problem of collective choice as one of the aggregation of individual preference orderings into a collective preference ordering has shaped most of the social choice literature and the way researchers have approached the problem. However, it is also widely recognized that, ultimately, preferences are only a representation of choice behavior and that therefore choice functions are the more fundamental object, at least from a behavioral perspective, see, for example, Wakker (1989). And in fact, already Arrow recognized that the requirement that a group having a collective preference is *prima facie* a strong assumption, and that it would be sufficient that a group be able to make collective *choices*. But, famously, he also showed that if these choices were consistent across menus in the sense of the Weak Axiom of Revealed Preference (WARP) they could be represented by a collective preference (Arrow, 1959). Thus, it seemed that the transition to the more fundamental and more general choice-function framework would not gain any mileage.

But WARP is obviously a strong condition and there is a large body of literature addressing various relaxations, starting with the seminal work by Gibbard (1969/2014), Sen (1969) and Mas-Colell and Sonnenschein (1972). Usually, the relaxations studied in the literature concern the consistency, or rationality, requirements imposed on collective choices, since the expansion of the *range* of an aggregator gives hope to escape the impossibility results by which social choice theory has been plagued. By contrast, the expansion of the *domain* of an aggregator (i.e., here, the set of individual preference profiles) seems to exacerbate the aggregation problem. But the interplay between the consistency (rationality) requirements at the individual and collective level is an intricate matter, and a natural starting point is to assume the *same* conditions at the individual and the collective level. This is what we do below.

The various relaxations of rationality considered in the literature are best described by their implications on choice behavior. Therefore, the most natural and general framework to study them is one of the aggregation of choice functions. Assuming the same choice consistency conditions at the individual and the collective level allows one to study the problem within the general *judgement aggregation* model developed by List and Pettit (2002), Dietrich (2007), Dietrich and List (2007), Nehring and Puppe (2002, 2010) and Dokow and Holzman (2010), among others. Within this model, a choice function can be described by specifying, for each menu A and each alternative $x \in A$, if x is choosable from A or not. This formulation lends itself to a natural *independence* requirement inspired by

Arrow's condition: that the collective decision whether or not x is choosable from menu A should depend only on the individual views about *this* issue. This condition is stronger than Arrow's original condition which (effectively) imposes independence for pairwise preference statements. As the latter correspond to choice behavior from binary menus, our stronger condition is even more systematic and, arguably, more natural.

Overview of Results

Evidently and unsurprisingly, the stronger independence condition cannot help escaping the Arrow impossibility. Indeed, our first main result shows that even on the domain of all choice functions that can be rationalized by some acyclic binary relation our choice independence condition implies dictatorship. This is in contrast to the case of Arrow's binary independence condition. The result may thus suggest that the 'non-impossibility' results by Gibbard (1969/2014), Sen (1969) and Mas-Colell and Sonnenschein (1972) are somewhat artificial and hinge on a particular way to formalize independent aggregation. Of course, this is not to deny the justification of weakening independence as much as possible in order to obtain possibility results. Our result simply hints at the fact that a natural and systematic formulation of choice independence already rules out all non-dictatorial aggregation rules under acyclic binary rationality (and thus *a forteriori* also under quasi-transitive rationality).

On the positive side, we show that path-independent choice functions (Plott, 1973) can be aggregated into a path-independent collective choice function under choice independence in a non-dictatorial fashion. While all such aggregation rules are oligarchic, they admit a very natural and simple interpretation in terms of *multiple selves*: It is well-known that a choice function is path-independent if and only if can be rationalized in terms of the union of the maximal elements of a finite collection of linear orderings (Aizerman and Malishevski, 1981). Interpreting these orderings as representing the multiple selves of the decision-maker, the choice independent, oligarchic aggregation rules can be simply described as taking the collection of all individual multiple selves as the multiple 'selves' of the group.

We then investigate if further relaxations of the rationality requirements on the choice functions enable possibility results and we characterize the anonymous aggregation rules ('quota rules') that satisfy a weak additional neutrality condition.

Finally, we show that stronger possibilities emerge by relaxing the general choice independence condition in a natural way to a condition of *independence across menus*. In particular, approval voting satisfies this condition while aggregation based on scoring rules does not.

Relation to the Literature

To the best of our knowledge, there is very little literature that explicitly addresses the aggregation of choice functions. An exception is Shelah (2005) who provides a very general impossibility result on the aggregation of singleton-valued choice functions under a symmetry condition but without any further rationality assumptions. However, this interesting and intricate result is unrelated to our present analysis since we do not assume single-valuedness, and this assumption is crucial in Shelah (2005).

7 The Model

Let X be a (possibly infinite) set of alternatives $x, y, z \in X$. Denote by $\mathcal{A} = 2^X \backslash \{\emptyset\}$ the collection of all non-empty decision problems (menus) $A \in \mathcal{A}$. A choice function (CF) is a mapping $c : \mathcal{A} \rightrightarrows X$ such that for all $A \in \mathcal{A}$, $c(A) \subseteq A$. Note that, in general, this allows for choice sets $c(A)$ to contain multiple alternatives.[63]

Let $\mathcal{C} = \{c : c \text{ is a CF}\}$ be the collection of all choice functions (on X) and $n \in \mathbb{N}$, $n \geq 2$, be the number of individuals in a group/society. An aggregation function $f : \mathcal{C}^n \to \mathcal{C}$ maps every *profile* of choice functions $(c_1, \ldots, c_n)$ into a social choice function $c = f(c_1, \ldots, c_n)$. We say that f is *consistent* on *domain* $\emptyset \neq \mathcal{D} \subseteq \mathcal{C}$ if $f(\mathcal{D}^n) \subseteq \mathcal{D}$.[64] Below, we consider aggregation on (sub-)domains $\mathcal{D}$ obtained by imposing (consistency) conditions on the choice functions under consideration.

As a minimal restriction, we demand that choice sets be non-empty when choosing from finite menus.

Finite Non-Emptiness:

$$|A| < \infty \implies c(A) \neq \emptyset. \tag{FNE}$$

We denote by $\mathcal{C}_{\text{fne}} = \{c \in \mathcal{C} : c \text{ satisfies (FNE)}\}$ the collection of all choice functions satisfying finite non-emptiness.

7.1 Rationalizable Choice Functions

Of particular interest in economics are choice functions that are *rationalizable* as maximal elements of some relation(s) $R \subseteq X \times X$. For any such R, call P its asymmetric component (that is $xPy \iff (xRy \& \neg yRx)$) and I its symmetric component ($xIy \iff (xRy \& yRx)$). We say that R is:

1. *acyclic* if P does not contain a cycle;

2. *quasi-transitive* if P is transitive;

3. a *weak order* if R is complete and transitive;

4. a *linear order* if R is an asymmetric weak order.

---

[63] Some authors refer to the general concept as a 'choice correspondence' and reserve the name 'choice function' for choice correspondences that contain (at most) a single alternative. In Part II of this Thesis, this assumption of singleton-valuedness is built into the definition of a choice function.

[64] In other words, when studying consistent aggregation functions on $\mathcal{D}$, it is without loss of generality to restrict the *co-domain* to $\mathcal{D}$ as well. Therefore, for the rest of Part III, we simply refer to $\mathcal{D}$ as the domain.

Note that these properties are ordered from weak to strong. That is, every linear order is a weak order, every weak order is quasi-transitive and every quasi-transitive R is acyclic.[65] The latter is a minimal condition if R is to rationalize a choice function $c \in \mathcal{C}_{\text{fne}}$ in terms of maximal elements seeing that $\max_A R := \{x \in A : yPx \text{ for no } y \in A\}$ is non-empty for all finite A if and only if R is acyclic.

A by now extensive literature on choice theory has characterized notions of rationalizability in terms of consistency requirements on choice functions. As a widely known example of such a result, recall that a choice function is rationalized by some weak order if and only if it satisfies the *Weak Axioms of Revealed Preference* (WARP). We list some of the most important and well-known conditions on choice functions here.

Weak Axiom of Revealed Preference (WARP):

$$\text{if } \{x, y\} \subseteq A \cap B, \text{ then } x \in c(A) \implies (y \in c(B) \implies x \in c(B)). \tag{WARP}$$

Contraction (Sen's α/Chernoff):

$$c(A \cup B) \cap A \subseteq c(A). \tag{α}$$

Expansion (Sen's β):

$$c(A \cup B) \cap A \neq \emptyset \implies c(A) \subseteq c(A \cup B). \tag{β}$$

γ-Expansion:

$$c(A) \cap c(B) \subseteq c(A \cup B). \tag{γ}$$

Aizerman-Expansion:

$$c(A \cup B) \subseteq A \implies c(A) \subseteq c(A \cup B). \tag{Aiz}$$

Path-Independence:

$$c(A \cup B) = c(c(A) \cup B). \tag{PI}$$

In general, WARP is equivalent to Contraction and Expansion. An interesting sub-case arises when c is singleton-valued (i.e., $|c(A)| = 1$ for all non-empty $c(A) \neq \emptyset$). In this case, WARP reduces to Contraction. Moreover, Expansion is stronger than both γ-Expansion and Aizerman-Expansion. We summarize in the following Lemma. A proof can be found, for example, in Moulin (1985).

Lemma 1. *Let $c \in \mathcal{C}_{fne}$.*

1. *c satisfies* (WARP) *if and only if it satisfies* (α) *and* (β).

[65] As an example of an acyclic relation that is not quasi-transitive, consider $R \subset \{x, y, z\}$ such that $xPyPzIx$. On the other hand, $zIyIxPy$ is both acyclic and quasi-transitive (but not a weak/linear order).

2. *c satisfies* (β) *only if it satisfies* (γ) *and* (Aiz).

3. *If* $|c(A)| \leq 1$ *for all* $A \in \mathcal{A}$*, then c satisfies* (α) *only if it satisfies* (β).

4. *c satisfies* (PI) *if and only if it satisfies* (α) *and* (Aiz).

Based on Moulin (1985), the following Lemma summarizes important results about the rationalizability of choice functions in the literature.

Lemma 2. *Let* $c \in \mathcal{C}_{fne}$.

1. *If and only if c satisfies* (α) *it is sub-rationalizable by some linear order* $R \subseteq X \times X$*:* $\max_A R \subseteq c(A)$ *for all* $A \in \mathcal{A}$.

2. *If and only if c satisfies* (α) *and* (γ) *it is rationalizable by some complete and acyclic* $R \subseteq X \times X$*:* $c(A) = \max_A R$ *for all* $A \in \mathcal{A}$.

3. *If and only if c satisfies* (α) *and* (Aiz) *it is pseudo-rationalizable: there exist linear orders* $R_1, \ldots, R_k \subseteq X \times X$ *such that* $c(A) = \bigcup_{j=1,\ldots,k} \max_A R_j$ *for all* $A \in \mathcal{A}$.

4. *If and only if c satisfies* (α), (γ) *and* (Aiz) *it is rationalizable by some complete and quasi-transitive* $R \subseteq X \times X$*:* $c(A) = \max_A R$ *for all* $A \in \mathcal{A}$*. Moreover, there exist linear order* $R_1, \ldots, R_k$ *such that R may be written as their Pareto relation:* $R_{Par} = \bigcap_{j=1,\ldots,k} R_j \cup \{(x,y) \in X \times X : (x,y),(y,x) \notin \bigcap_{j=1,\ldots,k} R_j\}$.

5. *If and only if c satisfies* (WARP) *it is rationalizable by some linear order* $R \subseteq X \times X$*:* $c(A) = \max_A R$ *for all* $A \in \mathcal{A}$.

The above results inform the following definitions:

- $\mathcal{C}_{\text{sub}} := \{c \in \mathcal{C}_{\text{fne}} : c \text{ is sub-rationalizable}\}$,
- $\mathcal{C}_{\text{acy}} := \{c \in \mathcal{C}_{\text{fne}} : c \text{ is rationalizable by some complete and acyclic } R \subseteq X \times X\}$,
- $\mathcal{C}_{\text{psd}} := \{c \in \mathcal{C}_{\text{fne}} : c \text{ is pseudo-rationalizable}\}$,
- $\mathcal{C}_{\text{Par}} := \{c \in \mathcal{C}_{\text{fne}} : c \text{ is Pareto-rationalizable}\}$,
- $\mathcal{C}_{\text{wo}} := \{c \in \mathcal{C}_{\text{fne}} : c \text{ is rationalizable by some weak order } R \subseteq X \times X\}$,
- $\mathcal{C}_{\text{lo}} := \{c \in \mathcal{C}_{\text{fne}} : c \text{ is rationalizable by some linear order } R \subseteq X \times X\}$.

Note that $\mathcal{C}_{\text{lo}} \subseteq \mathcal{C}_{\text{wo}} \subseteq \mathcal{C}_{\text{Par}} \subseteq \mathcal{C}_{\text{psd}}, \mathcal{C}_{\text{acy}} \subseteq \mathcal{C}_{\text{sub}} \subseteq \mathcal{C}_{\text{fne}}$.

7.2 Social Choice as Aggregation of $x \in c(A)$ Judgments

When aggregating a profile of choice functions into a *social* choice function, society is faced, for every menu $A \in \mathcal{A}$, $|A| \geq 2$ and every alternative $x \in A$, with the issue of whether it should (collectively) choose x from A.[66] More generally, both at the individual and at the collective level, a choice function may be thought of as a set of judgments on an agenda consisting of all such issues $x \in c(A)$.

Imposing consistency conditions on choice functions then translates into restrictions on what sets of judgments are considered feasible. For example, imposing (α) is inconsistent with simultaneously judging $x \in c(A)$ to be true and $x \in c(B)$ to be false when $x \in B \subset A$. Thus, consistency conditions on choice functions translate into logical dependencies (entailments) between issues. This approach allows us to employ the machinery developed in Nehring and Puppe (2002, 2010) to study what domains $\mathcal{D}$ entail dictatorial aggregation rules and when and what kind of non-dictatorial aggregation is possible.

Moreover, it suggests a natural notion of independence. Say that an aggregation rule is independent if the collective decision on whether some $x \in A$ is chosen from A may only depend on individual judgments on *this* issue alone. In other words, if two profiles of individual choice functions agree in terms of individual judgments on $x \in c_i(A)$, $i = 1, \ldots, n$, then the collective decision on whether $x \in c(A)$ must be the same for both profiles.

Independence: Consider any $A \in \mathcal{A}$ and $x \in A$. Let $c = f(c_1, \ldots, c_n)$ and $c' = f(c'_1, \ldots, c'_n)$. If, for all $i = 1, \ldots, n$, $x \in c_i(A) \iff x \in c'_i(A)$, then $x \in c(A) \iff x \in c'(A)$.

In the presence of independence, it is natural to demand that aggregation happen in a monotone fashion. That is to say that increased support among all individuals in favor of some choice $x \in c(A)$ should lead to it being chosen socially if it was before. Vice versa if more individuals reject $x \in c(A)$, x must not be chosen socially if it wasn't before. Thus, we strengthen independence to the following condition.

Monotone Independence: Consider any $A \in \mathcal{A}$ and $x \in A$. Let $c = f(c_1, \ldots, c_n)$ and $c' = f(c'_1, \ldots, c'_n)$. (i) If, for all $i = 1, \ldots, n$, $x \in c_i(A) \implies x \in c'_i(A)$, then $x \in c(A) \implies x \in c'(A)$; (ii) if for all $i = 1, \ldots, n$, $x \notin c_i(A) \implies x \notin c'_i(A)$, then $x \notin c(A) \implies x \notin c'(A)$.

As an example of a monotone and independent aggregation rule consider the social choice function resulting from *majority voting* on all issues. That is, we define $f_{maj} : \mathcal{C}^n \to \mathcal{C}$ such that, for all $A \in \mathcal{A}$, $x \in f_{maj}(c_1, \ldots, c_n)(A) := c_{maj}(A) \iff \frac{1}{n}|\{i : x \in c_i(A)\}| \geq n/2$.[67] However, as we show below, majority voting is inconsistent in general, even when only imposing (FNE).

[66] If A is a singleton, say $A = \{x\}$, the issue is trivial seeing that x must be chosen given (FNE).

[67] Note that, given this definition, we break ties in favor of inclusion of alternatives in the collective choice set.

Our notion of independence is stronger than that of Arrow. As the latter is an assumption on aggregation of weak orders, in order to compare the two, we must restrict attention to the domain $\mathcal{D} = \mathcal{C}_{\text{wo}}$. While Arrow-Independence imposes the independence concept for pairs of issues $(x \in c(\{x, y\}), y \in c(\{x, y\}))$[68], our notion of independence applies for each issue individually. Thus, it is stronger than Arrow Independence.

In analogy to Arrow's weak Pareto condition, we impose a weak unanimity assumption.

Unanimity: For all $c \in \mathcal{C}^n$, $f(c, \ldots, c) = c$.

While unanimity only requires that if all individual choice functions are the same, this be the collective choice function. In the presence of monotone independence it is equivalent to the stronger notion of *issue-wise* unanimity. That is, for all $A \in \mathcal{A}$ and all $x \in A$, if $x \in c_i(A)$ for all $i = 1, \ldots, n$, then $x \in c(A) = f(c_1, \ldots, c_n)(A)$.

We call an aggregation function satisfying unanimity and monotone independence an *Arrowian* aggregator.

[68] Arrow Independence requires that for any two profiles of weak orders $R, \ldots, R_n$ and $R'_1, \ldots, R'_n$ giving rise to the social orders R and R' respectively, (i) if, for all $i = 1, \ldots, n$, $xP_iy \iff xP'_iy$, then $xPy \iff xP'y$; (ii) if, for all $i = 1, \ldots, n$, $xI_iy \iff xI'_iy$, then $xIy \iff xI'y$. Note that on $\mathcal{C}_{\text{wo}}$, for all $x, y \in X$, $xPy \iff (x \in c(\{x, y\}) \& y \notin c(\{x, y\}))$ and $xIy \iff (x \in c(\{x, y\}) \& y \in c(\{x, y\}))$

8 (Im)Possibility of Arrowian Aggregation

Here we address the question of whether consistent monotone independent aggregation is possible on the different domains $\mathcal{D} \subseteq \mathcal{C}_{\text{fne}}$ defined above. Note that as the domain $\mathcal{D}$ is restricted further, aggregation needs to be consistent for a smaller set of individual choice functions (as more requirements are imposed) but, at the same time, needs to satisfy more stringent consistency conditions at the collective level.

From a majoritarian perspective, the space of possibilities is limited from the outset by the observation that ($x \in c(A)$-wise) majority voting is inconsistent on $\mathcal{D} = \mathcal{C}_{\text{fne}}$ (i.e. even without imposing further restrictions on choice functions) except for the special cases of three alternatives and two or four individuals.

Proposition 1. *Let $|X| \geq 3$. f_{maj} is consistent on $\mathcal{D} = \mathcal{C}_{fne}$ if and only if $|X| = 3$ and $n \in \{2, 4\}$.*

For example, consider the case of $n = 3$ individuals and let $A = \{x, y, z\} \subseteq X$. Consider a profile of choice functions $c_1, c_2, c_3 \in \mathcal{C}_{\text{fne}}$ such that $c_1(A) = \{a\}$, $c_2(A) = \{b\}$ and $c_3(A) = \{c\}$. Note that $f_{maj}(A) = \emptyset$. Thus, $c_{maj} = f_{maj}(c_1, c_2, c_3)$ does not satisfy (FNE), i.e., $f_{maj}(c_1, c_2, c_3) \notin \mathcal{C}_{\text{fne}}$.

When choice functions are rationalizable by some acyclic relation (satisfy contraction property (α) and extension property (γ)) all consistent Arrowian aggregators are necessarily dictatorial. The same holds when more (stringent) consistency conditions are imposed (quasi-transitive, weak or linear order rationalizable choice functions).

Theorem 3. *Every Arrowian aggregation rule f is consistent on $\mathcal{D} = \mathcal{C}_{acy}$ if and only if it is a dictatorship; that is there is some $j \in \{1, \dots, n\}$ such that for all $(c_1, \dots, c_n) \in \mathcal{C}^n_{acy}$:*

$$f(c_1, \dots, c_n) = c_j.$$

The same holds for $\mathcal{D} = \mathcal{C}_{Par}, \mathcal{C}_{wo}, \mathcal{C}_{lo}$.

On the other hand, when considering extension property (Aiz) instead of (γ) alongside (α), i.e. for path-independent choice functions pseudo-rationalizable by multiple linear orders ('selves'), possibilities emerge. While all consistent Arrowian aggregators are oligarchic, this allows for unanimity rule: for all $A \in \mathcal{A}$ and $x \in A$, $x \notin c(A)$ if and only if x is unanimously rejected by all individuals.

Theorem 4. *Every Arrowian aggregation rule f is consistent on $\mathcal{D} = \mathcal{C}_{psd}$ if and only if it is an oligarchic choice rule; that is, there exists some $\emptyset \neq M \subseteq \{1, \ldots, n\}$ such that, for all $(c_1, \ldots, c_n) \in \mathcal{C}^n_{psd}$ and all $A \in \mathcal{A}$:*

$$f(c_1, \ldots, c_n)(A) = \bigcup_{i \in M} c_i(A).$$

Note that if all c_i are pseudo-rationalized by $R^i_1, \ldots, R^i_{k_i}$, the collection of all 'selves' in society $\bigcup_{i=1,\ldots,n}\{R^i_1, \ldots, R^i_{k_i}\}$ allows one to define a pseudo-rationalizable collective choice function in the obvious fashion. The contribution of Theorem 4 is to show that this choice function can be enacted through Arrowian issue-wise aggregation; namely by letting $M = \{1, \ldots, n\}$ (which yields the unanimity rule). More generally, for every $M \subseteq \{1, \ldots, n\}$, the corresponding oligarchic rule yields the collective choice function pseudo-rationalized by $\bigcup_{i \in M}\{R^i_1, \ldots, R^i_{k_i}\}$.

Two important remarks are in order. First, the unanimity rule (more generally, every oligarchic rule) is Arrowian for all domains $\mathcal{D}$ we consider here. Yet as Theorem 3 shows, it is not consistent on $\mathcal{D} = \mathcal{C}_{\text{acy}}, \mathcal{C}_{\text{Par}}, \mathcal{C}_{\text{wo}}, \mathcal{C}_{\text{lo}}$. This is due to the fact that collective choice derived from the unanimity rule violates property (γ). For example, consider $X = \{a, b, c\}$ and two individuals $a \succ_1 b \succ_1 c$ and $c \succ_2 b \succ_2 a$. Unanimity rule yields $c(\{a, b\}) = \{a, b\}$ and $c(\{b, c\}) = \{b, c\}$ hence $b \in c(\{a, b\}) \cap c(\{b, c\})$. At the same time, $b \notin \{a, c\} = c(\{a, b, c\})$, in violation of property (γ).[69]

Second, the Pareto rule, according to which all alternatives are choosable from a given menu that are not strongly Pareto-dominated by some other alternative in it, is known to yield acyclic collective choices on $\mathcal{D} = \mathcal{C}_{\text{acy}}$.[70] However, it is not an Arrowian procedure in our model as it violates independence. To see this, reconsider our example for which $a \succ_1 b \succ_1 c$ and $c \succ_2 b \succ_2 a$. As no alternative is strongly Pareto-dominated in $\{a, b, c\}$, we have $b \in \{a, b, c\} = c(\{a, b, c\})$, although $b \notin c_1(\{a, b, c\})$ and $b \notin c_2(\{a, b, c\})$.[71] Now suppose the preference of the second individual changes so that $c \succ'_2 a \succ'_2 b$. As a strongly Pareto-dominates b we have $b \notin c(\{a, b, c\})$. However, the second individual has not actually changed her choice behavior for b in the menu $\{a, b, c\}$ seeing that still $b \notin c'_2(\{a, b, c\})$. Indeed, the second individual has not changed her overall choice behavior from $\{a, b, c\}$ as $\{c\} = c'_2(\{a, b, c\}) = c_2(\{a, b, c\})$. Thus, the Pareto rule even violates the weaker notion of 'independence across menus' which we introduce below and which, as we argue below, best captures Arrow's original notion in our framework. This shows that, while possibilities again emerge in the form of oligarchic rules, these produce choice functions that are pseudo-rationalizable but not guaranteed to be binary (rationalizable by some acyclic relation) in general. On the other hand, the Pareto rule and other 'oligarchic' rules in the sense of

[69] Similar arguments show that collective choices under any oligarchic rule violate (γ).
[70] If individual preferences are quasi-transitive, the Pareto rule yields quasi-transitive social choices (that is, it is consistent on $\mathcal{C}_{\text{Par}}$). If individual preferences are (weak or linear) orders, choosing all alternatives that are not weakly dominated, yields acyclic collective choices as well.
[71] In particular, the Pareto rule is not a unanimity rule in our more general choice function framework.

Gibbard (1969/2014) are not independent in our choice function setting.[72] These findings cast doubt on their status in the literature as representing 'possibilities' for independent aggregation (if one settles for acyclic social preferences).

We now restrict attention to finite X and consider the possibility of non-oligarchic rules (in which no single individual posses veto power) when dropping the extension properties (γ)/(Aiz) (while keeping/dropping (α)). We focus on rules that treat all individuals equally. Thus, the following condition is natural.

Anonymity: Let $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ be a permutation (of individuals). Then $f(c_1, \ldots, c_n) = f(c_{\pi(1)}, \ldots, c_{\pi(n)})$.

Anonymity requires that 'voter's names do not matter' in the sense that the collective choice rule be invariant to permuting (re-labeling) all individuals. That is, as mentioned above, it demands that all individuals be treated equally by the collective choice rule.

Menu-level Neutrality: Consider any $A \in \mathcal{A}$ such that $A = \{x_1, \ldots, x_m\}$ and any permutation $\pi : \{1, \ldots, m\} \to \{1, \ldots, m\}$. Let $c = f(c_1, \ldots, c_n)$ and $c' = f(c'_1, \ldots, c'_n)$. If, for all $i = 1, \ldots, n$ and all $j = 1, \ldots, m$, $x_j \in c_i(A) \iff x_{\pi(j)} \in c'_i(A)$, then for all $j = 1, \ldots, m$, $x_j \in c(A) \iff x_{\pi(j)} \in c'(A)$.

Thus, menu-level neutrality requires that all alternatives in any given menu are treated equal ('neutral') by the aggregation rule in the sense that permuting them results in choosing exactly the permuted originally chosen alternatives at the collective level. In other words, collective choice from any menu is invariant under a re-labeling of the alternatives. Taken together with monotone independence, this assumption implies that, at a given menu $A \in \mathcal{A}$, the rule determining whether any $x \in A$ be chosen collectively is the *same* for all alternatives in A.

Taken together, unanimity, monotone independence and anonymity require that aggregation, at any issue $x \in c(A)$, happens by setting an acceptance *quota* $0 < q_{x \in c(A)} < 1$ such that $x \in c(A) \iff \frac{1}{n}|\{i \in \{1, \ldots, n\} : x \in c_i(A)\}| \geq q_{x \in c(A)}$. Whenever some coalition $W \subseteq \{1, \ldots, n\}$ of individuals exceeds the quota $q_{x \in c(A)}$, we say that W is a *winning coalition* (for $x \in c(A)$). Given menu-level neutrality these quotas must be 'effectively equal' for all alternatives in some given menu in the sense that they imply the same winning coalitions for all $x \in A$.[73]

Theorem 5. *Every Arrowian, menu-level neutral and anonymous aggregation rule f is consistent on $\mathcal{C}_{fne}$ if and only if it is a quota-rule such that, for all $A \in \mathcal{A}$ with $|A| \geq 2$, the quotas for collectively choosing any x from A imply the same structure of winning coalitions and are such that (i) if $n \leq |A|$, then $0 < q_{x \in c(A)} \leq 1/n$; (ii) if $n > |A|$, then $0 < q_{x \in c(A)} \leq \frac{1}{|A|}(1 - \frac{r}{n}) + \frac{1}{n}\mathbb{1}(r \neq 0)$ where $r = n \mod |A|$.*

Theorem 5 implies that (maximal) consistent acceptance quotas at menu A approach $1/|A|$ for large societies (as $n/|A|$ grows large) and need to decrease in the size of A.

[72] In particular, this shows that 'oligarchies' in the sense of Gibbard (1969/2014) are not oligarchic in our framework.

[73] Note that - due to integer effects - a whole interval of quotas can induce the same set of winning coalitions.

For menus $|A| \geq n$, consistent quota rules reduce to unanimity rule on A for which each individual can veto not choosing any $x \in A$ collectively (i.e. rejecting $x \in c(A)$ requires unanimous consent).

When also imposing contraction consistency (α), every winning coalition at some menu A needs to be winning at all sub-menus $B \subseteq A$ to ensure (α) holds at the collective level. This implies that consistent quotas need to 'effectively'[74] decrease when moving to sub-menus. As all menus are sub-menus of the universal set, maximal consistent quotas are thus determined by $|X|$. Indeed, Theorem 5 implies that if $q_{x \in c(A)} = \bar{q}$ for all $A \in \mathcal{A}$ and $x \in A$, then $0 < \bar{q} \leq 1/|X|$ for large societies (as $n \to \infty$).

[74] That is, bar of any integer effects.

9 Relaxing Independence

The independence assumption imposed above is strong. It requires not only that aggregation is independent across different menus but also that it is independent across alternatives *within* any given menu. While we consider the former a natural condition of informational parsimony, the latter may be unnecessarily strong.[75] Thus, we consider the following weakening of independence.

Independence Across Menus: Consider any $A \in \mathcal{A}$. Let $c = f(c_1, \ldots, c_n)$ and $c' = f(c'_1, \ldots, c'_n)$. If, for all $i = 1, \ldots, n$, $c_i(A) = c'_i(A)$, then $c(A) = c'(A)$.

We would argue that independence across menus exactly captures Arrow's original condition in our generalized framework. Indeed, on $\mathcal{C}_{\text{wo}}$, independence across menus is equivalent to it. While Arrow's notion is often interpreted as stating that collective choices are not to change when removing (or adding) irrelevant alternatives from (to) a given menu, it simply states that collective choice from some *given* menu are not to change unless preferences over the alternatives in the menu change.[76] The demand that removing (adding) irrelevant alternatives does not change the collective choice is a condition on choice *across* menus. In the Arrow framework it is in fact guaranteed by WARP which Arrow implicitly imposes on collective choices by restricting the analysis to rationalizable social choice functions.

As an example of an aggregation rule that satisfies independence across menus but not our fully-fledged notion of independence consider 'approval voting' which, for every menu A, collectively chooses all alternatives with maximal approval by individuals. That is, $f_{\text{AV}}(c_1, \ldots, c_n)(A) = c_{\text{AV}}(A) := \text{argmax}_{x \in A} |\{i \in \{1, \ldots, n\} : x \in c_i(A)\}|$. On the other hand, Borda rule (more generally, all scoring rules) does not satisfy independence across menus seeing that calculating (Borda) scores at some menu $A \in \mathcal{A}$ relies on positional information to be revealed by individual choices from any pair of alternatives $x, y \in A$.[77]

[75] Note that, if collective choices are independent across menus, determining the collective choice set for some menu A only involves eliciting individual choice sets for A.

[76] Arrow himself defends his condition by reference to this interpretation (Arrow, 1963, p.26): "For example, suppose that an election system has been devised whereby each individual lists all the candidates in order of his preference and then, by a preassigned procedure, the winning candidate is derived from these lists. [...] Suppose that an election is held [...] and then one of the candidates dies. Surely the social choice should be made taking each of the individual's preference lists, blotting out completely the dead candidate's name, and considering only the orderings of the remaining names in going through the procedure of determining the winner." However, he seems to neglect that fact that, when one of the candidates dies, the menu of candidates to be chosen from is *not* the same as before. Thus, the condition that choice from some given menu does not depend on alternatives not in it does *not* imply that choices cannot change when removing (adding) irrelevant alternatives from (to) it (at least, not in the absence of WARP).

[77] Note that on $\mathcal{D} = \mathcal{C}_{\text{lo}}$, eliciting choices from all pairs of alternatives in some given $A \in \mathcal{A}$ reveals the individual's underlying linear order when restricted to A.

Proposition 2. *f_{AV} satisfies Independence Across Menus. It is consistent on $\mathcal{D} = \mathcal{C}_{fne}$ but inconsistent on $\mathcal{D} = \mathcal{C}_{sub}, \mathcal{C}_{acy}, \mathcal{C}_{psd}, \mathcal{C}_{Par}, \mathcal{C}_{wo}, \mathcal{C}_{lo}$.*

Proposition 2 states that approval voting is not (sub-/pseudo-)rationalizable in any of the ways considered here. To see this, note that it fails to satisfy the contraction property (α) even if all individual choice functions do. Indeed, let $\{x, y, z\} = A \subseteq X$ and suppose that c_1, c_2, c_3 are rationalized by the linear orders $x \succ_1 y \succ_1 z$, $y \succ_2 x \succ_2 z$ and $z \succ_3 y \succ_3 x$ respectively. Thus, $c_1(A) = x$, $c_2(A) = y$, $c_3(A) = z$ and $c_1(\{x, y\}) = x$ but $c_2(\{x, y\}) = c_3(\{x, y\}) = y$. Consequently, $x \in A = c_{\text{AV}}(A)$ but $x \notin \{y\} = c_{\text{AV}}(\{x, y\})$ failing (α).

Part IV

Consistent Rights on Property Spaces*

* This Part (including Appendix C) of the Thesis has appeared in the *Journal of Economic Theory* (Kretz, 2021) under the same title. The published article can be accessed at https://doi.org/10.1016/j.jet.2021.105323.

10 Introduction

In many aggregation problems, subgroups of agents have the right to predetermine certain *properties* of the aggregate. To respect rights, an aggregation rule must allow subgroups to enforce a property to which they hold a right if all subgroup members agree on it. In standard preference aggregation, for example, allowing individuals to fix those parts of the social ordering that falls within their private spheres corresponds to the social choice theoretic conception of *liberal rights* (Sen, 1970; Gibbard, 1974). More generally, when groups form collective beliefs or desires through aggregating judgments on a set of propositions, they may leave certain judgments to subgroups with *vested interests* or *expert knowledge* (Dietrich and List, 2008). Not least, in committees, delegations from special interest groups can often dismiss alternatives which fall short of some minimal criteria.

Yet, at least since Sen's (1970) famous 'liberal paradox' (Example 1 below), it is well understood that the rights given to different subgroups can be (jointly) inconsistent to the effect that no appealing aggregation function grants all of them at the same time. Going beyond impossibility results, we characterize when rights to properties are (in)consistent given that properties correspond to subsets of alternatives. Such *property spaces* (Nehring and Puppe, 2007, 2010, henceforth N&P) arise for a wide range of interesting applications. As in the examples above, properties may correspond to preference statements over fixed pairs of alternatives, to judgments on propositions (and their negations) or may be naturally suggested by the structure of alternatives. Our results generalize N&P as we show that their 'intersection property' not only characterizes monotone independent aggregators (equivalent to a particular type of rights) but serves to characterize consistency of rights more generally.

Before we preview our main results and discuss the relation to the existing literature on rights, we illustrate the problem of inconsistency in three examples.

Motivating Examples

Example 1. The Sen Liberal Paradox. Ann and Bob are owners of neighboring houses which can be painted either white (w) or yellow (y). Collectively, they are faced with four possible states: (w, w), (w, y), (y, y) and (y, w) – where first entries refer to the color of Ann's house. Suppose Ann and Bob have linear preference orders[78] over these states.

Liberals subscribe to the view that there be a protected private sphere within which individuals are free from interference. Arguably, if a collective ranking $\succ$ is to be found

[78] That is, complete, transitive and antisymmetric (preference) relations.

	$(y,y) \succ (w,y)$	$(w,y) \succ (w,w)$	$(w,w) \succ (y,y)$
Ann	✓	×	✓
Bob	×	✓	✓
	✓	✓	✓

Table 10.1: The Sen liberal paradox

on liberal grounds, Ann and Bob should be left alone to determine it over every pair of states which differ only with respect to their own house color. To formulate a *minimal* requirement, we may demand that they be free to decide $\succ$ over at least one such pair of states each. However, even in such minimal form, individual rights conflict with the equally natural requirement that $\succ$ respect unanimous preference statements (the Pareto condition; a right to society as a whole).

Indeed, suppose both Ann and Bob prefer their own house to be colorful (yellow walls), the other's not (white walls). When in conflict, their preference for a neutral-colored neighborhood prevails. That is, $(y,w) \succ_{Ann} (w,w) \succ_{Ann} (y,y) \succ_{Ann} (w,y)$ and $(w,y) \succ_{Bob} (w,w) \succ_{Bob} (y,y) \succ_{Bob} (y,w)$. Thus, by unanimous agreement, $(w,w) \succ (y,y)$. By minimal liberal rights, $(y,y) \succ_{Ann} (w,y) \implies (y,y) \succ (w,y)$ and $(w,y) \succ_{Bob} (w,w) \implies (w,y) \succ (w,w)$. Consequently, every minimally liberal and Paretian $\succ$ is cyclic. The 'liberal paradox' due to Sen (1970) is the fact that such cycles occur for every minimal assignment of rights if individual preferences are unrestricted.[79] For the present case, Table 10.1 visualizes Ann's (Bob's) (liberal) right by a solid (dashed) box and depicts the Pareto condition as a right held jointly by Ann and Bob (dotted boxes).

Note that our presentation departs from the orthodox view of the paradox as a fundamental incompatibility of Liberalism and Welfarism (in its arguably weakest form, Paretianism). Rather, we suggest to consider it in terms of an inconsistency of individual and collective rights. ■

Example 2. The Gibbard Liberal Paradox. Reconsider Example 1. If we drop minimality, individual rights are in fact internally inconsistent (even in the absence of the Pareto condition). To see this, suppose now that Bob is conformist to the effect that he always wants to match the color of his house to that of Ann's. Ann, on the other hand, is non-conformist. By principle, she prefers to paint her house in a different color than Bob's. Under full liberal rights, $(y,w) \succ_{Ann} (w,w)$ and $(w,y) \succ_{Ann} (y,y)$ imply that $(y,w) \succ (w,w)$ and $(w,y) \succ (y,y)$. At the same time, $(w,w) \succ (w,y)$ and $(y,y) \succ (y,w)$, seeing that $(w,w) \succ_{Bob} (w,y)$ and $(y,y) \succ_{Bob} (y,w)$. Combining, we have $(y,y) \succ (y,w) \succ (w,w) \succ (w,y) \succ (y,y)$. Thus, no acyclic $\succ$ can respect individual liberal

[79] See also Corollary 1 below.

	sales experience	techn. expertise	comm. skills	
HR	✓ (solid box)	✓ (dashed box)	×	a
Sales	✓ (solid box)	×	✓ (dotted box)	b
R&D	×	✓ (dashed box)	✓ (dotted box)	c
	✓	✓	✓	

Table 10.2: An inconsistent system of subcommittee rights

rights alone. The fact that, in general, such cycles cannot be avoided under full individual liberties for two agents (or more) is the 'Gibbard (liberal) paradox' (1974).[80] ■

Example 3. Departmental Rights in Hiring Committees. A company has to fill a job opening in Technical Sales by choosing one out of three candidates (a, b, c). Candidates differ with respect to their differential possession of three relevant qualifications: *sales experience* (candidates a and b), *technical expertise* (a and c) and *communication skills* (b and c). Each candidate uniquely corresponds to a set of qualifications (rows of Table 10.2).

Suppose the hiring committee is made up of members from Human Resources (HR), the Sales Department (Sales) and from Research and Development (R&D).[81] To each department, two qualifications are of particular interest. That is, conversely, for every qualification, there is a subcommittee of two groups with vested interests, say $\{HR \cup Sales\}$ for sales experience (solid boxes in Table 10.2), $\{HR \cup R\&D\}$ for technical expertise (dashed boxes) and $\{Sales \cup R\&D\}$ for communication skills (dotted boxes).

If every member is to cast a vote for one candidate, is it possible to always select an applicant while granting said subcommittees the right to insist on the respective qualifications? Suppose individual votes are homogeneous within departments and as given by the rows of Table 10.2. That is, HR collectively vote for candidate a, Sales for b and R&D for c. Then the rights under consideration imply that the selected candidate possess all three qualifications. As such a candidate does not exist, these rights are inconsistent. In contrast, no such inconsistent combination of qualifications is implied if some qualification can only be insisted upon by the *whole* committee. In other words, if we change the above system of rights to the effect that, for one of the qualifications, we only impose an ordinary unanimity condition, then rights are consistent. ■

[80] See also Proposition 3 below.
[81] We assume that all committee members are affiliated to one and only one department.

Overview of Results

What is the common feature that renders rights (in)consistent in the examples above? We show that consistent rights can be characterized in terms of a simple and well-known 'intersection property' when analyzed in the framework of (abstract) aggregation on property spaces developed by N&P. On property spaces, alternatives can be distinguished by means of (binary) properties. Due to restrictions imposed by logical or physical feasibility, properties are interdependent. In the preference aggregation setting of Examples 1 and 2 – where the set of alternatives is the set of linear orders, and properties correspond to pairwise preference statements – transitivity implies logical restrictions on properties. In the context of voting on candidates in Example 3 – where properties are qualifications – the availability of applicants puts physical constraints on their joint feasibility.

We show that a system of rights to properties is consistent if and only if every collection of groups holding rights over a minimally inconsistent (i.e., *critical*) family of properties has at least one common member (Intersection Property over Critical Families, IPC). A family is inconsistent if there is no alternative that possesses all properties in it. It is minimally so if all proper subfamilies are consistent. In Example 1, the family $\{(y, y) \succ (w, y), (w, y) \succ (w, w), (w, w) \succ (y, y)\}$ is critical.[82] Rights are inconsistent as the groups $\{Ann\}$, $\{Bob\}$ and $\{Ann, Bob\}$ fail to have a common member.[83] Likewise, in Example 3, *sales experience*, *technical expertise* and *communicative skills* make for a critical combination of properties, while no committee member belongs to all of the corresponding rights holding groups simultaneously: $(HR \cup Sales) \cap (HR \cup R\&D) \cap (Sales \cup R\&D) = \emptyset$.

We study when rights allow for a particularly natural way of aggregation that is monotone within and independent across properties (voting by properties). We prove that the characterizing condition of non-empty intersection continues to hold here under a suitably generalized concept of criticality (Intersection Property over Almost Critical Collections, IPAC). We derive tractable characterizations for important classes of property spaces developed in N&P, such as totally blocked spaces (trivial rights) and median spaces (independent rights). On semi-blocked spaces (see Nehring, 2006a), which include partial order aggregation and classification problems, every voter can be granted a minimal participation right if and only if every issue is decided independently via a unanimity rule.

Our work extends a fundamental result from N&P. They characterize monotone independent aggregation as voting by properties induced by some 'structure of winning coalitions' that satisfies IPC. Thus, expressed in terms of rights, a 'structure of winning coalitions' is a rights system which is *exhaustive* (i.e., maximally specified) to the extent that it is equivalent to a monotone independent aggregation procedure respecting it. Yet most of the rights systems one wishes to study are distinctly non-exhaustive (consider Example 1-3 above or

[82] It is inconsistent by transitivity. Every proper subset can be completed to a linear order, i.e., is consistent.

[83] Of course, the singletons $\{Ann\}$ and $\{Bob\}$ have empty intersection by themselves. However, the corresponding properties do not constitute a critical family.

any of the Examples below). Therefore, our results provide a crucial generalization. First, the consideration of non-exhaustive rights facilitates an analysis of rights *sui generis*. Second, unlike the case of exhaustive 'structures of winning coalitions', non-exhaustive rights engender distinct characterizations of when rights are consistent with *some* aggregation function (IPC, Theorem 6) and when they allow to be respected in monotone independent aggregation (IPAC, Theorem 7).

Relation to the Literature on Rights

Initiating the analysis of rights in economics, Sen (1970) adopted a *social choice theoretic* formulation of rights to pairwise (collective) preference statements. Drawing on this model, a large part of the early literature analyzed the robustness of the liberal paradox(es) to a weakening of rights (see, e.g., Gibbard, 1974; Blau, 1975; Kelly, 1976) and the Pareto condition (Sen, 1976; Coughlin, 1986) as well as to domain restrictions (see, e.g., Blau, 1975; Fine, 1975). More recently, Sen's paradox has been generalized to other settings. In the emerging field of judgment aggregation, Dietrich and List (2008) show that minimal group rights are incompatible with a unanimity condition when propositions are sufficiently logically connected.[84] Herzberg (2017) proves a version for probabilistic opinion pooling. It is interesting to note that, while Sen's paradox has traditionally been interpreted in terms of a fundamental incompatibility of liberalism (rights) and welfarism (the Pareto principle), our conceptualization allows to consider the Pareto principle – or, more generally, the unanimity condition – as a right held by society as a whole. Thus, Sen's paradox can also be understood as revealing a conflict between individual and collective rights.

On the other hand, the social choice theoretic formulation has met with conceptual opposition from several authors who have pointed out that the intuitive content of a (liberal) right is *not* to make individual rankings decisive for social preference. But for rights holders to have a strategy at their disposal which allows them to restrict collective choice to a subset of alternatives (see, e.g., Nozick, 1974; Bernholz, 1974; Gärdenfors, 1981; Sugden, 1985; Gaertner et al., 1992).[85] This alternative view has converged to analyzing rights in game forms[86] (see, e.g., Deb, 1994, 2004; Deb et al., 1997; Peleg, 1998; Fleurbaey and Van Hees, 2000; Boros et al., 2010). An important question in this literature is whether a system of rights is representable to the effect that the effectivity function[87] induced by some game form coincides with it.

While, to our best knowledge, contributions from both the social choice theoretic literature on rights and from judgment aggregation are limited to (im)possibility results, we provide a general characterization. As compared to the game form literature on rights –

[84] See Corollary 2 below.
[85] For example, following this line of argument, the right to one's own house color consists in the fact that one can go about painting it in every color one sees fit; thereby restricting the set of social states that can ensue. See also the continuation of Example 2 below.
[86] A game form is a game for which preferences are left unspecified.
[87] On effectivity functions, see also Moulin (1983); Peleg (2002).

where such results exist (see, e.g., Peleg, 1998) – introducing a property structure on the set of alternatives provides for an intuitive characterization in terms of (i) semantic interdependencies between the objects of rights and (ii) combinatorial characteristics of the corresponding rights subjects. Seeing that we (extensionally) define properties as subsets of alternatives, our model shares the basic intuition of this literature. At the same time, it differs in two respects. First, rights are *conjunctive* in our model. When individuals are part of several rights holding groups, they can exercise these rights simultaneously unless this implies enforcing an inconsistent combination of properties at the individual level. In particular, if the same group has rights to two properties, it can enforce their conjunction unless this is infeasible.[88] Second, our notion of representability (i.e., consistency) of rights differs. On the one hand, it is weaker to the effect that we study whether there is some game form that implements *at least* the considered rights (and potentially more). On the other hand, we consider representation by the restricted class of *voting* game forms. We show in Appendix C.1 that, given our notion of weak and conjunctive representation, this is without loss of generality.

The rest of Part IV is structured as follows. In Section 11, we introduce property spaces and define rights to properties. Section 12 presents our characterization of consistent rights. In Section 13, we characterize when rights are consistent with monotone independent aggregation (voting by properties). Section 14 concludes. Unless proofs are short and insightful, they are relegated to the Appendix.

[88] Undoubtedly, there are important rights which are non-conjunctive. For example, individuals have both the right to ride a bike and talk on a cell phone. However, there is no right to do both at the same time. Indeed, there is an obligation not to do it.

11 Rights on Property Spaces

Let X be some finite set of (abstract) objects, $|X| > 2$, and let $N = \{1, \ldots, n\}$ be a group of $n \geq 2$ individuals. We refer to every $x \in X$ as an *alternative* (or outcome) and to every $i \in N$ as a voter. If for every $i \in N$, $x_i \in X$, we say that $\boldsymbol{x} = (x_1, \ldots, x_n)$ is a profile (of votes). Thus, every $i \in N$ votes for exactly one alternative. An *aggregation function* is a mapping $f : X^n \to X$. It maps each profile $(x_1, \ldots, x_n) \in X^n$ to some feasible (collective) alternative $f(x_1, \ldots, x_n) \in X$.

11.1 Property Spaces

To turn the set of alternatives into a property space, we endow X with a property structure $\mathcal{P}$.

Property Space. We say that the ordered pair $(X, \mathcal{P})$ is a property space if and only if for all $P \in \mathcal{P} \subseteq 2^X$ and for all $y, z \in X$, $y \neq z$:

$$P \neq \emptyset, \qquad \text{(non-triviality)}$$

$$P^c := X \backslash P \in \mathcal{P}, \qquad \text{(negation-closedness)}$$

$$\exists\, Q \in \mathcal{P} : z \in Q,\ y \notin Q. \qquad \text{(separation)}$$

We refer to all $P, Q \in \mathcal{P}$ as *properties*.[89]

The intuition behind the construction of $\mathcal{P}$ is the following: Every property $P \in \mathcal{P}$ is identified with the subset of alternatives (note that $P \subseteq X$) which possess it. $\mathcal{P}$ is the collection of all properties such that (i) every $P \in \mathcal{P}$ is non-empty, i.e., there is some alternative that conforms to it. (ii) Every property in $P \in \mathcal{P}$ comes with a *complement* or *negation*, $P^c = X \backslash P \in \mathcal{P}$. This ensures that property membership is binary: either $x \in X$ belongs to P or to its complement P^c. When $P \in \mathcal{P}$, we refer to $\{P, P^c\}$ as a property-negation pair or an *issue*. (iii) The property structure is exhaustive to the effect that any two alternatives are distinguishable by at least one property.

Two comments may clarify the construction. First, properties are *extensionally* defined as subsets of alternatives. Thus, the set of alternatives X is *endowed* with a property structure. This roundabout way of defining properties in terms of alternatives, instead of alternatives in terms of properties, makes it possible to consider different property structures on the same set of underlying alternatives. While a particular property structure

[89] To highlight the natural connection of the property space framework with judgment aggregation theory (see Example 5 below), we depart from N&P's convention of labeling properties by $H \in \mathcal{H}$ and use $P, Q \in \mathcal{P}$ instead.

$\mathcal{P}$ might be natural on X, it is important to keep in mind that others are possible. Second, notwithstanding what we just mentioned, the above axioms can be easily seen to imply that for all $x \in X : \{x\} = \bigcap\{P \in \mathcal{P} : x \in P\}$. That is, every alternative is uniquely identified by the set of its constituent properties. In other words, once a property structure on X is fixed, we can conveniently think of any $x \in X$ as the collection of all the properties it possesses.[90]

We call every $\mathcal{F} \subseteq \mathcal{P}$ a family (of properties) and denote the set of all non-empty families by $\mathbb{F} = 2^{\mathcal{P}} \backslash \{\emptyset\}$. There is a natural notion of consistency on $\mathbb{F}$. Consider some family $\mathcal{F} = \{P_1, \ldots, P_r\} \in \mathbb{F}$. We say that $\mathcal{F}$ is *consistent* if and only if some alternative possesses all properties in $\mathcal{F}$: $\bigcap \mathcal{F} = \bigcap_{k=1,\ldots,r} P_k \neq \emptyset$. If and only if $\bigcap \mathcal{F} = \emptyset$, $\mathcal{F}$ is *inconsistent*. Clearly, every subfamily of a consistent family is itself consistent. To study dependencies between properties on $(X, \mathcal{P})$ it is instructive to consider families which are *minimally* inconsistent to the effect that removing any property yields a consistent family. We call such families *critical* and generically denote them by $\mathcal{G} \in \mathbb{F}$.

Critical Family. Let $\mathcal{G} \in \mathbb{F}$. $\mathcal{G}$ is critical if and only if $\bigcap \mathcal{G} = \emptyset$ and for all $P \in \mathcal{G} : \bigcap (\mathcal{G} \backslash \{P\}) \neq \emptyset$.

All property-negation pairs $\{P, P^c\} \in \mathbb{F}$ are critical. We refer to them as the *trivial* critical families.

Example 4. Linear Orders and SWFs. Let A be a set. Define $X_{Lin(A)} = \{> \subseteq A \times A :>$ is a linear order$\}$. For all distinct $a, b \in A$, let $P_{a>b} = \{> \in X_{Lin(A)} : a > b\}$ and note that $P^c_{a>b} = X \backslash P_{a>b} = P_{b>a}$. Denote by $\mathcal{P}_{Lin(A)}$ the set of all such properties, i.e., $\mathcal{P}_{Lin(A)} = \{P_{a>b}\}_{a \neq b \in A}$. $\left(X_{Lin(A)}, \mathcal{P}_{Lin(A)}\right)$ defines a property space. The non-trivial critical families are those produced by a preference cycle. That is, if $r \geq 3$ and all $a_1, \ldots, a_r \in A$ are distinct, then $\{P_{a_1>a_2}, \ldots, P_{a_{r-1}>a_r}, P_{a_r>a_1}\}$ is critical. These are the only non-trivial critical families. f is an aggregation function on $\left(X_{Lin(A)}, \mathcal{P}_{Lin(A)}\right)$ if and only if it is a social welfare function (SWF). ■

Example 5. Judgment Aggregation. Let L be a set of logical propositions such that if $p \in L$ then $\neg p \in L$, where $\neg p$ means "not p" (i.e., L is closed under logical negation). An agenda $Y \subseteq L$ is a set of propositions (and their negations) on which (collective) judgments have to be made. Every $A \subseteq Y$ is referred to as a *judgment set*. A is *complete* if and only if, for all $p \in Y$, $p \in A$ or $\neg p \in A$. In standard propositional logic, *consistency* of judgment sets can be defined in the usual way. For example, for the agenda $\{u, \neg u, v, \neg v, u \to v, \neg(u \to v)\}$, the judgment set $\{u, u \to v, v\}$ is consistent, while $\{\neg u, \neg(u \to v)\}$ is inconsistent.[91]

[90] Thus, in the presence of a property structure, individual votes $x_i \in X$ can also be thought of as votes on complete and consistent combinations of properties.

[91] For a general logic, consistency can be defined in terms of an entailment relation on $2^L \times L$ (cf. Dietrich, 2007). We also assume that the agenda does not contain any tautologies and contradictions, where $p \in L$ is a contradiction if p is inconsistent and a tautology if $\neg p$ is a contradiction.

Let Y be an agenda and define $X_Y = \{A \subseteq Y : A$ complete and consistent$\}$. For every $p \in Y$, let $P_p = \{A \in X_Y : p \in A\}$. Then $P_p^c = X_Y \backslash P_p = \{A \in X_Y : p \notin A\} = \{A \in X_Y : \neg p \in A\} = P_{\neg p}$. We define $\mathcal{P}_Y = \{P_p\}_{p \in Y}$ and note that $(X_Y, \mathcal{P}_Y)$ is a property space. On $(X_Y, \mathcal{P}_Y)$, every property corresponds to the judgment on some proposition and vice versa. A family of properties is consistent if and only if the corresponding set of propositions is consistent. ■

11.2 Rights

Rights System. A rights system is a correspondence $\mathcal{R} : \mathcal{P} \rightrightarrows 2^N \backslash \{\emptyset\}$.

For every property $P \in \mathcal{P}$, $\mathcal{R}(P)$ collects all (sub)groups[92] $G \subseteq N$ that have a right to it.[93] When group $G \subseteq N$ has a right to property P, it can enforce this property on aggregate. The right is *exercised* when all members $i \in G$ vote for some alternative which conforms to P (i.e., $\forall i \in N : x_i \in P$).[94] An aggregation function $f : X^n \to X$ *respects* the rights system $\mathcal{R}$ if and only if for all $P \in \mathcal{P}$ and all profiles $\boldsymbol{x} = (x_1, \ldots, x_n) \in X^n$:

$$G \in \mathcal{R}(P) \implies [G \subseteq \{i \in N : x_i \in P\} \implies f(\boldsymbol{x}) \in P]\,. \tag{R}$$

That is, f respects $\mathcal{R}$ if and only if it guarantees all properties P which are unanimously endorsed by some group G with a right to P. For example, requiring that $N \in \mathcal{R}(P)$ for all $P \in \mathcal{P}$ corresponds to a standard issue-wise unanimity condition on f. By (R), every group $G \in \mathcal{R}(P)$ can exercise its right to P irrespective of how agents $i \in N \backslash G$ vote.[95] We say that the rights system $\mathcal{R}$ is *consistent* if and only if there is *some* onto aggregation function that respects it. Consequently, $\mathcal{R}$ is consistent if and only if every possible scenario of joint rights exercise is compatible with some aggregate alternative. The scope for such scenarios, however, is limited by the need to vote consistently at the individual level (for all $i \in N$: $x_i \in X$).

We observe some immediate implications of (R). For rights system $\mathcal{R}$, define its (property-wise) monotone closure $\overline{\mathcal{R}}$ by $P \mapsto \overline{\mathcal{R}}(P) = \{G \subseteq N : G \supseteq G'$ for some $G' \in \mathcal{R}(P)\}$. Some $f : X^n \to X$ respects $\mathcal{R}$ if and only if it respects $\overline{\mathcal{R}}$. Intuitively, if some $G \subseteq N$ can force some property, then so can any group that is larger than G. Consequently, it is without loss of generality to simplify notation by restricting attention to

[92] Although every $G \subseteq N$ is a subgroup of N, we will also simply refer to it as a *group* for the rest of Part IV.

[93] We define rights systems as mappings from properties to collections of subgroups for notational convenience. Alternatively, we could model rights by means of correspondences $\widehat{\mathcal{R}} : 2^N \backslash \{\emptyset\} \rightrightarrows \mathcal{P}$ collecting, for every group $G \subseteq N$, the set of properties to which G has a right. Seeing that, for every such $\widehat{\mathcal{R}}$, $P \mapsto \widehat{\mathcal{R}}^{-1}(P) = \{G \in 2^N \backslash \{\emptyset\} : P \in \widehat{\mathcal{R}}(G)\}$ defines a rights system, our formulation is no less general.

[94] One might want to object that, in many contexts, a right to group $G \subseteq N$ refers to its ability to enforce a property under less stringent internal support, e.g., by simple group majority. Our formulation is without loss of generality in this regard, as such demands can be reformulated as rights to subgroups of G, e.g., rights to all majority subgroups of G.

[95] Note the familiarity with the concept of (α-)effectivity in game forms. We elaborate in Appendix C.1.

rights systems which are minimally specified in terms of superset inclusion. Second, note that if $\mathcal{R}$ grants G a right to P, there is nothing in (R) which excludes the possibility that an aggregation rule f respecting $\mathcal{R}$ *effectively* grants this right to a proper subgroup of G. For example, declaring any $i \in G$ a *local* dictator on the issue $\{P, P^c\}$ stays true to granting G a right to P according to (R). In this sense, $\mathcal{R}$ is consistent if there exists some aggregation function that grants *at least* the rights in $\mathcal{R}$.

We say that a rights system $\mathcal{R}$ is *trivial* if $\bigcap_{G\in\mathcal{R}(P),\, P\in\mathcal{P}} G \neq \emptyset$. Indeed, if $\mathcal{R}$ is trivial, there exists some individual $j \in N$ who is part of every rights holding group. Thus, making j a dictator ($\forall \boldsymbol{x} = (x_1, \ldots, x_n) \in X^n : f(\boldsymbol{x}) = x_j$) is (trivially) consistent with $\mathcal{R}$.

12 When Are Rights Consistent?

12.1 A Characterization

We are ready to state when some given rights system $\mathcal{R}$ is (in)consistent. To gain intuition, for every profile $\boldsymbol{x} = (x_1, \dots, x_n)$, consider the family of properties to which rights are exercised: $\mathcal{P}_{\mathcal{R}}(\boldsymbol{x}) = \{P \in \mathcal{P} : \{i \in N : x_i \in P\} \supseteq G \text{ for some } G \in \mathcal{R}(P)\}$.[96] Unless one of these families is inconsistent, we can define an onto aggregation function respecting $\mathcal{R}$ profile-wise by $\boldsymbol{x} \mapsto f(\boldsymbol{x}) \in \bigcap \mathcal{P}_{\mathcal{R}}(\boldsymbol{x})$. On the other hand, if one such family is inconsistent, there can be no $f : X^n \to X$ that respects rights.

Suppose $\boldsymbol{x}^\star = (x_1^\star, \dots, x_n^\star) \in X^n$ is such that $\mathcal{P}_{\mathcal{R}}(\boldsymbol{x}^\star)$ is inconsistent and consider a critical subfamily $\mathcal{G}$.[97] By definition, for each $P \in \mathcal{G}$, there exists some group $G_P \in \mathcal{R}(P)$ that endorses P unanimously: $x_i^\star \in P$ for all $i \in G_P$. However, there can be no $i \in N$ who is a member of *all* these groups at once, as this would imply voting inconsistently: $x_i^\star \in \bigcap_{P \in \mathcal{G}} P = \emptyset$. On the other hand, if there exist groups with empty intersection holding rights over $\mathcal{G}$, then there is some profile $\boldsymbol{x}^\star = (x_1^\star, \dots, x_n^\star)$ such that $\mathcal{G} \subseteq \mathcal{P}_{\mathcal{R}}(\boldsymbol{x}^\star)$, seeing that, for every $i \in N$, there is some $P \in \mathcal{G}$ such that $i \notin G_P$ (hence $\exists x_i^\star \in \bigcap \mathcal{G} \backslash \{P\} \neq \emptyset$). Thus, the following condition due to N&P characterizes consistency.

Intersection Property over Critical Families. $\mathcal{R} : \mathcal{P} \rightrightarrows 2^N \backslash \{\emptyset\}$ satisfies the *Intersection Property over Critical Families* (IPC) if and only if for all critical $\mathcal{G} = \{P_1, \dots, P_r\}$:

$$G_1 \in \mathcal{R}(P_1), \dots, G_r \in \mathcal{R}(P_r) \implies \bigcap_{k=1}^{r} G_k \neq \emptyset. \tag{IPC}$$

Theorem 6. *A rights system $\mathcal{R} : \mathcal{P} \rightrightarrows 2^N \backslash \{\emptyset\}$ is consistent if and only if $\mathcal{R}$ satisfies* (IPC).

Theorem 6 reduces the problem of consistency to an easily interpretable condition: Whenever rights are given to a critical set of properties, the respective rights holding groups must intersect to at least one common member. (IPC) provides a characterization in terms of (i) semantic dependencies between rights objects (properties to which rights are held) and (ii) the combinatorial characteristics of the corresponding rights subjects (groups holding rights). As the set of alternatives is *endowed* with a particular property structure, the exact interpretation of (IPC) hinges on the concrete application.

[96] Note that $\mathcal{P}_{\mathcal{R}}(\boldsymbol{x}) = \mathcal{P}_{\overline{\mathcal{R}}}(\boldsymbol{x})$.
[97] As is easily verified, every inconsistent family contains a critical subfamily.

Example 2 (ctd.). Using the construction in Example 4, $(X_1, \mathcal{P}_1) = \left(X_{Lin(A_1)}, \mathcal{P}_{Lin(A_1)}\right)$ with $A_1 = \{(w, w), (w, y), (y, w), (y, y)\}$ is a property space. In the social choice theoretic interpretation, if $\mathcal{R}_2$ respects full individual liberties on $(X_1, \mathcal{P}_1)$, we have, inter alia, $\{Ann\} \in \mathcal{R}_2\left(P_{(y,w)>(w,w)}\right) \cap \mathcal{R}_2\left(P_{(w,y)>(y,y)}\right)$ and $\{Bob\} \in \mathcal{R}_2\left(P_{(w,w)>(w,y)}\right) \cap \mathcal{R}_2\left(P_{(y,y)>(y,w)}\right)$.[98] As $\{P_{(y,w)>(w,w)}, P_{(w,w)>(w,y)}, P_{(w,y)>(y,y)}, P_{(y,y)>(y,w)}\}$ is critical, $\{Ann\} \cap \{Bob\} = \emptyset$ yields a violation of (IPC); $\mathcal{R}_2$ is inconsistent. Proposition 3 below contains the general statement of the Gibbard paradox.

However, as several authors have pointed out (see, e.g., Gaertner et al., 1992), the Gibbard paradox is counter-intuitive. In a more natural interpretation, individual liberal rights consist in the ability to paint one's house in the color one sees fit. Such rights are consistent. Indeed, consider the property space $(A_1, \widetilde{\mathcal{P}}_1)$, where $\widetilde{\mathcal{P}}_1 = \{P_A, P_A^c, P_B, P_B^c\}$ and $P_A = \{(w, w), (w, y)\}, P_B = \{(w, w), (y, w)\}$. Here, the properties refer directly to the color of Ann's (P_A if white) and Bob's (P_B if white) house. On $(A_1, \widetilde{\mathcal{P}}_1)$, the only critical families are the trivial ones. Thus, the rights system $\widetilde{\mathcal{R}}_2(P_A) = \widetilde{\mathcal{R}}_2(P_A^c) = \{\{Ann\}\}$ and $\widetilde{\mathcal{R}}_2(P_B) = \widetilde{\mathcal{R}}_2(P_B^c) = \{\{Bob\}\}$ satisfies (IPC) and is consistent. ■

Proposition 3. ***(cf. Gibbard, 1974, Theorem 1)*** *Let $A = A_0 \times A_1 \times \cdots \times A_n$ where A_0 is a set of public aspects and A_i are sets of aspects pertaining to the private sphere of individual $i \in N$. $|A_i| \geq 2$ for all $i \in N$. Consider $\left(X_{Lin(A)}, \mathcal{P}_{Lin(A)}\right)$. Let $\mathcal{R}$ be a rights system and suppose that, for all $i \in N$ and for all i-variants[99] $a, b \in A$, we have $\{i\} \in \mathcal{R}(P_{a>b}) \cap \mathcal{R}(P_{b>a})$. Then $\mathcal{R}$ is inconsistent for $n \geq 2$.*

Theorem 6 generalizes a fundamental result from N&P which is stated as Fact 2 below.[100] They characterize monotone independent aggregation as *voting by properties*[101] induced by some so-called 'structure of winning coalitions' satisfying (IPC). Viewed in terms of rights, a 'structure of winning coalitions' is simply a rights system which is maximally specified or *exhaustive* to the effect that, for all groups $G \subseteq N$ and properties $P \in \mathcal{P}$, it either affords G a right to P or else $N \backslash G$ a right to P^c.[102] If consistent, such exhaustive rights in effect *define* a natural aggregation procedure that is monotone (within properties) and independent (across properties) called voting by properties. In other words, if $\mathcal{R}$ is exhaustive, the demand that some aggregation function $f : X^n \to X$ respect rights (i.e., condition (R)) is equivalent to f being voting by properties induced by $\mathcal{R}$ satisfying (IPC).

Yet the rights systems studied in the literature are generally non-exhaustive (consider any of the Examples given this part of the Thesis). Theorem 1 shows that (IPC) continues to characterize consistency for non-exhaustive rights systems. However, while a consistent

[98] For the purpose of illustration, we denote individuals by names here instead of natural numbers.
[99] We say that $a, b \in A$ are i-variants if $\forall j \neq i : a_j = b_j$, where a_j is the projection of a on A_j.
[100] See Proposition 2.1 in Nehring and Puppe (2010); also Proposition 3.1 in Nehring and Puppe (2007).
[101] N&P refer to *voting by properties* as 'voting by issues' instead.
[102] To be precise, a 'structure of winning coalitions' as defined in N&P is an exhaustive rights system which is monotone: If group G has a right to P, then every superset of G does. However, as noted in Section 2.2 above, this distinction is immaterial as far as consistency is concerned: A rights system is consistent if and only if its (property-wise) monotone closure is.

exhaustive rights system *is* a monotone independent aggregation function in the aforementioned sense, Theorem 6 only ensures consistency with *some* aggregation function. Indeed, as we show in Section 12.3 below, there are interesting examples of non-exhaustive rights which are consistent yet cannot be respected in voting by properties. This raises the question of whether we can characterize when non-exhaustive rights are consistent in this stricter sense, i.e., when non-exhaustive rights can be consistently extended to an exhaustive system. Theorem 7 below provides such a characterization of *exhaustibly consistent* rights in terms of an analogous intersection property for a suitably generalized notion of criticality.

Before we develop the necessary theory in Section 13, we use Theorem 6 to relate the inconsistency of rights to structural properties of the underlying space and motivate Section 13 by discussing rights which are consistent but not exhaustibly so.

12.2 Structural Properties and a General Impossibility Result

Below, we present a general impossibility result (Proposition 4) for rights systems that grant property-wise unanimity rights (i.e., $N \in \bigcap_{P \in \mathcal{P}} \mathcal{R}(P)$). To this end, we first introduce structural characteristics of property spaces and rights systems based on entailments between properties.

Let $\mathcal{G} \in \mathbb{F}$ be critical and consider any $P \in \mathcal{G}$. The properties $\mathcal{G} \backslash \{P\}$ *entail* property P^c. Why? By criticality of $\mathcal{G}$, $\mathcal{G} \backslash \{P\}$ is consistent. That is, there exists some feasible $x \in \mathcal{G} \backslash \{P\}$. As $\bigcap \mathcal{G} = \emptyset$ and $x \in X = P \cup P^c$, we must have $x \in P^c$. In other words, every alternative consistent with $\mathcal{G} \backslash \{P\}$ must conform to P^c. Moreover, this entailment is *minimal* in the sense that no proper subset $\mathcal{G}' \subsetneq \mathcal{G} \backslash \{P\}$ entails P^c (seeing that $\mathcal{G}' \cup \{P\} \subsetneq \mathcal{G}$ is consistent by criticality of $\mathcal{G}$).

Minimal Entailment I. For $\mathcal{F} \in \mathbb{F}$, $P \in \mathcal{P}$ we define $\mathcal{F} \vdash P$ if and only if $\mathcal{F} \cup \{P^c\}$ is critical. When $\mathcal{F} \vdash P$, we say that $\mathcal{F}$ *minimally entails* P.

While $\vdash$ relates some property P to *families* of properties, at times, we are only interested in analyzing dependencies between P and some other property Q. Suppose $P \in \mathcal{F}$ and $\mathcal{F} \vdash Q$. Then *conditional* on the properties in $\mathcal{F} \backslash \{P\}$, P entails Q.

Conditional Entailment. $P \trianglerighteq Q$ if and only if there exists some $\mathcal{F} \in \mathbb{F}$ such that $P \in \mathcal{F}$ and $\mathcal{F} \vdash Q$. We denote by $\trianglerighteq^\star \subseteq \mathcal{P} \times \mathcal{P}$ the transitive closure of $\trianglerighteq$. When $P \trianglerighteq^\star Q$, we say that P *conditionally entails* Q. When we need to distinguish it from $\trianglerighteq^\star$, we refer to $\trianglerighteq$ as *direct conditional entailment.*

Note that $P \trianglerighteq Q$ if and only if there exists some critical $\mathcal{G} \in \mathbb{F}$ such that $\{P, Q^c\} \subseteq \mathcal{G}$. It follows that $P \trianglerighteq Q \iff Q^c \trianglerighteq P^c$ and $P \trianglerighteq^\star Q \iff Q^c \trianglerighteq^\star P^c$. We say that P *unconditionally entails* Q if and only if $\{P\} \vdash Q$. That is, P unconditionally entails Q if and only if $\{P, Q^c\}$ is critical. Thus, $\{P\} \vdash Q \iff P \subseteq Q$. By definition, unconditional entailment is direct. We say that $P, Q \in \mathcal{P}$ are *dependent* if and only if $P \trianglerighteq^\star Q^c$. P, Q are *directly* dependent if and only if $P \trianglerighteq Q^c$. That is, P, Q are directly dependent if and only if there exists some critical $\mathcal{G} \in \mathbb{F}$ such that $\{P, Q\} \subseteq \mathcal{G}$. We say that two *issues* $\{P, P^c\}, \{Q, Q^c\}$ are (directly) dependent if and only if we can find some (directly) dependent $\widehat{P}, \widehat{Q}$ such that $\widehat{P} \in \{P, P^c\}$ and $\widehat{Q} \in \{Q, Q^c\}$.

Median, Totally Blocked, Semi-blocked, Connected Property Spaces. Let $(X, \mathcal{P})$ be a property space. We say that $(X, \mathcal{P})$ is:

1. *median* if and only if all entailments are unconditional (cf. Nehring and Puppe, 2007, 2010),

2. *totally blocked* if and only if $\forall P, Q \in \mathcal{P} : P \trianglerighteq^\star Q$ (cf. Nehring and Puppe, 2007, 2010),

3. *semi-blocked* if and only if $(X, \mathcal{P})$ is not totally blocked and $\forall P, Q \in \mathcal{P}$: $[P \trianglerighteq^\star Q$ and $Q \trianglerighteq^\star P]$ or $[P \trianglerighteq^\star Q^c$ and $Q^c \trianglerighteq^\star P]$ (cf. Nehring, 2006a),

4. *connected* if and only if all issues are directly dependent (cf. Dietrich and List, 2008).

As is easily verified, $(X, \mathcal{P})$ is median if and only if all critical families $\mathcal{G} \in \mathbb{F}$ have length two ($|\mathcal{G}| = 2$). Thus, $(A_1, \widetilde{\mathcal{P}}_1)$ from Example 2 (ctd.) above is median.[103] $(X, \mathcal{P})$ is totally blocked if and only if all properties are dependent. For example, $(X_1, \mathcal{P}_1)$ (Example 2, ctd.) is totally blocked. More generally, on every finite set of alternatives, the space of linear orders – endowed with the property structure from Example 4 – is totally blocked. When using an analogous construction for properties, important examples of semi-blocked spaces include the partial orders (Example 9 below) and the equivalence relations (Example 10 below).

Independent, Autonomous Rights. Let $\mathcal{R}$ be a rights system on a property space $(X, \mathcal{P})$.

1. $\mathcal{R}$ is *independent* (*weakly independent*) if and only if there do not exist two disjoint groups $G, G' \subseteq N$ and two dependent (directly dependent) $P, Q \in \mathcal{P}$ such that $G \in \mathcal{R}(P)$ and $G' \in \mathcal{R}(Q)$.

2. $\mathcal{R}$ is *autonomous* (*weakly autonomous*) if and only if there exist two disjoint groups $G, G' \subseteq N$ and two distinct properties $P, Q \in \mathcal{P}$ such that $G \in \mathcal{R}(P) \cap \mathcal{R}(P^c)$ and $G' \in \mathcal{R}(Q) \cap \mathcal{R}(Q^c)$ (such that $G \in \mathcal{R}(P)$ and $G' \in \mathcal{R}(Q)$).

A rights system is (weakly) independent if and only if no two disjoint groups have rights to (directly) dependent properties. It is (weakly) autonomous if and only if two disjoint groups can each autonomously decide some distinct issue (property). If $\mathcal{R}$ is autonomous and weakly independent, disjoint groups can hold rights only over issues which are not directly dependent. When $(X, \mathcal{P})$ is connected, such issues do not exist. By consequence, no rights system can be autonomous and weakly independent at the same time.

[103] Median spaces play an important role in strategy-proof social choice. They admit strong possibility results for rich single-peaked domains (Nehring and Puppe, 2007). See the same reference for more examples of median spaces.

Fact 1. *Let $(X, \mathcal{P})$ be connected. There do not exist weakly independent and autonomous rights systems.*

The following proposition shows that failure of weak independence is sufficient for inconsistency when unanimity rights are granted to all properties. We use Fact 1 to derive two prominent impossibility results in the literature as corollaries. Corollary 1 establishes the 'liberal paradox' for group rights (cf. Example 1 above). Corollary 2 is its generalization to judgment aggregation: no (issue-wise) unanimous aggregation function can grant autonomous group rights on connected agendas.

Proposition 4. *Let $(X, \mathcal{P})$ be a property space and $\mathcal{R}$ be a rights system on $(X, \mathcal{P})$ such that $N \in \bigcap_{P \in \mathcal{P}} \mathcal{R}(P)$. Unless $\mathcal{R}$ is weakly independent, it is inconsistent.*

Corollary 1. ***(cf. Sen, 1976, A3)*** *Let A be some finite set. Consider $\left(X_{Lin(A)}, \mathcal{P}_{Lin(A)}\right)$. If for all $a \neq b \in A : N \in \mathcal{R}(P_{a>b})$ and there exist disjoint $G, G' \subseteq N$ such that, for some $c, c', d, d' \in A$, $G \in \mathcal{R}(P_{c>d}) \cap \mathcal{R}(P_{d>c})$ and $G' \in \mathcal{R}(P_{c'>d'}) \cap \mathcal{R}(P_{d'>c'})$, then $\mathcal{R}$ is inconsistent.*[104]

Corollary 2. ***(cf. Dietrich and List, 2008, Theorem 2)*** *Let Y be some agenda (of logical propositions). Suppose $(X_Y, \mathcal{P}_Y)$ is connected. If $N \in \bigcap_{p \in Y} \mathcal{R}(P_p)$ and there exist disjoint $G, G' \subseteq N$ such that, for some $p, q \in Y$, $G \in \mathcal{R}(P_p) \cap \mathcal{R}(P_{\neg p})$ and $G' \in \mathcal{R}(P_q) \cap \mathcal{R}(P_{\neg q})$, then $\mathcal{R}$ is inconsistent.*[105]

Without property-wise unanimity, however, (weak) independence is neither necessary nor sufficient for consistency. Indeed, when dropping the Pareto condition in Example 1 above, the remaining system of minimal liberal rights is not (weakly) independent but consistent. In Example 3 above, on the other hand, all rights holding groups are pairwise disjoint so that the system is trivially (weakly) independent but inconsistent. At the same time, independence completely characterizes consistency on median spaces. As an even stronger characterization holds for these spaces (see Proposition 5 below), we do not state this result here.

12.3 Consistency vs. Exhaustive Consistency

Theorem 1 shows (IPC) to be necessary and sufficient for rights to be respected by *some* aggregation function. A natural way to obtain an aggregation procedure is to decide each issue independently. In the presence of independence, monotonicity is the natural requirement that increased support for some property cannot lead to its complement being

[104] The result proven in Sen (1976, A3) is slightly more general to the effect that it shows that there can be no *social decision function* (i.e., no aggregation rule producing an acyclic social relation) that satisfies minimal group rights and the Pareto criterion.

[105] In analogy to footnote 104, Theorem 2 in Dietrich and List (2008) is slightly more general as it does not require collective judgment sets to be complete. It is possible to derive the full force of both impossibility results from an analogous statement of Theorem 1 for (non-empty) aggregation correspondences $F : X^n \rightrightarrows X$. We do not do so here to keep the analysis focused on aggregation functions.

accepted if it wasn't before. Apart from its long standing tradition in social choice theory beginning with Arrow (1963), (monotone) independence is a natural requirement in many contexts (for instance, see Example 6 below).[106] Moreover, while the very nature of rights may imply that aggregation happens asymmetrically across subgroups and issues, necessitating failures of anonymity and neutrality, monotone independence seems to stand in no immediate contradiction to rights per se.

Consequently, it is of great interest to study whether rights are *exhaustibly* consistent in the stronger sense of being respected by a monotone independent and onto aggregator. Unlike for exhaustive rights – 'structures of winning coalitions' in N&P – there are important examples of rights which are consistent but *not* exhaustibly so. For instance, this applies to minimal Sen rights (Example 1, ctd.) or minority rights in selecting members of a committee (Example 6). Another example involving majority rights in truth-functional judgment aggregation is developed alongside the theory of Section 13 (Example 7).

Example 1 (ctd.). If we drop the Pareto condition in Example 1, minimal Sen rights are consistent. At the same time, as a corollary of Arrow's Theorem (1963, ch. VIII), the only onto and monotone independent social welfare functions are the dictatorships. Thus, minimal Sen rights are not exhaustibly consistent. More generally, suppose that there are strictly more social alternatives than individuals ($|A| > n \geq 2$). Then there exists some consistent $\mathcal{R}$ granting minimal Sen rights. However, $\mathcal{R}$ is not exhaustibly consistent.[107] ■

Example 6. Committee Selection. Suppose a committee has to be elected from a set of candidates $\{1, \ldots, K\}$ subject to the constraint that at least k' and at most k'' candidates are selected (where $0 < k' \leq k'' < K$). Thus, $X_{(K;k',k'')} = \{C \subseteq \{1, \ldots, K\} : k' \leq |C| \leq k''\}$ is the set of feasible committees. If we let, for every $k \in \{1, \ldots, K\}$, $P_k = \{C \in X_{(K;k',k'')} : k \in C\}$ and define $\mathcal{P}_{(K;k',k'')} = \{P_k, P_k^c\}_{k=1,\ldots,K}$, then $\left(X_{(K;k',k'')}, \mathcal{P}_{(K;k',k'')}\right)$ is a property space such that property k refers to whether candidate k is elected to the committee (Nehring and Puppe, 2010). Monotone independent aggregation on $\left(X_{(K;k',k'')}, \mathcal{P}_{(K;k',k'')}\right)$ amounts to the natural conception that a committee can be selected by voting on each candidate separately and that additional votes can never remove a candidate from the committee.

To safeguard their rights and interests, minority subgroups may demand representation in the committee. That is, minorities may demand the right to elect 'their' candidate to the committee.[108] If $G_1, \ldots, G_m \subseteq N$ are m such (possibly disjoint) minority groups and $k_1, \ldots, k_m$ the corresponding candidates, we consider the rights system $\mathcal{R}$ given by

[106] From a normative point of view, Nehring and Puppe (2007) show that monotone independence is closely linked to strategy-proofness of voting rules: on rich single-peaked domains, a social choice function is strategy-proof if and only if it is monotone independent.

[107] Since $(X_{Lin(A)}, \mathcal{P}_{Lin(A)})$ is totally blocked, this follows from Proposition 5 below.

[108] Such rights *are* given in reality. The Māori, a people indigenous to New Zealand, may serve as a point in case. New Zealand's national parliament reserves a number of designated 'Māori seats' for representatives elected by voters of Māori descent (see, e.g., Geddis, 2006).

$\mathcal{R}(P_{k_j}) = \{G_j\}$ for $j = 1, \ldots, m$ and $\mathcal{R}(P) = \emptyset$ else. As long as $m \leq k''$, $\mathcal{R}$ is consistent.[109] Intuitively, if the number of minority candidates is less than the maximal size of the committee, each can be granted a right to membership. At the same time, if at least two minorities are disjoint – more generally, unless $\mathcal{R}$ is trivial – this requires decisions on the other candidates to be made *depending* on whether said rights have been exercised. For example, in the simple case when $m = k''$ minorities elect 'their' candidate to the committee, all remaining candidates have to be declined irrespective of how many votes they receive. Thus, $\mathcal{R}$ is inconsistent with monotone independent aggregation, i.e., not exhaustibly consistent.[110] ■

[109] This can be seen by noting that the only critical families containing only non-negated properties are those containing exactly $k'' + 1$ of them; i.e., $\mathcal{G} = \{P_{k_1}, \ldots, P_{k_{k''+1}}\}$ for some pairwise distinct $k_1, \ldots, k_{k''+1} \in \{1, \ldots, K\}$. As $\mathcal{R}$ grants rights for at most $m < k'' + 1$ properties, (IPC) holds trivially.

[110] The non-trivial critical families are those described in footnote 109 as well as those containing exactly $K - k' + 1$ negated properties (i.e., $\mathcal{G} = \{P^c_{k_1}, \ldots, P^c_{k_{K-k'+1}}\}$ for some pairwise distinct $k_1, \ldots, k_{K-k'+1} \in \{1, \ldots, K\}$). As a result, $\left(X_{(K;k',k'')}, \mathcal{P}_{(K;k',k'')}\right)$ is totally blocked (see also Nehring and Puppe, 2010). Inconsistency with monotone independent aggregation follows from Proposition 5 below.

13 Consistent Rights in Voting by Properties

On property spaces, an aggregation function $f : X^n \to X$ is independent if and only if issues are decided separately and independently of each other to the effect that the collective decision on some issue can change only if some individual changed her vote on it. Once independence is enacted, monotonicity corresponds to the natural requirement that increased support for some property cannot lead to its complement being accepted when it wasn't before. On property spaces, imposing both independence and monotonicity is equivalent to the following condition.

Monotone Independence. $f : X^n \to X$ is monotone independent if and only if for all $P \in \mathcal{P}$ and all $\boldsymbol{x} = (x_1, \ldots, x_n), \boldsymbol{y} = (y_1, \ldots, y_n) \in X^n$:

$$[\forall i \in N : x_i \in P \implies y_i \in P] \implies [f(\boldsymbol{x}) \in P \implies f(\boldsymbol{y}) \in P]\,. \tag{MI}$$

Monotone independence plays a crucial role in abstract aggregation theory as it allows for a unified characterization of all onto aggregators as *voting by properties* (Nehring and Puppe, 2010).[111] We recall this result as Fact 2 below.

13.1 Monotone Independent Aggregation as Voting by Properties

A rights system $\mathcal{R} : \mathcal{P} \rightrightarrows 2^X\backslash\{\emptyset\}$ is *exhaustive* if and only if, for all $P \in \mathcal{P}$, either $G \in \mathcal{R}(P)$ or $N\backslash G \in \mathcal{R}(P^c)$. If and only if $\mathcal{R}'$ is some rights system such that, for all $P \in \mathcal{P}$, $\mathcal{R}(P) \subseteq \mathcal{R}'(P)$, we say that $\mathcal{R}'$ *extends* $\mathcal{R}$. Thus, $\mathcal{R}'$ extends $\mathcal{R}$ if and only if it grants all rights in $\mathcal{R}$ and possibly more. $\mathcal{R}'$ is *monotone* if and only if it is equal to its (property-wise) monotone closure, $P \mapsto \overline{\mathcal{R}'}(P) = \{G \subseteq N : G \supseteq G' \text{ for some } G' \in \mathcal{R}'(P)\}$, i.e., if and only if $\mathcal{R}' = \overline{\mathcal{R}'}$. We note that every consistent exhaustive rights system $\mathcal{R}'$ is monotone.[112]

Voting by Properties. Given some exhaustive rights system $\mathcal{R}'$, we define the correspondence $F_{\mathcal{R}'} : X^n \rightrightarrows X$ by

$$(x_1, \ldots, x_n) \mapsto F_{\mathcal{R}'}(x_1, \ldots, x_n) = \bigcap \{P \in \mathcal{P} : \{i \in N : x_i \in P\} \in \mathcal{R}'(P)\}.$$

We refer to $F_{\mathcal{R}'}$ as voting by properties (induced by the exhaustive rights system $\mathcal{R}'$).

Thus, separately for each property $P \in \mathcal{P}$, $F_{\mathcal{R}'}$ accepts P iff the groups of individuals voting for P have a right to it. As shown in Nehring and Puppe (2010) – who refer

[111] Note that N&P refer to *voting by properties* as 'voting by issues' instead; cf. footnote 101.

[112] Indeed, suppose that there exist some $P \in \mathcal{P}$ and some $N \supseteq G' \supsetneq G$ such that $G \in \mathcal{R}(P)$, $G' \notin \mathcal{R}(P)$. As $\mathcal{R}$ is exhaustive, $\mathcal{R}(P^c) \ni N\backslash G' \subseteq N\backslash G$, in violation of (IPC).

to exhaustive rights systems as 'structures of winning coalitions' – $\mathcal{F}_{\mathcal{R}'}$ is a monotone independent aggregation function if and only if $\mathcal{R}'$ satisfies (IPC). Vice versa, every monotone independent aggregation function is voting by properties induced by some consistent exhaustive $\mathcal{R}'$.

Fact 2. ***(see Nehring & Puppe, 2010, Proposition 2.1)*** *An onto* $f : X^n \to X$ *satisfies* (MI) *if and only if it is voting by properties* $F_{\mathcal{R}'}$ *induced by some exhaustive* $\mathcal{R}' : \mathcal{P} \rightrightarrows 2^N \backslash \{\emptyset\}$ *satisfying* (IPC).[113]

13.2 Rights in Voting by Properties

Before we return to the question of whether some given (possibly *non-exhaustive*) rights system $\mathcal{R}$ can be respected in monotone independent aggregation, we reappraise the notion of a right in the context of voting by properties. If and only if $\mathcal{R}$ is respected in voting by properties, there exists some consistent exhaustive $\mathcal{R}'$ such that $F_{\mathcal{R}'}$ respects $\mathcal{R}$. By (R) and the definition of $F_{\mathcal{R}'}$, this is the case if and only if, for all $P \in \mathcal{P}$, $G \in \mathcal{R}(P) \implies (\forall G' \subseteq N, G' \supseteq G : G' \in \mathcal{R}'(P))$. As $\mathcal{R}'$ is necessarily monotone, this reduces to the requirement that it extend $\mathcal{R}$.

Fact 3. *Let* $\mathcal{R}$ *be a rights system. There exists some onto* $f : X^n \to X$ *satisfying* (MI) *and* (R) *if and only if there exists some consistent exhaustive rights system* $\mathcal{R}'$ *such that for all* $P \in \mathcal{P}$*:*

$$\mathcal{R}(P) \subseteq \mathcal{R}'(P). \tag{R*}$$

Consequently, $\mathcal{R}$ is consistent with monotone independent aggregation if and only if it can be consistently *extended* to some exhaustive rights system $\mathcal{R}'$. In this case, we say that $\mathcal{R}$ is *exhaustibly consistent*. As Theorem 7 below shows, $\mathcal{R}$ is exhaustibly consistent if and only if it satisfies an intersection property in the spirit of (IPC) for a generalized concept of criticality. To gain intuition for why (IPC) alone is insufficient to guarantee existence of a consistent exhaustive extension, we note the following fact about consistent exhaustive rights.

Fact 4. *Let* $\mathcal{R}'$ *be an exhaustive rights system satisfying* (IPC). *If* $P, Q_1, \ldots, Q_r \in \mathcal{P}$ *are such that* $\{Q_1, \ldots, Q_r\} \vdash P$*, then*

$$G_1 \in \mathcal{R}'(Q_1), \ldots, G_r \in \mathcal{R}'(Q_r) \implies \bigcap_{k=1,\ldots,r} G_k \in \mathcal{R}'(P).$$

In words, every consistent exhaustive rights system $\mathcal{R}'$ is intersection-closed under minimal entailment. When we ask whether rights system $\mathcal{R}$ can be consistently extended to some exhaustive $\mathcal{R}'$, we have to keep in mind the restrictions which are implicitly put on

[113] Nehring and Puppe derive the result for monotone independent and unanimous $f : X^n \to X$. In general, unanimity implies ontoness but not vice versa. However, in the presence of monotonicity, both are equivalent.

	u	v	$u \wedge v$
voter no. 1	✓	✓	✓
voter no. 2	✓	×	×
voter no. 3	×	✓	×
	✓	✓	×

Table 13.1: A discursive dilemma

such extensions by way of Fact 4. We illustrate this point for majority rights in judgment aggregation.

Example 7. Discursive Dilemma. Suppose that some committee (of odd size) has to reach collective judgments on a conjunctive agenda, where the conclusion $c \leftrightarrow u \wedge v$ is endorsed if and only if the premises u and v are. Here, the *discursive dilemma* (Pettit, 2001) consists in the fact that majority voting on the premises is inconsistent with the majority judgment on the conclusion in general. And, vice versa, direct voting on the conclusion is incompatible with majority judgments on the premises. In other words, granting majority rights both on the premises and the conclusion is inconsistent. See Table 13.1 for an example with three voters. Boxes depict rights held by the different majorities od voters: $\{1, 2\}$ (solid), $\{1, 3\}$ (dashed) and $\{2, 3\}$ (dotted).[114]

How about majority rights on the premises alone? This *is* a consistent assignment of rights seeing that the conclusion can simply be made depending on majority judgments on the premises (*premise based procedure*). Yet, it is not exhaustibly consistent: majority rights on the premises are inconsistent with *any* monotone independent method of voting on the conclusion.[115] Indeed, consider $(X_{Y_7}, \mathcal{P}_{Y_7})$ for $Y_7 = \{u, \neg u, v, \neg v, u \wedge v, \neg(u \wedge v)\}$ (cf. Example 5). Let $\mathcal{R}_7(P) = \{G \subseteq N : |G| > n/2\}$ for $P = P_u, P_{\neg u}, P_v, P_{\neg v}$, and $\mathcal{R}(P) = \emptyset$ for $P = P_{u \wedge v}, P_{\neg(u \wedge v)}$. Suppose there exists some consistent exhaustive extension $\mathcal{R}'_7$. If $i \in N$, there exist non-empty and disjoint $G, G' \subseteq N \backslash \{i\}$ such that $|G \cup \{i\}|, |G' \cup \{i\}| > n/2$. By way of Fact 4, $\{P_u, P_v\} \vdash P_{u \wedge v}$ implies that $(G \cup \{i\}) \cap (G' \cup \{i\}) = \{i\} \in \mathcal{R}'_7(P_{u \wedge v})$. Thus, $\mathcal{R}'_7$ violates (IPC) over the critical family $\{P_{\neg u}, P_{u \wedge v}\}$ (e.g., $(G \cup G') \cap \{i\} = \emptyset$), a contradiction.[116] ■

[114] The original example known as the *doctrinal paradox* (Kornhauser and Sager, 1986, 1993) refers to a court of three judges assessing individually whether a defendant owes damages to a plaintiff (the conclusion) by evaluating whether the contract was valid and whether the defendant was in breach of it (the two premises).

[115] Note that, given majority rights on the premises, an aggregation function is fully monotone independent if and only if it is monotone independent on the conclusion.

[116] Too see that such G, G' exist, we simply split up $N \backslash \{i\}$ in two groups of equal size. For n even, every two-member set can be obtained as the intersection of two strict majorities. A parallel argument, combined with the fact that $\{P_{u \wedge v}\} \vdash P_u, P_v$ yields a violation of (IPC) over the critical family $\{P_u, P_v, P_{\neg(u \wedge v)}\}$ provided that $n \geq 4$. If $n = 2$, majority rights are simply unanimity rights and thus respected by any *unanimity rule* (see Section 13.5.2 for a definition).

13.3 Minimal Entailment and Almost Critical Collections

As in Example 7 above, a violation of (IPC) for every consistent exhaustive extension can be *implicit* in $\mathcal{R}$ when the intersection property fails to hold over a collection of properties which are *almost* critical to the effect that they imply some critical family by minimal entailment. However, such violations might appear only after taking into account entailments at higher orders. Also, some properties may be simultaneously involved in different entailments. Consequently, violations of (IPC) can be jointly implied by *multisets* over $\mathcal{P}$.

A multiset is a generalization of a set which allows for members to appear any finite number of times. For example, if $P, Q \in \mathcal{P}$ then $\{P, Q\}$, $\{P, P, Q\}$ and $\{Q, Q, Q\}$ are multisets over $\mathcal{P}$.[117] We refer to any non-empty multiset over $\mathcal{P}$ as a *collection* (of properties).

Collection. We define $\mathbb{C} = \{\mathcal{C} : \mathcal{C}$ is a non-empty multiset over $\mathcal{P}\}$ and call every $\mathcal{C} \in \mathbb{C}$ a *collection* (of properties) from $\mathcal{P}$.

As $\mathcal{P}$ is finite, every $\mathcal{C} \in \mathbb{C}$ is finite.[118] Thus, we can write $\mathcal{C} = \{P_1, \ldots, P_r\}$ for some $P_1, \ldots, P_r \in \mathcal{P}$, where, possibly, $P_k = P_l$ for $k \neq l$. The concept of a collection generalizes that of a family: $\mathbb{F} \subseteq \mathbb{C}$. We define a union-operator $\sqcup$ on $\mathbb{C}$ as follows. Let $\mathcal{C} = \{P_1, \ldots, P_r\}$, $\mathcal{C}' = \{Q_1, \ldots, Q_{r'}\} \in \mathbb{C}$. Then $\mathcal{C} \sqcup \mathcal{C}' = \{P_1, \ldots, P_r, Q_1, \ldots, Q_{r'}\}$. For $s \geq 3$ and $\mathcal{C}_1, \ldots, \mathcal{C}_s \in \mathbb{C}$, we define $\bigsqcup_{l=1,\ldots,s} \mathcal{C}_l$ inductively based on the binary case. We say that $\{\mathcal{C}_1, \ldots, \mathcal{C}_s\}$ is a partition of $\mathcal{C}$ if and only if $\mathcal{C} = \bigsqcup_{l=1,\ldots,s} \mathcal{C}_l$. To formalize our arguments from above, we introduce a generalized minimal entailment relation on $\mathbb{C}$.

Minimal Entailment II. Let $\mathcal{C}, \mathcal{C}' = \{P_1, \ldots, P_r\} \in \mathbb{C}$. We define $\mathcal{C} \Vdash \mathcal{C}'$ if and only if, for all $k = 1, \ldots, r$, there exists some $\mathcal{F}_k \in \mathbb{F}$ such that $\mathcal{F}_k \vdash P_k$ and $\mathcal{C} = \bigsqcup_{k=1,\ldots,r} \mathcal{F}_k$. We denote by $\Vdash^\star$ the transitive closure of $\Vdash$.

$\mathcal{C} \in \mathbb{C}$ minimally entails $\mathcal{C}' \in \mathbb{C}$ if and only if (i) every property in $\mathcal{C}'$ is minimally entailed (in the sense of $\vdash$) by some family and (ii) these families form a partition of $\mathcal{C}$. Note that $\Vdash$ generalizes $\vdash$. Indeed, for every $\mathcal{F} \in \mathbb{F}$ and $Q \in \mathcal{P}$, $\mathcal{F} \vdash Q \iff \mathcal{F} \Vdash \{Q\}$. For $j \geq 2$, we write $\mathcal{C} \Vdash^j \mathcal{C}'$ if and only if there exist $\mathcal{C}_1, \ldots, \mathcal{C}_{j-1}$ such that $\mathcal{C} \Vdash \mathcal{C}_{j-1} \Vdash \ldots \Vdash \mathcal{C}_1 \Vdash \mathcal{C}'$ and let $\Vdash^1 = \Vdash$. Note that $\mathcal{C} \Vdash^\star \mathcal{C}' \iff (\exists j \in \mathbb{N} : \mathcal{C} \Vdash^j \mathcal{C}')$.

Almost Critical Collections. Let $j \in \mathbb{N}$. We say that $\mathcal{C} \in \mathbb{C}$ is *almost critical* (*j-critical*) if and only if $\mathcal{C} \Vdash^\star \mathcal{G}$ ($\mathcal{C} \Vdash^j \mathcal{G}$) for some critical $\mathcal{G} \in \mathcal{F}$.

A collection is almost critical if and only if a critical family of properties can be deduced by repeated minimal entailments. As $\Vdash$ is reflexive, every critical family $\mathcal{F} \in \mathbb{F} \subseteq \mathbb{C}$ is almost critical. Moreover, j-criticality implies j'-criticality for any $j' > j$ (see Lemma 8 in Appendix C.3).

[117] Formally, a multiset over some universe U is a mapping $M : U \to \mathbb{N}_0$ which assigns to every member of $u \in U$ a *multiplicity* $M(u)$. A multiset is non-empty if $supp(M) = \{u \in U : M(u) > 0\} \neq \emptyset$. To facilitate the exposition we will stick to the informal notation introduced above. For example, if $U = \{u, u', u''\}$ we will write $M = \{u, u', u'\}$ instead of $M : U \to \mathbb{N}_0$, $M(u) = 1, M(u') = 2, M(u'') = 0$. Moreover, if $M, M' : U \to \mathbb{N}_0$ are such that for all $u \in U : M(u) \leq M'(u)$, we write $M \subseteq M'$.

[118] A multiset M is finite if $supp(M)$ is finite. In this case, we can define $|M| = \sum_{u \in supp(M)} M(u)$. Over a finite universe, every multiset is finite.

13.4 A Characterization

Reconsider Example 7 from above. We have $\{P_u, P_v\} \vdash P_{u\wedge v}$, thus $\{P_{\neg u}, P_u, P_v\} \Vdash \{P_{\neg u}, P_{u\wedge v}\}$. As the latter is critical, $\{P_{\neg u}, P_u, P_v\}$ is 1-critical; a fortiori, almost critical. The implicit failure of (IPC) over $\{P_{\neg u}, P_{u\wedge v}\}$ (for consistent exhaustive extensions) surfaces as a violation of a corresponding intersection property over the 1-critical family $\{P_{\neg u}, P_u, P_v\}$ for $\mathcal{R}_7$. For example, if $n = 3$, we have $\{1,3\}\cap\{1,2\}\cap\{2,3\} = \emptyset$. More generally, if for some $i \in N$, G, G' are chosen as in Example 7 above, we have: $G\cup\{i\} \in \mathcal{R}_7(P_{\neg u})$, $G' \cup \{i\} \in \mathcal{R}_7(P_u)$ and $G \cup G' \in \mathcal{R}_7(P_v)$ but $(G \cup \{i\}) \cap (G' \cup \{i\}) \cap (G \cup G') = \emptyset$.

The following condition excludes such implied inconsistencies at any order $j \in \mathbb{N}$ of j-criticality. As Theorem 7 below shows, it is necessary and sufficient in order for a consistent exhaustive extension to exist.

Intersection Property Over Almost Critical Collections. A rights system $\mathcal{R}$ satisfies the *Intersection Property over Almost Critical Collections* (IPAC) on $(X, \mathcal{P})$ if and only if for every almost critical collection $\mathcal{C} = \{P_1, \ldots, P_r\} \in \mathbb{C}$:

$$G_1 \in \mathcal{R}(P_1) \cup \{N\}, \ldots, G_r \in \mathcal{R}(P_r) \cup \{N\} \implies \bigcap_{k=1}^{r} G_k \neq \emptyset. \qquad \text{(IPAC)}$$

In analogy to (IPC), (IPAC) demands that every collection of groups holding rights to an *almost* critical collection of properties must intersect to at least one common member. Note how (IPAC) takes account of the fact that every onto and monotone independent aggregation function is property-wise unanimous (i.e., $N \in \bigcap_{P\in\mathcal{P}} \mathcal{R}'(P)$ for every consistent exhaustive extension $\mathcal{R}'$). As every critical family is almost critical, (IPAC) is stronger than (IPC).

To vindicate the intuition that (IPAC) exactly excludes those violations of (IPC) that are implicit by minimal entailment for consistent exhaustive extensions, we introduce the minimal-entailment closure of some rights system. For $\mathcal{R} : \mathcal{P} \rightrightarrows 2^N\backslash\{\emptyset\}$, let $P \mapsto C^1(\mathcal{R})(P) = \{G \subseteq N : G = \bigcap_{k=1,\ldots,r} G_k$ for some $G_1 \in \mathcal{R}(Q_1), \ldots, Q_r \in \mathcal{R}(P_r)$ such that $\{Q_1, \ldots, Q_r\} \vdash P\}$ and, inductively, for $j \geq 2$, define $P \mapsto C^j(\mathcal{R})(P) = C^1(C^{j-1}(\mathcal{R}))(P)$. $C^1(\mathcal{R})$ extends $\mathcal{R}$ so as to include groups which are implied to have rights in consistent exhaustive extensions by Fact 4. In other words, $C^1(\mathcal{R})$ closes $\mathcal{R}$ with respect to $\vdash$. In the same fashion $C^j(\mathcal{R})$ closes $C^{j-1}(\mathcal{R})$ inductively for all $j \geq 2$. Lastly, for every $P \in \mathcal{P}$, let $C^\star(\mathcal{R})(P) = \bigcup_{j\in\mathbb{N}} C^j(\mathcal{R})(P)$. Then $C^\star(\mathcal{R})$ is the closure of $\mathcal{R}$ with respect to chains of minimal entailments of any length.[119] As Theorem 7 below shows, requiring $\mathcal{R}$ to satisfy (IPAC) is equivalent to imposing (IPC) on $C^\star(\mathcal{R})$.

Theorem 7. *Let $\mathcal{R} : \mathcal{P} \rightrightarrows 2^N\backslash\{\emptyset\}$. The following are equivalent:*

1. *There exists some onto $f : X^n \to X$ satisfying* (MI) *and* (R).

[119] Note that for all $j \in \mathbb{N}$, $C^j(\mathcal{R})(P) \subseteq C^{j+1}(\mathcal{R})(P) \subseteq 2^N$; hence $\lim_{j\to\infty} C^j(\mathcal{R})(P) = \bigcup_{j\in\mathbb{N}} C^j(\mathcal{R})(P) = C^\star(\mathcal{R})(P)$.

2. *There exists some exhaustive* $\mathcal{R}' : \mathcal{P} \rightrightarrows 2^N \backslash \{\emptyset\}$ *satisfying* (IPC) *and* ($\mathrm{R}^\star$).

3. $\mathcal{R}$ *satisfies* (IPAC).

4. $C^\star(\mathcal{R})$ *satisfies* (IPC).

As Theorem 7 shows (IPAC) is what characterizes consistency with monotone independent aggregation in general. Unlike for the case of exhaustive rights – for which Fact 4 implies that $C^\star(\mathcal{R}) = \mathcal{R}$; hence (IPC) and (IPAC) coincide – the intersection property needs to be extended to hold over all almost critical families. Generalizing N&P, our results not only allow to consider non-exhaustive rights but also to differentiate between simple and exhaustive consistency. While (IPC) is equivalent to (simple) consistency (Theorem 6), (IPAC) characterizes when rights can be respected in voting by properties.

13.5 Possibilities and Impossibilities on Special Domains

To check (IPAC) for some given rights system $\mathcal{R}$, we need to investigate rights holding groups over every almost critical collection. To this end, it is of considerable interest to understand the size and structure of the class of almost critical collections. Not surprisingly, it depends on the very structure of the property space $(X, \mathcal{P})$ under consideration. We focus on some special domains that are of particular interest.

13.5.1 Totally Blocked and Median Spaces

Generally speaking, the size of the class of almost critical collections tends to increase with the complexity of the agenda, i.e., with the degree of inter-dependencies between properties. At the one extreme, when $(X, \mathcal{P})$ is totally blocked, for every collection of properties, there exists some almost critical collection that contains it. Consequently, the only rights systems consistent with voting by properties are the *trivial* ones (satisfying $\bigcap_{G \in \mathcal{R}(P),\, P \in \mathcal{P}} G \neq \emptyset$). On median spaces, at the other extreme, only the critical families are almost critical. Thus, (IPAC) reduces to (IPC). Here, rights are (exhaustibly) consistent if and only if they are independent (or weakly independent, seeing that both notions of independence coincide on median spaces). We summarize in the following proposition.

Proposition 5. *Let* $(X, \mathcal{P})$ *be a property space.*

1. *If* $(X, \mathcal{P})$ *is totally blocked, every collection of properties is contained in some almost critical collection. A rights system* $\mathcal{R}$ *is exhaustibly consistent if and only if it is trivial.*

2. *If* $(X, \mathcal{P})$ *is median, the almost critical collections are exactly the critical families. If* $\mathcal{R}$ *is a rights system, the following are equivalent:*

   a) $\mathcal{R}$ *is consistent.*

b) $\mathcal{R}$ is exhaustibly consistent.

c) $\mathcal{R}$ is independent.

d) $\mathcal{R}$ is weakly independent.

Example 8. Points on the Real Line. Let $a_1, \ldots, a_r \in \mathbb{R}$, $r \geq 2$ be such that $a_1 < a_2 < \cdots < a_r$. Define $X_8 = \{a_1, \ldots, a_r\}$, and, for each $k = 1, \ldots, r-1$, define $P_k = \{a \in X_8 : a \leq a_k\}$ as well as $P_k^c = X_8 \backslash P_k = \{a \in X_8 : a > a_k\}$. For $\mathcal{P}_8 = \{P_k, P_k^c\}_{k=1,\ldots,r-1}$, $(X_8, \mathcal{P}_8)$ is a property space (cf. Nehring and Puppe, 2007, Example 1). As $\mathcal{G} \in \mathbb{F}$ is critical if and only if $\mathcal{G} = \{P_k, P_{k'}^c\}$ for some $k \leq k'$, $(X_8, \mathcal{P}_8)$ is median.

Consider the rights system $\mathcal{R}_8$ such that, for all $k = 1, \ldots, r-1$, $\mathcal{R}(P_k) = \mathcal{R}(P_k^c) = \{G \subseteq N : |G| > n/2\}$. As all rights holding groups intersect, $\mathcal{R}_8$ is (weakly) independent. Thus, by Proposition 5, it is consistent with monotone independent aggregation. Indeed, selecting the median vote (respectively, the lower/greater of the two median votes when n is even) defines an onto, monotone independent aggregation function that respects $\mathcal{R}_8$. ■

13.5.2 Semi-Blocked Spaces

On semi-blocked spaces, the conditional entailment relation, $\trianglerighteq^\star$, induces a distinctive structure for the set of properties. We can partition $\mathcal{P}$ into two sets of properties all mutually conditionally entailing each other, $\mathcal{P}^+$ and $\mathcal{P}^-$, such that all $P \in \mathcal{P}^+$ conditionally entail every $Q \in \mathcal{P}^-$ but not vice versa. This first part of Fact 5 below was shown in Nehring (2006a) (to keep the exposition self-contained, we give a proof in the Appendix). Figure 13.1 depicts the resulting dependence structure in a graph such that $P \trianglerighteq^\star Q$ if and only if vertex Q can be reached from vertex P via some directed path. For the second part of Fact 5, we show that the almost critical collections relate to this structure as follows. Every almost critical collection contains at most one element from $\mathcal{P}^-$. Conversely, every property $P \in \mathcal{P}^-$ and every collection from $\mathcal{P}^+$ can be jointly embedded in some almost critical collection.

Fact 5. *Suppose $(X, \mathcal{P})$ is semi-blocked. There exist disjoint $\mathcal{P}^+, \mathcal{P}^- \subsetneq \mathcal{P}$ such that $\mathcal{P}^+ \cup \mathcal{P}^- = \mathcal{P}$ and, for all $P \in \mathcal{P}$:*

1. $P \in \mathcal{P}^+ \iff P^c \in \mathcal{P}^-$
2. $P \in \mathcal{P}^- \implies \forall Q \in \mathcal{P}^- : P \trianglerighteq^\star Q$
3. $P \in \mathcal{P}^+ \implies \forall Q \in \mathcal{P} : P \trianglerighteq^\star Q$.

If $\widetilde{\mathcal{C}} \in \mathbb{C}$ is almost critical on $(X, \mathcal{P})$, then it contains at most one element from $\mathcal{P}^-$. Moreover, for every multiset $\mathcal{C}$ over $\mathcal{P}^+$ and every $P \in \mathcal{P}^-$, there exists some almost critical $\widetilde{\mathcal{C}} \supseteq \mathcal{C} \sqcup \{P\}$.

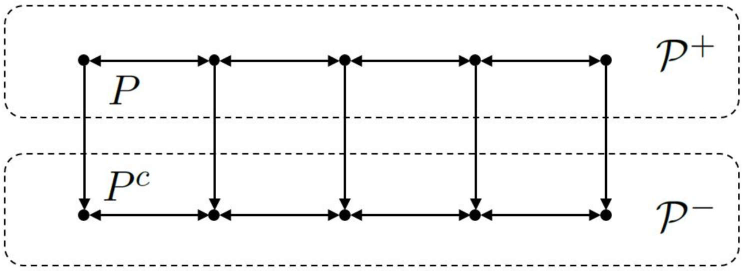

Figure 13.1: Conditional entailment structure on a semi-blocked space

As an immediate consequence, a rights system $\mathcal{R}$ is consistent with voting by properties on a semi-blocked space $(X, \mathcal{P})$ if and only if for all $G' \in \bigcup_{P \in \mathcal{P}^-} \mathcal{R}(P)$:

$$\left(\bigcap_{G \in \mathcal{R}(P),\, P \in \mathcal{P}^+} G \right) \cap G' \neq \emptyset.$$ [120]

A fortiori, $\mathcal{R}$ is consistent with voting by properties only if $\bigcap_{G \in \mathcal{R}(P),\, P \in \mathcal{P}^+} G \neq \emptyset$. If an exhaustibly consistent $\mathcal{R}$ affords rights on $\mathcal{P}^+$, it must be *locally* trivial to the effect that there be some non-empty subgroup of individuals who belong to all rights holding groups on $\mathcal{P}^+$. It follows that exhaustive consistency is impossible if rights are autonomous or if majority rights are granted on some $P \in \mathcal{P}^+$ (see Example 7, ctd.).

Example 7 (ctd.). Note that we have $P_{u \wedge v} \trianglerighteq^\star P_u, P_v \trianglerighteq^\star P_{u \wedge v}$ and $P_{\neg(u \wedge v)} \trianglerighteq^\star P_{\neg u}, P_{\neg v} \trianglerighteq^\star P_{\neg(u \wedge v)}$. Seeing that $P_u \trianglerighteq^\star P_{\neg v}$, $(X_7, \mathcal{P}_7)$ is semi-blocked with $\mathcal{P}_7^- = \{P_{\neg u}, P_{\neg v}, P_{\neg(u \wedge v)}\}$. If $\mathcal{R}(\widetilde{P}) \supseteq \{G \subseteq N : |G| > n/2\}$ for some $\widetilde{P} \in \{P_u, P_v, P_{u \wedge v}\}$, then $\bigcap_{G \in \mathcal{R}(P),\, P \in \mathcal{P}^+} G \subseteq \bigcap \{G \subseteq N : |G| > n/2\} = \emptyset$ (provided that $n > 2$, cf. footnote 116). As a result, as soon as majority rights are granted on the conclusion or on either of the premises, rights are not exhaustibly consistent. In light of this, in a more fundamental sense, the discursive dilemma is the fact that majority rights of such kind are inconsistent with monotone independent aggregation. ■

Thus, the room for non-trivial, exhaustibly consistent rights on semi-blocked spaces is limited. Indeed, if voters are additionally assumed to be minimally relevant in the sense that $\mathcal{R}$ guarantees that every voter is pivotal for some issue and profile of votes, then $\mathcal{R}$ is consistent with voting by properties if only if it is the unanimity rule with default $\mathcal{P}^-$;

[120] Necessity is obvious. To show sufficiency, we verify (IPAC). Let $\mathcal{C} = \{P_1, \ldots, P_r\} \in \mathbb{C}$ be almost critical and $G_1 \in \mathcal{R}(P_1), \ldots, G_r \in \mathcal{R}(P_r)$. If $\mathcal{C}$ is a multiset over $\mathcal{P}^+$, we have $\bigcap_{k=1,\ldots,r} G_k \supseteq \bigcap_{G \in \mathcal{R}(P),\, P \in \mathcal{P}^+} G \neq \emptyset$. If $\mathcal{C}$ contains exactly one element from $\mathcal{P}^-$, let $P_{k'}$, $k' \in \{1, \ldots, r\}$, be that element. We have $\bigcap_{k=1,\ldots,r} G_k = \left(\bigcap_{k=\in\{1,\ldots,r\}, k \neq k'} G_k\right) \cap G_{k'} \supseteq \left(\bigcap_{G \in \mathcal{R}(P),\, P \in \mathcal{P}^+} G\right) \cap G_{k'} \neq \emptyset$. Thus, (IPAC) holds and $\mathcal{R}$ is exaustibly consistent.

that is, if and only if, for all $\boldsymbol{x} = (x_1, \ldots, x_n) \in X^n$ and all $P \in \mathcal{P}$,

$$f(\boldsymbol{x}) \in P \iff \begin{cases} \exists i \in N : x_i \in P & \text{for } P \in \mathcal{P}^- \\ \forall i \in N : x_i \in P & \text{for } P \in \mathcal{P}^+ \end{cases}.$$

Minimal Relevance for (voter) $i \in N$. $\mathcal{R}$ satisfies minimal relevance for (voter) $i \in N$ if and only if

$$\exists P_i \in \mathcal{P}\, \exists G_i \subseteq N : i \in G_i \in \mathcal{R}(P_i) \text{ and } N \backslash G_i \cup \{i\} \in \mathcal{R}(P_i^c). \tag{MR-i}$$

Proposition 6. *Let $(X, \mathcal{P})$ be semi-blocked and suppose $\mathcal{R}$ satisfies* (MR-i) *for all $i \in N$. If some onto $f : X^n \to X$ satisfies* (MI) *and* (R)*, then f is the unanimity rule with default $\mathcal{P}^-$; where $\mathcal{P}^-$ is as defined in Fact 5.*

Two interesting examples of semi-blocked spaces are the partial orders (Example 9 below) and the equivalence relations equipped with the natural property structure of pairwise equivalence (Example 10 below). For partial order aggregation, where it can be shown that $\mathcal{P}^-_{PO(A)} = \{P^c_{a \geq b}\}_{a \neq b \in A}$, minimal relevance rights engender widespread incomparability of (social) alternatives (seeing that the default partial order $\bigcap \mathcal{P}^-_{PO(A)}$ is the empty relation). In the context of classification problems (i.e., of aggregation of equivalence relations), Proposition 6 shows that the unique monotone independent operator which guarantees that each characteristic classification is relevant is the *meet operator*.

Example 9. Partial Orders. For a finite set A, let $X_{PO(A)} = \{\geq \subseteq A \times A : \geq$ is a partial order[121]$\}$. For all $a \neq b \in A$, define $P_{a \geq b} = \{\geq \in X_{PO(A)} : a \geq b\}$. Unlike for the linear orders in Example 4, we do not have $P^c_{a \geq b} = P_{b \geq a}$ in general, seeing that a partial order is not necessarily complete. Let $\mathcal{P}_{PO(A)} = \{P_{a \geq b}, P^c_{a \geq b}\}_{a \neq b \in A}$. $\big(X_{PO(A)}, \mathcal{P}_{PO(A)}\big)$ is a semi-blocked property space with $\mathcal{P}^-_{PO(A)} = \{P^c_{a \geq b}\}_{a \neq b \in A}$. ■

Example 10. Classification Problems. Let A be some set. $X_{Equiv(A)} = \{\sim \subseteq A \times A :$ $\sim$ is an equivalence relation[122]$\}$. For each $a \neq b \in A$, define $P_{a \sim b} = \{\sim \in X_{Equiv(A)} : a \sim b\}$. When $\mathcal{P}_{Equiv(A)} = \{P_{a \sim b}, P^c_{a \sim b}\}_{a \neq b \in A}$, $\big(X_{Equiv(A)}, \mathcal{P}_{Equiv(A)}\big)$ is a property space. Moreover, it is semi-blocked with $\mathcal{P}^+_{Equiv(A)} = \{P_{a \sim b}\}_{a \neq b \in A}$ and $\mathcal{P}^-_{Equiv(A)} = \{P^c_{a \sim b}\}_{a \neq b \in A}$.

In classification problems, every $i \in \mathbb{N}$ is best understood not as a voter but as an attribute or characteristic (respectively, a conceptual perspective) that classifies a set of objects (cf. Fishburn and Rubinstein, 1986). For example, dogs might be classified according to sex, breed, size etc. The problem of aggregation thus consists in merging these characteristic classifications. By Proposition 6, the only monotone independent aggregator for which each attribute classification is relevant (as defined by (MR-i)) is the meet operator. ■

[121] That is, $\geq$ is reflexive, transitive and antisymmetric.
[122] That is, $\sim$ is reflexive, transitive and symmetric.

14 Conclusion

We have provided a novel characterization of consistent rights in terms of semantic interdependencies between properties as rights objects and combinatorial characteristics of the corresponding rights subjects. We have shown that consistent rights can be characterized by means of a simple condition when alternatives differ in terms of properties: whenever rights are given to a combination of properties that is critical (minimally inconsistent), the corresponding rights holding groups must have at least one common member (Intersection Property over Critical Families, IPC). Under property-wise unanimity, rights are consistent only if they are weakly independent (no rights to directly dependent properties for disjoint subgroups).

We have demonstrated that the condition of non-empty intersection must be extended to hold over almost critical (i.e., minimally entailing some critical family) multisets (Intersection Property over Almost Critical Collections, IPAC) to characterize when rights are exhaustibly consistent in the sense of being respected by some onto and monotone independent aggregation function (voting by properties). On totally blocked spaces (where all properties are mutually dependent), rights are exhaustibly consistent if and only if they are trivial (i.e., can be respected by some dictatorship rule); on median spaces (where conditional entailment coincides with subsethood), if and only if they are (weakly) independent. On semi-blocked spaces, minimal relevance rights for all voters pin down monotone independent aggregation functions to a unanimity rule with fixed default.

Our results generalize Nehring and Puppe (2007, 2010) who characterize monotone independent aggregators as voting by properties induced by 'structures of winning coalitions' satisfying IPC. In the interpretation put forth by our analysis, a 'structure of winning coalitions' is a rights system which is exhaustive (maximally specified) to the effect that it defines a (monotone independent) aggregation procedure. Thus, our work provides an analysis of rights sui generis by allowing for non-exhaustive rights and by deriving distinct characterizations for rights being respected by some aggregation function (consistent rights: IPC) and those being respected in onto and monotone independent aggregation (exhaustibly consistent rights: IPAC).

Part V

Premise vs. Conclusion Based Judgment Aggregation with Correlated Voters

15 Introduction

Democracy as a decision-making ideal can be reasoned for on at least two grounds. First, procedural democrats would base their argument on desirable properties of the democratic *process* per se. Second, on a more outcome-based stance, epistemological democrats maintain that democracy is a good ideal as it tends to produce correct answers on questions of unknown truth-value (see, e.g., List and Goodin, 2001). In this latter vein, Condorcet's (1785/ 2014) classical Jury Theorem states that group judgments reached on some proposition by majority rule (i) are correct with probability approaching one as the number of voters tends to infinity (*asymptotic part*), and (ii) become more reliable as the group size (number of voters) is increased (*finite part*); given that individuals vote independently and are more likely to be correct than not.

When collective judgments need to be made on several logically inter-connected propositions, however, invoking Condorcet's Jury Theorem proposition-by-proposition runs into the problem that proposition-wise majority judgments may be (jointly) inconsistent. This fact was first observed in a jurisprudential setting where legal doctrine dictates that a conclusion (the defendant owes damages to the plaintiff) be judged correct if and only if two premises are (the contract is valid and the defendant was in breach of it) and has since become known as the *doctrinal paradox* (Kornhauser and Sager, 1986, 1993).[123] In general, when the agenda consists of a set of premises and a conclusion, two particularly simple procedures to remedy the consistency problem while safeguarding the majoritarian ideal are: (1) to derive the collective conclusion judgment truth-functionally based on the premise majority judgments (the *Premise Based Procedure*, PBP), (2) to consider majority judgment on the conclusion directly (the *Conclusion Based Procedure*, CBP). From a procedural point of view, choosing between the PBP and CBP constitutes a *discursive dilemma* (Pettit, 2001). On the one hand, the collective conclusion derived from premise majorities (i.e. the result for the PBP) may be inconsistent with how a majority of voters judges the conclusion. On the other hand, said conclusion majority (i.e. the result for the CBP) can generally not be consistently reasoned for based on premise majorities. In this part of the thesis, we study how the two procedures fare on epistemological grounds by considering their respective probability of yielding a correct conclusion judgment (which we also refer to as their respective *reliability*) on general truth-functional agendas (several, logically independent premises; one conclusion). Unlike other studies we are aware of, we allow for (positive and homogeneous) correlation between voters. To this end, we develop

[123] For more details, see also Example 7 and Footnote 114 in Part IV.

a tractable model that has voter competence $p \in [0,1]$ and inter-voter correlation coefficient $\rho \in [0,1]$ as sole parameters. In the model, voters are correlated through their joint dependence on some (latent) binary random variable which we interpret as representing a state of high/good or low/bad information.

We analyze both the asymptotic (the number of voters approaches infinity) and finite-sample properties of both procedures. When voters are competent and vote independently, the PBP yields correct judgments with asymptotic certainty across all scenarios no matter the agenda. In contrast, the asymptotic behavior of the CBP is uncertain. For some agendas, it may hinge on the concrete scenario under consideration. In some of them, the CBP generates judgments that are wrong with asymptotic certainty unless voters' competence is sufficiently close to one. This uncertainty is masked by only considering weighted behavior (across all possible scenarios). In finite samples, none of the two procedures guaranteesthat the probability of a correct conclusion judgment increases as the number of voter rises in general (i.e. across all agendas and scenarios). However, the outlined asymptotic properties imply that while the reliability of the PBP must increase eventually (for large numbers of voters), the CBP's reliability may decrease monotonically to zero. At the same time, for a given agenda, scenario and competence level, the CBP can outperform the PBP for any finite number of voters. When voters are correlated, the picture that emerges is more complex. Here, we show that the PBP is asymptotically no better (and no worse) than a single voter for moderate levels of competence. Crucially, this result holds for every (truth-functional) agenda (and across all scenarios). In finite samples, adding more voters to the group may not improve the reliability of the PBP. This can be driven by correlation-related effects or – as in the absence of correlation – by the fact that more accurate premise majority judgments need not translate into more accurate conclusion judgments. This also implies that the effect of correlation can itself be ambiguous. While correlation reduces the accuracy of any premise majority judgment, this may translate – rather paradoxically – into a higher reliability for the conclusion. Not surprisingly, the CBP both is asymptotically fallible and may grow less reliable as the number of voters grows in general. Yet – in analogy to the uncorrelated case – there may exist scenarios in which the CBP outperforms the PBP. Interestingly, this is not only true in finite samples but may hold asymptotically for some levels of voter competence.

Relation to the Literature

In its basic form, the Jury Theorem due to Condorcet (1785/2014) relies on voters being (unconditionally) independent and competent (correct with probability greater than one half). Independence is a troubling assumption when voters influence each other, have access to common sources of information etc. Moreover, not all voters may be competent in all circumstances. As a consequence, an extensive body of literature has formed that relaxes these assumptions. For a survey of alternative Jury Theorems, the reader is referred to Dietrich and Spiekermann (2019, 2021). Here, we only highlight some of the work that is most closely related to our analysis.

As pointed out by Dietrich and Spiekermann (2013), judgments can only reasonably be defended to be independent when conditioned on common causes. As some causes may have detrimental effects on accurate judgments (e.g., misleading evidence in a court case, cognitive biases or environmental influences known to adversely affect decisions), this makes conditional competence a problematic assumption (which is, however, required to salvage the asymptotic part of the Condorcet Jury Theory). Consequently, their *competence-sensitive* Jury Theorem only retains the non-asymptotic part. As we show below, even the latter is in danger when decisions are made on a truth-functional agenda of propositions. This is due to the fact that (i) a higher majority accuracy on the premises need not translate into an increased likelihood of a correct conclusion judgment, and (ii) their result breaks down when the distribution of conditional voter competence does not satisfy a (rather strong) symmetry assumption.[124] Conceptually, their work is related to Ladha (1993) who analyses *exchangeable* voters to employ de Finetti's Theorem. We stick to this exchangeable structure. While our analysis dispenses with the symmetry assumptions on the distribution of the latent random variable used in these works, we assume instead that the latent random variable is itself binary (reflecting either a state of high/good or low/bad information) to keep the analysis tractable.

List (2005) extends the epistemic analysis to the emerging field of Judgment Aggregation, which considers *sets* of judgments on multiple, logically interconnected, propositions (see, e.g. Dietrich, 2007). Bovens and Rabinowicz (2006) compare the PBP and the CBP on a simple truth-functional agenda that features two (logically independent) premises and a conclusion that is logically equivalent to the conjunction of the premises. They show that the *Premise Based Procedure* is universally superior if a correct conclusion needs to be reached for the right reasons. When one is interested in correct conclusion judgments for whatever reason, the CBP may be more reliable for moderately competent voters (and smaller group sizes). Hartmann and Sprenger (2012) generalize the first result by showing that the PBP is best at 'tracking the truth' (i.e., judging *all* propositions on the agenda correctly) among a class of aggregators that includes the CBP on general truth-functional agendas.[125] On the other hand, Hartmann et al. (2010) extend the latter line of inquiry to the class of distance-based procedures (which include both the PBP and CBP as special cases) on general conjunctive agendas. While the PBP (distance parameter $t = 0$) remains asymptotically infallible, procedures that put more weight on the conclusion judgment

[124] To be precise, they assume that voter competence 'tends to exceed 0.5': for all $0 < \epsilon \leq 0.5$, the probability that individual conditional voter competence is at level $0.5 + \epsilon$ is larger than that at level $0.5 - \epsilon$. This subsumes the conditions developed by Ladha (1993). However, this assumption is inconsistent, for instance, with the simple binary case that voter competence is, say, 0.8 and 0.4 both with probability 1/2.

[125] To be precise, they consider the class of aggregators which are 'unbiased' to the effect that they treat acceptance/rejection decisions symmetrically (hence there is no 'bias' to accept or reject). While this condition might be reasonable on procedural grounds, it may seem less compelling for epistemic reasons.

(higher values of $t > 0$) are superior for small groups of moderately competent voters (the CBP corresponds to $t = \infty$).

Our work extends this analysis but differs in at least two important respects. First, and foremost, while all of the above studies assume voters to be (unconditionally) competent and independent, we propose a simple and tractable model that features correlated voters which are competent on average (but may be incompetent in a state of low/inadequate information). Second, while the literature that is closest to our work (Bovens and Rabinowicz, 2006; Hartmann et al., 2010) has only considered conjunctive agendas, we allow for a general truth-functional conclusion. Moreover, we argue that it is insightful to conduct the analysis separately for every possible true state of the world (truth scenario). Thus allows us to uncover uncertainty that is disguised when weighting all scenarios with equal prior probability as previous studies did. In particular, we show that in the uncorrelated case – which parallels that of the cited literature – the reliability of the PBP and the CBP can be decomposed into a majority function (reflecting the insights of the classic Condorcet Jury Theorem) and a scenario function (capturing how premise competences translate into correct conclusion judgments for a given truth-scenario and agenda). Both approaches differ only to the extent that their respective reliability is a reverse concatenation of these two functions. This facilitates the analysis both from the intuitive and the analytical perspective.

Lastly, as first pointed out in this context by Austen-Smith and Banks (1996), even if group members share a common interest, truth-telling need not be an equilibrium (of an adequately defined game). That is, in general, voters cannot be relied upon to reveal their private information. While this can be addressed in terms of a design problem regarding the aggregation procedure to be used (for recent example, see, e.g., Bozbay et al., 2014; de Clippel and Eliaz, 2015; Bozbay, 2019), we do not address this issue here. We note, however, that problems with insincere voting may not be robust when voters are motivated by *expressive* concerns (that is, by a desire to express their true opinion; see, e.g., Schuessler, 2000) in addition to instrumental concerns (that is, how their votes affect outcomes). This is due to the fact that every single voter's probability of being pivotal (in majority voting) is small in large groups (see, e.g., Dietrich and Spiekermann, 2019).

16 The Model

Consider an agenda consisting of $K \in \mathbb{N}$ independent premises and one (truth-functional) conclusion. For this purpose, let a *(truth) scenario* or *state (of the world)* $x \in X := \{0,1\}^K$ be an assignment of truth values x^k to premises $k = 1, \ldots, K$, where $x^k = 1$ if premise k is true and $x^k = 0$ if it is false. Then, every agenda under consideration is represented by some function $F : X \to \{0,1\}$ that maps every scenario to a truth assignment for the conclusion.[126] For example, the agenda consisting of two premises P_1 and P_2 and the conclusion $C \leftrightarrow (P_1 \leftrightarrow P_2)$ corresponds to $F(0,0) = F(1,1) = 1$ and $F(1,0) = F(0,1) = 0$.

We interpret states $x \in \{0,1\}^K$ as reflecting the *objective* truth about the premise (and hence conclusion) judgments under consideration. Yet this truth is unknown to voters $i \in \mathbb{N}$. We assume that, for every premise $k = 1, \ldots, K$, voter i receives a noisy (Bernoulli) signal

$$\tilde{v}_i^k = x^k \boxplus \tilde{e}_i^k \in \{0,1\}, \tag{16.1}$$

where $\boxplus$ is addition on the set $\{0,1\}$ such that $0 \boxplus 0 = 1 \boxplus 1 = 0$ and $0 \boxplus 1 = 1 \boxplus 0 = 1$.[127] Thus, in assessing premise k, voter i makes the random (Bernoulli) error $\tilde{e}_i^k \in \{0,1\}$. If $\tilde{e}_i^k = 0$, signal $\tilde{v}_i^k$ recovers the true premise state x^k; that is, voter i is *correct* about premise k. If $\tilde{e}_i^k = 1$, on the other hand, voter i is *wrong* about premise k. Given voter i receives signals $\tilde{v}_i = (\tilde{v}_i^1, \ldots, \tilde{v}_i^k, \ldots, \tilde{v}_i^K)$, she then deduces the conclusion signal $F(\tilde{v}_i)$.

16.1 Error Model

In the following, we study errors that are identically distributed and independent across premises but may be correlated across voters (for a given premise).[128] Technically, we assume that (i) for every $k = 1, \cdots K$, $(\tilde{e}_i^k)_{i \in \mathbb{N}}$ is an exchangeable sequence of (identically but not necessarily independently distributed) errors;[129] and that (ii) the *sequences* $(\tilde{e}_i^k)_{i \in \mathbb{N}}, k = 1, \ldots, K$ are identically distributed and jointly independent.[130] For any given

[126] As the conclusion is truth-functionally determined by the premises, every agenda is *uniquely* represented by some such function F.

[127] That is, $(\{0,1\}.\boxplus)$ is an Abelian group.

[128] Seeing that $\tilde{e}_i^k$ can be *defined* as $\tilde{e}_i^k := \tilde{v}_i^k \boxminus x^k$, equation (16.1) is perfectly general. All assumptions on the joint distribution of voter signals $\tilde{v}_i^k$ can thus be equivalently put as assumptions on the joint distribution of errors $\tilde{e}_i^k$.

[129] A sequence of random variables $\tilde{y}_1, \tilde{y}_2, \ldots$ is exchangeable iff for every finite permutation σ, the distribution of the permuted sequence $\tilde{y}_{\sigma(1)}, \tilde{y}_{\sigma(2)}, \ldots$ is the same as that of the original sequence.

[130] We say that two *sequences* $(\tilde{y}_i)_{i \in \mathbb{N}}$ and $(\tilde{z}_i)_{i \in \mathbb{N}}$ are independent if, for all $i_1, \ldots, i_n \in \mathbb{N}$, the random vectors $(\tilde{y}_{i_1}, \ldots, \tilde{y}_{i_n})$ and $(\tilde{z}_{i_1}, \ldots, \tilde{z}_{i_n})$ are independent. Two random vectors $(\tilde{y}_1, \ldots, \tilde{y}_n)$, $(\tilde{z}_1, \ldots, \tilde{z}_n)$ are said to be independent if $\mathbb{P}(\tilde{y}_1 = y_1, \ldots, \tilde{y}_n = y_n, \tilde{z}_1 = z_1, \ldots, \tilde{z}_n = z_n) = \mathbb{P}(\tilde{y}_1 = y_1, \ldots, \tilde{y}_n = y_n)\mathbb{P}(\tilde{z}_1 = z_1, \ldots, \tilde{z}_n = z_n)$ for all realizations y, z. Note that this does exclude that the component random variables

premise, denote any single voter's probability of being correct by p and let $0 \leq \rho \leq 1$ be the positive correlation of errors across voters.[131] As any two $\left(\tilde{e}_i^k\right)_{i\in\mathbb{N}}$, $\left(\tilde{e}_i^l\right)_{i\in\mathbb{N}}$ are identically distributed, both p and ρ must be constant across premises. That is, for all $k = 1, \ldots, K$ and all $i \neq j \in \mathbb{N}$:

$$p = p_i^k := \mathbb{P}(\tilde{v}_i^k = x^k) = \mathbb{P}(\tilde{e}_i^k = 0)$$

and

$$\rho = \rho_{i,j}^k := \frac{CoV(\tilde{e}_i^k, \tilde{e}_j^k)}{\sigma_{\tilde{e}_i^k}\sigma_{\tilde{e}_j^k}} = \frac{CoV(\tilde{e}_i^k, \tilde{e}_j^k)}{p(1-p)}.$$

To make the analysis tractable, we assume that voters' errors are given by

$$\tilde{e}_i^k = (1 - \tilde{m}_i^k)\tilde{b}_i^k + \tilde{m}_i^k\tilde{c}^k, \tag{16.2}$$

where $\tilde{b}_i^k, \tilde{m}_i^k, \tilde{c}^k$ are Bernoulli random variables with $\tilde{b}_i^k, \tilde{c}^k \sim B(1-p)$ and $\tilde{m}_i^k \sim B(\sqrt{\rho})$ such that $(\tilde{b}_i^k, \tilde{m}_i^k, \tilde{c}^k)_{i\in\mathbb{N},k=1,\ldots,K}$ are jointly independent. Thus, individual errors $\tilde{e}_i^k$ are 'mixtures' of i.i.d. errors $\tilde{b}_i^k$ and an error $\tilde{c}^k$ that is common to all voters. Both errors have the same accuracy $\mathbb{P}(b_i^k = 0) = \mathbb{P}(c^k = 0) = p$. We may interpret $\tilde{c}^k$ as a common information source (on premise k) access to which fully explains the dependence structure between individual signals. That is, when conditioned on $\tilde{c}^k$, individual errors (respectively, signals) as defined by Equation (16.2) are independent.

Seeing that

$$p + \sqrt{\rho}(1-p) = \mathbb{P}(\tilde{e}_i^k = 0|\tilde{c}^k = 0) \geq \mathbb{P}(\tilde{e}_i^k = 0|\tilde{c}^k = 1) = p - \sqrt{\rho}p \tag{16.3}$$

we define

$$p_H := p + \sqrt{\rho}(1-p) \text{ and } p_L := p - \sqrt{\rho}p \tag{16.4}$$

to be the individual probabilities of being correct in case of *high* ($\tilde{c}^k = 0$) and *low* ($\tilde{c}^k = 1$) collective competence. Thus, $\tilde{c}^k$ moves all individual competences for premise k above or below the average level

$$\mathbb{P}(\tilde{e}_i^k = 0) = pp_H + (1-p)p_L = p.$$

To verify that the structure prescribed by equation (16.2) is in accordance with the assumptions made above, note that for all $k = 1, \ldots, K$ and all $i \neq j \in \mathbb{N}$, e_i^k, e_j^k are i.i.d. Bernoulli

of $\tilde{y}$ depend on each other, and likewise for $\tilde{z}$. Joint independence refers to the generalization of these concepts to the respective joint distributions.

[131] Note that p and ρ are necessarily *constant* across voters for any fixed k since $(\tilde{e}_i^k)_{io\in\mathbb{N}}$ is exchangeable.

random variables ($e_i^k, e_j^k \sim B(1-p)$) such that

$$\begin{aligned} CoV(\tilde{e}_i^k, \tilde{e}_j^k) &= \mathbb{E}[\mathbb{E}[\tilde{e}_i^k \tilde{e}_j^k | \tilde{c}^k]] - \mathbb{E}[\tilde{e}_i^k]\mathbb{E}[\tilde{e}_j^k] \\ &= p\mathbb{P}(\tilde{e}_i^k = \tilde{e}_j^k = 1 | \tilde{c}^k = 0) + (1-p)\mathbb{P}(\tilde{e}_i^k = \tilde{e}_j^k = 1 | \tilde{c}^k = 1) - (1-p)^2 \\ &= p(1-p)^2(1-\sqrt{\rho})^2 + (1-p)\left((1-p) + \sqrt{\rho}p\right)^2 - (1-p)^2 \\ &= \rho p(1-p). \end{aligned}$$

Moreover, joint independence of all $\tilde{b}_i^k, \tilde{m}_i^k, \tilde{c}^k$ implies joint independence of the error sequences $\left(\tilde{e}_i^1\right)_{i\in\mathbb{N}}, \ldots, \left(\tilde{e}_i^K\right)_{i\in\mathbb{N}}$.

In technical terms, $\tilde{c}^k$ is referred to as a *latent* random variable with the help of which the joint distribution of the exchangeable sequence $(\tilde{e}_i^k)_{i\in N}$ can be expressed as a 'mixture' of that of i.i.d. sequences $(\tilde{e}_i^k | c^k)_{i\in\mathbb{N}}$. More generally, De Finetti's Theorem states that such latent random variables taking values in $[0,1]$ exist for any exchangeable sequence. Consequently, equation (16.2) adds two restrictions to the general case: (i) the latent random variables are themselves *binary*; (ii) the probability of high collective competence ($\mathbb{P}(\tilde{c}^k = 0)$) is equal to the average individual competence p. This is shown in Appendix D.1.

16.2 State-Dependent Probability of Correct Conclusion Judgments

While the error model above prescribes a joint probability distribution over events specifying which individuals are correct on what premise judgments, we are ultimately interested in the probability of correct conclusion judgments. Given an agenda represented by $F : X \to \{0,1\}$ and probabilities $q_1, \ldots, q_K$ of correct premise-k signals, the probability of deriving a correct conclusion judgment from *independent* premise signals is given by

$$\hat{f}_x(q_1, \ldots, q_K) := \sum_{\substack{v\in\{0,1\}^K:\\ F(v)=F(x)}} \prod_{k=1}^{K} q_k^{1-|x_k - v_k|}(1-q_k)^{|x_k - v_k|}. \tag{16.5}$$

Seeing that $\hat{f}_x(q_1, \ldots, q_K)$ is determined by the truth scenario $x \in X$, we refer to $\hat{f}_x$ as the *scenario function*. We note that $\hat{f}_x$ is affine in all of its arguments.[132] If all premise accuracies are equal (if $q = q_1 = q_2 = \cdots = q_K$), then $\hat{f}_x$ reduces to the one-dimensional scenario function

$$f_x(q) := \hat{f}_x(q, \ldots, q) = \sum_{\substack{v\in\{0,1\}^K:\\ F(v)=F(x)}} q^{K-||x-v||_1}(1-q)^{||x-v||_1}.^{133} \tag{16.6}$$

[132] See Lemma 15 in the Appendix.
[133] Where $||\cdot||$ is the 1-norm on $\{0,1\}^K$. That is, for all $y \in \{0,1\}^K$, $||y||_1 = \sum_{k=1}^{K} |y_k|$.

For example, consider the agenda given by two premises P_1 and P_2 and the conclusion $C \leftrightarrow (P_1 \wedge P_2)$ and represented by $F : \{0,1\}^2 \to \{0,1\}$ such that $F(x) = 1 \iff x = (1,1)$. Figure 16.1 plots f_x for all states $x = (0,0), (0,1), (1,0), (1,1)$. In state $x = (0,0)$ (where $F(x) = 0$), a correct conclusion will be reached unless both premise signals are wrong. That is, $f_{x=(0,0)}(q) = 1 - (1-q)^2$. In scenario $x = (1,1)$ (where $F(x) = 1$), on the other hand, a correct conclusion emerges only if all premises are judged correctly; hence $f_{x=(1,1)}(q) = q^2$. Lastly, when $x = (1,0)$ or $x = (0,1)$ (and $F(x) = 0$), the correct conclusion is reached unless the true premise is rightly judged to be correct while the false premise is wrongly judged to be true. Thus, $f_x(q) = 1 - q(1-q)$ for $x = (1,0), (0,1)$. Note in particular that competent premise judgments may not result in competent conclusion judgments. To see this, reconsider state $x = (1,1)$. If voter $i \in \mathbb{N}$ receives independent premise signals that are correct with probability $p > 0.5$, she is correct on the conclusion with probability $f_{x=(1,1)}(p) = p^2$. Thus, for $p \in (0.5, 1/\sqrt{2})$, she is competent on the premises ($p > 0.5$) but not on the conclusion ($f_{x=(1,1)}(p) < 0.5$).

As another example consider the conclusion $C \leftrightarrow (P_1 \leftrightarrow P_2)$. Here, in each of the states, the correct conclusion is reached if and only if both premises are either judged correctly or incorrectly. That is, $f_x(q) = q^2 + (1-q)^2$ for $x = (0,0), (0,1), (1,0), (1,1)$. Figure 16.2 depicts the scenario function for this agenda.

We note some immediate properties of $\hat{f}_x(\cdot, \ldots, \cdot)$ and $f_x(\cdot)$ that hold no matter what the agenda and independent of the concrete state x under consideration. First, note that all $\hat{f}_x$ (and hence f_x) are (infinitely) continuously differentiable. Second, $\hat{f}_x(1, \ldots, 1) = f_x(1) = \prod_{k=1}^{K} 1^{1-|x_k - x_k|} 0^{|x_k - x_k|} = 1$. Third, for all $x \in \{0,1\}^K$, let $\bar{x}$ be such that for all $k = 1, \ldots, K$, $\bar{x}_k = 1$ if $x_k = 0$ and $\bar{x}_k = 0$ if $x_k = 1$. Then, $\hat{f}_x(0, \ldots, 0) = f_x(0) = 1$ if $F(\bar{x}) = F(x)$ and $\hat{f}_x(0, \ldots, 0) = f_x(0) = 0$ else (i.e. if $F(\bar{x}) \neq F(x)$). Moreover, $\hat{f}_x(0.5, \ldots, 0.5) = f_x(0.5) = \sum_{v \in \{0,1\}^K : F(v) = F(x)} 0.5^K = \frac{|\{v \in \{0,1\}^K : F(v) = F(x)\}|}{2^K}$.

16.3 The Premise and Conclusion Based Procedure

We study two particularly simple procedures to combine voters' judgments into a collective conclusion judgment. The Premise Based Procedure (PBP) derives the latter truth-functionally from premise-wise majority judgments while the Conclusion Based Approach (CBP) constructs it directly from the majority of voters' conclusion judgment. Consider any finite amount $n \geq 1$ of voters. As voters' signals are exchangeable we may consider the n voters first in line without any loss of generality. For $i = 1, \ldots, n$, let $v_i \in \{0,1\}^K$ be a realization of the random vector $\tilde{v}_i = (\tilde{v}_i^1, \ldots, \tilde{v}_i^K)$ where all $\tilde{v}_i^k$ are subject to equation (16.2) for $k = 1, \ldots, K$. Moreover, for all $k = 1, \ldots, K$, let $\tilde{v}_{[n]}^k = (\tilde{v}_i^k)_{i=1,\ldots,n}$ denote the vector containing the first n voters' premise signals. For judgments $y_1, \ldots, y_n \in \{0,1\}$, let $m(y_1, \ldots, y_n)$ denote the majority judgment (which exists and is unique for odd n). Given voters' realizations $v_1, \ldots, v_n$, the PBP determines the truth value of the conclusion truth-functionally based on majority judgments on the

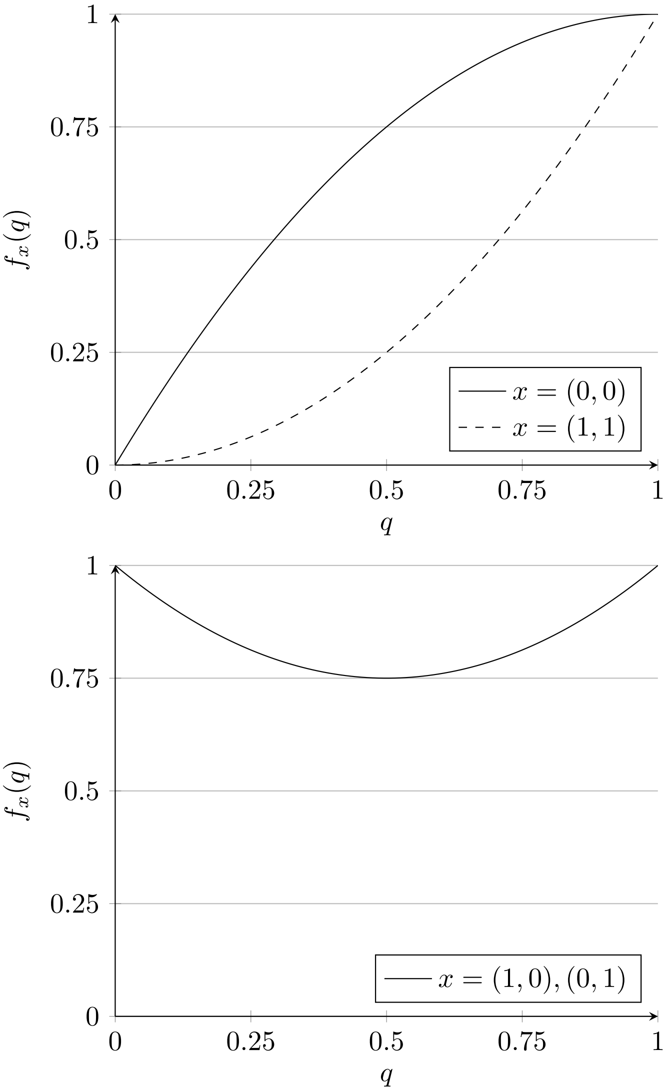

Figure 16.1: Scenario function f_x for $C \leftrightarrow (P_1 \wedge P_2)$

premises: $F(m(v_1^1, \dots, v_n^1), \dots, m(v_1^K, \dots, v_n^K))$. In scenario $x \in \{0,1\}^K$, the probability of uncovering the true conclusion state $F(x)$ is thus given by

$$\begin{aligned} P_{\rho,n}^x(p) :=& \mathbb{P}(F(m(\tilde{v}_{[n]}^1), \dots, m(\tilde{v}_{[n]}^K)) = F(x)) \\ =& \mathbb{P}(\tilde{v}_1 = v_1, \dots, \tilde{v}_n = v_n : F(m(v_1^1, \dots, v_n^1), \dots, m(v_1^K, \dots, v_n^K)) = F(x)). \end{aligned} \quad (16.7)$$

On the other hand, the CBP reaches a judgment on the conclusion by direct majority voting on voters' conclusion signals $F(v_1), \dots, F(v_n)$. Here, the probability of reaching the correct conclusion $F(x)$ in scenario x is given by

$$\begin{aligned} C_{\rho,n}^x(p) :=& \mathbb{P}(m(F(\tilde{v}_1), \dots, F(\tilde{v}_n)) = F(x)) \\ =& \mathbb{P}(\tilde{v}_1 = v_1, \dots, \tilde{v}_n = v_n : m(F(v_1^1, \dots, v_1^K), \dots, F(v_n^1, \dots, v_n^K)) = F(x)). \end{aligned} \quad (16.8)$$

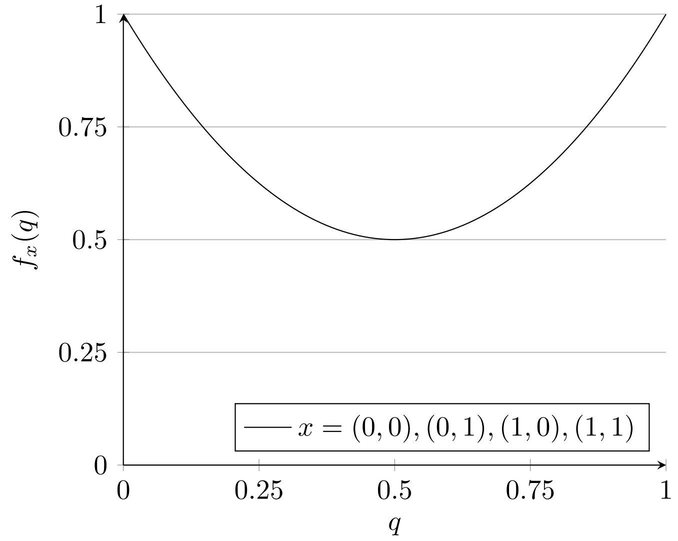

Figure 16.2: Scenario function f_x for $C \leftrightarrow (P_1 \leftrightarrow P_2)$

We also refer to $P^x_{\rho,n}$ and $C^x_{\rho,n}$ as the reliability of the Premise and Conclusion Based Procedure respectively. Note that these reliabilities are inherently scenario-dependent. Given an agenda represented by $F : X \to \{0,1\}$, scenario $x \in X$ and parameters ρ and n, they are functions in voters' competence p.

If some *a priori* probability distribution over truth scenarios $\pi \in \Delta\{0,1\}^K$ exists, it is possible to consider the weighted reliability functions

$$P_{\rho,n}(p) = \sum_{x \in \{0,1\}^K} \pi(x) P^x_{\rho,n} \text{ and } C_{\rho,n}(p) = \sum_{x \in \{0,1\}^K} \pi(x) C^x_{\rho,n}.$$

However, seeing that the essence of the epistemic exercise is to form judgments on propositions of unknown truth value, it is unclear where such a priori probabilities would originate exactly. Therefore, our analysis below will not pre-suppose a prior probability distribution over truth scenarios.

17 Independent Voters

For every premise $k = 1, \ldots, K$, the probability that the majority of n independent (voter) signals of accuracy p is correct is given by (assuming n is odd)

$$g_n(p) := \sum_{m>n/2} \binom{n}{m} p^m (1-p)^{n-m}. \tag{17.1}$$

As the properties of g_n depend only the group size n but are independent of the agenda and the truth scenario, we refer to it as the *majority function*. Note that for every finite $n \in \mathbb{N}$, $g_n(\cdot)$ is (infinitely) continuously differentiable, $g_n(0) = 0$, $g_n(0.5) = 0.5$ and $g_n(1) = 1$. Moreover, as $n \to \infty$ (see, e.g., Grofman et al., 1983),

$$g_n(p) \downarrow 0 \text{ for } p \in (0, 0.5) \text{ and } g_n(p) \uparrow 1 \text{ for } p \in (0.5, 1). \tag{17.2}$$

These properties capture the two statements of the classical *Condorcet Jury Theorem*: For competent voters ($p > 0.5$), the probability that majority voting on a single proposition yields the correct result is increasing in group size (*finite part*) and converges to 1 (*asymptotic part*) as the group grows infinitely large. Figure 17.1 plots the majority function for $n = 1$, $n = 5$ and $n = 51$ voters (left) as well as the limit step function $g_\infty(\cdot) := \lim_{n\to\infty} g_n(\cdot)$ (right).

For $\rho = 0$, equation (16.2) reduces to a model of i.i.d. errors $\tilde{e}_i^k$. All voter signals $\tilde{v}_i^k$ are i.i.d. both across premises and voters. Thus, majorities being correct on the premises $k = 1, \ldots, K$ are independent events, each of which occurs with probability $g_n(p)$. Consequently, the PBP's probability of reaching the correct conclusion in scenario x is given by

$$P^x_{\rho=0,n}(p) = f_x(g_n(p)). \tag{17.3}$$

The CBP, on the other hand, relies on majority voting on voters' conclusion signals. In scenario x, voter i receives a correct such conclusion signal with probability $f_x(p)$. As voters' conclusion signals are independent (seeing that they are functions of jointly independent random variables), the probability that the CBP yields a correct result is given by

$$C^x_{\rho=0,n}(p) = g_n(f_x(p)). \tag{17.4}$$

Generally speaking, aggregation happens along two dimensions: voters and propositions. First, for a given proposition, we can aggregate dispersed voter information into a collective (majority) signal. Second, any set of premise signals $v \in \{0,1\}^K$ logically translates into a

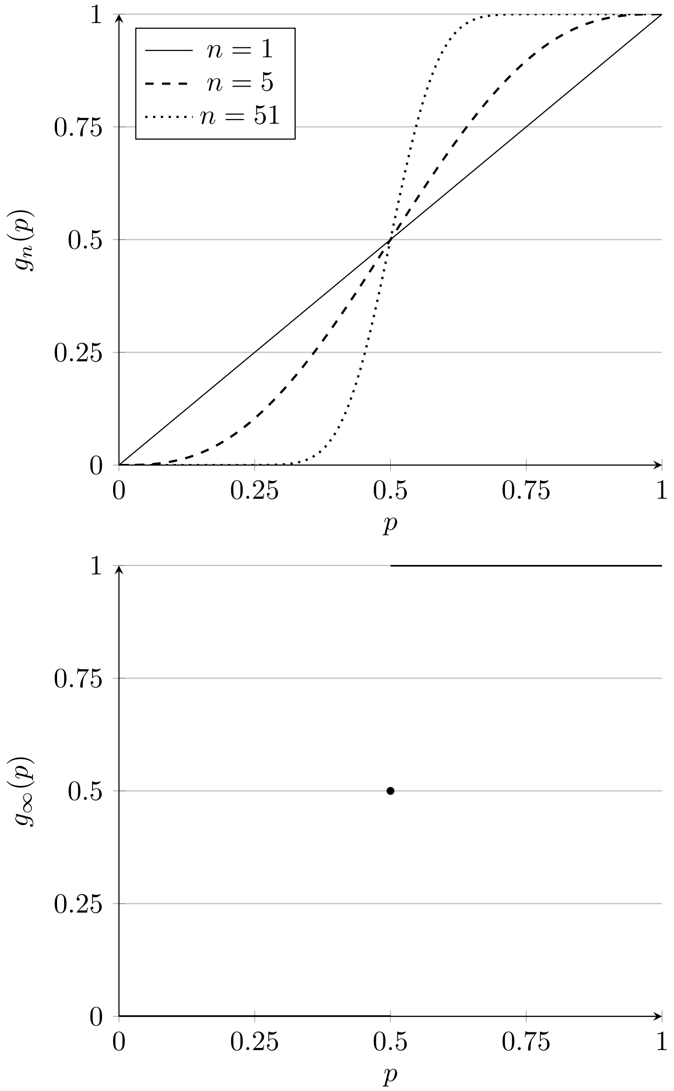

Figure 17.1: Majority function $g_n(\cdot)$ for $n = 1$, $n = 5$ and $n = 51$ (left) and for $n = \infty$ (right)

conclusion signal $F(v)$. The PBP and the CBP differ in that they aggregate information along these two dimensions in opposite order. The PBP pools voter information at the premise level before logically combining these majority premise signals into a conclusion. The CBP has voters combine premise information into a conclusion signal at the individual level first. Then, voters' conclusion signals are aggregated into a collective judgment (by majority voting). Thus, formally speaking, the reliability of both approaches can be expressed as reverse compositions of the majority function $g_n(\cdot)$ and the scenario function $f_x(\cdot)$ (for an example, see Figures 17.3 and 17.4 further below).

17.1 Asymptotic Analysis

We begin by analyzing the performance of both approaches as the number of competent ($p > 0.5$) voters approaches infinity. Consider the PBP first. Since, no matter the agenda $F : X \to \{0, 1\}$ nor the scenario $x \in X$, the scenario function $f_x(\cdot)$ is continuous, we have $\lim_{n\to\infty} P^x_{\rho=0,n}(p) = f_x(g_\infty(p))$. As the probability of a correct majority judgment on any premise converges to 1 ($g_\infty(p) = 1$ for $p \in (0.5, 1]$, cf. Equation (17.2)), the PBP reaches the correct conclusion with asymptotic certainty.

Generally speaking, the same is not true for the CBP, however. At every competence level $p \in (0.5, 1]$, $g_n(f_x(p))$ converges to 1 if only if $f_x(p) > 0.5$. If, for truth scenario $x \in X$, $p > 0.5$ but $f_x(p) < 0.5$, then the CBP will reach the wrong conclusion with asymptotic certainty even though voters are competent on the premises. As an example, consider a conjunctive agenda with two premises: $C \leftrightarrow (P_1 \wedge P_2)$; that is, $F(x) = 1 \iff x = (1, 1)$. If both premises are true ($x = (1, 1)$), we have $f_{x=(1,1)}(p) = p^2 < 0.5 < p$ for moderately competent voters such that $p \in (0.5, 1/\sqrt{2})$. Consequently, $\lim_{n\to\infty} C^{x=(1,1)}_{\rho=0,n}(p) = 0$ at all such p. Generally speaking, the asymptotic part of a CJT holds for the CBP in scenario $x \in X$ if and only if it is *truth-conducive* to the effect that competent premise judgments translate into competent conclusion judgments. Formally, say that scenario $x \in X$ is truth-conducive for the agenda represented by $F : X \to \{0, 1\}$ if and only if $f_x(p) > 0.5$ for all $p \in (0.5, 1)$. We say that the agenda (represented by F) is truth-conducive if and only if all $x \in X$ are truth-conducive for F.

Proposition 7. *For every agenda (represented by) $F : X \to \{0, 1\}$ and all $x \in X$:*

1. $\lim\limits_{n\to\infty} P^x_{\rho=0,n}(p) = 1$ *for all* $p \in (0.5, 1)$,
2. $\lim\limits_{n\to\infty} C^x_{\rho=0,n}(p) = 1$ *for all* $p \in (0.5, 1)$ *iff* x *is truth-conducive.*

As the truth scenario $x \in X$ is unknown, asymptotic infallibility for groups of competent voters is guaranteed for the CBP if and only if the *agenda* is truth-conducive. Such agendas do exist. As a simple example, consider an equivalence agenda with two premises ($C \leftrightarrow (P_1 \leftrightarrow P_2)$) as introduced above: $F(x) = 1 \iff (x = (0, 0)$ or $x = (1, 1))$. Here, recall that, for all $x \in \{0, 1\}^2$, $f_x(p) = p^2 + (1 - p)^2 \geq 0.5$ (cf. Figure 16.2 above); thus, F is truth-conducive. In contrast, asymptotic infallibility is guaranteed for the PBP *no matter* the agenda. Thus, premise based aggregation is asymptotically non-inferior (and situationally strictly better, cf. our discussion above) to its conclusion-based counterpart for competent and independent voters.[134]

[134] At the same time, for a given agenda and scenario, the CBP may outperform the PBP when also considering non-competent voters. For example, reconsider our conjunctive agenda with two premises. If both premises are false ($x = (0, 0)$), we have $f_{x=(0,0)}(p) = 1 - (1 - p)^2 > 0.5 \iff p > 1 - 1/\sqrt{2}$. Hence, $\lim_{n\to\infty} C^{x=(0,0)}_{\rho=0,n}(p) = 1 > 0 = f_{x=(0,0)}(0) = \lim_{n\to\infty} P^{x=(0,0)}_{\rho=0,n}(p)$ for $p \in (1 - 1/\sqrt{2}, 0.5)$.

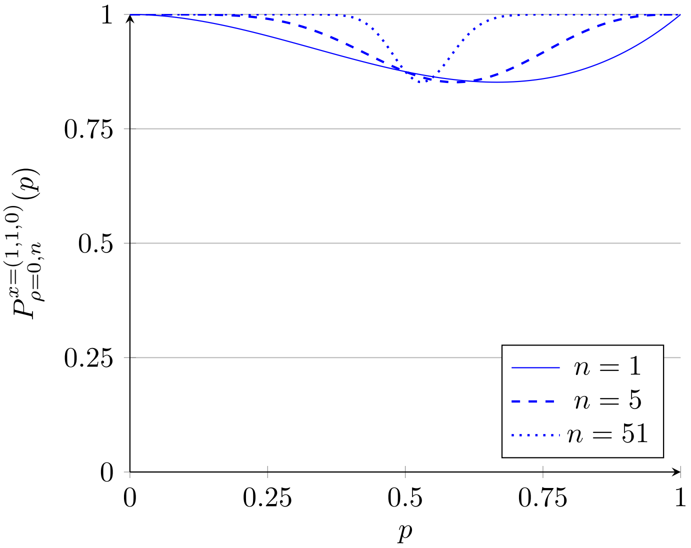

Figure 17.2: Reliability of the PBP in scenario $x = (1, 1, 0)$ on the agenda $C \leftrightarrow (P_1 \wedge P_2 \wedge P_3)$

17.2 The Finite Case

Does the asymptotic superiority of the PBP carry over to finite samples of voters? As we will see, this is not the case in general. First, note that the finite part of a CJT holds for the CBP under the exact same conditions as its asymptotic counterpart: as $g_n(\cdot)$ is increasing in n on $(0.5, 1)$ and decreasing in n on $(0, 0.5)$ (cf. Equation (17.2)), $C^x_{\rho=0,n}(p) = g_n(f_x(p))$ is increasing in n if and only if $f_x(p) > 0.5$. That is, the reliability of the CBP increases in group size for all $p \in (0.5, 1)$ if and only if scenario $x \in X$ is truth-conducive.

Second, consider the reliability of the PBP for competence levels $p \in (0.5, 1)$ in scenario $x \in X$. As $g_n(p)$ is increasing in n, $P^x_{\rho=0,n}(p) = f_x(g_n(p))$ is increasing in n if $f_x(\cdot)$ is increasing on $(0.5, 1)$; that is, if increased competence at evaluating premises correctly translates into higher competence at evaluating the conclusion correctly. In this case, we say that truth scenario $x \in X$ induces *increasing competence*. If all truth scenarios $x \in X$ on the agenda (represented by) $F : X \to \{0, 1\}$ induce increasing competence, we say that the agenda induces increasing competence.

Proposition 8. *For every agenda (represented by $F : X \to \{0, 1\}$) and all $x \in X$:*

1. *$P^x_{\rho=0,n}(p)$ is increasing in n for all $p \in (0.5, 1)$ and all odd $n \geq 1$ if x induces increasing competence,*
2. *$C^x_{\rho=0,n}(p)$ is increasing in n for all $p \in (0.5, 1)$ and all odd $n \geq 1$ iff x is truth-conducive.*

For example, consider a conjunctive agenda with three premises $(C \leftrightarrow (P_1 \wedge P_2 \wedge P_3))$; that is, $F(x) = 1 \iff x = (1, 1, 1)$. Figure 17.2 depicts the scenario function for $x = (1, 1, 0)$ (solid blue) alongside the reliability of the PBP for a group size of $n = 5$ (dashed blue) and $n = 51$ (dotted blue). The scenario function $f_{x=(1,1,0)}$ is decreasing on $(0.5, 2/3)$ and takes on its minimum at $q = 2/3$. Intuitively speaking, when the first two premises are factually correct, the probability of a false conclusion is maximal when the probability of judging the first two premises correctly while judging the last premise incorrectly is maximal. Here the scenario $x = (1, 1, 0)$ does not induce increasing competence. As a consequence, the PBP becomes less reliable when increasing the number of voters from one to five to fifty-one for competence levels p that are above but close to one half.[135] Note that $x = (1, 1, 0)$ *is* truth-conducive, however. Thus, the CBP grows more reliable as the number of voters increases.[136]

As another example, consider an agenda with three premises such that $F(1, 1, 1) = F(0, 0, 0) = F(1, 0, 0) = F(0, 1, 0) = 1$ and $F(x) = 0$ else. In state $x = (1, 1, 1)$, we have $f_{x=(1,1,1)}(q) = q^3 + (1 - q)^3 + 2q(1 - q)^2$. Note that $f_{x=(1,1,1)}(0.5) = 0.5$, $f_{x=(1,1,1)}(0) = f_{x=(1,1,1)}(1) = 1$, $f_{x=(1,1,1)}$ is decreasing on $(0, (1 + \sqrt{7})/6)$, takes on its minimum $f_{x=(1,1,1)}((1 + \sqrt{7})/6) < 0.5$ at $q = (1 + \sqrt{7})/6 \approx 0.61$ and is increasing on $((1 + \sqrt{7})/6), 1)$. Thus, $x = (1, 1, 1)$ is neither truth-conducive nor does it induce increasing competence. Here the finite part of a Jury Theorem fails for both procedures (for competence levels above but close to one half). However, the PBP's reliability increases in the number of voters eventually, that is for large enough odd $n \geq n_p$ where $n_p = \min\{n \in \mathbb{N} : g_n(p) \geq (1 + \sqrt{7})/6\}$.[137] In contrast, for competence levels above but close to 0.5, the reliability of the CBP is decreasing in n for all n and approaches 0 in the limit.

Lastly, as an example of an agenda that features a scenario which induces increasing competence but is not truth-conducive, consider the state $x = (1, 1)$ on a conjunctive agenda with two premises (cf. Figure 16.1, left).

PBP vs. CBP

While the analysis above establishes finite-sample and asymptotic properties of the PBP and CBP in isolation, we now turn to the question of how they fare in direct comparison. While Proposition 7 shows that the PBP is non-inferior to the CBP in the limit, this need not be borne out by all truth scenarios in finite samples. Consider our conjunctive agenda with two premises from above. Here, in finite samples with $n > 1$, the CBP outperforms the PBP for all $p \in (0.5, 1)$ in all scenarios except $x = (1, 1)$. Figures 17.3 and 17.4 (left) illustrate the case of $n = 5$ voters. Thus, it is no surprise that the CBP emerges as the winner when weighting all scenarios with equal prior probability (cf. Bovens

[135] The same is true for scenarios $x = (1, 0, 1)$ and $x = (0, 1, 1)$, of course.

[136] It is an open question whether there exists a whole *agenda* that is truth-conducive but does not induce increasing competence. Note that, in our present example, while scenario $x = (1, 1, 0)$ is truth-conducive, the agenda as a whole is *not*. For example, in scenario $x = (1, 1, 1)$, we have $f_{x=(1,1,1)}(q) = q^3$.

[137] Note, however, that n_p grows indefinitely large as $p \downarrow 0.5$.

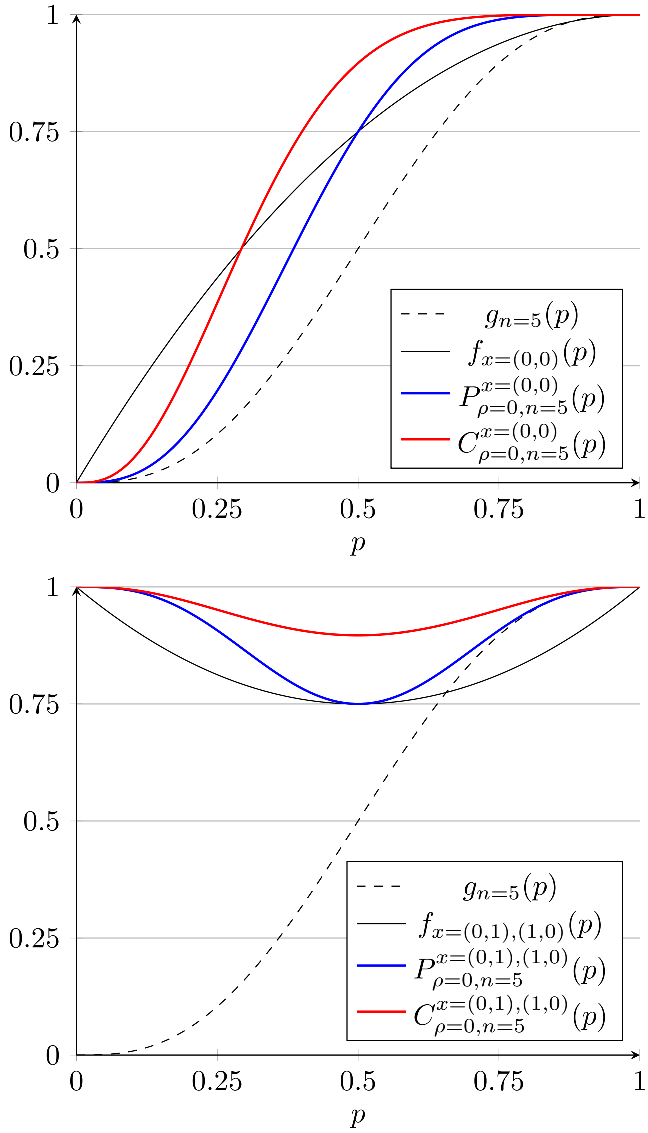

Figure 17.3: The CBP (red) outperforms the PBP (blue) in scenarios $x = (0,0)$ (left) and $x = (1,0),(0,1)$ (right) on the agenda $C \leftrightarrow (P_1 \wedge P_2)$

and Rabinowicz, 2006).[138] However, it must be stressed that weighted behavior disguises the fact that the CBP's reliability is much more dependent on the truth scenario when compared to the PBP. Indeed, the CBP performs well in scenarios that feature high levels of voter competence anyway (in scenarios $x = (0,0),(1,0),(0,1)$), while it performs poorly in scenarios that are not truth-conducive (in scenario $x = (1,1)$). Thus, the rationality of choosing the CBP over PBP critically relies on (i) the prior probability of the non-truth-conducive scenario $x = (1,1)$ being not too high and (ii) a sufficient 'tolerance' for

[138] Note that the states $x = (0,0)$ and $x = (1,1)$ are symmetric to the effect that when both are weighted with equal (prior) probability, differences in reliability between the two procedures exactly cancel out. Thus, comparisons of weighted reliability are driven entirely by the scenarios $x = (1,0),(0,1)$ in both of which the CBP outperforms the PBP.

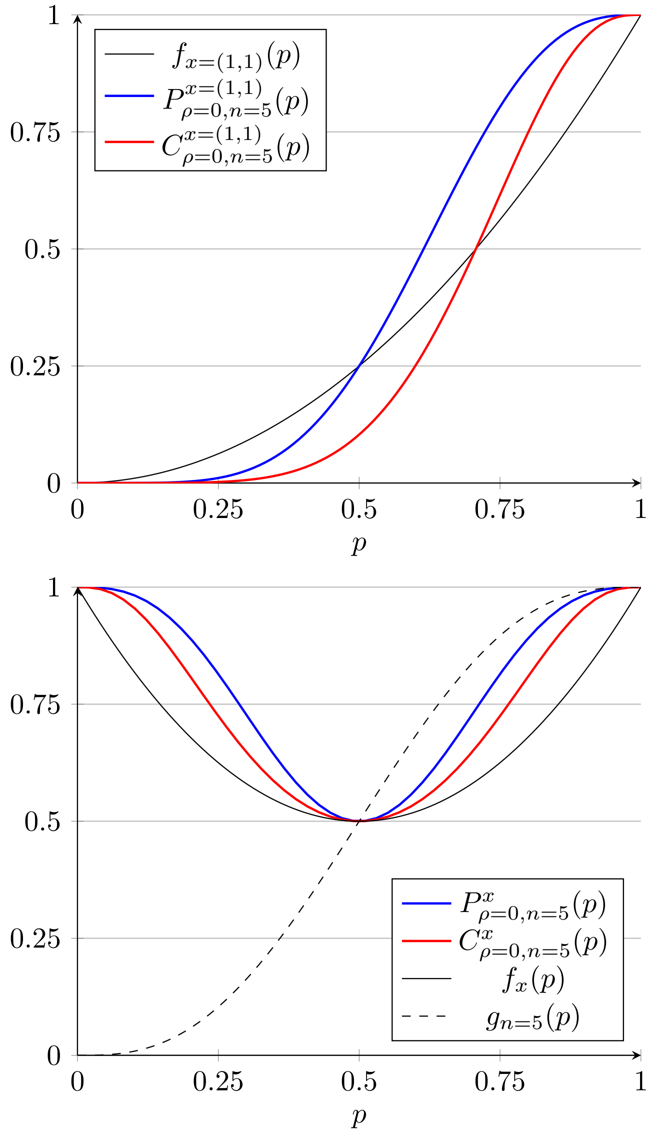

Figure 17.4: The PBP (blue) outperforms CBP (red) in scenario $x = (1,1)$ on the agenda $C \leftrightarrow (P_1 \wedge P_2)$ (left) and in all scenarios on the agenda $C \leftrightarrow (P_1 \leftrightarrow P_2)$ (right) for $n = 5$ voters

risk or uncertainty. Indeed, if both procedures are compared based on worst-case behavior (a max-min rationale reflecting 'infinite' aversion to uncertainty), scenario $x = (1,1)$ is pivotal: $\min_{x\in\{0,1\}^2} P^x_{\rho=0,n}(p) = P^{x=(1,1)}_{\rho=0,n}(p) > C^{x=(1,1)}_{\rho=0,n}(p) = \min_{x\in\{0,1\}^2} C^x_{\rho=0,n}(p)$ for all $p \in (0.5,1)$ and all odd $n > 1$.

At the same time there exist agendas for which the PBP is more reliable in *all* truth scenarios. As a simple example, reconsider the equivalence agenda with two premises and remember that $f_x(q) = q^2 + (1-q)^2$. Figure 17.4 (right) shows the case of $n = 5$ voters. However, results are qualitatively similar for all odd $n > 1$.

18 Correlated Voters

Reconsider Equation (16.2) and recall that voters' signals on premises $k = 1, \ldots, K$ are correlated only due to their joint dependence on the latent random variables $\tilde{c}^1, \ldots, \tilde{c}^K$. When conditioning on them, signals are independent both across voters and premises. That is, errors $(\tilde{e}_i^k | \tilde{c}^k)_{i \in \mathbb{N}}$ are i.i.d. sequences such that (cf. Equations (16.3) and (16.4)):

$$\mathbb{P}(\tilde{e}_i^k = 0 | \tilde{c}^k = 0) = p_H = p + \sqrt{\rho}(1-p)$$

and

$$\mathbb{P}(\tilde{e}_i^k = 0 | \tilde{c}^k = 1) = p_L = p - \sqrt{\rho} p.$$

Let $y = (y_1, \ldots, y_K) \in \{0,1\}^K$ be a joint realization of the (latent) random variables $\tilde{c}_1, \ldots, \tilde{c}_K$. For premise $k = 1, \ldots, K$, a majority of voters is correct on it with probability $g_n(p_H^{1-y_k} p_L^{y_k})$ (that is, with probability $g_n(p_H)$ if $y_k = 0$ and $g_n(p_L)$ if $y_k = 1$).[139] As these are conditionally independent events, the PBP is correct with probability

$$P_{\rho,n}^x(p) = \sum_{y \in \{0,1\}^K} \mathbb{P}(\tilde{c}^1 = y_1, \ldots, \tilde{c}^K = y_K) \cdot \hat{f}_x(g_n(p_H^{1-y_1} p_L^{y_1}), \ldots, g_n(p_H^{1-y_K} p_L^{y_K})). \quad (18.1)$$

At the same time, each individual voter's conditional probability of being correct on the conclusion (her conditional conclusion competence) is given by $\hat{f}_x(p_H^{1-y_1} p_L^{y_1}, \ldots, p_H^{1-y_K} p_L^{y_K})$. As all voters' conclusion signals are conditionally independent (as functions of conditionally independent random variables), the CBP is correct with probability

$$C_{\rho,n}^x(p) = \sum_{y \in \{0,1\}^K} \mathbb{P}(\tilde{c}^1 = y_1, \ldots, \tilde{c}^K = y_K) \cdot g_n(\hat{f}_x(p_H^{1-y_1} p_L^{y_1}, \ldots, p_H^{1-y_K} p_L^{y_K})). \quad (18.2)$$

Note that, conditional on the (latent) random variables $\tilde{c}_1, \ldots, \tilde{c}_K$, the two procedures differ only in terms of the order of applying the group majority $g_n(\cdot)$ vs. the scenario function $\hat{f}_x(\cdot, \ldots, \cdot)$ – in analogy to the uncorrelated case.

As $\tilde{c}^1, \ldots, \tilde{c}^K$ are jointly independent, the error sequences $(\tilde{e}_i^k)_{i \in \mathbb{N}}$ are independent across premises. For every premise $k = 1, \ldots, K$, with probability p (resp., $1-p$), individual competences are raised (lowered) to p_H (resp., p_L). Thus, the probability of a correct premise majority is given by $pg_n(p_H) + (1-p)g_n(p_L)$. As this probability is identical for all premises and the corresponding majority events are independent, we have the following result.

[139] Note that, unless all realizations $y_1, \ldots, y_K$ are identical, these probabilities vary across premises.

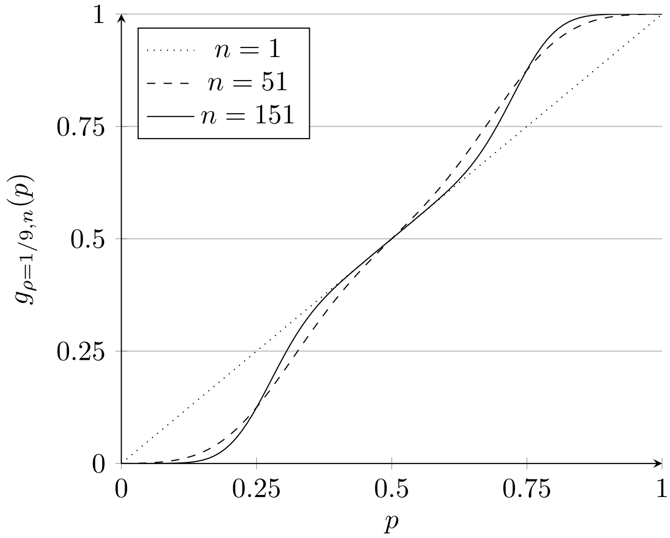

Figure 18.1: Majority function $g_{\rho,n}$ for $\rho = 1/9$ and different population sizes n

Proposition 9. *Equation* (18.1) *reduces to:* $P^x_{\rho,n}(p) = f_x(pg_n(p_H) + (1-p)p_L)$.

Technically speaking, the result is due to the fact that (i) the state function $\hat{f}_x(\cdot, \cdots, \cdot)$ is linear in all of its arguments (cf. Lemma 15 in the Appendix) and (ii) joint independence of $\tilde{c}^1, \cdots, \tilde{c}^K$ implies $\mathbb{P}(\tilde{c}^1 = y_1, \ldots, \tilde{c}^K = y_K) = p^{K-||y||_1}(1-p)^{||y||_1}$. In contrast, as $g_n(\cdot)$ is non-linear, $C^x_{\rho,n}(p)$ cannot be simplified along the lines of Proposition 9.

Due to its importance to the PBP, we define

$$g_{\rho,n}(p) := pg_n(p + \sqrt{\rho}(1-p)) + (1-p)g_n((1-\sqrt{\rho})p), \tag{18.3}$$

so that $P^x_{\rho,n}(p) = f_x(g_{\rho,n}(p))$.

18.1 Asymptotic Analysis

We start by considering the PBP. Clearly, the asymptotic behavior of $P^x_{\rho,n}(\cdot)$ is driven by that of $g_{\rho,n}(\cdot)$. Now, at every $p \in (0.5, 1]$, $g_{\rho,n}(p)$ is the weighted average of g_n evaluated at $p_L = p - \sqrt{\rho}p < p$ and $p_H = p + \sqrt{\rho}(1-p) > p$. As $n \to \infty$, g_n can only take on the values 0 or 1 (except at 0.5, cf. Figure 17.1). Since $0.5 < p < p_H$, we always have $g_n(p_H) \to 1$. Now if $0.5 < p_L$, then $g_n(p_L) \to 1$ hence $g_{\rho,n}(p) \to 1$. On the other hand, if $p_L < 0.5$, we have $g_{\rho,n}(p) \to (1-p)\cdot 0 + p \cdot 1 = p$. Lastly, if $0.5 = p_L$, then $g_n(p) \to p + (1-p)/2$. Note that we have $p_L = (1-\sqrt{\rho})p \leq 0.5 \iff p \leq 0.5/(1-\sqrt{\rho}) = 0.5 \cdot \big(1 + \sqrt{\rho}/(1-\sqrt{\rho})\big)$. This case always obtains when correlation is sufficiently high; namely when $0.25 \leq \rho \leq 1$.

The following Proposition is immediate (seeing that f_x is continuous):

Proposition 10. *Let* $p \in (0.5, 1]$*. For* $0 < \rho < \frac{1}{4}$*, let* $p^*_\rho := \frac{1}{2(1-\sqrt{\rho})}$*. Then, as* $n \to \infty$*,*

$$P^x_{\rho,n}(p) = f_x(g_{\rho,n}(p)) \to \begin{cases} 1 & p^*_\rho < p \leq 1 \\ f_x\left(p^*_\rho + (1-p^*_\rho)/2\right) & p = p^*_\rho \\ f_x(p) & 0.5 < p < p^*_\rho \end{cases}.$$

For $\frac{1}{4} \leq \rho \leq 1$*, we have* $\lim_{n\to\infty} P^x_{\rho,n}(p) = f_x(p)$ *for all* $p \in (0.5, 1]$*.*

Unlike in the uncorrelated case, the PBP does not satisfy the asymptotic part of a Jury Theorem. As $p^*_\rho = 1/(2(1-\sqrt{\rho})) > 0.5$ for every $\rho > 0$, it fails for competence levels close to 0.5. For all such $p \in (0.5, p^*_\rho)$, the asymptotic reliability of the PBP equals that of a single voter.

Consider the CBP next. Again, in the limit $n \to \infty$, $g_n(p)$ takes on the value 1 if $p > 0.5$, the value 0.5 if $p = 0.5$ and the value 0 if $p < 0.5$. So the probability of a correct CBP conclusion is given by the total probability that latent random variables produce conditional voter competences above 0.5 for the conclusion (plus one half times the total probability that they produce conditional voter competences of exactly 0.5 for the conclusion). That is,

$$\begin{aligned} C^x_{\rho,n}(p) = & \sum_{\substack{y\in\{0,1\}^K: \\ \hat{f}_x(p_H^{1-y_1} p_L^{y_1},\ldots,p_H^{1-y_K} p_L^{y_K})>0.5}} p^{K-||y||_1}(1-p)^{||y||_1} \\ + 0.5\cdot & \sum_{\substack{y\in\{0,1\}^K: \\ \hat{f}_x(p_H^{1-y_1} p_L^{y_1},\ldots,p_H^{1-y_K} p_L^{y_K})=0.5}} p^{K-||y||_1}(1-p)^{||y||_1}. \end{aligned}$$

Note that unless $p = 1$, all possible realizations $y_1, \ldots, y_K$ of the (latent) random variables $\tilde{c}_1, \ldots, \tilde{c}_K$ have strictly positive probability. Thus, Equation 18.2 implies that the CBP is asymptotically infallible in scenario $x \in X$ and for competence level $p \in (0.5, 1]$ if and only if all corresponding conditional conclusion competences of voters exceed 0.5. This is stated in the following Proposition.

Proposition 11. *Let* $p \in (0.5, 1]$*. Then* $C^x_{\rho,n}(p) \to 1$ *as* $n \to \infty$ *if and only if we have:* $\min_{y\in\{p_L,p_H\}^K} \hat{f}_x(y_1, \ldots, y_K) > 0.5$.

As in the case of uncorrelated voters, the asymptotic part of a Jury Theorem does not hold for the CBP.

In analogy to the case of uncorrelated voters, the CBP's asymptotic infallibility depends not only on voters' competence p and the correlation ρ but critically also on the concrete scenario $x \in X$ under consideration (note its presence in Proposition 11); hence, in particular, on the agenda. In contrast, Proposition 10 shows that this is not true for PBP. Whether, at some given level of voter competence p, it produces correct conclusion judgments with asymptotic certainty depends only the correlation ρ. The PBP is asymptotically infallible for all p such that $2p \leq 1/(1-\sqrt{\rho})$ (and for no p if $\rho \geq 1/4$).

While the PBP is asymptotically non-inferior to the CBP in the uncorrelated case no matter the scenario and agenda, this is not true any more in the presence of correlation. For example, reconsider the scenario $x = (0,0)$ for the agenda $C \leftrightarrow (P_1 \wedge P_2)$. Here, we have $\hat{f}_{x=(0,0)}(q_1, q_2) = 1-(1-q_1)(1-q_2)$. As $\hat{f}_{x=(0,0)}$ is increasing in both q_1 and q_2, it is minimal when q_1 and q_2 are. Therefore, at every $p \in (0.5, 1]$, the CBP asymptotically yields the correct result with certainty if and only if $\hat{f}_{x=(0,0)}(p_L, p_L) = f_{x=(0,0)}(p_L) = 1-(1-p(1-\sqrt{\rho}))^2 > 0.5$. That is, if and only if $p > (1-1/\sqrt{2})/(1-\sqrt{\rho})$. For example, when $\rho = 1/9$, this holds for all $p \in (0.5, 1]$.[140] In contrast, $p^*_{\rho=1/9} = 3/4$. Consequently, we have for all $p \in (0.5, 0.75)$: $\lim_{n\to\infty} C^{x=(0,0)}_{\rho=1/9,n}(p) = 1 > f_{x=(0,0)}(p) = 1-(1-p)^2 \lim_{n\to\infty} P^{x=(0,0)}_{\rho=1/9,n}(p)$.

On the other hand, for scenario $x = (1,1)$, we have $\hat{f}_{x=(1,1)}(q_1, q_2) = q_1 q_2$ which is increasing in both q_1 and q_2. Therefore, at every $p \in (0.5, 1]$, the CBP asymptotically yields the correct result with certainty if and only if $\hat{f}_{x=(1,1)}(p_L, p_L) = f_{x=(1,1)}(p_L) = p^2(1-\sqrt{\rho})^2 > 0.5$. That is, if and only if $p > 1/(\sqrt{2}(1-\sqrt{\rho}))$. Such $p \in (0.5, 1]$ exist only if $\rho < (3-2\sqrt{2})/2 < 0.1$. For example, if $\rho = 1/9$, we have $\lim_{n\to\infty} C^{x=(1,1)}_{\rho=1/9,n}(p) < 1$ for all $p \in (0.5, 1]$. In particular, for all $p \in (p^*_{\rho=1/9}, 1] = (0.75, 1]$, we have $\lim_{n\to\infty} C^{x=(1,1)}_{\rho=1/9,n}(p) < 1 = \lim_{n\to\infty} P^{x=(1,1)}_{\rho=1/9,n}(p)$. Moreover, note that, at every $p \in (0.5, 1]$, $\hat{f}_{x=(1,1)}$ is maximal when both (latent) random variables realize to $y_1 = y_2 = 0$ so as to increase voters' conditional competences for both premises to p_H. Consequently, if $\hat{f}_{x=(1,1)}(p_H, p_H) = f_{x=(1,1)}(p_H) < 0.5$, the CBP asymptotically yields the wrong result with certainty. For $0 < \rho < (\sqrt{2}-1)^2 \approx 0.17$, this is the case if and only if $p < (1/\sqrt{2}-\sqrt{\rho})/(1-\sqrt{\rho})$.[141] For example, if $\rho = 1/9$, $(1/\sqrt{2}-\sqrt{\rho})/(1-\sqrt{\rho}) \approx 0.56$. Consequently, we have for all $p \in (0.5, 0.56)$: $\lim_{n\to\infty} C^{x=(1,1)}_{\rho=1/9,n}(p) = 0 < p^2 = f_{x=(1,1)} = \lim_{n\to\infty} P^{x=(1,1)}_{\rho=1/9,n}(p)$.

As another example, recall that on the agenda $C \leftrightarrow (P_1 \leftrightarrow P_2)$, we have in all states $x \in \{0,1\}^2$: $\hat{f}(q_1, q_2) = q_1 q_2 + (1-q_1)(1-q_2)$. For all $q_1 = q_2 \neq 0.5$, we have $\hat{f}(q_1, q_2) > 0.5$. Thus, the CBP yields asymptotically infallible conclusion judgments at voters' competence level $p \in (0.5, 1]$ if and only if $p_L p_H + (1-p_L)(1-p_H) > 0.5$. That is, if and only if $p > 0.5 + \sqrt{\rho}/(2(1-\sqrt{\rho})) = 1/(2(1-\sqrt{\rho})) = p^*_\rho$. Moreover, if $p < p^*_\rho$, the CBP yields the correct conclusion with asymptotic probability $p^2 + (1-p)^2$ (the probability that $y_1 = y_2$). As $p^2 + (1-p)^2 = f_x(p)$, the CBP and the PBP are asymptotically equivalent on the agenda $C \leftrightarrow (P_1 \leftrightarrow P_2)$.[142]

So while the PBP's asymptotic reliability is reduced to that of a single voter for competence levels below a threshold that only depends on the correlation ρ, the CBP's asymptotic behavior is uncertain to the extent that it depends on the unknown truth scenario $x \in X$

[140] When $\rho = 1/9$, we have $(1-1/\sqrt{2})/(1-\sqrt{\rho}) \approx 0.44$.

[141] For $\rho \geq (\sqrt{2}-1)^2$, we have $\hat{f}_{x=(1,1)}(p_H, p_H) \geq 0.5$ for all $p \in (0.5, 1]$.

[142] At $p = p^*_\rho$, $C^x_{\rho,n}(p)$ approaches $0.5 \cdot (1 + p^2 + (1-p)^2)$ while the $P^x_{\rho,n}(p)$ approaches $0.5 \cdot (1 + p^2)$. Hence, strictly speaking, both approaches are asymptotically equivalent except at $p = p^*_\rho$ where the CBP is superior.

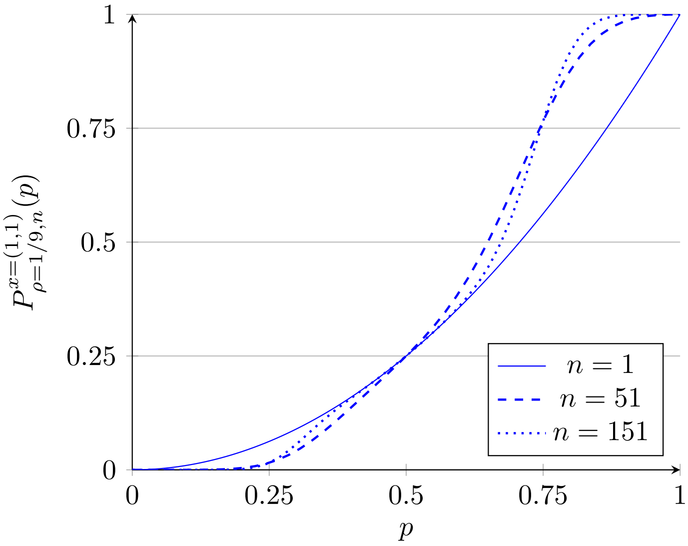

Figure 18.2: Reliability of the PBP in scenario $x = (1,1)$ on the agenda $C \leftrightarrow (P_1 \wedge P_2)$ for correlation $\rho = 1/9$

(hence, in particular, the agenda). This is familiar from our analysis for independent voters further above. However, interestingly, some of the variance in the CBP's reliability is reduced by correlation. Intuitively speaking, the CBP's reliability is sensitive to the underlying truth scenario because it depends on whether, in a given scenario, individual competences are such that voters reach the correct conclusion with probability larger than one half. Incorporating correlation through dependence of premise signals on latent random variables introduces more dispersion in terms of these conditional conclusion competences. So in cases when the CBP is wrong on the conclusion without correlation, there might now be some probability that it is right given correlation. The opposite holds for situations when the CBP is correct in the absence of correlation. Here, the presence of correlation might entail that the CBP is wrong with some probability. In this sense, correlation is a moderating force on the CBP. In contrast, the PBP relies on aggregating information at the premise level. Here, the presence of correlation makes premise majorities less reliable. As a consequence, individual competences have to exceed a critical value to guarantee correct premise judgments in the limit (note that for $p \in (0.5, p^*_\rho)$, $\lim_{n\to\infty} g_{\rho>0,n}(p) = p < 1 = \lim_{n\to\infty} g_{\rho=0,n}(p)$). Consequently, correlation tends to moderate differences in asymptotic behavior between the PBP and the CBP. At the same time, the main trade-off when considering both procedures remains the same. While the PBP is not asymptotically infallible in the presence of correlation, its reliability at a given level of voter competence is known. On the other hand, the CBP can produce asymptotically more reliable judgments in *some* scenarios and less reliable ones in others, so choosing it over the PBP necessitates bearing the uncertainty (or risk, in the case of known prior probabilities on truth scenarios) this brings.

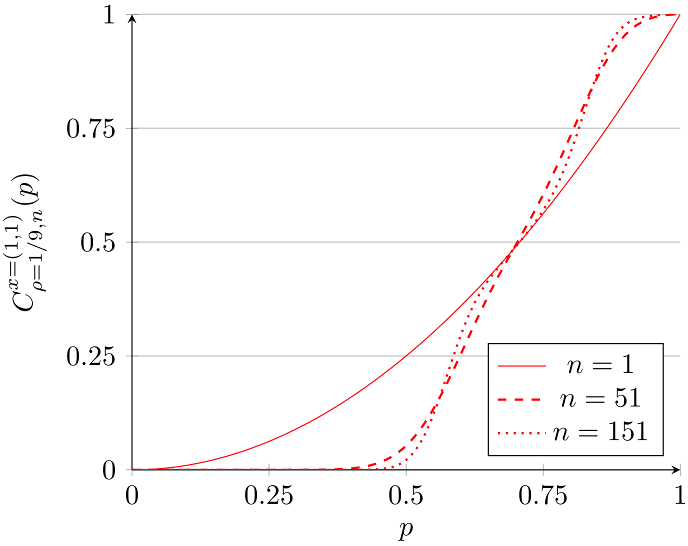

Figure 18.3: Reliability of the CBP in scenario $x = (1, 1)$ on the agenda $C \leftrightarrow (P_1 \wedge P_2)$ for correlation $\rho = 1/9$

18.2 The Finite Case

As we pointed out above, the behavior of the PBP depends largely on that of the group majority function in the presence of correlation $g_{\rho,n}(p)$. Interestingly, $g_{\rho,n}(p)$ may not be monotonically increasing with respect to voter population size n. Figure 18.1 illustrates the probability of a correct majority (premise) judgment $g_{\rho,n}(\cdot)$ for correlation coefficient $\rho = 0.1$ and different population sizes $n = 1$ (dotted), $n = 5$ (dashed) and $n = 51$ (solid). Note how for moderately competent voters, the reliability of majority judgments increases from $n = 1$ to $n = 5$ but decreases from $n = 5$ to $n = 51$. To see why this is the case, note that $g_{\rho,n}(p) = pg_n(p_H) + (1-p)g_n(p_L)$ is a convex combination of $g_n(p_L)$ and $g_n(p_H)$. As $p_H > p > 0.5$, $g_n(p_H)$ is increasing in n; unless $p_L > 0.5$ (that is, $p > p^*_\rho$), however, $g_n(p_L)$ is decreasing in n (cf. Equation 17.2). This implies that $g_{\rho,n}(p)$ is monotonically increasing in n for all $p \in (p^*_\rho, 1]$ but 'hump-shaped' for all $p \in (0.5, p^*_\rho)$ (given that $\rho \leq 1/4$; if $\rho > 1/4$, it is 'hump-shaped' for all $p \in (0.5, 1)$).[143]

This non-monotonicity (for competence levels close enough to 0.5) is in contrast to Dietrich and Spiekermann (2013) who salvage the non-asymptotic part of single-proposition Jury Theorems. The reason is that our distribution over voters' competence levels – p_H with probability p and p_L with probability $1-p$ – does not have their property of 'tendency to exceed 0.5' – which requires the probability mass at every $0.5+\epsilon$ to exceed that of $0.5-\epsilon$.

[143] For $p > 0.5$, p_H is closer to 1 than p_L is to 0 (note that $1 - p_H = (1-p)(1-\sqrt{\rho}) < p(1-\sqrt{\rho}) = p_L$). As p gets closer to one of the end, more of of the increase (decrease) of $g_n(p)$ in n happens for smaller n (that is, the rate of increase/decrease $(g_{n+1}(p) - g_n(p))/(g_\infty(p) - g_1(p))$ is more strongly decaying in n as p is closer to one of the ends; also cf. Figure 17.1). This implies that for small n, $g_n(p_H)$ dominates and $g_{\rho,n}(p)$ is increasing in n, while for large n, $g_n(p_H)$ becomes dominant and $g_{\rho,n}(p)$ eventually decreases in n.

Our model fails to have that property as p_L and p_H are generally not symmetric around 0.5; that is, $p_H - 0.5 \neq 0.5 - p_L$.

Not surprisingly, the PBP fails to grow more reliable as the number of voters increases. In the presence of correlation, this has two potential causes. First, analogous to the case of uncorrelated voters, if scenario $x \in X$ does not induce increasing competence (that is, if $f_x(\cdot)$ is not increasing on $(0.5, 1]$), then a higher probability of judging the premises correctly may not translate into an increased probability of judging the conclusion correctly. This means that $P^x_{\rho,n}(p) = f_x(g_{\rho,n}(p))$ may fail to increase in the number of voters n even if $g_{\rho,n}(p)$ does. Second, if scenario $x \in X$ does induce increasing competence, then $P^x_{\rho,n}(p)$ is increasing in n if and only $g_{\rho,n}(p)$ is. As we observed above, however, $g_{\rho,n}(p)$ fails to grow in n (for all n) for competence levels close enough to 0.5 (or when correlation $\rho \geq 1/4$). For this reason, non-monotonicity in the number of voters is a generic property of the PBP in our model of correlated voters. For example, Figure 18.2 plots $P^x_{\rho=1/9,n}(p)$ in scenario $x = (1, 1)$ for a conjunctive agenda with two premises and $n = 1$, $n = 51$ and $n = 151$ voters. Here, we have $p^*_{\rho=1/9} = 3/4$. Note how $P^{x=(1,1)}_{\rho=1/9,n}(p)$ increases from $n = 1$ (solid blue) to $n = 51$ (dashed blue) but decreases from $n = 51$ to $n = 151$ (dotted blue) for all $p \in (0.5, 0.75)$.

Consider the CBP next. Its reliability $C^x_{\rho,n}(p)$ in scenario $x \in X$ is a (probability-)weighted average of $g_n(\cdot)$ evaluated at all possible conditional conclusion competences $\hat{f}_x(p_H^{1-y_1} p_L^{y_1}, \ldots, p_H^{1-y_K} p_L^{y_K})$. Unless all of them are above (below) 0.5 at a given competence level p, $C^x_{\rho,n}(p)$ will not be monotonically increasing (decreasing) in n. In other words, the CBP satisfies the finite part of a Jury Theorem at some $p \in (0.5, 1]$ if and only if it satisfies the asymptotic part (cf. Proposition 11). As in our discussion for the asymptotic part above, this critically depends on the concrete scenario $x \in X$ under consideration. For example, reconsider the conjunctive agenda with two premises. Figure 18.3 plots $C^x_{\rho=1/9,n}(p)$ in scenario $x = (1, 1)$ for $n = 1$ (solid red), $n = 51$ (dashed red) and $n = 151$ (dotted red) voters. Clearly, the CBP fails to grow more reliable for all $p \in (0.5, 1]$ in the scenario under consideration. On the other hand, recall that in scenario $x = (0, 0)$ the CBP is asymptotically correct with certainty for all $p \in (0.5, 1]$ hence also grows more reliable as n increases.

PBA vs. CBA

Table 18.1 summarizes the characterizing conditions for the finite and asymptotic part of a Jury Theorem to hold for both the PBP and the CBP. As regards the finite part for the PBA the stated conditions are sufficient conditions.[144]

[144] We conjecture that these conditions are necessary as well. However, this is not immediately obvious. Consider the case of independent voters. Intuitively, if $f_x(p'') \leq f_x(p')$ for some $p'' > p' > 0.5$ and p is such that $g_n(p) = p'$, then increasing n to $n' > n$ such that $g_{n'}(p) = p''$ would yield the desired result. However, increasing n leads to *discrete* increases in $g_n(p)$ hence such n' may not exist at a given p. On the other hand, note that said discrete increases get arbitrarily small as n grows. Moreover, for every

	finite	infinite
$\rho = 0$	**pba:**[s] f_x increasing on $(p, 1]$ **cba:** $f_x(p) > 0.5$	**pba:** $p > 0.5$ **cba:** $f_x(p) > 0.5$
$\rho > 0$	**pba:**[s] $p > p^*_\rho$ and f_x increasing on $(p, 1]$ **cba:** $\min_{y \in \{p_L, p_H\}^K} \hat{f}_x(y) > 0.5$	**pba:** $p > p^*_\rho$ **cba:** $\min_{y \in \{p_L, p_H\}^K} \hat{f}_x(y) > 0.5$

Table 18.1: Characterizing ([s] sufficient) conditions for whether the (in)finite part of a CJT holds at $p \in (0.5, 1]$ given correlation $0 \leq \rho \leq 1$, $p^*_\rho = 1/(2(1 - \sqrt{\rho}))$, $p_L = (1 - \sqrt{\rho})p$ and $p_H = \sqrt{\rho} + (1 - \sqrt{\rho})p$

As in the case of independent voters, the asymptotic properties of the PBP are independent of the concrete scenario and agenda under consideration. However, in the presence of correlation, the PBP fails to uncover the true conclusion with asymptotic certainty in general. For it to hold, voters' (average) competence needs to surpass a critical level which increases in the extent of correlation. At competence levels below it or if correlation is too great, the PBP is asymptotically equivalent to a single voter. For example, if correlation $\rho = 1/4$, the PBP asymptotically behaves like a single voter at all competence levels. As in this case half of all voters' premise signals are determined by the corresponding latent random variable, which yields correct evidence with the same probability with which single voters judge the premise correctly, this is hardly surprising.

Again, as for independent voters, the CBP's asymptotic reliability critically depends on the truth scenario $x \in X$ and the agenda that is being analyzed. For some agendas, there are scenarios for which the CBP satisfies the asymptotic part of a Jury Theorem; namely those scenarios $x \in X$ in which all possible combinations of conditional competences for the premises lead to competent conclusion judgments (for example, see our discussion above for $\rho = 1/9$ in scenario $x = (0, 0)$ on the agenda $C \leftrightarrow (P_1 \wedge P_2)$). However, such scenarios do not exist for every agenda. For example, reconsider the agenda $C \leftrightarrow (P_1 \leftrightarrow P_2)$ when correlation is $\rho = 1/9$. Recall that, here, the scenario function $\hat{f}_x(q_1, q_2)$ is identical for all scenarios $x \in \{0, 1\}^2$ and such that the CBP yields asymptotically infallible conclusions only for competence levels $p \in (0.75, 1]$. In this case, the same is true for the PBP. Thus, as for the case no correlation, the two approaches exhibit the same asymptotic properties. Interestingly, however, correlation may change how the two approaches compare in a given finite sample of voters. Figure 18.4 plots the reliability of both procedures for $n = 51$ voters. While, the PBP (blue) is superior to the CBP (red) at all competence levels given voters are uncorrelated (solid), the CBP is superior for competence levels closer to 0.5 when correlation $\rho = 1/9$ is present (dashed).

$n \in \mathbb{N}$, there exists some $p \in (0.5, 1]$ such that $g_n(p) = p'$ (seeing that $g_n(0.5) = 0.5$, $g_n(1) = 1$ and g_n is continuous).

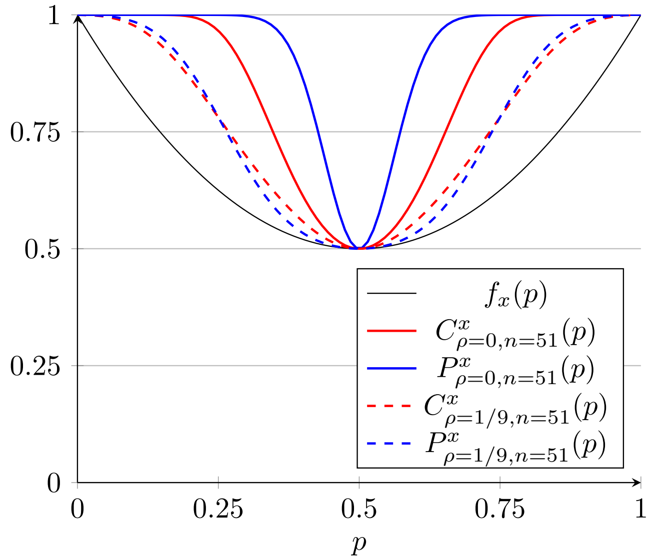

Figure 18.4: PBP (blue) vs. CBP (red) on the agenda $C \leftrightarrow (P_1 \leftrightarrow P_2)$ for correlation $\rho = 0$ (solid) and $\rho = 1/9$ (dashed) and $n = 51$ voters

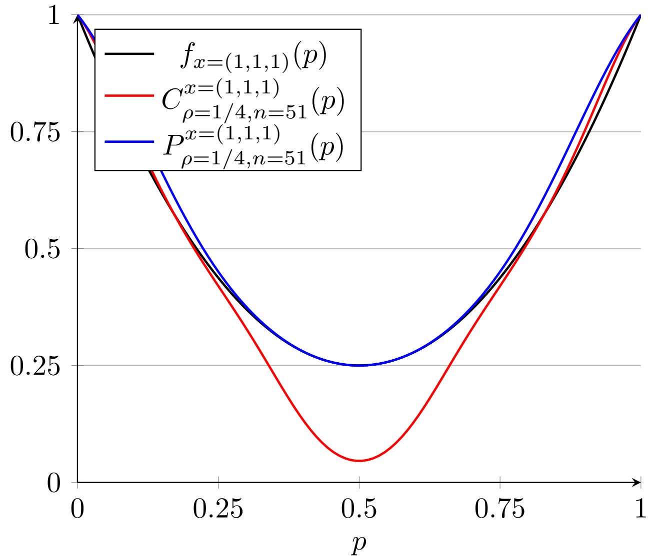

Figure 18.5: PBP (blue) vs. CBP (red) in scenario $x = (1, 1, 1)$ of the agenda $C \leftrightarrow (P_1 \leftrightarrow P_2 \leftrightarrow P_3)$ for correlation $\rho = 1/4$ and $n = 51$ voters

At the same time, there exist scenarios in which the PBP (remains) superior at all competence levels. As an example, Figure 18.5 plots the reliability of both procedures in scenario $x = (1, 1, 1)$ on the agenda $C \leftrightarrow (P_1 \leftrightarrow P_2 \leftrightarrow P_3)$ for correlation $\rho = 1/4$ and $n = 51$ voters.

19 Conclusion

We compared the (epistemic) reliability of the Premise Based Procedure (PBP) and the Conclusion Based Procedure (CBP) on truth-functional agendas with correlated voters. We modeled correlation through the presence of binary latent random variables at the premise level which were assumed to raise/lower the conditional competence (that is, the probability of being correct) on premises above/below the average level reflecting common access to a (binary) source of information that provides factually correct/incorrect evidence. Conditional on the latent random variables, voters were assumed to be independent. Our analysis was conducted separately for each possible true state of the world (truth scenario) of a given agenda and did not (necessarily) assume a prior probability distribution on those states.

In the absence of correlation the PBP is guaranteed to yield the correct conclusion judgment for competent voters with asymptotic (that is, when the number of voters tends to infinity) certainty no matter the truth scenario or the agenda under consideration. In general, this is not true for the CBP. While it may be asymptotically infallible and outperform the PBP for finitely many voters in a lot of truth scenarios of a given agenda, it can lead to wrong conclusions with asymptotic certainty and be severely outperformed by the PBP in finite samples in other truth scenarios. This makes both the finite-sample and asymptotic properties of the CBP highly uncertain. Importantly, this uncertainty may be disguised when considering weighted behavior that relies on assigning all truth scenarios the same prior probability given that the CBP performs well in a majority of them.

Broadly speaking, this comparative result carries over to the case of correlated voters. In the presence of correlation, the PBP is not guaranteed to be asymptotically infallible in general but requires voters' competence to be above a threshold level (strictly larger than one half and depending on the extent of correlation). As for the uncorrelated case, this result holds true for every truth scenario on every agenda. Likewise, the CBP's aptness as an epistemic procedure for correlated voters is highly sensitive to the truth scenario under consideration. Interestingly, however, some of the variance in its reliability is moderated by the presence of correlation. Moreover, while a choice between the CBP and the PBP pits the CBP against an asymptotically infallible procedure in the case of independent voters, this no longer holds for correlated voters. Ultimately, a decision between the procedures would depend on the agenda under consideration and societal attitudes towards uncertainty or – given a prior distribution on truth scenarios – the combined probability of scenarios in which the CBP performs badly and societal risk attitudes.

Our analysis relies on the simplifying assumption that latent random variables are binary and yield correct evidence with the same probability with which voters judge premises correctly. We believe that studying if and to what extent our findings carry over to more general distributional assumptions would be both an interesting and important task for future research. Our analysis suggests that the reliability of both approaches critically hinges on the probability with which latent random variables raise voters' competence above one half as so to enable a classical Jury Theorem to be at work (at the premise or conclusion level).

Appendix

A Appendix to Part II

A.1 Relation to Temptation Models of Self-Control

In the axiomatic treatment of self-control problems pioneered by Gul and Pesendorfer (2001) self-control problems result from temptations which must be resisted at the choice stage. When choosing some alternative $x \in A$, the decision maker incurs a self-control cost from foregoing the most tempting alternative in A as measured in terms of a temptation utility. Their model and two prominent generalizations thereof are special cases of our self-control-cost representation

$$U(A) = \max_{x \in A} u(x) - C(x, A) \tag{$\star\star$}$$

when (1) the choice-stage utility v is interpreted as temptation and (2) self-control costs depend only on maximal temptations in A (and the associated shortfall in temptation utility).

1. **Gul and Pesendorfer (2001)**
   develop the base case in which costs are given by the temptation utility foregone by not choosing the most tempting alternative:

   $$C(x, A) = \max_{y \in A} v(y) - v(x).$$

2. **Takeoka (2008); Noor and Takeoka (2010)**
   allow for marginal costs to be increasing. That is, they consider

   $$C(x, A) = \phi\left(\max_{y \in A} v(y) - v(x)\right)$$

   for some strictly increasing and *convex* function $\phi(\cdot) \geq 0$ with $\phi(0) = 0$.

3. **Noor and Takeoka (2015)**
   consider menu-dependent costs

   $$C(x, A) = \psi\left(\max_{y \in A} v(y)\right)\left(\max_{y \in A} v(y) - v(x)\right)$$

   for some (weakly) increasing $\psi(\cdot) \geq 0$ such that $\psi(l) > 0$ for all $l > \min_{x \in X} v(x)$.

We verify the properties of a self-control-cost function (cf. Definition 2) for $C(x, A)$ as defined above.

1. **Gul and Pesendorfer (2001)**
   As
   $$C(x, A) = \max_{y \in A} v(y) - v(x)$$
   is a special case of both Takeoka (2008); Noor and Takeoka (2010) and Noor and Takeoka (2015), the proof is included there.

2. **Takeoka (2008); Noor and Takeoka (2010)**
   Let
   $$C(x, A) = \phi\left(\max_{y \in A} v(y) - v(x)\right)$$
   for some strictly increasing and *convex* function $\phi(\cdot) \geq 0$ with $\phi(0) = 0$.

   a) Clearly, $\phi(\max_{y \in A} v(y) - v(x)) > 0$ only if $\max_{y \in A} v(y) - v(x) > 0$; thus, $v(y) > v(x)$ for some $y \in A$. Vice versa, if there exists some $y \in A$ such that $v(y) - v(x)$, then $\max_{y \in A} v(y) - v(x) > 0$. As ϕ is strictly increasing and $\phi(0) = 0$, we must have $C(x, A) = \phi(\max_{y \in A} v(y) - v(x)) > 0$.

   b) As ϕ is non-decreasing, we have $\phi(\max_{y \in A} v(y) - v(x)) \leq \phi(\max_{y \in A \cup B} v(y) - v(x))$.

   c) We show that $C(x, x \cup A) \leq C(y, A)$ whenever $v(x) > v(y)$ which is sufficient. Consider two cases. If $v(x) \geq \max_{z \in A} v(z)$, then $C(x, x \cup A) = \phi(0) = 0$ and the claim is obvious. Else, $\max_{z \in x \cup A} v(z) = \max_{z \in A} v(z)$. As ϕ is non-decreasing, $C(x, x \cup A) = \phi(\max_{z \in A} v(z) - v(x)) \leq \phi(\max_{z \in A} v(z) - v(y)) = C(y, A)$.

3. **Noor and Takeoka (2015)**
   Let
   $$C(x, A) = \psi\left(\max_{y \in A} v(y)\right)\left(\max_{y \in A} v(y) - v(x)\right)$$
   for some non-decreasing $\psi(\cdot) \geq 0$.

   a) Clearly, $\psi(\max_{y \in A} v(y))(\max_{y \in A} v(y) - v(x)) > 0$ only if $\max_{y \in A} v(y) - v(x) > 0$; thus, $v(y) > v(x)$ for some $y \in A$. Vice versa, if there exists some $y \in A$ such that $v(y) - v(x)$, then $\max_{y \in A} v(y) > v(x) \geq \min_{z \in X} v(z)$. Thus, $\psi(\max_{y \in A} v(y)) > 0$. Consequently, $C(x, A) = \psi(\max_{y \in A} v(y))(\max_{y \in A} v(y) - v(x)) > 0$.

   b) As ψ is non-decreasing, $\psi(\max_{y \in A} v(y))(\max_{y \in A} v(y) - v(x)) \leq \psi(\max_{y \in A \cup B} v(y))$ $(\max_{y \in A} v(y) - v(x)) \leq \psi(\max_{y \in A \cup B} v(y))(\max_{y \in A \cup B} v(y) - v(x))$.

   c) We show that $C(x, x \cup A) \leq C(y, A)$ whenever $v(x) > v(y)$ which is sufficient. Consider two cases. If $v(x) \geq \max_{z \in A} v(z)$, then $C(x, x \cup A) = \psi(\max_{z \in A} v(z)) \cdot 0 = 0$ and the claim is obvious. Else, $\max_{z \in x \cup A} v(z) =$

$\max_{z\in A} v(z)$. Thus, $C(x, x\cup A) = \psi(\max_{z\in A} v(z))(\max_{z\in A} v(z) - v(x)) \leq \psi(\max_{z\in A} v(z))(\max_{z\in A} v(z) - v(y)) = C(y, A)$.

Noor and Takeoka (2010) consider our more general representation ($\star\star$) in their introduction but give no axiomatic foundation for it. However, they do axiomatize what they call a 'general self-control' model where $C(x, A) = \tau\left(x, \max_{y\in A} v(y)\right)$ for some $\tau(\cdot,\cdot)$ that is weakly increasing in its second argument and satisfies: (i) $\tau(x, v(y)) > 0 \implies v(x) < v(y)$; (ii) $[u(x) > u(y)$ and $v(x) < v(y)] \implies \tau(x, v(y)) > 0$. This does not define a self-control-cost function in the sense of Definition 2. While, like the models presented above, it is more specific than our model to the effect that it allows costs only to depend on the temptation maximum (instead of all temptations in menu A), it puts less restrictions on the structure of costs at any given menu. For example, it allows for the possibility that $\tau(x, \max_{z\in A} v(z)) < \tau(z, \max_{z\in A} v(z))$ for $\{x, y\} \in A$ even if $x \gtrless y$ (that is, if $v(x) < v(y)$ and $u(x) > u(y)$). Thus, the less tempting alternative x may incur a smaller self-control cost (even though $\max_{z\in A} v(z) - v(x) > \max_{z\in A} v(z) - v(y)$). We would argue that this is inconsistent with the intuition of self-control costs being caused by resisting temptation.

While sharing the axiomatic approach, the temptation models above differ from our work as they are developed in a lottery setting and impose the Independence Axiom.[145] We do not need to invoke the expected utility assumption. A notable exception in this regard is Gul and Pesendorfer (2006). They axiomatize a generic temptation model in a finite choice setting. Under a regularity condition, they derive a 'strict and generic' representation $U(A) = \zeta(\max_{x\in A} \omega(x), \max_{y\in A} v(y))$ where $\zeta(\cdot,\cdot)$ is strictly increasing in the first and strictly decreasing in the second argument. Thus, we can equivalently write $U(A) = \max_{x\in A} \zeta(\omega(x), \max_{y\in A} v(y))$. Again, this may be put in the form of ($\star\star$) by letting $C(x, A) = \zeta(\omega(x), v(x)) - \zeta\left(\omega(x), \max_{y\in A} v(y)\right)$ and noting that $u(x) = \tau(\omega(x), v(x))$. As for the general model in Noor and Takeoka (2010), however, this does not define a self-control-cost function (cf. Definition 2). Again, it is more specific than our model to the effect that costs may depend only on maximal temptations but less demanding on the cost structure at a given menu A such that – somewhat counter-intuitively – costs may be smaller for a less tempting alternative.[146]

Our self-control-cost representation ($\star\star$) lends itself to an indirect utility interpretation where menu A is evaluated by $U(A) = \max_{x\in A} u(x, A)$, the utility received from the best alternative according to $u(x, A) = u(x) - C(x, A)$. Gul and Pesendorfer (2001) point out that this allows to model a desire for commitment (and the underlying problems of self-control) without the need to invoke time-inconsistent preferences. Rather, self-control problems are explainable in terms of preferences being menu-dependent. The experienced

[145] Further examples include Dekel et al. (2009) and Stovall (2010) who generalize G&P so as to include subjective uncertainty about temptations; Stovall (2018) additionally introduces subjective uncertainty about commitment utility. Since we model choice-stage behavior in terms of choice *functions*, our model does not incorporate uncertainty about choices.

[146] Specifically, when $\{x, y\} \subseteq A$ such that $x \gtrless y$, we may have $C(x, A) = \zeta(\omega(x), v(x)) - \zeta\left(\omega(x), \max_{z\in A} v(z)\right) < \zeta(\omega(y), v(y)) - \zeta\left(\omega(y), \max_{z\in A} v(z)\right) = C(y, A)$ if $\omega(x) < \omega(y)$.

utility of choosing x from some menu A (which presents the decision maker with self-control problems) is less than that of receiving x choice-free (i.e. from the commitment menu $\{x\}$). This creates a desire for commitment even if the decision maker's (menu-dependent) preferences remain unchanged as time passes.

In contrast, our PTSC representation ($\star$) has a natural interpretation in terms of dynamic inconsistency. We may think of the decision maker as if anticipating a change of preference from u to v. Yet, unlike for Strotz models (Strotz, 1955; Gul and Pesendorfer, 2005), PTSC agents may not just sit idly by but can constrain their own choices (made according to v) to conform to a previously made plan. As Theorem 2 above shows, this interpretation is indistinguishable from the time-consistent (indirect utility) version in terms of the behavioral observables in our model (menu preferences and choice from menus). Thus, temptation models of self-control are equally consistent with a dynamically inconsistent decision maker optimally planning to self-control. It is interesting to note, however, that a prediction that is implicit in our PTSC representation is that observed choices from menus should purely reflect choice-stage preferences (as represented by v) when the planning-stage is absent. In other words, choice behavior may systematically differ when decision makers have previously had the chance to plan (optimal self-control) as compared to a situation when they are not (no self-control). This would be irreconcilable with the alternative view of menu-dependent yet time-consistent preferences.

While most temptation models do not model choice from menus explicitly, the indirect utility interpretation for preference over menus $U(A) = \max_{x \in A} u(x, A)$ suggests choice according to $u(x, A)$. For example, in Gul and Pesendorfer (2001), where $u(x, A) = u(x) + v(x) - \max_{y \in A} v(y)$, the decision maker can be thought of as choosing according to $u(x) + v(x)$ (a compromise between commitment and temptation utility). Thus, as choice is rationalizable by a standard (menu-independent) utility function, it satisfies WARP. The convex (Takeoka, 2008; Noor and Takeoka, 2010) and menu-dependent (Noor and Takeoka, 2015) models allow for violations of WARP. However, both generalizations only allow for cost specifications that feature cost differentials which are increasing in menu size. As a consequence, these models can produce failures of WARP related to decreasing self-control but cannot incorporate those connected to increasing self-control.

To make these statements precise, let $x \gtrless y$. We say that choice function $c(\cdot)$ exhibits *decreasing self-control* if there exist $A, B \in \mathcal{A}$ with $\{x, y\} \subseteq A$ such that $x = c(A)$ and $y = c(A \cup B)$. While x is chosen from menu A (using costly self-control), the decision maker loses self-control at the larger menu $A \cup B$. Note that this constitutes a violation of WARP (Sen's Condition α). Analogously, we say that $c(\cdot)$ exhibits *increasing self-control* if there exist $A, B \in \mathcal{A}$ with $\{x, y\} \subseteq A$ such that $y = c(A)$ and $x = c(A \cup B)$. Here, the decision maker gains self-control as she moves from A to the larger menu $A \cup B$. Again, this violates WARP. As we argue in the main text the latter violation is of particular behavioral interest as it reflects an exercise of self-control that is positively responsive to the *magnitude* or *stakes* of the decision problem. Intuitively, when there is little at stake, the returns to self-control might be too small to put up the effort. As the stakes increase, however, so do the benefits associated with self-control eventually making it worthwhile. As we write in the main text above, such a *magnitude* effect is empirically well established in intertemporal settings with choice between a smaller, sooner reward and a larger, later reward, for example.

To see that the aforementioned temptation models are unable to capture this, consider the cost differential for two alternatives such that $x \gtrless y$; that is, $u(x) > u(y)$ and $v(x) < v(y)$. For the base model in Gul and Pesendorfer (2001), it is given by $\max_{z \in A} v(z) - v(x) - (\max_{z \in A} v(z) - v(y)) = v(y) - v(x)$, a constant that is independent of the menu. Thus, either x is the better choice globally (if $u(x) - u(y) > v(y) - v(x)$) or else y is (hence no violation of WARP is possible). For the convex and menu-dependent model, the cost differential is $\phi(\max_{z \in A} v(z) - v(x)) - \phi(\max_{z \in A} v(z) - v(y))$ and $\psi(\max_{z \in A} v(z))(v(y) - v(x))$ respectively. Due to the convexity of ϕ and the non-decreasingness of ψ both increase when alternatives are added to menu A. In contrast, our definition of a self-control-cost function in representation ($\star\star$) allows for choices to exhibit increasing self-control. For a particularly simple example, consider the case of a fixed cost $k > 0$. We discuss this case in detail and consider several applications to intertemporal choice in Section 4 above. Importantly, a fixed of self-control are inconsistent with the temptation models above. Note that –considered as a function over foregone temptation utility– it is discontinuous at 0. This also introduces a non-convexity.

A.2 Proof of Theorem 7

A.2.1 Sufficiency of Axioms 0.1-3

We define for all $x, y \in X$: $x \geq y :\iff [x \gg y \textit{ or } y \gtrless x \textit{ or } x = y]$. Remember that $x \gg y :\iff x \sim \{x, y\} \succ y$ and $y \gtrless x :\iff y \succ \{x, y\} \succsim x$.

Lemma 3. *$\geq$ is a linear order on X.*

Proof. By definition, $\gg$ and $\gtrless$ are asymmetric and irreflexive. The identity relation $=$ is symmetric and reflexive. Thus, $x \succ y \iff [y \gtrless x$ or $x \gg y]$ and $x = y \iff x = y$ (where $\succ$ and $=$ denote the asymmetric and symmetric part of $\geq$). By consequence, $\geq$ is anti-symmetric.

As $\gg$, $\gtrless$ and $=$ are transitive, so is their disjunction $\geq$.

Lastly, to see completeness, let $x, y \in X$. If $x = y$, then $x \geq y$. Else, $x \neq y$. As the restriction of $\succ$ to singletons is a linear order, we have $x \succ y$ or $y \succ x$. W.l.o.g. consider the first case. By Axiom 2, $\{x, y\} \succsim y$. If $x = c(\{x, y\})$, then $x \succsim \{x, y\}$ by Axiom 1. Else if $y = c(\{x, y\})$, then $x \succ y \succsim \{x, y\}$ by Axiom 1. Thus, $x \succsim \{x, y\}$ in both cases. Consequently, we have $x \succsim \{x, y\} \succsim y$. As $x \succ y$, we have either $x \sim \{x, y\} \succ y$ or $x \succ \{x, y\} \succsim y$. Thus, $x \succ y$ or $y \succ x$. □

As $\geq$ is a linear order on X, there exists a utility function $v : X \to \mathbb{R}$ representing it.

Lemma 4. *There exists a utility function $U : \mathcal{A} \to \mathbb{R}$ representing $\succsim$ such that for all $x, y \in X$ and all $B, C \in \mathcal{A}$ for which $y \gtrless x$, $y = c(y \cup C)$ and $B \subseteq C$ hold, we have $U(x) - U(x \cup B) \leq U(y) - U(y \cup C)$.*

Proof. As $\succsim$ is a weak order and $\mathcal{A}$ is finite, there exists some $\widetilde{U} : \mathcal{A} \to \mathbb{R}$ representing it.[147] Defining $U(A) = \exp(\gamma \cdot \widetilde{U}(A))$, we show that the desired property holds if $\gamma > 0$ is appropriately chosen. Note that U – it being a positive monotone transformation of $\widetilde{U}$ – represents $\succsim$.

Let $x, y \in X$ and $B, C \in \mathcal{A}$ be as stated above. By Axiom 1, $U(y) \geq U(y \cup C)$.

Consider two cases: (1) If $\widetilde{U}(x) - \widetilde{U}(x \cup B) \leq 0$, we have $U(x) - U(x \cup B) \leq 0 \leq U(y) - U(y \cup C)$ for every $\gamma > 0$. In this case, let $\gamma_{x,y,B,C} = 1 > 0$. (2) Else $\widetilde{U}(x) - \widetilde{U}(x \cup B) > 0$. By Axiom 2, $x \gtrless z$ for some $z \in B \subseteq C$; thus, $y \gtrless x \gtrless z$. By transitivity (Axiom 0.2), $y \gtrless z$. By Axiom 1, $\widetilde{U}(y \cup C) \leq \widetilde{U}(\{y, z\}) < \widetilde{U}(y)$. Consequently, there exists some $\bar{U} \in \mathbb{R}$ such that $\widetilde{U}(y \cup C) < \bar{U} < \widetilde{U}(y)$ and $\widetilde{U}(x) < \bar{U}$. For every $\gamma > 0$, we have $U(y) - U(y \cup C) > \exp(\gamma \widetilde{U}(y)) - \exp(\gamma \bar{U})$. Moreover, strict monotonicity, convexity and continuous differentiability of $\exp(\gamma \cdot)$ imply $\exp(\gamma \widetilde{U}(y)) - \exp(\gamma \bar{U}) \geq \gamma \exp(\gamma \bar{U})(\widetilde{U}(y) - \bar{U})$ and $\gamma \exp(\gamma \widetilde{U}(x))(\widetilde{U}(x) - \widetilde{U}(x \cup B)) \geq U(x) - U(x \cup B)$. Thus, $U(y) - U(y \cup C) > U(x) - U(x \cup B)$ if

$$\gamma \exp(\gamma \bar{U})(\widetilde{U}(y) - \bar{U}) \geq \gamma \exp(\gamma \widetilde{U}(x))(\widetilde{U}(x) - \widetilde{U}(x \cup B))$$
$$\iff \gamma \geq \frac{1}{\bar{U} - \widetilde{U}(x)} \ln\left(\frac{\widetilde{U}(x) - \widetilde{U}(x \cup B)}{\widetilde{U}(y) - \bar{U}}\right).$$

Let $\gamma_{x,y,B,C} = \max\left\{\frac{1}{\bar{U} - \widetilde{U}(x)} \ln\left(\frac{\widetilde{U}(x) - \widetilde{U}(x \cup B)}{\widetilde{U}(y) - \bar{U}}\right), 1\right\} > 0$.

Thus, the desired property holds for every $\gamma \geq \max_{x,y,B,C} \gamma_{x,y,B}, C > 0$ where x, y, B, C are as stated in the lemma. □

For every $x \in X$, define $u(x) = U(x)$ where U is as stated in Lemma 4. For all $P \in \mathcal{A}$, define $x_P = \max(P, v) = \operatorname{argmax}_{y \in P} v(y)$. We use Lemma 4 to define a planning-cost function.

Lemma 5. *Let $\succsim$ be represented by U as given through Lemma 4. For all $(P \subseteq A, A) \in \mathcal{A} \times \mathcal{A}$, let $\mathcal{Y}(P, A) = \{(y, C \supseteq \{z \in A \backslash P : v(z) > v(x_P)\}) \in X \times \mathcal{A} : [y \gtrless x_P$ or $y = x_P]$ and $y = c(y \cup C)\}$ and define*

$$\kappa(P, A) = \begin{cases} K & \mathcal{Y}(P, A) = \emptyset \\ \min\limits_{(y,C) \in \mathcal{Y}(P,A)} (U(y) - U(y \cup C)) & \mathcal{Y}(P, A) \neq \emptyset \end{cases}$$

where $K > \max\limits_{A,B \in \mathcal{A}} |U(A) - U(B)| \geq 0$.

Then, κ is a planning-cost function.

[147] For example, let $U(A) = |WT(A)|$; where $WT(A) = \{B \in \mathcal{A} : A \succsim B\} \subseteq \mathcal{A}$.

Proof. Let $\emptyset \neq P \subseteq A \in \mathcal{A}$. Note that for all $(y, C) \in \mathcal{Y}(P, A)$: $U(y) - U(y \cup C) \geq 0$ by Axiom 1. Thus, $\kappa(P, A) \geq 0$. We verify the properties of a planning-cost function:

1. First, suppose that $\{z \in A\backslash P : v(z) > v(x_P)\} = \emptyset$. As $c(\{x_P\}) = x_P$, we have $(x_P, \emptyset) \in \mathcal{Y}(P, A)$. Hence $\kappa(P, A) \leq U(x_P) - U(x_P) = 0$. Second, suppose there exists some $z \in A\backslash P$ such that $v(z) > v(x_P)$. If $\mathcal{Y}(P, A) = \emptyset$, the claim is immediate. Else, consider any $(y, C) \in \mathcal{Y}(P, A)$. Note that $v(y) \leq v(x_P) < v(z)$. By Axiom 3, we cannot have $u(y) < u(z)$. Thus, $u(y) > u(z)$. Seeing that $y \gtrless z \in B$, we have by Axiom 1: $U(y) > U(\{y, z\}) \geq U(y \cup C)$ hence $U(y) - U(y \cup C) > 0$. As this holds for all $(y, C) \in \mathcal{Y}(P, A)$, we have $\kappa(A, P) = \min_{(y,C)\in\mathcal{Y}(P,A)}(U(y) - U(y \cup C)) > 0$.

2. If $x_P \in A$, then $x_{P\cap A} = x_P$. If $\mathcal{Y}(P, A \cup B) = \emptyset$, then $\kappa(P \cap A, A) \leq K = \kappa(P, A \cup B)$. Else, consider any $(y, C) \in \mathcal{Y}(P, A \cup B)$. Then, $y = x_P = x_{P\cap A}$ or $y \gtrless x_P = x_{P\cap A}$, $y = c(y \cup C)$ and $C \supseteq \{z \in (A \cup B)\backslash P : v(z) > v(x_P)\} \supseteq \{z \in A\backslash(P \cap A) : v(z) > v(x_{P\cap A})\}$ seeing that $A\backslash(P \cap A) = A\backslash P \subseteq (A \cup B)\backslash P$. Thus, $(y, C) \in \mathcal{Y}(P \cap A, A)$. Consequently, $\kappa(P \cap A, A) = \min_{(y,C)\in\mathcal{Y}(P\cap A,A)}(U(y) - U(y \cup C)) \leq \min_{(y,C)\in\mathcal{Y}(P,A\cup B)}(U(y) - U(y \cup C)) = \kappa(P, A \cup B)$.

3. Note that $(A \cup B)\backslash(P \cup B) = A\backslash B$.

   a) If $u(x_P) = u(x_{P\cup B})$, then $x_{P\cup B} = x_P$. We have $\mathcal{Y}(P \cup B, A \cup B) = \mathcal{Y}(P, A)$; thus, $\kappa(P \cup B, A \cup B) = \kappa(P, A)$. Else, $u(x_{P\cup B}) < u(x_P)$ implies $x_{P\cup B} \neq x_P$. Thus, we must have $x_P \gtrless x_{P\cup B}$. As $\{z \in (A\cup B)\backslash(P\cup B) : v(z) > v(x_{P\cup B})\} \subseteq \{z \in A\backslash P : v(z) > v(x_P)\}$, we have $\mathcal{Y}(P \cup B, A \cup B) \supseteq \mathcal{Y}(P, A)$. Consequently, $\kappa(P\cup B, A\cup B) = \min_{(y,C)\in\mathcal{Y}(P\cup B,A\cup B)} U(y)-U(y\cup B) \leq \min_{(y,C)\in\mathcal{Y}(P,A)} U(y)-U(y \cup B) = \kappa(P, A)$.

   b) As $u(x_{P\cup B}) > u(x_P)$, we have $x_{P\cup B} \neq x_P$. Consequently, $v(x_{P\cup B}) > v(x_P)$. If $\mathcal{Y}(A, P) = \emptyset$, then $\kappa(P \cup B, A \cup B) \leq K = \kappa(P, A)$; thus, $\kappa(P \cup B, A \cup B) - \kappa(P, A) \leq 0 < U(x_{P\cup B})-U(x_P)$. Else, let $(y^*, B^*) = \text{argmax}_{(y,B)\in\mathcal{Y}(P,A)} U(y)-U(y \cup B)$. We show that there exists some $(y', B') \in \mathcal{Y}(P \cup B, A \cup B)$ s.t. $U(y') - U(y' \cup B') - (U(y^*) - U(y^* \cup B^*)) \leq U(x_{P\cup B}) - U(x_P)$.

   Clearly, $v(x_{P\cup B}) > v(x_P) \geq v(y^*)$. If $y^* \gtrless x_{P\cup B}$, then $(y^*, B^*) \in \mathcal{Y}(P \cup B, A \cup B)$ and the claim holds trivially (noting that $U(x_{P\cup B}) - U(x_P) > 0$). Else, we have $x_{P\cup B} \gg y^*$. By Axiom 3, $U(x_{P\cup B} \cup y^* \cup B^*) > U(y^* \cup B^*)$; thus, $x_{P\cup B} = c(x_{P\cup B} \cup y^* \cup B^*)$ (by Axiom 1). $U(x_{P\cup B} \cup y^* \cup B^*) - U(y^* \cup B^*) > 0$ implies that $U(x_{P\cup B})-U(y^*)-(U(x_{P\cup B}\cup y^*\cup B^*)-U(y^*\cup B^*)) < U(x_{P\cup B})-U(y^*) \leq U(x_{P\cup B}) - U(x_P)$. Finally, let $y' = x_{P\cup B}$ and $B' = y^* \cup B^*$.

   Lastly, the claim implies that $\min_{(y,B)\in\mathcal{Y}(P\cup B,A\cup B)}(U(y)-U(y\cup B))-(U(y^*)-U(y^* \cup B^*)) \leq U(y') - U(y' \cup B') - (U(y^*) - U(y^* \cup B^*)) < U(x_{P\cup B}) - U(x_P)$. That is, $\kappa(P \cup B, A \cup B) - \kappa(P, A) < U(x_{P\cup B}) - U(x_P)$.

□

Let κ be as defined in Lemma 5 and consider any $A \in \mathcal{A}$. Let $x^\star = c(A)$ and define $P^\star = \{y \in A : v(y) \leq v(x^\star)\}$. By definition, $x_{P^\star} = x^\star = c(A)$. As $(x^\star, A \backslash \{x^\star\}) \in \mathcal{Y}(P^\star, A)$, we have $\kappa(P^\star, A) \leq U(x^\star) - U(A)$. Thus, $u(x_{P^\star}) - \kappa(P^\star, A) \geq U(A)$. Now consider any $P \subseteq A$ and any $(y, C) \in \mathcal{Y}(P, A)$. As $y = c(y \cup C)$ and $C \supseteq \{z \in A \backslash P : v(z) > v(x_P)\}$, Axiom 3 implies that $\{z \in A \backslash P : v(z) > v(x_P)\} = \{z \in A \backslash P : v(z) > v(x_P) \text{ and } u(z) < u(x_P)\}$. Using the fact that $y = x_P$ or $y \gtreqless x_P$, $y = c(y \cup C)$ and $C \supseteq \{z \in A \backslash P : x_P \gtreqless z\}$ together with Lemma 4: $U(x_P) - (U(y) - U(y \cup C)) \leq U(x_P) - (U(x_P) - U(x_P \cup \{z \in A \backslash P : x_P \gtreqless z\})) = U(x_P \cup \{z \in A \backslash P : x_P \gtreqless z\})$. Now Axiom 2 implies $U(x_P \cup \{z \in A \backslash P : x_P \gtreqless z\}) \leq U(A)$. We obtain $u(x_P) - \kappa(P, A) = U(x_P) - \min_{(y,C) \in \mathcal{Y}(P,A)} U(y) - U(y \cup C) \leq U(A)$. Hence $P^\star$ solves $(\star)$ and $U(A) = u(x_{P^\star}) - \kappa(P^\star, A) = \max_{P \subseteq A} u(x_P) - \kappa(P, A)$.

A.2.2 Necessity of Axioms 0.1–3

Suppose that $\succsim$ is represented by $U(A) = \max_{P \subseteq A} u(x_P) - \kappa(P, A)$ where κ is a planning-cost function and that $c(A) = x_{P^\star}$ for some solution $P^\star \subseteq A$. Note that $U(\{x\}) = u(x)$ for all $x \in X$. Thus, u represents the restriction of $\succsim$ to singletons; that is, $x \succsim y \iff u(x) \geq u(y)$. Note that the first property in the definition of a planning-cost function implies that $U(x) > U(\{x, y\}) \geq U(y) \iff [u(x) > u(y) \text{ and } v(x) < v(y)]$ as well as that $U(x) = U(\{x, y\}) > U(y) \iff [u(x) > u(y) \text{ and } v(x) > v(y)]$. Thus, $x \geq y \iff: [x \gg y \text{ or } y \gtreqless x \text{ or } x = y] \iff v(x) \geq v(y)$.

Axiom 0.1: $\succsim$ is represented by some utility function, therefore a weak order. As the restriction of $\succsim$ to singletons is represented by some strictly increasing $u : X \to \mathbb{R}$, it is a linear order.

Axiom 0.2: Let $x \gtreqless y$ and $y \gtreqless z$. Then $u(x) > u(y) > u(z)$ and $v(z) > v(y) > v(x)$. Hence $u(x) > u(z)$ and $v(z) > v(x)$; therefore, $x \gtreqless z$. Let $x \gg y$ and $y \gg z$. Then $u(x) > u(y) > u(z)$ and $v(x) > v(y) > v(z)$. Hence $u(x) > u(z)$ and $v(x) > v(z)$; therefore $x \gg z$.

Axiom 1: Let $c(A \cup B) = x \in A$. Suppose $P^\star \subseteq A \cup B$ solves $(\star)$ such that $x = x_{P^\star}$. Then, $x = x_{P^\star \cap A}$ and $P^\star \cap A \subseteq A$. By the second property of a planning-cost function: $\kappa(P^\star \cap A, A) = \kappa(P^\star \cap A, (A \cup B) \cap A) \leq \kappa(P^\star, A \cup B)$. Consequently, $U(A) \geq U(x) - \kappa(P^\star \cap A, A) \geq U(x) - \kappa(P^\star, A \cup B) = U(A \cup B)$.

Axiom 2: Let $P^\star$ solve $(\star)$ such that $x_{P^\star} = c(A)$ and assume that for all $y \in B$: $u(x) < u(y)$ or $v(x) > v(y)$ (note that we may assume w.l.o.g. that $x \notin B$). Define $P' = P^\star \cup B \subseteq A \cup B$. Then $x_{P'} = x_{P^\star}$ or $u(x_{P'}) > u(x_{P^\star})$. Using the third property of a planning-cost function, we have $U(A \cup B) \geq u(x_{P'}) - \kappa(P', A \cup B) = u(x_{P'}) - \kappa(P^\star \cup B, A \cup B) \geq$ [148] $u(x_{P^\star}) - \kappa(P^\star, A) = U(A)$.

Axiom 3: Let $P^\star \subseteq A$ solve $(\star)$ such that $x_{P^\star} = c(A)$ and suppose $x \gg x_{P^\star}$, i.e. $u(x) > u(x_{P^\star})$ and $v(x) > v(x_{P^\star})$. Seeing that $x = \text{argmax}_{y \in x \cup P^\star} v(y)$, we have $U(x \cup A) \geq$

[148] If $x_{P'} = x_{P^\star}$, we use part a) of the third property of a planning-cost function; if $u(x_{P'}) > u(x_{P^\star})$, we employ part b).

$u(x) - \kappa(x \cup P^\star, x \cup A) > u(x_{P^\star}) - \kappa(P^\star, A) = U(A)$ (where the last strict inequality uses property 3.b) of planning-cost function κ).

A.3 Proof of Theorem 2

Let κ be a planning-cost function. For all $A \in \mathcal{A}$ and $x \in A$, let $P_x = \{y \in A : v(y) \leq v(x)\}$ and note that $x_{P_x} = x$. We define $C(x, A) = \kappa(P_x, A)$. We show that C is indeed a self-control-cost function:

1. We have $C(x, A) > 0 \iff \kappa(P_x, A) > 0 \iff \exists y \in A \backslash P_x : v(y) > v(x) \iff \exists y \in A : v(y) > v(x)$.

2. First, using $\{y \in A \cup B : v(y) > v(x)\} \subseteq P_x \cup B$: $C(x, A) = \kappa(P_x, A) \leq \kappa(\{y \in A \cup B : v(y) > v(x)\}, A \cup B) = C(x, A \cup B)$ (seeing that $P_x = \{y \in A \cup B : v(y) > v(x)\} \cap A$). Second, suppose that for all $y \in B : v(y) > v(x)$ (again, assume w.l.o.g. that $y \neq x$). Then, $C(x, A) = \kappa(P_x, A) = \kappa(P_x \cup B, A \cup B) = \kappa(\{y \in A \cup B : v(y) > v(x)\}, A \cup B) = C(x, A \cup B)$.

3. Let $y \in A$ and suppose $v(x) > v(y)$. If $u(x) < u(y)$, we have $C(y, A) = \kappa(P_y, A) \geq \kappa(x \cup P_y, x \cup A) \geq \kappa(\{z \in x \cup A : v(z) \leq v(x)\}, x \cup A) = C(x, x \cup A)$. Else if $u(x) > u(y)$, we have $C(x, x \cup A) - C(y, A) = \kappa(\{z \in x \cup A : v(z) \leq v(x)\}, x \cup A) - \kappa(P_y, A) \leq \kappa(x \cup P_y, x \cup A) - \kappa(P_y, A) < u(x) - u(y)$.

Suppose $P^\star$ solves $\max_{P \subseteq A} u(x_P) - \kappa(P, A)$ and let $x^\star = x_{P^\star}$. Assume that there exists some $y \in A$ such that $u(y) - C(y, A) > u(x^\star) - C(x^\star, A)$. Then $u(y) - \kappa(P_y, A) > u(x^\star) - \kappa(P^\star, A)$ contradicts optimality of $P^\star$. Vice versa, if $x^\star$ solves $\max_{x \in A} u(x) - C(x, A)$, let $P^\star = P_{x^\star}$. Assume that there exists some $P \subseteq A$ such that $u(x_P) - \kappa(P, A) > u(x^\star) - \kappa(P^\star, A)$. We must have $P \subseteq \{y \in A : v(y) \leq v(x_P)\} = P_{x_P}$. Thus, $u(x_P) - C(x_P, A) = u(x_P) - \kappa(P_{x_P}, A) \geq u(x_P) - \kappa(P, A) > u(x^\star) - \kappa(P_{x^\star}, A) = u(x^\star) - C(x^\star, A)$ contradicting optimality of $x^\star$.

Lastly, if $P^\star$ solves $\max_{P \subseteq A} u(x_P) - \kappa(P, A)$ and $x^\star = x_{P^\star}$, we have $\max_{P \subseteq A} u(x_P) - \kappa(P, A) = u(x^\star) - \kappa(P^\star, A) = u(x^\star) - C(x^\star, A) = \max_{x \in A} u(x) - C(x, A)$.

A.4 Characterizing When The Single-Kinked Value Function (4.13) is a Solution to Equations (4.9)–(4.12)

Below, we identify necessary and sufficient conditions for the value function to exhibit a single kink separating a self-control and no-self-control region in the infinite-horizon case (cf. Equation (4.13)). The kink makes it necessary to consider two candidate solutions when optimal future wealth is close to $\bar{w}$; in this case, $c_{\xi,NSC}$ and $c_{\xi,SC}$ simultaneously satisfy the FOCs (4.14) and (4.15). Yet when w is small (close to zero), optimal consumption is given by $c_{\xi,NSC}$; when w is large ($w \gg \bar{w}$) it is given by $c_{\xi,SC}$. This follows from the fact that in these cases only one of the FOCs has a solution. Moreover, a (corner) solution for which future wealth sits exactly at $\bar{w}$ is never optimal (see main text).

Moreover, using (4.16), note that indirect utilities

$$\begin{aligned}&\left[(1-\delta)c^{\sigma}_{\xi,NSC}+\delta(b_{NSC}R(w-c_{\xi,NSC}))^{\sigma}\right]^{\frac{1}{\sigma}}\\=&\left[(1-\delta)\mu(\xi,b_{NSC})^{\sigma}+\delta(b_{NSC}R(1-\mu(\xi,b_{NSC}))^{\sigma}\right]^{\frac{1}{\sigma}}\cdot w\end{aligned}$$

and

$$\begin{aligned}&\left[(1-\delta)c^{\sigma}_{\xi,SC}+\delta(-a+b_{SC}R(w-c_{\xi,SC}))^{\sigma}\right]^{\frac{1}{\sigma}}\\=&\left[(1-\delta)\mu(\xi,b_{SC})^{\sigma}+\delta(Rb_{SC}(1-\mu(\xi,b_{SC}))^{\sigma}\right]^{\frac{1}{\sigma}}\cdot\left[-\frac{a}{b_{SC}R}+w\right]\end{aligned}$$

are linearly increasing in w. At the same time, the cost of self-control is fixed at $k>0$. Thus, optimal self-control behavior is such that when w is large (tends to $+\infty$), the (utility) benefits of self-control are large and eventually surpass k; when w is small (tends to 0), (utility) benefits of self-control are small and below k. Consequently, for small enough wealth levels w, the decision maker chooses not to self-control hence $U(w)=V(w)$. As the solution to the no-self-control problem (4.11) for small w is given by $c_{\delta,NSC}$, we must have

$$\begin{aligned}&b_{NSC}\cdot w=\left[(1-\delta)\mu(\beta\delta,b_{NSC})^{\sigma}+\delta(b_{NSC}R(1-\mu(\beta\delta,b_{NSC})))^{\sigma}\right]^{\frac{1}{\sigma}}\cdot w\\\iff\ &b_{NSC}=\left[(1-\delta)\mu(\beta\delta,b_{NSC})^{\sigma}+\delta(b_{NSC}R)^{\sigma}(1-\mu(\beta\delta,b_{NSC}))^{\sigma}\right]^{\frac{1}{\sigma}}.\end{aligned}\tag{A.1}$$

Analogously, if w is sufficiently large, self-control is worthwhile to the decision maker (hence $U(w)+k=W(w)$) and the solution to the self-control problem (4.10) is given by $c_{\beta,SC}$; hence

$$k-a+b_{SC}\cdot w=\left[(1-\delta)\mu(\delta,b_{SC})^{\sigma}+\delta\left(b_{SC}R(1-\mu(\delta,b_{SC}))\right)^{\sigma}\right]^{\frac{1}{\sigma}}\cdot\left[-\frac{a}{b_{SC}R}+w\right].$$

Matching coefficients, we must have

$$b_{SC}=\left[(1-\delta)\mu(\delta,b_{SC})^{\sigma}+\delta(b_{SC}R)^{\sigma}(1-\mu(\delta,b_{SC}))^{\sigma}\right]^{\frac{1}{\sigma}}\tag{A.2}$$

and

$$k-a=-\frac{a}{R}\implies a=\frac{R}{R-1}k.\tag{A.3}$$

Using Equation (4.17) above, we observe that for all $0 < \xi \leq 1$ and $b > 0$:

$$\begin{aligned}
&(1-\beta)\mu(\xi,b)^\sigma + \xi(bR(1-\mu(\xi,b)))^\sigma \\
=&(1-\beta)\mu(\xi,b)^\sigma\left[1+\frac{\xi}{1-\beta}(bR)^\sigma\left(\frac{1-\mu(\xi,b)}{\mu(\xi,b)}\right)^\sigma\right] \\
=&(1-\beta)\mu(\xi,b)^\sigma\left[1+\left(\frac{\xi}{1-\beta}\right)^\gamma(bR)^{\gamma-1}\right] \\
=&(1-\beta)\mu(\xi,b)^{\sigma-1}. \qquad \text{(A.4)}
\end{aligned}$$

We use this to simplify (A.2) further. As $\frac{\sigma-1}{\sigma} = \frac{1}{1-\gamma}$ and $\frac{1}{\sigma} = -\frac{\gamma}{1-\gamma}$, we obtain

$$b_{SC} = (1-\delta)^{-\frac{\gamma}{1-\gamma}}\mu(\delta, b_{SC})^{\frac{1}{1-\gamma}}.$$

Thus,

$$b_{SC}^{1-\gamma} = (1-\delta)^{-\gamma}\frac{\left(\frac{\delta}{1-\delta}\right)^{-\gamma}(b_{SC}R)^{1-\gamma}}{1+\left(\frac{\delta}{1-\delta}\right)^{-\gamma}(b_{SC}R)^{1-\gamma}}$$

hence

$$\beta^\gamma R^{\gamma-1} = 1-\mu(\delta, b_{SC}). \qquad \text{(A.5)}$$

Solving for b_{SC}, we obtain

$$\begin{aligned}
b_{SC} =& \frac{1}{R}\left(\frac{\delta}{1-\delta}\right)^{\frac{\gamma}{1-\gamma}}\left[R(R\delta)^{-\gamma}-1\right]^{\frac{1}{1-\gamma}} \\
=&(1-\delta)^{-\frac{\gamma}{1-\gamma}}\left[1-R^{-1}(R\delta)^\gamma\right]^{\frac{1}{1-\gamma}}. \qquad \text{(A.6)}
\end{aligned}$$

Note that as long as $R > 1$ and $\delta R \leq 1$, we have $b_{SC} > 0$ for all $\gamma > 0$.

In similar fashion, we can rewrite (A.1) as

$$\begin{aligned}
b_{NSC} &= (1-\delta)^{-\frac{\gamma}{1-\gamma}}\mu(\beta\delta, b_{NSC})\left[1+\frac{\delta}{1-\delta}(b_{NSC}R)^{1-\frac{1}{\gamma}}\left(\frac{1-\mu(\beta\delta,b_{NSC})}{\mu(\beta\delta,b_{NSC})}\right)^{1-\frac{1}{\gamma}}\right]^{-\frac{\gamma}{1-\gamma}} \\
&= (1-\delta)^{-\frac{\gamma}{1-\gamma}}\mu(\beta\delta, b_{NSC})\left[1+\frac{\delta}{1-\delta}\left(\frac{\beta\delta}{1-\delta}\right)^{\gamma-1}(b_{NSC}R)^{\gamma-1}\right]^{-\frac{\gamma}{1-\gamma}} \\
&= (1-\delta)^{-\frac{\gamma}{1-\gamma}}\mu(\beta\delta, b_{NSC})\mu(\alpha, b_{NSC})^{\frac{\gamma}{1-\gamma}}
\end{aligned}$$

where α is such that $\alpha := \beta^{1-\frac{1}{\gamma}}\delta$. Simplifying, we obtain

$$1 = \delta^{-\gamma}R^{1-\gamma}\frac{1}{\left(1+\left(\frac{\beta\delta}{1-\delta}\right)^{-\gamma}(b_{NSC}R)^{1-\gamma}\right)^{1-\gamma}\left(1+\left(\frac{\alpha}{1-\delta}\right)^{-\gamma}(b_{NSC}R)^{1-\gamma}\right)^{\gamma}},$$

thus

$$(1-\mu(\beta\delta, b_{NSC}))^{1-\gamma}(1-\mu(\alpha, b_{NSC}))^{\gamma} = \delta^{\gamma}R^{\gamma-1}. \tag{A.7}$$

Comparing (A.5) and (A.7), we note that

$$(1-\mu(\delta, b_{SC})) = (1-\mu(\beta\delta, b_{NSC}))^{1-\gamma}(1-\mu(\alpha, b_{NSC}))^{\gamma}. \tag{A.8}$$

As $\alpha = \beta^{-\frac{1}{\gamma}}\beta\delta > \delta$, we have $1-\mu(\beta\delta, b_{NSC}) < 1-\mu(\alpha, b_{NSC})$ (cf. Equation (4.17)). Thus, $1-\mu(\delta, b_{SC}) = (1-\mu(\beta\delta, b_{NSC}))\left(\frac{1-\mu(\alpha, b_{NSC})}{1-\mu(\beta\delta, b_{NSC})}\right)^{\gamma} > 1-\mu(\beta\delta, b_{NSC})$.

Optimal decision under no-self-control ($\beta\delta$-discounting)

Above, we identified two candidate solutions for problem (4.12): $c_{\beta\delta,NSC}$ and $c_{\beta\delta,SC}$. Note that $c_{\beta\delta,SC}$ is optimal if and only if

$$\begin{aligned}
&[(1-\delta)(\mu(\beta\delta, b_{NSC}))^{\sigma} + \beta\delta(b_{NSC}R(1-\mu(\beta\delta, b_{NSC})))^{\sigma}]^{\frac{1}{\sigma}} \cdot w \\
\leq &[(1-\delta)\mu(\beta\delta, b_{SC})^{\sigma} + \beta\delta(b_{SC}R(1-\mu(\beta\delta, b_{SC})))^{\sigma}]^{\frac{1}{\sigma}} \cdot \left[-\frac{a}{b_{SC}R} + w\right].
\end{aligned}$$

Using (A.4), we obtain

$$\begin{aligned}
\frac{a}{b_{SC}R}\mu(\beta\delta, b_{SC})^{\frac{1}{1-\gamma}} &\leq \left[\mu(\beta\delta, b_{SC})^{\frac{1}{1-\gamma}} - \mu(\beta\delta, b_{NSC})^{\frac{1}{1-\gamma}}\right] \cdot w \\
\iff w &\geq \frac{a}{b_{SC}R}\frac{\mu(\beta\delta, b_{SC})^{\frac{1}{1-\gamma}}}{\mu(\beta\delta, b_{SC})^{\frac{1}{1-\gamma}} - \mu(\beta\delta, b_{NSC})^{\frac{1}{1-\gamma}}} \\
&= \frac{a}{b_{SC}R}\frac{1}{1-\left(\frac{\mu(\beta\delta, b_{NSC})}{\mu(\beta\delta, b_{SC})}\right)^{\frac{1}{1-\gamma}}} := \bar{w}_{\beta\delta}.
\end{aligned} \tag{A.9}$$

Seeing that $\mu(\beta\delta, b_{SC}) > \mu(\beta\delta, b_{NSC}) \iff \gamma < 1$, we note that $\bar{w}_{\beta\delta} > \frac{a}{b_{SC}R} > 0$.

Thus, we have

$$V(w) = \begin{cases} b_{NSC} \cdot w & \text{if } w < \bar{w}_{\beta\delta} \\ (1-\delta)^{-\frac{\gamma}{1-\gamma}}\mu(\beta\delta, b_{SC})\mu(\alpha, b_{SC})^{\frac{1}{1-\gamma}}\left[-\frac{a}{b_{SC}R} + w\right] & \text{if } w \geq \bar{w}_{\beta\delta} \end{cases}, \tag{A.10}$$

where $(1-\delta)^{-\frac{\gamma}{1-\gamma}}\mu(\beta\delta, b_{SC})\mu(\alpha, b_{SC})^{\frac{1}{1-\gamma}} > b_{NSC}$. Note that $\bar{w}_{\beta\delta}$ is the cut-off when discounting with $\beta\delta$. However, V evaluates the optimal $\beta\delta$-based decision by discounting with δ. Thus, V is discontinuous at $\bar{w}_{\beta\delta}$ with $V(\bar{w}_{\beta\delta}) > b_{NSC}\bar{w}_{\beta\delta}$.

Optimal decision under self-control (δ-discounting)

Analogous to the no-self-control case above, $c_{\delta,SC}$ is optimal in (4.10) if and only if

$$\begin{aligned}&[(1-\delta)(\mu(\delta,b_{NSC}))^{\sigma}+\delta(b_{NSC}R(1-\mu(\delta,b_{NSC})))^{\sigma}]^{\frac{1}{\sigma}}\cdot w\\ \le&[(1-\delta)\mu(\delta,b_{SC})^{\sigma}+\delta(b_{SC}R(1-\mu(\delta,b_{SC})))^{\sigma}]^{\frac{1}{\sigma}}\cdot\left[-\frac{a}{b_{SC}R}+w\right].\end{aligned}$$

Using (A.4) and (A.2), we obtain

$$\begin{aligned}\frac{a}{R}&\le\left[b_{SC}-(1-\delta)^{\frac{\gamma}{\gamma-1}}\mu(\delta,b_{NSC})^{\frac{1}{1-\gamma}}\right]\cdot w\\ \iff w&\ge\frac{a}{b_{SC}R}\frac{\mu(\delta,b_{SC})^{\frac{1}{1-\gamma}}}{\mu(\delta,b_{SC})^{\frac{1}{1-\gamma}}-\mu(\delta,b_{NSC})^{\frac{1}{1-\gamma}}}\\ &=\frac{a}{b_{SC}R}\frac{1}{1-\left(\frac{\mu(\delta,b_{NSC})}{\mu(\delta,b_{SC})}\right)^{\frac{1}{1-\gamma}}}:=\bar{w}_{\delta}.\end{aligned}\tag{A.11}$$

Again, as $\mu(\delta,b_{SC})>\mu(\delta,b_{NSC})\iff\gamma<1$, we note that $\bar{w}_\delta>\frac{a}{b_{SC}R}>0$.

Thus, we have

$$W(w)-k=\begin{cases}-k+(1-\delta)^{-\frac{\gamma}{1-\gamma}}\mu(\delta,b_{NSC})^{\frac{1}{1-\gamma}}\cdot w & \text{if } w\le\bar{w}_\delta\\ -a+b_{SC}\cdot w & \text{if } w>\bar{w}_\delta\end{cases},\tag{A.12}$$

where $a>k$ and $(1-\delta)^{-\frac{\gamma}{1-\gamma}}\mu(\delta,b_{NSC})^{\frac{1}{1-\gamma}}<b_{SC}$. Thus, W is kinked upwards at $\bar{w}_\delta$.

Note that

$$\begin{aligned}&\bar{w}_\delta\le\bar{w}_{\beta\delta}\\ \iff&\left(\frac{\mu(\delta,b_{NSC})}{\mu(\delta,b_{SC})}\right)^{\frac{1}{1-\gamma}}\le\left(\frac{\mu(\beta\delta,b_{NSC})}{\mu(\beta\delta,b_{SC})}\right)^{\frac{1}{1-\gamma}}.\end{aligned}$$

Moreover, for all $\xi\in(0,1)$:

$$\frac{\mu(\xi,b_{NSC})}{\mu(\xi,b_{SC})}=\left(\frac{b_{NSC}}{b_{SC}}\right)^{1-\gamma}\frac{1+\left(\frac{\xi}{1-\delta}\right)^{-\gamma}(b_{SC}R)^{1-\gamma}}{1+\left(\frac{\xi}{1-\delta}\right)^{-\gamma}(b_{NSC}R)^{1-\gamma}}.\tag{A.13}$$

As $b_{SC}>b_{NSC}$, this term is strictly decreasing in ξ when $\gamma<1$ and strictly increasing in ξ when $\gamma>1$. Consequently, $\left(\frac{\mu(\xi,b_{NSC})}{\mu(\xi,b_{SC})}\right)^{\frac{1}{1-\gamma}}$ is strictly decreasing in ξ and we have $\bar{w}_\delta<\bar{w}_{\beta\delta}$. Intuitively, as the decision maker places higher weight on the future when self-controlling, she will accrue enough savings to induce self-control tomorrow at a lower critical wealth level ($\bar{w}_\delta$) than when having no-self-control ($\bar{w}_{\beta\delta}$).

Remember that $\alpha := \beta^{1-\frac{1}{\gamma}}\delta$. Thus, by construction, $\frac{\mu(\delta,b)}{1-\mu(\delta,b)} = \left(\frac{\mu(\alpha,b)}{1-\mu(\alpha,b)}\right)^{\gamma}\left(\frac{\mu(\beta\delta,b)}{1-\mu(\beta\delta,b)}\right)^{1-\gamma}$ for all $b > 0$.

Lemma 6. *For all $b > 0$, $0 < \beta < 1$ and $\gamma > 0$, $\gamma \neq 1$, we have $\mu(\delta,b)^{\frac{1}{1-\gamma}} > \mu(\beta\delta,b)\mu(\alpha,b)^{\frac{\gamma}{1-\gamma}}$.*

Proof. To begin with, note that for all real numbers $r > s > 0$ and every $\lambda \in (0,1)$, we have $s < s^{1-\lambda}r^{\lambda} < r$.

For $0 < \gamma < 1$, we have $\alpha > \delta$. By definition of α, it holds that $\delta = (\beta\delta)^{1-\gamma}\alpha^{\gamma}$. Moreover, as $\ln\left(1 + x^{-\gamma}(1-\delta)^{\gamma}(bR)^{1-\gamma}\right)$ is a strictly convex function in $\ln(x)$,[149] we have

$$\left(1+\left(\frac{\beta\delta}{1-\delta}\right)^{-\gamma}(bR)^{1-\gamma}\right)^{1-\gamma}\left(1+\left(\frac{\alpha}{1-\delta}\right)^{-\gamma}(bR)^{1-\gamma}\right)^{\gamma} > 1+\left(\frac{\delta}{1-\delta}\right)^{-\gamma}(bR)^{1-\gamma}.$$

Seeing that

$$\mu(\delta,b) = \frac{1-\mu(\delta,b)}{(1-\mu(\beta\delta,b))^{1-\gamma}(1-\mu(\alpha,b))^{\gamma}}\mu(\beta\delta,b)^{1-\gamma}\mu(\alpha,b)^{\gamma}$$

and

$$\frac{1-\mu(\delta,b)}{(1-\mu(\beta\delta,b))^{1-\gamma}(1-\mu(\alpha,b))^{\gamma}} = \frac{\left(1+\left(\frac{\beta\delta}{1-\delta}\right)^{-\gamma}(bR)^{1-\gamma}\right)^{1-\gamma}\left(1+\left(\frac{\alpha}{1-\delta}\right)^{-\gamma}(bR)^{1-\gamma}\right)^{\gamma}}{1+\left(\frac{\delta}{1-\delta}\right)^{-\gamma}(bR)^{1-\gamma}}$$
$$> 1,$$

we have $\mu(\delta,b)^{\frac{1}{1-\gamma}} > \mu(\beta\delta,b)\mu(\alpha,b)^{\frac{\gamma}{1-\gamma}}$.

For $\gamma > 1$, we have $\beta\delta < \alpha = \beta^{1-\frac{1}{\gamma}}\delta < \delta$ and

$$\frac{\mu(\alpha,b)}{1-\mu(\alpha,b)} = \left(\frac{\mu(\delta,b)}{1-\mu(\delta,b)}\right)^{\frac{1}{\gamma}}\left(\frac{\mu(\beta\delta,b)}{a-\mu(\beta\delta,b)}\right)^{1-\frac{1}{\gamma}}.$$

Analogous to the above case, we obtain $\mu(\alpha,b) > \mu(\delta,b)^{\frac{1}{\gamma}}\mu(\beta\delta,b)^{1-\frac{1}{\gamma}}$. Thus, $\mu(\delta,b)^{\frac{1}{1-\gamma}} > \mu(\beta\delta,b)\mu(\alpha,b)^{\frac{\gamma}{1-\gamma}}$. □

Self-control vs. no self-control

U as given by Equation (4.13) is indeed a solution to the Bellman equations (4.9)-(4.11) if the decision maker optimally chooses to forego self-control for all $w < \bar{w} < \bar{w}_{\beta\delta}$ while self-controlling for all $w \geq \bar{w} \geq \bar{w}_{\delta}$. Given Equations (A.10) and (A.12) we derived for V and W above, this is the case if and only if (i) $\bar{w}_{\delta} \leq \bar{w}$ and (ii) $\bar{w} < \bar{w}_{\beta\delta}$ and $V(w) \leq -a + b_{SC} \cdot w$ for all $w \geq \bar{w}_{\beta\delta}$. To see this, note that – by definition – $\bar{w}_{\delta} \leq \bar{w} < \bar{w}_{\beta\delta}$ means that

[149] Note that $\frac{d\ln\left(1+\exp(-\gamma\ln(x))(1-\delta)^{\gamma}(bR)^{1-\gamma}\right)}{d\ln(x)} = -\gamma\frac{\exp(-\gamma\ln(x))(1-\delta)^{\gamma}(bR)^{1-\gamma}}{1+\exp(-\gamma\ln(x))(1-\delta)^{\gamma}(bR)^{1-\gamma}}$ is strictly increasing in $\ln(x)$.

$W(w) - k \geq V(w) \iff w \geq \bar{w}$ for all $w \in [\bar{w}_\delta, \bar{w}_{\beta\delta}]$. As $W(w) - k$ is linear with a single kink at $\bar{w}_\delta$ and $W(0) - k = -k < 0 = V(0)$, $\bar{w}_\delta \leq \bar{w}$ implies that $W(w) - k \leq V(w)$ for all $0 \leq w \leq \bar{w}_\delta$ as well. Lastly, as $(1-\delta)^{-\frac{\gamma}{1-\gamma}} \mu(\beta\delta, b_{SC}) \mu(\alpha, b_{SC})^{\frac{1}{1-\gamma}} < b_{SC}$, we have $-a + b_{SC} \cdot w \geq V(w)$ for all $w \geq \bar{w}_{\beta\delta}$ if and only if $-a + b_{SC} \cdot \bar{w}_{\beta\delta} \geq V(\bar{w}_{\beta\delta})$. Moreover, the latter can hold only if $\bar{w} < \bar{w}_{\beta\delta}$. Indeed, if $\bar{w} \geq \bar{w}_{\beta\delta}$, then $V(\bar{w}_{\beta\delta}) > b_{NSC} \cdot \bar{w}_{\beta\delta} \geq -a + b_{SC} \cdot \bar{w}_{\beta\delta}$ seeing that V exhibits an upward jump at $w = \bar{w}_{\beta\delta}$, cf. Equation (A.10).

Consequently, Condition (ii) can be equivalently stated as $-a + b_{SC} \cdot \bar{w}_{\beta\delta} \geq V(\bar{w}_{\beta\delta})$. Using Equations (A.6) and (A.9), that is,

$$
\begin{aligned}
-a + b_{SC} \cdot \bar{w}_{\beta\delta} &\geq \left[-\frac{a}{b_{SC} R} + \bar{w}_{\beta\delta} \right] \left[(1-\delta)\mu(\beta\delta, b_{SC})^\sigma + \delta(1 - \mu(\beta\delta, b_{SC}))^\sigma (R b_{SC})^\sigma \right]^{\frac{1}{\sigma}} \\
\iff \left(-\frac{a}{b_{SC} R} \right) (1-\delta)^{-\frac{\gamma}{1-\gamma}} \mu(\delta, b_{SC})^{\frac{1}{1-\gamma}} & \left[R - \frac{1}{1 - \left(\frac{\mu(\beta\delta, b_{NSC})}{\mu(\beta\delta, b_{SC})} \right)^{\frac{1}{1-\gamma}}} \right] \\
\geq \left(-\frac{a}{b_{SC} R} \right) & \left(- \frac{\left(\frac{\mu(\beta\delta, b_{NSC})}{\mu(\beta\delta, b_{SC})} \right)^{\frac{1}{1-\gamma}}}{1 - \left(\frac{\mu(\beta\delta, b_{NSC})}{\mu(\beta\delta, b_{SC})} \right)^{\frac{1}{1-\gamma}}} \right) (1-\delta)^{-\frac{\gamma}{1-\gamma}} \mu(\beta\delta, b_{SC}) \mu(\alpha, b_{SC})^{\frac{\gamma}{1-\gamma}} \\
\iff R - (R-1) \left(\frac{\mu(\beta\delta, b_{SC})}{\mu(\beta\delta, b_{NSC})} \right)^{\frac{1}{1-\gamma}} & \geq \left(\frac{\mu(\beta\delta, b_{SC})^{1-\gamma} \mu(\alpha, b_{SC})^\gamma}{\mu(\delta, b_{SC})} \right)^{\frac{1}{1-\gamma}}.
\end{aligned}
$$

Using Equation (A.13), Condition (i) can be expressed as

$$
\begin{aligned}
\frac{a}{b_{SC} R} \frac{1}{1 - \left(\frac{\mu(\delta, b_{NSC})}{\mu(\delta, b_{SC})} \right)^{\frac{1}{1-\gamma}}} &\leq \frac{a}{b_{SC} - b_{NSC}} \\
\iff \frac{1}{R} \frac{1}{b_{SC} - b_{NSC} \left(\frac{1 + \left(\frac{\delta}{1-\delta} \right)^{-\gamma} (b_{SC} R)^{1-\gamma}}{1 + \left(\frac{\delta}{1-\delta} \right)^{-\gamma} (b_{NSC} R)^{1-\gamma}} \right)^{\frac{1}{1-\gamma}}} &\leq \frac{1}{b_{SC} - b_{NSC}}.
\end{aligned}
$$

Finally, we note that both conditions are satisfied for the parameter combination $\beta = 0.1$, $\delta = 0.9$, $\gamma = 0.8$, $k = 0.1$ and $R = 1.03$ considered in the main text for which $b_{SC} \approx 0.04774$, and $b_{NSC} \approx 0.00395$. Moreover, we have $a \approx 3.43333$ and $\bar{w} \approx 78.41362$ in this case.

B Appendix to Part III

B.1 Proofs for Chapter 8

Proof of Proposition 1

We start with the special case that $|X| = 3$ and $n \in \{2,4\}$. Consider menus of two alternatives first. As all individuals choose at least one alternative from them, there must be a weak majority for at least one alternative. Now consider the universal menu X. If $n = 2$, every single individual forms a weak majority by herself. If $n = 4$, at least one alternative must be chosen be chosen by two (or more) individuals forming a weak majority. Thus, all collective choice sets are non-empty.

Now let $|X| \geq 3$ and consider some menu $\{x, y, z\} := A \in \mathcal{A}_{\geq 2}$. The case of $n = 3$ individuals is covered in the main text. Let $n \geq 5$ and $k = n \mod 3$. If $k = 0$, let $(c_1, \dots, c_n) \in \mathcal{C}^n_{\text{fne}}$ be such that $c_i(A) = x$, $c_i(A) = y$ and $c_i(A) = z$ for $n/3$ individuals each. If $k = 1$, let $(c_1, \dots, c_n) \in \mathcal{C}^n_{\text{fne}}$ be such that $c_i(A) = x$, $c_i(A) = y$ for $n/3$ individuals each and $c_i(A) = z$ for $(n-1)/3 + 1$ individuals. If $k = 2$, let $(c_1, \dots, c_n) \in \mathcal{C}^n_{\text{fne}}$ be such that $c_i(A) = x$ for $n/3$ individuals and $c_i(A) = y$ and $c_i(A) = z$ for $(n-2)/3+1$ individuals each. Note that the share of voters in support of each alternative in A is thus bounded by $\frac{\frac{n-1}{3}+1}{n} = \frac{1}{3} + \frac{2}{3n} \underbrace{\leq}_{n \geq 5} \frac{1}{3} + \frac{2}{15} = \frac{7}{15} < \frac{1}{2}$. Consequently, $c_{\text{maj}}(A) = \emptyset$.

Proof of Theorems 3, 4 and 5

We analyze the aggregation of choice functions as a judgment aggregation problem on a property space. This methodology was developed in Nehring and Puppe (2002); Nehring (2006a); Nehring and Puppe (2010). Our results are applications of the general characterization results obtained therein.

Let $\emptyset \neq \mathcal{D} \subseteq \mathcal{C}$ be a domain of choice functions such that for all $A \in \mathcal{A}_{\geq 2} := \{A \in \mathcal{A}, |A| \geq 2\}$ and $x \in A$ there exist $c, c' \in \mathcal{D}$ such that $x \in c(A)$ and $x \notin c(A)$.[150] For example, $\mathcal{D} = \mathcal{C}_{\text{fne}}$, $\mathcal{D} = \mathcal{C}_{\text{psd}}$ or $\mathcal{D} = \mathcal{C}_{\text{wo}}$. For all $A \in \mathcal{A}_{\geq 2}$ and all $x \in A$, define $H_{x|A} := \{c \in \mathcal{D} : x \in c(A)\}$ and $H^c_{x|A} := \mathcal{D} \backslash H_{x|A}$ (note that every issue $H_{x|A}, H^c_{x|A}$ partitions $\mathcal{D}$). Thus, $H_{x|A}$ corresponds to the *property* that x is chosen from A. Let $\mathcal{H} = \{H_{x|A}, H^c_{x|A} : x \in A \in \mathcal{A}_{\geq 2}\}$ be the collection of all properties. As every choice

[150] This is a minimal richness condition requiring that no issue $x \in c(A)$ is pre-determined. However, this assumption is not crucial. Alternatively, we can simply ignore issues $x \in c(A)$ which *all* choice functions agree on.

function is identified with a *unique* combination of properties, $(\mathcal{D}, \mathcal{H})$ defines a *property space*.

We say that a family (of properties) $\mathcal{G} \subseteq \mathcal{H}$ is *inconsistent* if $\bigcap \mathcal{G} = \emptyset$ (*consistent* if $\bigcap \mathcal{G} \neq \emptyset$). $\mathcal{G}$ is *critical* if it is *minimally* inconsistent; that is, $\bigcap \mathcal{G} = \emptyset$ and for all $G \in \mathcal{G}$, $\bigcap (\mathcal{G} \backslash \{G\}) \neq \emptyset$. The critical families capture the dependency structure between properties. If $H, G^c \in \mathcal{G}$ and $\mathcal{G}$ is critical, then property H *conditionally entails* property G (seeing that – conditional on the other properties in the family – if $x \in H$, then $x \in G$); we write $H \geq_0 G$ and let $\geq$ denote the transitive closure of $\geq_0$ and let $\equiv$ be the symmetric part of $\geq$. We note that $\geq_0$ (thus, $\geq$) is complementation-adapted; that is, if $H \geq_0 G$, then $G^c \geq_0 H^c$. Moreover, as all families $\{H, H^c\}$ are trivially critical, $\geq_0$ (thus, $\geq$) is reflexive.

Lemma 7. *Let* $\emptyset \neq \mathcal{D} \subseteq \mathcal{C}$.

1. *Suppose all* $c \in \mathcal{D}$ *satisfy* (FNE) *and* (α). *Then for all* $A, B \in \mathcal{A}_{\geq 2}$, $x \in A$, $y \in B$, $x \neq y$: $H^c_{x|A} \geq H_{y|B}$.

2. *Suppose all* $c \in \mathcal{D}$ *satisfy* (γ) *and* (α). *Then for all* $A, B \in \mathcal{A}_{\geq 2}$, $x \in A$, $H_{x|A} \geq H^c_{x|A}$.

3. *Suppose all* $c \in \mathcal{D}$ *satisfy* (Aiz) *and* (α). *Then for all* $A, B \in \mathcal{A}_{\geq 2}$, $x \in A$ *and* $y \in B$: $H_{x|A} \geq H_{y|B}$. *Thus,* $H_{x|A} \equiv H_{y|B}$ *and* $H^c_{x|A} \equiv H^c_{y|B}$.

Proof. 1. By (α), $H^c_{x|A} \geq_0 (H^c_{x|A \cup B})$. By (FNE), $H^c_{x|A \cup B} \geq_0 H_{y|A \cup B}$. Again, by (α), $H_{y|A \cup B} \geq_0 H_{y|B}$.

2. Suppose first that $A \neq X$. By (γ), $H_{x|A} \geq_0 H^c_{x|X \backslash A \cup x}$ (conditional $H^c_{x|X}$). By (α), $H^c_{x|(X \backslash A) \cup \{x\}} \geq_0 H^c_{x|X}$. Again, by (γ), $H^c_{x|X} \geq_0 H^c_{x|A}$ (conditional on $H_{x|(X \backslash A) \cup \{x\}}$). Now, if $A = X$, let $B \subsetneq X$. By (α) and what we just showed, $H_{x|X} \geq_0 H_{x|B} \geq H^c_{x|B} \geq_0 H^c_{x|X}$.

3. Suppose first that $y \in B \backslash A$. By (Aiz), $H_{x|A} \geq_0 H_{y|A \cup B}$ (conditional on $\{H^c_{y_1|A \cup B}, \ldots, H^c_{y_k|A \cup B}, H^c_{x|A \cup B}\}$ for $(B \backslash A) \backslash \{y\} = \{y_1, \ldots, y_k\}$). By (α), $H_{y|A \cup B} \geq_0 H^c_{y|B}$. Else $y \in A \cap B$. Then, by (α), $H_{x|A} \geq_0 H_{x|A \backslash \{y\}}$ and, by (Aiz), $H_{x|A \backslash \{y\}} \geq_0 H_{y|A \cup B}$ (conditional on $\{H^c_{y_1|A \cup B}, \ldots, H^c_{y_k|A \cup B}, H^c_{x|A \cup B}\}$ for $B \backslash A = \{y_1, \ldots, y_k\}$). Now, again by (α), $H_{y|A \cup B} \geq_0 H_{y|B}$.

   Consequently, $H_{x|A} \geq H_{y|B}$ and $H_{y|B} \geq H_{x|A}$; thus, $H_{x|A} \equiv H_{y|B}$. By complementation-adaptedness, $H^c_{x|A} \equiv H^c_{y|B}$.

□

Proof of Theorem 3

Consider $\mathcal{D} = \mathcal{C}_{\text{acy}}$. Note that all $c \in \mathcal{D}$ satisfy (FNE), (α) and (γ). Let $x \in A \in \mathcal{A}_{\geq 2}$, $y \in B \in \mathcal{A}_{\geq 2}$. (i) Assume $x \neq y$ first. By Lemma 7, parts 1 and 2, $H^c_{x|A} \geq H_{y|B} \geq H^c_{y|B}$. By complementation-adaptedness, $H_{y|b} \geq H_{x|A}$. (ii) Now let $x = y$. Then pick any $z \neq x$

and some $C \in \mathcal{A}_{\geq 2}$ such that $z \in C$. By what was just shown, $H^c_{x|A} \geq H^c_{z|C} \geq H^c_{y|B}$. Again, by complementation-adaptedness, $H_{y|B} \geq H_{x|A}$. In total, we conclude that $H_{x|A} \equiv H_{y|B}$ and $H^c_{x|A} \equiv H_{y|B}$. Moreover, again by Lemma 7, parts 1 (and what we just showed for the case $x = y$) and 2, $H^c_{x|A} \geq H_{x|A} \geq H^c_{x|A}$; thus, $H^c_{x|A} \equiv H^c_{x|A}$. Consequently, $\mathcal{D} = \mathcal{C}_{\text{acy}}$ is totally blocked. Note that putting additional restrictions on choice functions introduces additional conditional entailments without removing existing one. Thus, all $\mathcal{D} = \mathcal{C}_{\text{Par}}, \mathcal{C}_{\text{wo}}, \mathcal{C}_{\text{lo}}$ are totally blocked. The results follows from (Nehring and Puppe, 2010, Theorem 1).

Proof of Theorem 4

Consider $\mathcal{D} = \mathcal{C}_{\text{psd}}$. Note that all $c \in \mathcal{D}$ satisfy (FNE),(α) and (Aiz). Let $x \in A \in \mathcal{A}_{\geq 2}$, $y \in B \in \mathcal{A}_{\geq 2}$. By Lemma 7, part 3, $H_{x|A} \equiv H_{y|B}$ and $H^c_{x|A} \equiv H^c_{y|B}$. Moreover, letting $z \neq x$ and $z \in C \in \mathcal{A}_{\geq 2}$ and using the result just obtained together with part 1 of Lemma 7, we have $H^c_{x|A} \geq H_{z|C} \geq H_{y|B}$. At the same time, seeing that all critical families contain at most one un-negated property, $H_{x|A} \not\geq H^c_{y|B}$. Thus, $\mathcal{D} = \mathcal{C}_{\text{psd}}$ is semi-blocked. By Nehring (2006a), an Arrowian aggregation rule is consistent on $\mathcal{C}_{\text{psd}}$ if and only if it an *oligarchy*. That is there exists some $M \subseteq \{1, \ldots, n\}$ such that for $(c_1, \ldots, c_n) \in \mathcal{C}^n_{\text{psd}}$ and all $A \in \mathcal{A}_{\geq 2}$, all $x \in A$: $x \in f(c_1, \ldots, c_n)(A) \iff \exists i \in M : x \in c_i(A)$. Thus, $f(c_1, \ldots, c_n)(A) = \bigcup_{i \in M} c_i(A)$.

Proof of Theorem 5

The equivalence of anonymous Arrowian aggregation to the existence of quotas $q_{x \in c(A)}$ is established in (Nehring and Puppe, 2010, Proposition 2.2).[151] Menu-level neutrality requires that, for some given menu $A \in \mathcal{A}_{\geq 2}$, all 'local' aggregation rules applied to the issues $x \in c(A)$ are identical. That is, the structure of winning coalitions must be the same for all such issues.

Now consider any $A \in \mathcal{A}_{\geq 2}$.

(i) Suppose $n \leq |A|$. If, for all $x \in A$, $0 < q_{x \in c(A)} \leq 1/n$, then each single individual forms a winning coalition for all $x \in c(A)$. As all individual choice sets $c_i(A)$ are non-empty, so is the collective choice set $c(A)$. On the other hand, if $q_{x \in c(A)} > 1/n$ for some $x \in A$, then $q_{x \in c(A)} > 1/n$ for all $x \in A$ (as all winning coalitions need to be the same). Consider some $(c_1, \ldots, c_n) \in \mathcal{C}^n_{\text{fne}}$ such that each individual chooses one (and only one) distinct alternative from A. Then $c(A) = \emptyset$.

[151] Nehring and Puppe (2010) consider quotas q_H, q_{H^c} such that H resp. H^c are accepted iff the fraction of voters supporting H resp. H^c is *strictly* greater than q_H and q_{H^c} respectively to treat both H and H^c equally. We only demand that support in favor of accepting x as collectively choosable from A *weakly* exceeds the quota $q_{x \in c(A)}$. On the other hand, the fraction of voters not choosing x from A needs to *strictly* exceed $q_{x \notin c(A)} := 1 - q_{x \in c(A)}$ to ban x from $c(A)$. However, except for very special cases (e.g., when $q_{x \in c(A)} = 1/2$ and n is even), these formulations are equivalent.

(ii) Suppose $n > |A|$ and let $r = n \mod |A|$. Suppose that $0 < q_{x\in c(A)} \leq \frac{1}{|A|}(1 - \frac{r}{n}) + \frac{1}{n}\mathbb{1}(r \neq 0)$ for all $x \in A$. All individual choice sets $c_i(A)$ are non-empty. Thus, if $r = 0$, there exists some alternative $x \in A$ such that at least fraction $1/|A|$ of all individuals choose x from A. If $r > 0$, there exists some alternative $x \in A$ such that at least fraction $\frac{1}{|A|}(1-\frac{r}{n})+\frac{1}{n}$ of all individuals choose x from A. Hence $x \in c(A)$ in both cases and $c(A) \neq \emptyset$. Conversely, suppose there exists some $x \in A$ such that $q_{x\in c(A)} > \frac{1}{|A|}(1 - \frac{r}{n}) + \frac{1}{n}\mathbb{1}(r \neq 0)$. Let $x_1, \ldots, x_k \in X$ be such that $A = \{x_1, \ldots, x_k\}$. Consider a profile $(c_1, \ldots, c_n) \in \mathcal{C}^n_{\text{fne}}$ such that $c_i(A) = x_j$, $j = 1, \ldots, n - r$ for $(n - r)/|A|$ individuals each and $c_i(A) = x_j$, $j = n - r + 1, \ldots, n$ for $(n - r)/|A| + 1$ individuals each (note that the total sum of individuals is thus n). Thus, the fraction of individuals voting for $x \in c(A)$ is less than $\frac{1}{|A|}(1 - \frac{r}{n}) + \frac{1}{n}\mathbb{1}(r \neq 0)$ for each $x \in A$. As all structures of winning coalition are the same, this implies that $c(A) = \emptyset$.

B.2 Proof for Chapter 9

The proof of Proposition 2 is given in the main text.

C Appendix to Part IV

C.1 Relation to Effectivity Functions and Game Forms

An effectivity function is a mapping $E : 2^N \backslash \{\emptyset\} \rightrightarrows 2^X \backslash \{\emptyset\}$ such that (i) $E(N) = 2^X \backslash \{\emptyset\}$ and (ii) $X \in E(G)$ for all $G \subseteq N$. For every group G, $E(G)$ lists all subsets of outcomes for which G is *effective*. Given some game form $\Gamma = (N, (S_i)_{i \in N}, g)$ – where N is the set of players, S_i is player i's set of strategies and onto $g : \times_{i \in N} S_i \to X$ maps profiles of strategies to outcomes – we say that $\emptyset \neq G \subseteq N$ is α-effective for $Y \subseteq X$ if there exists some $s_G \in \times_{i \in G} S_i$ such that for all $s_{N \backslash G} \in \times_{i \in N \backslash G} S_i$: $g(s_G, s_{N \backslash G}) \in Y$. A game form gives rise to an (α)-effectivity function $E_\Gamma : 2^N \backslash \{\emptyset\} \rightrightarrows 2^X \backslash \{\emptyset\}$ defined by $2^N \backslash \{\emptyset\} \ni G \mapsto E_\Gamma(G) = \{Y \subseteq X : G \text{ is } \alpha\text{-effective for } Y\}$. An important question in the study of effectivity functions is when an effectivity function E can be represented by some game form Γ in the sense that $E_\Gamma = E$. We state a basic result from Peleg (1998).

Fact 6. *Let $E : 2^N \backslash \{\emptyset\} \rightrightarrows 2^X \backslash \{\emptyset\}$ be an effectivity function. Then E can be represented by some game form Γ (i.e., $E_\Gamma = E$) if and only if for all $G, G' \in 2^N \backslash \{\emptyset\}$:*

$$\text{(super-additive)} \quad (Y \in E(G), Y' \in E(G') \text{ and } G \cap G' = \emptyset) \implies Y \cap Y' \in E(G \cup G'),$$

$$\text{(monotone)} \quad (Y \in E(G), X \supseteq Y' \supseteq Y) \implies Y' \in E(G).$$

As any representable effectivity function E is monotone by Fact 6, it is often insightful to consider its *basis*, the ($\subseteq$-)smallest effectivity function such that its monotone closure equals E. That is, for every $G \in 2^N \backslash \{\emptyset\}$, let $basis(E)(G) = \{Y \subseteq X : Y \in E(G) \text{ and } Y' \in E(G) \text{ for no } Y' \subsetneq Y\}$. $basis(E)$ contains the essential information about a monotone E in the sense that E is representable if and only if $basis(E)$ is super-additive.

There is a close connection between effectivity functions and rights systems as we defined them in this paper. For every $\mathcal{R} : \mathcal{P} \rightrightarrows 2^N \backslash \emptyset$, the inverse correspondence defined by $2^N \backslash \{\emptyset\} \ni G \mapsto \mathcal{R}^{-1}(G) = \{P \in \mathcal{P} : G \in \mathcal{R}(P)\}$ is a mapping $\mathcal{R}^{-1} : 2^N \backslash \{\emptyset\} \rightrightarrows \mathcal{P} \subseteq 2^X \backslash \{\emptyset\}$. Indeed, recall that properties are extensionally defined as subsets of the underlying set X. If $P \in \mathcal{R}^{-1}(G)$, the intuition that G can restrict the eventual choice to come from $P \subseteq X$ parallels that of the effectivity function approach. Due to our distinct setup, however, $\mathcal{R}^{-1}$ is not itself an effectivity function.

There are further important differences. First, conceptionally, our model is *semantic*. That is, we conceptualize rights in terms of subsets which have (respectively, are given) a meaning as and through properties. Second, our definition of respect for rights implies a *conjunctive* notion of rights. When individuals are part of several rights holding groups they can generally exercise these rights simultaneously unless this implies forcing an inconsistent combination of subsets (i.e., properties) at the individual level. In particular, if the same group G has rights to two properties P and Q, then G is effective for $P \cap Q$ under any aggregation function that respects these rights unless the conjunction of these properties is infeasible (i.e, $P \cap Q = \emptyset$).[152] This is in line with the intuition that groups and individuals have several rights which can be exercised at the same time. By contrast, for general effectivity functions, an effectivity set should be thought of as arising from a comprehensive exercise of rights by individuals or groups respectively.

Lastly, we analyze rights under a concept of representation that deviates from the general case in two respects. (i) We consider representation by the restricted class of *voting* game forms. That is, game forms for which each player's set of strategies is equal to the set of outcomes (alternatives). Indeed, $\mathcal{R}$ is consistent if and only if there exists some onto $f : X^n \to X$ satisfying (R) such that $(N, (S_i = X)_{i \in N}, f)$ defines a (voting) game form (cf. Proposition 12 below). (ii) Our concept of respect for rights is weaker than the game form notion of representability in the sense that if f respects $\mathcal{R}$ then f generally respects rights systems $\mathcal{R}'$ which extend $\mathcal{R}$ (i.e., such that $\mathcal{R}'(P) \supseteq \mathcal{R}(P)$ for all $P \in \mathcal{P}$).[153]

To formally explore the connection to property space rights, we define an analogous notion for effectivity functions. We say that Γ weakly represents E if and only if, for all $G \in 2^N \backslash \{\emptyset\}$, $E(G) \subseteq E_\Gamma(G)$. If some Γ weakly represents E, we say that E is weakly representable. The following *Intersection Property for Effectivity Functions* (IPE) is necessary and sufficient for weak representation by some game form.

Intersection Property for Effectivity Functions. We say that an effectivity function $E : 2^N \backslash \{\emptyset\} \rightrightarrows 2^X \backslash \{\emptyset\}$ satisfies the Intersection Property for Effectivity Functions (IPE) if and only if for all pairwise disjoint $G_1, \ldots, G_r \in 2^N \backslash \{\emptyset\}$:

$$Y_1 \in E(G_1), \ldots, Y_r \in E(G_r) \implies \bigcap_{k=1}^{r} Y_k \neq \emptyset. \tag{IPE}$$

Fact 7. *Let $E : 2^N \backslash \{\emptyset\} \rightrightarrows 2^X \backslash \{\emptyset\}$ be an effectivity function. The following are equivalent:*

1. *E is weakly representable.*
2. *There exists some monotone and super-additive $\overline{E}$ that extends basis(E) (such that, for all $G \in 2^N \backslash \{\emptyset\}$: basis$(E)(G) \subseteq E(G) \subseteq \overline{E}(G)$).*
3. *E satisfies* (IPE).

Proof. Equivalence of 1. and 2. follows immediately from Fact 6. We show equivalence of 2. and 3.

[152] In our model of rights, for group G to exercise a right to $P \cap Q$, all of its members $i \in G$ must submit a feasible view (vote) $x_i \in P \cap Q$. If $P \cap Q = \emptyset$ such a right does not exist in the sense that it can never be exercised in an individually feasible way.

[153] This is particularly obvious in Chapter 13 where we show that a rights system $\mathcal{R}$ is consistent with voting by properties if and only if there exists some consistent exhaustive rights system that *extends* $\mathcal{R}$.

First, suppose E satisfies (IPE). We show that (IPE) guarantees the existence of a smallest monotone and super-additive extension of E which we define, for all $G \in 2^N \backslash \{\emptyset\}$, by $G \mapsto \overline{E}(G) = \{Y \subseteq X : Y \supseteq \bigcap_{k=1,\dots,r} Y_k$ where $Y_1 \in E(G_1), \dots, Y_r \in E(G_r)$ for some partition
$G_1, \dots, G_r$ of $G\}$. Clearly, for all $G \in 2^N \backslash \{\emptyset\}$, $basis(E)(G) \subseteq E(G) \subseteq \overline{E}(G)$ and, by (IPE), $\emptyset \notin \overline{E}(G)$. Moreover, for all $G \in 2^N \backslash \{\emptyset\}$, $\overline{E}(G) \supseteq E(G)$ implies $X \in \overline{E}(G)$. In particular, $E(N) = 2^N \backslash \{\emptyset\}$. Thus, $\overline{E}$ is an effectivity function. Monotonicity of $\overline{E}$ is obvious. We verify that it is also super-additive. Consider two disjoint and non-empty $G, G' \subseteq N$ and suppose $Y \in \overline{E}(G)$, $Y' \in \overline{E}(G')$. Then there exist partitions $G_1, \dots, G_r$ of G, $G'_1, \dots, G'_{r'}$ of G' and $Y_1 \in E(G_1), \dots, Y_r \in E(G_r)$, $Y'_1 \in E(G'_1), \dots, Y'_{r'} \in E(G'_{r'})$ such that $Y \supseteq \bigcap_{k=1,\dots,r} Y_k$ and $Y' \supseteq \bigcap_{k=1,\dots,r'} Y'_k$. Clearly, $X \supseteq Y \cap Y' \supseteq \left(\bigcap_{k=1,\dots,r} Y_k\right) \cap \left(\bigcap_{k=1,\dots,r'} Y'_k\right)$. As $G \cap G' = \emptyset$, $G_1, \dots, G_r, G'_1, \dots, G'_{r'}$ partition $G \cup G'$. Consequently, we have $Y \cap Y' \in \overline{E}(G \cup G')$.

To prove the reverse implication, it suffices to note that the super-additivity property defined above for pairs of disjoint subsets generalizes to countable collections of pairwise disjoint subsets by induction. Thus, if pairwise disjoint $G_1, \dots, G_r$ give rise to a violation of (IPE), then $\emptyset \in E(G)$ for $G = \bigcup_{k=1,\dots r} G_k \subseteq N$ by super-additivity, contradicting the definition of an effectivity function. □

When rights are rights to (combinations of) properties (i.e., when $basis(E)$ only contains subsets that are combinations of properties on $(X, \mathcal{P})$), the following proposition shows that the restriction to voting game forms is without loss of generality as long as we consider weak and conjunctive representation of rights. In this case, checking whether some effectivity function E is representable by some voting game form is equivalent to analyzing consistency of $\mathcal{R}_E$, the rights system induced by it on $(X, \mathcal{P})$. For all $G \in 2^N \backslash \{\emptyset\}$, let

$$\mathcal{R}_E^{-1}(G) = \bigcup \{\widehat{\mathcal{P}} \subseteq \mathcal{P} : \bigcap \widehat{\mathcal{P}} \in basis(E)(G)\}$$

and define, for each $P \in \mathcal{P}$, $\mathcal{R}_E(P) = \{G \in 2^N \backslash \{\emptyset\} : P \in \mathcal{R}_E^{-1}(G)\}$.[154]

Conjunctive Extension. For two effectivity functions $E, \overline{E}$, we say that $\overline{E}$ *conjunctively extends* E if and only if $\overline{E}$ extends E and we have for all $G_1, \dots, G_r \in 2^N \backslash \{\emptyset\}$ and all $Y_1 \in E(G_1), \dots, Y_r \in E(G_r)$:

$$\left(\forall i \in \bigcup_{k=1,\dots r} G_k : \bigcap_{k: i \in G_k} Y_k \neq \emptyset\right) \implies \bigcap_{k=1,\dots,r} Y_k \in \overline{E}\left(\bigcup_{k=1,\dots r} G_k\right).$$

We say that some $Y \subseteq X$ is ($\mathcal{P}$-)convex if and only if there is some $\mathcal{P}_Y \subseteq \mathcal{P}$ such that $Y = \bigcap \mathcal{P}_Y$. Note that, given the convention $\bigcap \emptyset = X$, the comprehensive set X is convex.

[154] Note that $(\mathcal{R}_E)^{-1} = \mathcal{R}_E^{-1}$. That is, $\mathcal{R}_E^{-1}$ is indeed the inverse of R_E thus defined.

$E : 2^N \backslash \{\emptyset\} \rightrightarrows 2^X$ is ($\mathcal{P}$-)convex valued if and only if for every $G \in 2^N \backslash \{\emptyset\}$: $Y \in E(G)$ implies that G is ($\mathcal{P}$-)convex.

Proposition 12. *Let $(X, \mathcal{P})$ be a property space and let $E : 2^N \backslash \{\emptyset\} \rightrightarrows 2^X \backslash \{\emptyset\}$ be an effectivity function with $(\mathcal{P}-)$convex valued basis. The following are equivalent:*

1. *E is weakly represented by some voting game form $\Gamma = (N, (S_i = X)_{i \in N}, f)$, where $f : X^n \to X$ respects $\mathcal{R}_E$.*

2. *There exists some monotone and super-additive effectivity function $\overline{E}$ that extends $basis(E)$ conjunctively.*

3. *$\mathcal{R}_E$ is consistent.*

Proof. We prove 1. $\implies$ 2. $\implies$ 3. $\implies$ 1.

1. $\implies$ 2. Let $\overline{E} = E_\Gamma$. By Fact 6, $\overline{E}$ is monotone and super-additive. Clearly, $\overline{E}$ extends E; a fortiori, it extends $basis(E)$. We verify that it does so conjunctively. Let $G_1, \ldots, G_r \in 2^N \backslash \{\emptyset\}$ and $Y_1 \in basis(E)(G_1), \ldots, Y_r \in basis(E)(G_r)$. We have, for some $\mathcal{P}_{Y_1}, \ldots, \mathcal{P}_{Y_r} \subseteq \mathcal{P}$, $Y_1 = \bigcap \mathcal{P}_{Y_1}, \ldots, Y_r = \mathcal{P}_{Y_r}$. Thus, for all $k = 1, \ldots, r$, $\mathcal{P}_{Y_k} \subseteq \mathcal{R}_E^{-1}(G_k) \iff (\forall P \in \mathcal{P}_{Y_k} : G_k \in \mathcal{R}_E(P))$. Suppose that, for all $i \in \bigcup_{k=1,\ldots,r} G_k$, $\bigcap_{k: i \in G_k} Y_k \neq \emptyset$, i.e., there exist $x_i^\star \in \bigcap_{k: i \in G_k} \bigcap \mathcal{P}_{Y_k}$. Then, for all $(x_i^\star)_{i \in N \backslash G} \in X^{(n-|G|)}$, all $k = 1, \ldots, r$ and all $P \in \mathcal{P}_{Y_k}$, $\{i \in N : x_i^\star \in P\} \supseteq G_k$. As f respects $\mathcal{R}_E$, we have $f(x_1^\star, \ldots, x_r^\star) \in \bigcap_{k=1,\ldots,r} \bigcap \mathcal{P}_{Y_k} = \bigcap_{k=1,\ldots,r} Y_k$. Thus, $\bigcap_{k=1,\ldots,r} Y_k \in E_\Gamma \left(\bigcup_{k=1,\ldots,r} G_k \right) = \overline{E} \left(\bigcup_{k=1,\ldots,r} G_k \right)$.

2. $\implies$ 3. We prove the contraposition. Suppose $\mathcal{R}_E$ is inconsistent. By Theorem 6, (IPC) is violated. That is, there exist some critical $\{P_1, \ldots, P_r\} \subseteq \mathcal{P}$ and $G_1 \in \mathcal{R}_E(P_1), \ldots, G_r \in \mathcal{R}_E(P_r)$ such that $\bigcap_{k=1,\ldots,r} G_k = \emptyset$. Thus, for all $i \in \bigcup_{k=1,\ldots,r} G_k$, $\bigcap_{k: i \in G_k} P_k \neq \emptyset$. Moreover, for $k = 1, \ldots, r$, there exist $\mathcal{P}_k \subseteq \mathcal{P}$ such that $P_k \in \mathcal{P}_k$ and $\bigcap \mathcal{P}_k \in basis(E)(G_k)$. Seeing that, for every effectivity function $\overline{E}$, $\bigcap_{k=1,\ldots,r} \bigcap \mathcal{P}_k \subseteq \bigcap_{k=1,\ldots,r} P_k = \emptyset \notin \overline{E} \left(\bigcup_{k=1,\ldots,r} G_k \right)$, a conjunctive extension does not exist.

3. $\implies$ 1. There exists some onto $f : X^n \to X$ that respects $\mathcal{R}_E$. Clearly, $\Gamma = (N, (S_i = X)_{i \in N}, f)$ defines a game form. We show that it weakly represents E. Let $G \in 2^N \backslash \{\emptyset\}$ and $Y \in E(G)$. There exists some $Y' \subseteq Y$ with $Y' \in basis(E)(G)$. Thus, $Y' = \bigcap \mathcal{P}_{Y'}$ for some $\mathcal{P}_{Y'} \subseteq \mathcal{P}$. By definition, we have $\mathcal{P}_{Y'} \subseteq \mathcal{R}_E^{-1}(G)$; i.e., for all $P \in \mathcal{P}_{Y'}$, $G \in \mathcal{R}_E(P)$. For all $i \in G$, let $x_i^\star \in \bigcap \mathcal{P}_{Y'}$. We have for all $(x_i^\star)_{i \in N \backslash G} \in X^{(n-|G|)}$ and all $P \in \mathcal{P}_{Y'}$: $\{i \in N : x_i^\star \in P\} \supseteq G$. Thus, as f respects $\mathcal{R}_E$, $f(x_1^\star, \ldots, x_r^\star) \in P$. This is, for all $(x_i^\star)_{i \in N \backslash G} \in X^{(n-|G|)}$, $f(x_1^\star, \ldots, x_r^\star) \in \bigcap \mathcal{P}_{Y'} = Y'$. Hence $Y' \in E_\Gamma(G)$. As E_Γ is monotone (cf. Fact 6), we have $Y \supseteq Y' \in E_\Gamma(G)$. □

Finally, we note that for every effectivity function on X, there is some property structure $\mathcal{P}$ such that $basis(E)$ is convex-valued on $(X, \mathcal{P})$. Indeed, for every $x \in X$, let $P_x = \{x\}$. Define $\mathcal{P}_X = \{P_x, P_x^c\}_{x \in X}$ and call $(X, \mathcal{P}_X)$ the *discrete* property space. On $(X, \mathcal{P}_X)$, every subset $Y \subseteq X$ is convex, seeing that $Y = \bigcap_{x \notin Y} P_x$.

C.2 Proofs for Chapter 12

C.2.1 Proof of Theorem 6

$\Longleftarrow$: Let (IPC) hold. For each $\boldsymbol{x}(x_1, \ldots, x_n) \in X^n$, define $\mathcal{P}_{\mathcal{R}}(\boldsymbol{x}) = \{P \in \mathcal{P} : \{i \in N : x_i \in P\} \supseteq G$ for some $G \in \mathcal{R}(P)\}$. If, for every $\boldsymbol{x} \in X^n$, we can define $f(\boldsymbol{x}) = y_x$ for some $y_x \in \bigcap \mathcal{P}_{\mathcal{R}}(\boldsymbol{x})$, f is an aggregation function and respects $\mathcal{R}$ by construction. (By convention, we let for $\mathcal{P}_{\mathcal{R}}(\boldsymbol{x}) = \emptyset$, $\bigcap \mathcal{P}_{\mathcal{R}}(\boldsymbol{x}) = X$.) Since, for all $\tilde{x} \in X$, $\tilde{x} \in \bigcap \mathcal{P}_{\mathcal{R}}((\tilde{x}, \ldots, \tilde{x}))$, we can also find a unanimous (a fortiori, onto) f that respects rights in this case.

Assume the aforementioned is not possible, that is, suppose there exists some $\boldsymbol{x} = (x_1, \ldots, x_n) \in X^n$ s.t. $\mathcal{P}_{\mathcal{R}}(\boldsymbol{x})$ is inconsistent ($\bigcap \mathcal{P}_{\mathcal{R}}(\boldsymbol{x}) = \emptyset$ and $\mathcal{P}_{\mathcal{R}}(\boldsymbol{x}) \neq \emptyset$). Then there exists some critical $\mathcal{G} \subseteq \mathcal{P}_{\mathcal{R}}(\boldsymbol{x})$. Let $P_1, \ldots, P_r \in \mathcal{P}$ be such that $\mathcal{G} = \{P_1, \ldots, P_r\}$. By definition of $\mathcal{P}_{\mathcal{R}}(\boldsymbol{x})$, for each $k = 1, \ldots, r$, there exists some $G_k \in \mathcal{R}(P_k)$ s.t. $G_k \subseteq \{i \in N : x_i \in P_k\}$. By (IPC), $\bigcap_{k=1,\ldots,r} G_k \neq \emptyset$. Hence there exists some $i \in N$ such that, for all $k = 1, \ldots, r$, $x_i \in P_k$. Consequently, $x_i \in \cap \mathcal{G} = \emptyset$, a contradiction.

$\Longrightarrow$: Suppose (IPC) does not hold. That is, suppose there exist some critical family $\mathcal{G} = \{P_1, \ldots, P_k\}$ and groups $G_1 \in \mathcal{R}(P_1), \ldots, G_k \in \mathcal{R}(P_k)$ such that $\bigcap_{k=1,\ldots,r} G_k = \emptyset$. For all $i \in \bigcup_{k=1,\ldots,r} G_k$, consider $\mathcal{P}_i^{\mathcal{G}} = \{P_k \in \mathcal{G} : i \in G_k\}$, the family of all properties P_k in $\mathcal{G}$ for which i is part of group G_k. Evidently, $\mathcal{P}_i^{\mathcal{G}} \subseteq \mathcal{G}$. As $\bigcap_{k=1,\ldots,r} G_k = \emptyset$, for each $i \in \bigcup_{k=1,\ldots,r} G_k$, there exists some $k_i \in \{1, \ldots, r\}$ such that $P_{k_i} \notin \mathcal{P}_i^{\mathcal{G}}$. Thus, by criticality of $\mathcal{G}$, all $\mathcal{P}_i^{\mathcal{G}}$ are consistent. That is, for all $i \in \bigcup_{k=1,\ldots,r} G_k$, there exist $x_i^\star \in \bigcap \mathcal{P}_i^{\mathcal{G}}$. For all $i \in N \backslash (\bigcup_{k=1,\ldots,r} G_k)$, let $x_i^\star \in X$ be arbitrary. Now suppose some $f : X^n \to X$ respects $\mathcal{R}$. By construction, for all $k = 1, \ldots, r$, $\{i \in N : x_i^\star \in P_k\} \supseteq G_k$. By (R), $f(\boldsymbol{x}^\star) \in \bigcap_{k=1,\ldots,r} P_k = \emptyset$, a contradiction.

C.2.2 Proof of Proposition 3

There exist distinct $a, b, c, d \in A$ such that a and b, c and d are 1-variants; b and c, a and d are $2-$variants. To see this, note that $n \geq 2$ and, for all $i = 1, \ldots, n$, $|A_i| \geq 2$. Consequently, there exist distinct $a_1, a_1' \in A_1$ and distinct $a_2, a_2' \in A_2$. For all $i \in \{0, 1, \ldots, n\} \backslash \{1, 2\}$, fix $a_i \in A_i$. Then $a = (a_0, a_1, a_2, \ldots, a_n)$, $b = (a_0, a_1', a_2, \ldots, a_n)$, $c = (a_0, a_1', a_2', \ldots, a_n)$ and $d = (a_0, a_1, a_2', \ldots, a_n)$ is a possible choice.

By assumption, $\{1\} \in \mathcal{R}(P_{a>b}) \cap \mathcal{R}(P_{c>d})$ and $\{2\} \in \mathcal{R}(P_{b>c}) \cap \mathcal{R}(P_{d>a})$. As $\{P_{a>b}, P_{b>c}, P_{c>d}, P_{d>a}\}$ is critical, $\{1\} \cap \{2\} = \emptyset$ implies a violation of (IPC). By Theorem 6, $\mathcal{R}$ is inconsistent.

C.2.3 Proof of Fact 1

Assume that $\mathcal{R}$ is weakly independent. Suppose there exist distinct $P, Q \in \mathcal{P}$ and disjoint $G, G' \subseteq N$ such that $G \in \mathcal{R}(P) \cap \mathcal{R}(P^c)$ and $G' \in \mathcal{R}(Q) \cap \mathcal{R}(Q^c)$. As $(X, \mathcal{P})$ is connected, the issues $\{P, P^c\}$ and $\{Q, Q^c\}$ are directly dependent. Thus, there exist directly dependent

$\widehat{P} \in \{P, P^c\}$, $\widehat{Q} \in \{Q, Q^c\}$, contradicting weak independence. Consequently, $\mathcal{R}$ is not autonomous.

C.2.4 Proof of Proposition 4, Corollaries 1 and 2

Proposition 4

If $\mathcal{R}$ is not weakly independent, there exist disjoint $G, G' \subseteq N$ and directly dependent $P, Q \in \mathcal{P}$ such that $G \in \mathcal{R}(P)$ and $G' \in \mathcal{R}(Q)$. By direct dependence, $\{P, Q\} \subseteq \mathcal{G}$ for some critical $\mathcal{G} \subseteq \mathcal{P}$. Letting, for all $\widehat{P} \in \mathcal{G}\backslash\{P, Q\}$, $G_{\widehat{P}} = N \in \mathcal{R}(\widehat{P})$, we have $G \cap G' \cap \left(\bigcap_{\widehat{P} \in \mathcal{G}\backslash\{P,Q\}} G_{\widehat{P}}\right) = G \cap G' \cap N = \emptyset$, in violation of (IPC). Thus, $\mathcal{R}$ is not consistent.

Corollary 1

We have $\{P_{c>d}, P_{d>c}\} = \{P_{c>d}, P^c_{c>d}\}$ and $\{P_{c'>d'}, P_{d'>c'}\} = \{P_{c'>d'}, P^c_{c'>d'}\}$. Thus, $\mathcal{R}$ is autonomous. If $(X_{Lin(A)}, \mathcal{P}_{Lin(A)})$ is connected, we can use Fact 1 and Proposition 4 to prove the claim.

We show that $(X_{Lin(A)}, \mathcal{P}_{Lin(A)})$ is connected. Let $\{P_{a>b}, P_{b>a}\}, \{P_{c>d}, P_{d>c}\} \subseteq \mathcal{P}$ be any pair of issues. We have $a \neq b$ and $c \neq d$. If the issues are identical, they are trivially directly dependent. Suppose the issues are distinct. Then there are two possible cases: either all a, b, c, d are distinct or exactly one pair of elements are equal (if two pairs of elements are equal we have identical issues again).

Case 1: a, b, c, d are distinct. The family $\{P_{a>b}, P_{b>c}, P_{c>d}, P_{d>a}\}$ is critical; hence $\{P_{a>b}, P_{b>a}\}$ and $\{P_{c>d}, P_{d>c}\}$ are directly dependent.

Case 2: One pair of elements is equal. W.l.o.g., assume that $b = c$ (the proof for $b = d$, $a = c$ and $a = d$ is analogous). Then the family $\{P_{a>b}, P_{b>d}, P_{d>a}\} = \{P_{a>b}, P_{c>d}, P_{d>a}\}$ is critical, and thus, $\{P_{a>b}, P_{b>a}\}$ and $\{P_{c>d}, P_{d>c}\}$ are directly dependent.

Corollary 2

Seeing that $\{P_p, P_{\neg p}\} = \{P_p, P^c_p\}$ and $\{P_q, P_{\neg q}\} = \{P_q, P^c_q\}$, $\mathcal{R}$ is autonomous. The claim follows by Fact 1 and Proposition 4.

C.3 Some Lemmas

We establish some basic properties of $\Vdash^j$ and $\Vdash^\star$.

Lemma 8. *Let $\mathcal{C}, \mathcal{C}', \mathcal{C}'' \in \mathbb{C}$ and $j, j' \in \mathbb{N}$.*

1. *$\Vdash^j$ and $\Vdash^\star$ are reflexive.*

2. $(\mathcal{C} \Vdash^j \mathcal{C}', \mathcal{C}' \Vdash^{j'} \mathcal{C}'') \implies \mathcal{C} \Vdash^{(j+j')} \mathcal{C}''$.

3. *If $j' > j$, then $\mathcal{C} \Vdash^j \mathcal{C}' \implies \mathcal{C} \Vdash^{j'} \mathcal{C}'$. Thus, if $\mathcal{C}$ is j-critical, it is j'-critical.*

4. $\mathcal{C} \Vdash^j \mathcal{C}' = \{P_1, \ldots, P_r\} \iff (\forall k = 1, \ldots, r\, \exists \mathcal{C}_k \in \mathbb{C} : \mathcal{C}_k \Vdash^j \{P_k\}$ *and* $\mathcal{C} = \bigsqcup_{k=1,\ldots,r} \mathcal{C}_k)$,
$\mathcal{C} \Vdash^\star \mathcal{C}' = \{P_1, \ldots, P_r\} \iff (\forall k = 1, \ldots, r\, \exists \mathcal{C}_k \in \mathbb{C} : \mathcal{C}_k \Vdash^\star \{P_k\}$ *and* $\mathcal{C} = \bigsqcup_{k=1,\ldots,r} \mathcal{C}_k)$

Proof. 1. Let $\mathcal{C} = \{P_1, \ldots, P_r\} \in \mathbb{C}$. As $\{P, P^c\}$ is critical for all $P \in \mathcal{P}$, for $k = 1, \ldots, r$, $\mathbb{F} \ni \{P_k\} \vdash P_k$ and $\sqcup_{k=1,\ldots,r}\{P_k\} = \mathcal{C}$. Consequently, $\mathcal{C} \Vdash \mathcal{C}$. A fortiori, $\mathcal{C} \Vdash^\star \mathcal{C}$. Consider $j \geq 2$. For $l = 1, \ldots, j-1$, set $\mathcal{C}_l = \mathcal{C}$. We have: $\mathcal{C} \Vdash \mathcal{C}_{j-1} \Vdash \ldots \Vdash \mathcal{C}_1 \Vdash \mathcal{C}$, i.e., $\mathcal{C} \Vdash^j \mathcal{C}$. As $\mathcal{C}$ was chosen arbitrarily, the assertions follow.

2. Suppose $j, j' \geq 2$. There exist $\mathcal{C}_1, \ldots, \mathcal{C}_{j-1} \in \mathbb{C}$ such that $\mathcal{C} \Vdash \mathcal{C}_{j-1} \Vdash \ldots \Vdash \mathcal{C}_1 \Vdash \mathcal{C}'$ and $\mathcal{C}'_1, \ldots, \mathcal{C}'_{j'-1} \in \mathbb{C}$ such that $\mathcal{C}' \Vdash \mathcal{C}'_{j'-1} \Vdash \ldots \Vdash \mathcal{C}'_1 \Vdash \mathcal{C}''$. Letting $\mathcal{C}'_{j'} = \mathcal{C}', \mathcal{C}'_{j'+1} = \mathcal{C}_1, \ldots, \mathcal{C}'_{j'+j-1} = \mathcal{C}_{j-1}$ we have $\mathcal{C} \Vdash \mathcal{C}'_{j+j'-1} \Vdash \ldots \Vdash \mathcal{C}'_1 \Vdash \mathcal{C}''$. Hence $\mathcal{C} \Vdash^{(j+j')} \mathcal{C}''$. The proof is analogous when $j = 1$ or $j' = 1$.

3. Note that $j' - j \in \mathbb{N}$. By part 1, $\mathcal{C} \Vdash^{(j'-j)} \mathcal{C}$. By part 2, $\mathcal{C} \Vdash^{(j'-j)} \mathcal{C}, \mathcal{C} \Vdash^j \implies \mathcal{C} \Vdash^{j'} \mathcal{C}$.

4. For $\Vdash^j$, we prove the claim by induction over $j \in \mathbb{N}$.

Induction basis: $j = 1$. The claim holds by definition of $\Vdash^1 = \Vdash$.

Induction step: $j \rightsquigarrow j+1$. Assume the assertion holds for $\Vdash^j$ and consider $\mathcal{C}, \mathcal{C}' \in \mathbb{C}$ such that $\mathcal{C} \Vdash^{(j+1)} \mathcal{C}'$. There exist $\mathcal{C}_1, \ldots, \mathcal{C}_j \in \mathbb{C}$ such that $\mathcal{C} \Vdash \mathcal{C}_j \Vdash \ldots \Vdash \mathcal{C}_1 \Vdash \mathcal{C}'$. That is, $\mathcal{C} \Vdash^j \mathcal{C}_1 \Vdash \mathcal{C}' = \{P_1, \ldots, P_r\}$. By definition of $\Vdash$, there exist $\mathcal{F}_1, \ldots, \mathcal{F}_r \in \mathbb{F}$ such that, for all $k = 1, \ldots, r$, $\mathcal{F}_k \Vdash \{P_k\}$ and $\bigsqcup_{k=1,\ldots,r} \mathcal{F}_k = \mathcal{C}_1$. Let $s_1, \ldots, s_r \in \mathbb{N}$ and $P^1_1, \ldots, P^1_{s_1}, \ldots, P^r_1, \ldots, P^r_{s_r} \in \mathcal{P}$ be such that $\mathcal{F}_1 = \{P^1_1, \ldots, P^1_{s_1}\}, \ldots, \mathcal{F}_r = \{P^r_1, \ldots, P^r_{s_r}\}$. By (the inductive) assumption, there are $\mathcal{C}^1_1, \ldots, \mathcal{C}^1_{s_1}, \ldots, \mathcal{C}^r_1, \ldots, \mathcal{C}^r_{s_r} \in \mathbb{C}$ such that, for $k = 1, \ldots, r$, $l = 1, \ldots, s_k$, $\mathcal{C}^k_l \Vdash^j \{P^k_l\}$ and $\mathcal{C} = \bigsqcup_{k=1,\ldots,r,l=1,\ldots,s_k} \mathcal{C}^k_l$. Letting $\mathcal{C}^1 = \bigsqcup_{l=1,\ldots,s_1} \mathcal{C}^1_l, \ldots, \mathcal{C}^r = \bigsqcup_{l=1,\ldots,s_r} \mathcal{C}^r_l$, we have $\mathcal{C} = \bigsqcup_{k=1,\ldots,r,l=1,\ldots,s_k} \mathcal{C}^k_l = \bigsqcup_{k=1,\ldots,r} \bigsqcup_{l=1,\ldots,s_k} \mathcal{C}^k_l = \bigsqcup_{k=1,\ldots,r} \mathcal{C}^k$. Additionally, for $k = 1, \ldots, r$, $\mathcal{C}^k \Vdash^j \mathcal{F}_k \Vdash \{P_k\}$ (by Lemma 9, part 3 below); hence $\mathcal{C}^k \Vdash^{(j+1)} \{P_k\}$ (by part 2).

If $\mathcal{C} \Vdash^\star \mathcal{C}'$ we have $\mathcal{C} \Vdash^{j^\star} \mathcal{C}'$ for some $j^\star \in \mathbb{N}$. There exist $\mathcal{C}_1, \ldots, \mathcal{C}_r \in \mathbb{C}$ such that $\mathcal{C} = \sqcup_{k=1,\ldots,r}\mathcal{C}_k$ and, for all $k = 1, \ldots, r$, $\mathcal{C}_k \Vdash^{j^\star} \{P_k\}$; hence $\mathcal{C}_k \Vdash^\star \{P_k\}$.

□

Lemma 9. *Let* $j, r \in \mathbb{N}, r \geq 2$ *and* $\mathcal{C}_1, \ldots, \mathcal{C}_r, \mathcal{C}'_1, \ldots, \mathcal{C}'_r \in \mathbb{C}$.

1. *If* $\mathcal{C}_1 \Vdash^j \mathcal{C}'_1$ *and* $\mathcal{C}_2 \Vdash^j \mathcal{C}'_2$, *then* $(\mathcal{C}_1 \sqcup \mathcal{C}_2) \Vdash^j (\mathcal{C}'_1 \sqcup \mathcal{C}'_2)$.

2. *If* $\mathcal{C}_1 \Vdash^\star \mathcal{C}'_1$ *and* $\mathcal{C}_2 \Vdash^\star \mathcal{C}'_2$, *then* $(\mathcal{C}_1 \sqcup \mathcal{C}_2) \Vdash^\star (\mathcal{C}'_1 \sqcup \mathcal{C}'_2)$.

3. *If* $\mathcal{C}_1 \Vdash^j \mathcal{C}'_1, \ldots, \mathcal{C}_r \Vdash^j \mathcal{C}'_r$, *then* $\bigsqcup_{k=1,\ldots,r} \mathcal{C}_k \Vdash^j \bigsqcup_{k=1,\ldots,r} \mathcal{C}'_k$.

4. *If* $\mathcal{C}_1 \Vdash^\star \mathcal{C}'_1, \ldots, \mathcal{C}_r \Vdash^\star \mathcal{C}'_r$, *then* $\bigsqcup_{k=1,\ldots,r} \mathcal{C}_k \Vdash^\star \bigsqcup_{k=1,\ldots,r} \mathcal{C}'_k$.

Proof. 1. We show the claim by induction over $j \in N$.

Induction basis: $j = 1$. Let $\mathcal{C}'_1 = \{P^1_1, \ldots, P^1_{r_1}\}, \mathcal{C}'_2 = \{P^2_1, \ldots, P^2_{r_2}\}$. As $\mathcal{C}_1 \Vdash \mathcal{C}'_1, \mathcal{C}_2 \Vdash \mathcal{C}'_2$, there are $\mathcal{C}^1_1, \ldots, \mathcal{C}^1_{r_1}, \mathcal{C}^2_1, \ldots, \mathcal{C}^2_{r_2} \in \mathbb{C}$ such that: for $k = 1, \ldots, r_1$, $\mathcal{C}^1_k \Vdash \{P^1_k\}$ and $\mathcal{C}_1 = \bigsqcup_{k=1,\ldots,r_1} \mathcal{C}^1_k$; for $k = 1, \ldots, r_2$, $\mathcal{C}^2_k \Vdash \{P^2_k\}$ and $\mathcal{C}_2 = \bigsqcup_{k=1,\ldots,r_2} \mathcal{C}^2_k$. As $(\bigsqcup_{k=1,\ldots,r_1} \mathcal{C}^1_k) \sqcup (\bigsqcup_{k=1,\ldots,r_2} \mathcal{C}^2_k) = \mathcal{C}_1 \sqcup \mathcal{C}_2$, we have $(\mathcal{C}_1 \sqcup \mathcal{C}_2) \Vdash (\mathcal{C}'_1 \sqcup \mathcal{C}'_2)$.

Induction step: $j \rightsquigarrow j+1$. Assume the claim holds for $\Vdash^j$. As $\mathcal{C}_1 \Vdash^{(j+1)} \mathcal{C}'_1$, $\mathcal{C}_2 \Vdash^{(j+1)} \mathcal{C}'_2$, there are $\mathcal{C}^1_1, \ldots, \mathcal{C}^1_j, \mathcal{C}^2_1, \ldots, \mathcal{C}^2_j \in \mathbb{C}$ such that $\mathcal{C}_1 \Vdash \mathcal{C}^1_j \Vdash \ldots \Vdash \mathcal{C}^1_1 \Vdash \mathcal{C}'_1$ and $\mathcal{C}_2 \Vdash \mathcal{C}^2_j \Vdash \ldots \Vdash \mathcal{C}^2_1 \Vdash \mathcal{C}'_2$. We have $\mathcal{C}^1_j \Vdash^j \mathcal{C}'_1$ and $\mathcal{C}^2_j \Vdash^j \mathcal{C}'_2$. By (the inductive) assumption and the induction basis: $\mathcal{C}_1 \sqcup \mathcal{C}_2 \Vdash \mathcal{C}^1_j \sqcup \mathcal{C}^2_j \Vdash^j \mathcal{C}'_1 \sqcup \mathcal{C}'_2$. Thus, by Lemma 8, part 2, $\mathcal{C}_1 \sqcup \mathcal{C}_2 \Vdash^{(j+1)} \mathcal{C}'_1 \sqcup \mathcal{C}'_2$.

2. There exist $j, j' \in \mathbb{N}$ such that $\mathcal{C}_1 \Vdash^j \mathcal{C}'_1$, $\mathcal{C}_2 \Vdash^{j'} \mathcal{C}'_2$. W.l.o.g., let $j \geq j'$. If $j = j'$, $\mathcal{C}_2 \Vdash^j \mathcal{C}'_2$. By Lemma 8, part 3, the same holds when $j > j'$. By part 1, we have $(\mathcal{C}_1 \sqcup \mathcal{C}_2) \Vdash^j (\mathcal{C}'_1 \sqcup \mathcal{C}'_2)$. A fortiori, $(\mathcal{C}_1 \sqcup \mathcal{C}_2) \Vdash^\star (\mathcal{C}'_1 \sqcup \mathcal{C}'_2)$.

3. By induction over $r \in \mathbb{N}$.

   Induction basis: $r = 2$. See part 1.

   Induction step: $r \rightsquigarrow r+1$. Assume the the claim holds $r \in \mathbb{N}, r \geq 2$. We have $\bigsqcup_{k=1,\ldots,r+1} \mathcal{C}_k = (\bigsqcup_{k=1,\ldots,r} \mathcal{C}_k) \sqcup \mathcal{C}_{r+1}$. By (the inductive) assumption, $\bigsqcup_{k=1,\ldots,r} \mathcal{C}_k \Vdash^j \bigsqcup_{k=1,\ldots,r} \mathcal{C}'_k$. Thus, by part 1, $\bigsqcup_{k=1,\ldots,r+1} \mathcal{C}_k = (\bigsqcup_{k=1,\ldots,r} \mathcal{C}_k) \sqcup \mathcal{C}_{r+1} \Vdash^j (\bigsqcup_{k=1,\ldots,r} \mathcal{C}'_k) \sqcup \mathcal{C}'_{r+1} = \bigsqcup_{k=1,\ldots,r+1} \mathcal{C}'_k$.

4. By induction over $r \in \mathbb{N}$.

   Induction basis: $r = 2$. See part 2.

   Induction step: $r \rightsquigarrow r+1$. Assume the claim holds for $r \in \mathbb{N}, r \geq 2$. We have $\bigsqcup_{k=1,\ldots,r+1} \mathcal{C}_k = (\bigsqcup_{k=1,\ldots,r} \mathcal{C}_k) \sqcup \mathcal{C}_{r+1}$. By (the inductive) assumption, $\bigsqcup_{k=1,\ldots,r} \mathcal{C}_k \Vdash^\star \bigsqcup_{k=1,\ldots,r} \mathcal{C}'_k$. Thus, by part 2, $\bigsqcup_{k=1,\ldots,r+1} \mathcal{C}_k = (\bigsqcup_{k=1,\ldots,r} \mathcal{C}_k) \sqcup \mathcal{C}_{r+1} \Vdash^\star (\bigsqcup_{k=1,\ldots,r} \mathcal{C}'_k) \sqcup \mathcal{C}'_{r+1} = \bigsqcup_{k=1,\ldots,r+1} \mathcal{C}'_k$.

□

If $\emptyset \neq \mathcal{C} \subseteq \mathcal{C}' \in \mathbb{C}$, w.l.o.g., there exist $P_1, \ldots, P_r, \ldots, P_s \in \mathcal{P}$ (with $1 \leq r \leq s$) such that $\mathcal{C} = \{P_1, \ldots, P_r\}$ and $\mathcal{C}' = \{P_1, \ldots, P_r, \ldots, P_s\}$. We define $\mathcal{C}' \backslash \mathcal{C} = \{P_{r+1}, \ldots, P_s\}$ (hence $\mathcal{C} \backslash \mathcal{C} = \emptyset$). Now for arbitrary $\mathcal{C}, \mathcal{C}' \in \mathbb{C}$, we define $\mathcal{C}' \backslash \emptyset = \mathcal{C}'$ and $\mathcal{C}' \backslash \mathcal{C} = \mathcal{C}' \backslash (\mathcal{C}' \sqcap \mathcal{C})$, where $\sqcap$ is the natural intersection operator on $\mathcal{C}$ (such that, e.g., $\{u, u', u'\} \sqcap \{u', u', u''\} = \{u', u'\}$ on the universe $U = \{u, u', u''\}$).

Lemma 10. *Let $j \in \mathbb{N}$, and $\mathcal{C} = \{P_1, \ldots, P_r\} \in \mathbb{C}$ be j-critical. Then, for $k = 1, \ldots, r$, $\mathcal{C} \backslash \{P_k\} = \{P_1, \ldots, P_{k-1}, P_{k+1}, \ldots, P_r\} \Vdash^{(2j+1)} \{P^c_k\}$. Thus, if $\mathcal{C}$ is almost critical, $\mathcal{C} \backslash \{P_k\} \Vdash^\star \{P^c_k\}$.*

Proof. We prove the assertion by induction over $j \in \mathbb{N}$.

Induction basis: $j = 1$. Note that, if $\mathcal{G} = \{Q_1, \ldots, Q_s\} \in \mathbb{F}$ is critical, we have, for $l = 1, \ldots, s$, $\mathcal{G}\backslash\{Q_l\} \vdash Q_l^c$ or, equivalently, $\mathcal{G}\backslash\{Q_l\} \Vdash \{Q_l^c\}$. Now, as $\mathcal{C}$ is 1-critical, $\mathcal{C} \Vdash \mathcal{G}$ for some critical $\mathcal{G} \in \mathbb{F}$. Let $Q_1, \ldots, Q_s \in \mathcal{P}$ be such that $\mathcal{G} = \{Q_1, \ldots, Q_s\}$. There exist $\mathcal{F}_1, \ldots, \mathcal{F}_s \in \mathbb{F}$ such that: (i) for $l = 1, \ldots, s$, $\mathcal{F}_l \vdash Q_l$ and (ii) $\bigsqcup_{l=1,\ldots,s} \mathcal{F}_l = \mathcal{C}$. By (ii), for every $k = 1, \ldots, r$, there exists some $1 \leq l_k \leq s$ such that $P_k \in \mathcal{F}_{l_k}$. By Lemma 9, part 3, we have $\mathcal{C}\backslash\mathcal{F}_{l_k} = \bigsqcup_{l\in\{1,\ldots,s\}\backslash\{l_k\}} \mathcal{F}_l \Vdash \mathcal{G}\backslash\{Q_{l_k}\}$. By reflexivity (Lemma 8, part 1), $\mathcal{F}_{l_k}\backslash\{P_k\} \Vdash \mathcal{F}_{l_k}\backslash\{P_k\}$. $\mathcal{F}_{l_k} \vdash Q_{l_k}$ implies that $\mathcal{F}_{l_k} \cup \{Q_{l_k}^c\}$ is critical. Thus, using Lemma 9, part 1, we obtain $\mathcal{C}\backslash\{P_k\} = (\mathcal{C}\backslash\mathcal{F}_{l_k}) \sqcup (\mathcal{F}_{l_k}\backslash\{P_k\}) \Vdash (\mathcal{G}\backslash\{Q_{l_k}\}) \sqcup (\mathcal{F}_{l_k}\backslash\{P_k\}) \Vdash \{Q_{l_k}^c\} \sqcup (\mathcal{F}_{l_k}\backslash\{P_k\}) = (\mathcal{F}_{l_k} \cup \{Q_{l_k}^c\})\backslash\{P_k\} \Vdash \{P_k^c\}$. By Lemma 8, part 2, $\mathcal{C}\backslash\{P_k\} \Vdash^3 \{P_k^c\}$.

Induction step: $j \rightsquigarrow j + 1$. Assume the claim holds for $j \in \mathbb{N}$. If $\mathcal{C}$ is $(j + 1)$-critical, there exist $\mathcal{C}_1, \ldots, \mathcal{C}_j \in \mathbb{C}$ and some critical $\mathcal{G} \in \mathbb{F}$ such that $\mathcal{C} \Vdash \mathcal{C}_j \Vdash \ldots \Vdash \mathcal{C}_1 \Vdash \mathcal{G}$. Clearly, $\mathcal{C}_j$ is j-critical. Let $Q_1, \ldots, Q_s \in \mathcal{P}$ be such that $\mathcal{C}_j = \{Q_1, \ldots, Q_s\}$. As $\mathcal{C} \Vdash \mathcal{C}_j$, there exist $\mathcal{F}_1, \ldots, \mathcal{F}_s \in \mathbb{F}$ such that: (i) for all $l = 1, \ldots, s$, $\mathcal{F}_l \vdash Q_l$ and (ii) $\bigsqcup_{l=1,\ldots,s} \mathcal{F}_l = \mathcal{C}$. By (ii), there exists some $1 \leq l_k \leq s$ such that $P_k \in \mathcal{F}_{l_k}$. By Lemma 9, part 3, we have $\mathcal{C}\backslash\mathcal{F}_{l_k} = \bigsqcup_{l\in\{1,\ldots,s\}\backslash\{l_k\}} \mathcal{F}_l \Vdash \mathcal{C}_j\backslash\{Q_{l_k}\}$. By (the inductive) assumption $\mathcal{C}_j\backslash\{Q_{l_k}\} \Vdash^{2j+1} \{Q_{l_k}^c\}$. By reflexivity (Lemma 8, part 1), $\mathcal{F}_{l_k}\backslash\{P_k\} \Vdash \mathcal{F}_{l_k}\backslash\{P_k\}$ and $\mathcal{F}_{l_k}\backslash\{P_k\} \Vdash^j \mathcal{F}_{l_k}\backslash\{P_k\}$. As $\mathcal{F}_{l_k} \vdash Q_{l_k}$, $\mathcal{F}_{l_k} \cup \{Q_{l_k}^c\}$ is critical. Thus, using Lemma 9, part 1, we obtain $\mathcal{C}\backslash\{P_k\} = (\mathcal{C}\backslash\mathcal{F}_{l_k}) \sqcup (\mathcal{F}_{l_k}\backslash\{P_k\}) \Vdash (\mathcal{C}_j\backslash\{Q_{l_k}\}) \sqcup (\mathcal{F}_{l_k}\backslash\{P_k\}) \Vdash^{2j+1} \{Q_{l_k}^c\} \sqcup (\mathcal{F}_{l_k}\backslash\{P_k\}) = (\mathcal{F}_{l_k} \cup \{Q_{l_k}^c\})\backslash\{P_k\} \Vdash \{P_k^c\}$. By Lemma 8, part 2, $\mathcal{C}\backslash\{P_k\} \Vdash^{2j+1+2} \{P_k^c\} = \mathcal{C}\backslash\{P_k\} \Vdash^{2(j+1)+1} \{P_k^c\}$.

Lastly, $\mathcal{C}$ is almost critical if and only if it is j'-critical for some $j' \in \mathbb{N}$. Thus, $\mathcal{C}\backslash\{P_k\} \Vdash^{(2j'+1)} \{P_k^c\}$. A fortiori, $\mathcal{C}\backslash\{P_k\} \Vdash^\star \{P_k^c\}$. □

We recall the definition of minimal entailment closure. For $\mathcal{R} : \mathcal{P} \rightrightarrows 2^N\backslash\{\emptyset\}$, define $P \mapsto C^1(\mathcal{R})(P) = \{G \subseteq N : G = \bigcap_{k=1,\ldots,r} G_k$ for some $G_1 \in \mathcal{R}(Q_1), \ldots, G_r \in \mathcal{R}(Q_r)$ such that $\{Q_1, \ldots, Q_r\} \vdash P\}$. For $j \geq 2$, we define $C^j(\mathcal{R})$ inductively by $P \mapsto C^j(\mathcal{R})(P) = C^1(C^{j-1}(\mathcal{R}))(P)$. Lastly, for every $P \in \mathcal{P}$, $C^\star(\mathcal{R})(P) = \bigcup_{j\in\mathbb{N}} C^j(\mathcal{R})(P)$.

Lemma 11. *Let $\mathcal{R} : \mathcal{P} \rightrightarrows 2^N\backslash\{\emptyset\}$, $P \in \mathcal{P}$ and $j \in \mathbb{N}$.*

1. *$G \in C^j(\mathcal{R})(P)$ if and only if there exist $r \in \mathbb{N}$, $Q_1, \ldots, Q_r \in \mathcal{P}$ and $G_1 \in \mathcal{R}(Q_1), \ldots, G_r \in \mathcal{R}(Q_r)$ such that $\{Q_1, \ldots, Q_r\} \Vdash^j \{P\}$ and $\bigcap_{k=1,\ldots,r} G_k = G$.*

2. *$G \in C^\star(\mathcal{R})(P)$ if and only if there exist $r \in \mathbb{N}$, $Q_1, \ldots, Q_r \in \mathcal{P}$ and $G_1 \in \mathcal{R}(Q_1), \ldots, G_r \in \mathcal{R}(Q_r)$ such that $\{Q_1, \ldots, Q_r\} \Vdash^\star \{P\}$ and $\bigcap_{k=1,\ldots,r} G_k = G$.*

Proof. 1. We show the claim by induction over $j \in \mathbb{N}$.

Induction basis: $j = 1$. The claim holds by definition of $C^1(\mathcal{R})$.

Induction step: $j \rightsquigarrow j + 1$. Assume the claim holds for $j \in \mathbb{N}$. We have

$$
\begin{aligned}
G \in \mathcal{C}^{(j+1)}(\mathcal{R})(P) &\iff \big(G \in C^1(C^j(\mathcal{R}))(P)\big)\\
&\overset{IB}{\iff} (\exists r \in \mathbb{N} \exists Q_1, \dots, Q_r \in \mathcal{P} \exists G_1 \in \mathcal{C}^j(\mathcal{R})(Q_1), \dots,\\
&\qquad G_r \in \mathcal{C}^j(\mathcal{R})(Q_r) : G = \bigcap_{k=1,\dots,r} G_k \text{ and } \{Q_1, \dots, Q_r\} \Vdash \{P\})\\
&\overset{IA}{\iff} (\exists r \in \mathbb{N} \exists Q_1, \dots, Q_r \in \mathcal{P} \exists G_1, \dots, G_r \in 2^N \backslash \{\emptyset\} :\\
&\qquad G = \bigcap_{k=1,\dots,r} G_k,\ \{Q_1, \dots, Q_r\} \Vdash \{P\} \text{ and } \forall k = 1, \dots, r\\
&\qquad \exists s_1, \dots, s_k \in \mathbb{N} \exists Q_1^k, \dots, Q_{s_k}^k \in \mathcal{P} \exists G_1^k \in \mathcal{R}(Q_1^k), \dots,\\
&\qquad G_{s_k}^k \in \mathcal{R}(Q_{s_k}^k) : G_k = \bigcap_{l=1,\dots,s_k} G_l^k \text{ and } \{Q_1^k, \dots, Q_{s_k}^k\} \Vdash^j \{Q_k\})\\
&\iff (\exists r \in \mathbb{N} \exists s_1, \dots, s_r \in \mathbb{N} \exists Q_1^1, \dots, Q_{s_1}^1, \dots, Q_1^r, \dots, Q_{s_r}^r \in \mathcal{P}\\
&\qquad \exists G_1^1 \in \mathcal{R}(Q_1^1), \dots, G_{s_1}^1 \in \mathcal{R}(Q_{s_1}^1), \dots, G_1^r \in \mathcal{R}(Q_1^r), \dots,\\
&\qquad G_{s_r}^r \in \mathcal{R}(Q_{s_r}^r) : G = \bigcap_{\substack{k=1,\dots,r\\ l=1,\dots,s_k}} G_l^k \text{ and } \{Q_1^1, \dots, Q_{s_r}^r\} \Vdash^{j+1} \{P\});
\end{aligned}
$$

where the last equivalence uses Lemma 9, part 3.

2. We have, for all $P \in \mathcal{P}$:

$$
\begin{aligned}
G \in C^\star(\mathcal{R})(P) &\iff \Big(\exists j^\star \in \mathbb{N} : G \in C^{j^\star}(\mathcal{R})(P)\Big)\\
&\overset{1.}{\iff} (\exists j^\star \in \mathbb{N}\, \exists r \in \mathbb{N}\, \exists Q_1, \dots, Q_r \in \mathcal{P}\, \exists G_1 \in \mathcal{R}(Q_1), \dots,\\
&\qquad G_k \in \mathcal{R}(Q_k) : \{Q_1, \dots, Q_r\} \Vdash^{j^\star} \{P\} \text{ and } \bigcap_{k=1,\dots,r} G_k = G)\\
&\iff (\exists r \in \mathbb{N}\, \exists Q_1, \dots, Q_r \in \mathcal{P}\, \exists G_1 \in \mathcal{R}(Q_1), \dots, G_k \in \mathcal{R}(Q_k) :\\
&\qquad \{Q_1, \dots, Q_r\} \Vdash^\star \{P\} \text{ and } \bigcap_{k=1,\dots,r} G_k = G).
\end{aligned}
$$

□

Lemma 12. *Let $\mathcal{R} : \mathcal{P} \rightrightarrows 2^N \backslash \{\emptyset\}$ be exhaustive and satisfy* (IPC) *(i.e., be a consistent exhaustive rights system). For all $j \in \mathbb{N}$, $Q_1, \dots, Q_r, P \in \mathcal{P}$ and $G_1 \in \mathcal{R}(Q_1), \dots, G_r \in \mathcal{R}(Q_r)$:*

1. $\{Q_1, \dots, Q_r\} \Vdash^j \{P\} \implies \bigcap_{k=1,\dots,r} G_k \in \mathcal{R}(P)$,

2. $\{Q_1, \dots, Q_r\} \Vdash^\star \{P\} \implies \bigcap_{k=1,\dots,r} G_k \in \mathcal{R}(P)$.

Proof. 1. By induction over $j \in \mathbb{N}$.

Induction basis: $j = 1$. We have $\{Q_1, \dots, Q_r\} \Vdash \{P\} \iff \{Q_1, \dots, Q_r\} \vdash P \iff \{Q_1, \dots, Q_r, P^c\}$ is critical. If $\bigcap_{k=1,\dots,r} G_k = N$, $N \backslash \Big(\bigcap_{k=1,\dots,r} G_k\Big) = \emptyset \notin$

$\mathcal{R}(P^c) \implies \bigcap_{k=1,\ldots,r} G_k = N \in \mathcal{R}(P)$ as $\mathcal{R}$ is exhaustive. If $\bigcap_{k=1,\ldots,r} G_k \neq N$, suppose for a contradiction that $\bigcap_{k=1,\ldots,r} G_k \notin \mathcal{R}(P)$. As $\mathcal{R}$ is exhaustive, $\bigcap_{k=1,\ldots,r} G_k \notin \mathcal{R}(P) \implies N \backslash \left(\bigcap_{k=1,\ldots,r} G_k\right) \in \mathcal{R}(P^c)$ yielding a violation of (IPC) over critical $\{Q_1, \ldots, Q_r, P^c\}$.

Induction step: $j \rightsquigarrow j+1$. Assume the claim holds for $j \in \mathbb{N}$. If $\{Q_1, \ldots, Q_r\} \Vdash^{(j+1)} \{P\}$, there exist $\mathcal{C}_1, \ldots, \mathcal{C}_j \in \mathbb{C}$ such that $\{Q_1, \ldots, Q_r\} \Vdash \mathcal{C}_j \Vdash \ldots \Vdash \mathcal{C}_1 \Vdash \{P\}$. Hence $\{Q_1, \ldots, Q_r\} \Vdash^j \mathcal{C}_1 \Vdash \{P\}$. Let $P_1, \ldots, P_s \in \mathcal{P}$ be such that $\mathcal{C}_1 = \{P_1, \ldots, P_s\}$. By Lemma 8, part 4, there exist $\mathcal{C}^1, \ldots, \mathcal{C}^s \in \mathbb{C}$ such that $\bigsqcup_{l=1,\ldots,s} \mathcal{C}^l = \{Q_1, \ldots, Q_r\}$ and, for $l = 1, \ldots, s$, $\mathcal{C}^l \Vdash^j \{P_l\}$. W.l.o.g., for $l = 1, \ldots, s$, $\mathcal{C}^l = \{Q_{r_{l-1}+1}, \ldots, Q_{r_l}\}$ for some $r_0 = 0 < r_1 < r_2 < \cdots < r_{s-1} < r = r_s$. By (the inductive) assumption, for all $l = 1, \ldots, s$, $\bigcap_{k=r_{l-1}+1,\ldots,r_l} G_k \in \mathcal{R}(P_l)$. Thus, by the induction basis, $\bigcap_{k=1,\ldots,r} G_k = \bigcap_{l=1,\ldots,s} \left(\bigcap_{k=r_{l-1}+1,\ldots,r_l} G_k\right) \in \mathcal{R}(P)$.

2. $\{Q_1, \ldots, Q_r\} \Vdash^\star \{P\} \iff \left(\{Q_1, \ldots, Q_r\} \Vdash^{j^\star} \{P\} \text{ for some } j^\star \in \mathbb{N}\right)$. Hence the claim follows from part 1.

□

Lemma 13. *Let $\mathcal{R} : \mathcal{P} \rightrightarrows 2^N \backslash \{\emptyset\}$ satisfy* (IPAC). *For all $P \in \mathcal{P}$, $G \in 2^N \backslash \{\emptyset\}$, define*

$$\mathcal{P} \ni \widehat{P} \mapsto \mathcal{R}_{(G,P)}(\widehat{P}) = \begin{cases} \mathcal{R}(\widehat{P}) & \text{if } \widehat{P} \neq P \\ \mathcal{R}(\widehat{P}) \cup \{G\} & \text{if } \widehat{P} = P \end{cases}.$$

If $G \in 2^N \backslash \{\emptyset\}$, $P \in \mathcal{P}$ are such that $G \notin \mathcal{R}(P)$ and $N \backslash G \notin \mathcal{R}(P^c)$ then $\mathcal{R}_{(G,P)}$ or $\mathcal{R}_{(N\backslash G,P^c)}$ satisfies (IPAC).

Proof. Suppose that both $\mathcal{R}_{(G,P)}$ and $\mathcal{R}_{(N\backslash G,P^c)}$ violate (IPAC). That is, there exist almost critical $\mathcal{C}_1 = \{Q_1^1, \ldots, Q_r^1\} \in \mathbb{C}$ and $\mathcal{C}_2 = \{Q_1^2, \ldots, Q_s^2\} \in \mathbb{C}$ as well as $G_1^1 \in \mathcal{R}_{(G,P)}(Q_1^1), \ldots, G_r^1 \in \mathcal{R}_{(G,P)}(Q_r^1)$ and $G_1^2 \in \mathcal{R}_{(N\backslash G,P^c)}(Q_1^2), \ldots, G_s^2 \in \mathcal{R}_{(N\backslash G,P^c)}(Q_s^2)$ such that (i) $\bigcap_{k=1,\ldots,r} G_k^1 = \emptyset$ and (ii) $\bigcap_{l=1,\ldots,s} G_l^2 = \emptyset$. Unless, for some $1 \le k' \le r$, $Q_{k'}^1 = P$ and $G_{k'}^1 = G$, $\mathcal{R}$ violates (IPAC), in contradiction with our assumption. Analogously, for some $1 \le l' \le s$, we must have $Q_{l'}^2 = P^c$ and $G_{l'}^2 = N \backslash G$. For all $k \in \{1, \ldots, r\} \backslash \{k'\}$ and all $l \in \{1, \ldots, s\} \backslash \{l'\}$, define:

$$\widehat{G}_k^1 = \begin{cases} G_k^1 & \text{if } G_k^1 \in \mathcal{W}(Q_k^1) \\ N & \text{if } G_k^1 = G \text{ and } Q_k^1 = P \end{cases} \text{ and } \widehat{G}_l^2 = \begin{cases} G_l^2 & \text{if } G_l^2 \in \mathcal{W}(Q_l^2) \\ N & \text{if } G_l^2 = N \backslash G \text{ and } Q_l^2 = P^c \end{cases}.$$

Thus, for all $k \neq k'$, $\widehat{G}_k^1 \in \mathcal{R}(Q_k^1) \cup \{N\}$ and, for all $l \neq l'$, $\widehat{G}_l^2 \in \mathcal{R}(Q_l^2) \cup \{N\}$. By (i), $\bigcap_{k=1,\ldots,r,k\neq k'} \widehat{G}_k^1 \subseteq N \backslash G$. By (ii), $\bigcap_{l=1,\ldots,s,l\neq l'} \widehat{G}_l^2 \subseteq G$. If we show that $\mathcal{C}_1 \backslash \{Q_{k'}^1\} \sqcup \mathcal{C}_2 \backslash \{Q_{l'}^2\}$ is almost critical, we deduce that $\mathcal{R}$ violates (IPAC), a contradiction.

To complete the proof, we show that $\mathcal{C}_1 \backslash \{Q_{k'}^1\} \sqcup \mathcal{C}_2 \backslash \{Q_{l'}^2\}$ is indeed almost critical. As $\mathcal{C}_1, \mathcal{C}_2$ are almost critical, we can use Lemma 10 to conclude that $\mathcal{C}_1 \backslash \{Q_{k'}^1\} \Vdash^\star \{(Q_{k'}^1)^c\} =$

$\{P^c\}$ and $\mathcal{C}_2\backslash\{Q^2_{l'}\} \Vdash^\star \{(Q^2_{l'})^c\} = \{(P^c)^c\} = \{P\}$. By Lemma 9, part 2, $(\mathcal{C}_1\backslash\{Q^1_{k'}\}) \sqcup (\mathcal{C}_2\backslash\{Q^2_{l'}\}) \Vdash^\star \{P^c\} \sqcup \{P\} = \{P, P^c\}$. As $\{P, P^c\}$ is critical, $\mathcal{C}_1\backslash\{Q^1_{k'}\} \sqcup \mathcal{C}_2\backslash\{Q^2_{l'}\}$ is almost critical. □

Lemma 14. *Let $P, Q \in \mathcal{P}$.*

1. $P \trianglerighteq^\star Q \iff (\exists \mathcal{C} \in \mathbb{C} : P \in \mathcal{C} \Vdash^\star \{Q\})$. *Moreover, $\{P\} \Vdash^\star \{Q\}$ only if $P \subseteq Q$.*
2. *If $P \trianglerighteq^\star Q$, $Q \trianglerighteq^\star P$ and $P \neq Q$, then there exists some collection $\mathcal{C}' \in \mathbb{C}$ such that $|\mathcal{C}'| \geq 2$ and $P \in \mathcal{C}' \Vdash^\star \{Q\}$.*

Proof. 1. We have

$$\begin{aligned}
P \trianglerighteq^\star Q &\iff \exists r \in \mathbb{N} \exists Q_1, \ldots, Q_r \in \mathcal{P} : P \trianglerighteq Q_1 \trianglerighteq \cdots \trianglerighteq Q_r \trianglerighteq Q \\
&\iff \exists r \in \mathbb{N} \exists Q_1, \ldots, Q_r \in \mathcal{P} \exists \mathcal{F}, \mathcal{F}_1, \ldots, \mathcal{F}_r \in \mathbb{F} : \\
&\qquad P \in \mathcal{F} \vdash Q_1, Q_1 \in \mathcal{F}_1 \vdash Q_2, \ldots, Q_r \in \mathcal{F}_r \vdash Q \\
&\iff \exists r \in \mathbb{N} \exists Q_1, \ldots, Q_r \in \mathcal{P} \exists \mathcal{F}, \mathcal{F}_1, \ldots, \mathcal{F}_r \in \mathbb{F} : \\
&\qquad P \in \mathcal{F} \sqcup \left(\bigsqcup_{k=1,\ldots r} \mathcal{F}_k \backslash \{Q_k\} \right) \Vdash^\star \{Q\} \\
&\iff \exists \mathcal{C} \in \mathbb{C} : P \in \mathcal{C} \Vdash^\star \{Q\}.
\end{aligned}$$

By non-triviality of $\mathcal{P}$, all properties are individually consistent. Thus, for every $\widehat{P}, \widehat{Q} \in \mathcal{P}$, it holds $\widehat{P} \subseteq \widehat{Q} \iff \left(\left\{\widehat{P}, \widehat{Q}^c\right\} \text{ is critical}\right) \iff \left\{\widehat{P}\right\} \vdash \widehat{Q} \iff \left\{\widehat{P}\right\} \Vdash \left\{\widehat{Q}\right\}$. Now suppose that $\{P\} \Vdash^\star \{Q\}$. We must have $\{P\} \Vdash \{Q_1\} \Vdash \ldots \Vdash \{Q_r\} \Vdash \{Q\}$ for some $Q_1, \ldots, Q_r \in \mathcal{P}$. Thus, $P \subseteq Q_1 \subseteq \cdots \subseteq Q_r \subseteq Q$.

2. By part 1, there exist $\mathcal{C}_1, \mathcal{C}_2 \in \mathbb{C}$ such that $Q \in \mathcal{C}_1 \Vdash^\star P$ and $P \in \mathcal{C}_2 \Vdash^\star Q$. As $P \neq Q$, we have $|\mathcal{C}_1| \geq 2$ or $|\mathcal{C}_2| \geq 2$. If $|\mathcal{C}_2| \geq 2$, take $\mathcal{C}' = \mathcal{C}_2$. Otherwise, $\{P\} \Vdash^\star Q$. Then, by Lemma 9, part 4, $\{P\} \sqcup (\mathcal{C}_1\backslash\{Q\}) \Vdash^\star \mathcal{C}_1 \Vdash^\star \{P\} \Vdash^\star \{Q\}$. Thus, take $\mathcal{C}' = \{P\} \sqcup (\mathcal{C}_1\backslash\{Q\})$.

□

C.4 Proofs for Chapter 13

C.4.1 Proof of Fact 3

The proof is contained in the main text.

C.4.2 Proof of Fact 4

We have $\{Q_1, \ldots, Q_r\} \vdash P \iff \{Q_1, \ldots, Q_r\} \Vdash \{P\}$. The claim follows by Lemma 12.

C.4.3 Proof of Theorem 7

Fact 3 and Theorem 6 imply equivalence of 1 and 2. We complete the proof by showing $3 \Longrightarrow 2 \Longrightarrow 4 \Longrightarrow 3$.

Proof of $3 \Longrightarrow 2$

If $\mathcal{R}$ is exhaustive, take $\mathcal{R}' = \mathcal{R}$. Suppose $\mathcal{R}$ is not exhaustive. That is, there exist some $G \in 2^N \backslash \{\emptyset\}$ and some $P \in \mathcal{P}$ such that $G \notin \mathcal{R}(P)$ and $N \backslash G \notin \mathcal{R}(P^c)$. By Lemma 13, there exists some rights system $\mathcal{R}^{(1)}$ satisfying (IPAC) which extends $\mathcal{R}$ such that $G \in \mathcal{R}^{(1)}(P)$ or $N \backslash G \in \mathcal{R}^{(1)}(P^c)$. If $\mathcal{R}^{(1)}$ is exhaustive, take $\mathcal{R}' = \mathcal{R}^{(1)}$. Otherwise use Lemma 13 again to extend $\mathcal{R}^{(1)}$ to some $\mathcal{R}^{(2)}$ satisfying (IPAC) such that $G' \in \mathcal{R}^{(2)}(Q)$ or $N \backslash G' \in \mathcal{R}^{(2)}(Q^c)$ for some $G' \in 2^N \backslash \{\emptyset\}$ and $Q \in \mathcal{P}$ such that $G' \notin \mathcal{R}^{(1)}(Q)$ and $N \backslash G \notin \mathcal{R}^{(1)}(Q^c)$. When continued, this procedure will produce some exhaustive $\mathcal{R}'$ satisfying (IPAC) after a finite amount of steps. This is true because there are only finitely many pairs $(G, H) \in 2^N \backslash \{\emptyset\} \times \mathcal{P}$; seeing that both N and $\mathcal{P} \subseteq 2^X$ are finite.

Proof of $2 \Longrightarrow 4$

Let $\mathcal{R}'$ be some exhaustive rights system satisfying (IPC) and ($\mathrm{R}^\star$). Assume for a contradiction that $\mathcal{C}^\star(\mathcal{R})$ does not satisfy (IPC). That is, there exist some critical $\{P_1, \ldots, P_r\} \in \mathbb{F}$ and $G_1 \in \mathcal{C}^\star(\mathcal{R})(P_1), \ldots, P_r \in \mathcal{C}^\star(\mathcal{R})(P_r)$ such that $\bigcap_{k=1,\ldots,r} G_k = \emptyset$. By Lemma 11, part 2, for $k = 1, \ldots, r$, there exist $s_k \in \mathbb{N}$, $\{Q_1^k, \ldots, Q_{s_k}^k\} \in \mathbb{C}$ and $G_1^k \in \mathcal{R}(Q_1^k) \subseteq \mathcal{R}'(Q_1^k), \ldots, G_{s_k}^k \in \mathcal{R}(Q_{s_k}^k) \subseteq \mathcal{R}'(Q_{s_k}^k)$ such that $\{Q_1^k, \ldots, Q_{s_k}^k\} \Vdash^\star P_k$ and $G_k = \bigcap_{l=1,\ldots,s_k} G_l^k$. By Lemma 12, part 2, for all $k = 1, \ldots, r$, $G_k = \bigcap_{l=1,\ldots,s_k} G_l^k \in \mathcal{R}'(P_k)$. Thus, $\mathcal{R}'$ violates (IPC), a contradiction.

Proof of $4 \Longrightarrow 3$

Let $\mathcal{C}^\star(\mathcal{R})$ satisfy (IPC). Assume for a contradiction that $\mathcal{R}$ does not satisfy (IPAC). There exist some almost critical $\{P_1, \ldots, P_r\} \in \mathbb{C}$ and $G_1 \in \mathcal{R}(P_1), \ldots, G_r \in \mathcal{R}(P_r)$ such that $\bigcap_{k=1,\ldots,r} G_k = \emptyset$. As $\{P_1, \ldots, P_r\}$ is almost critical, there exists some critical $\mathcal{G} = \{Q_1, \ldots, Q_s\} \in \mathbb{F}$ such that $\{P_1, \ldots, P_r\} \Vdash^\star \mathcal{G}$. By Lemma 8, part 4, there exist $\mathcal{C}_1 \Vdash^\star Q_1, \ldots, \mathcal{C}_s \Vdash^\star Q_s$ with $\{P_1, \ldots, P_r\} = \sqcup_{l=1,\ldots,s} \mathcal{C}_l$. W.l.o.g., for $l = 1, \ldots, s$, $\mathcal{C}_l = \{P_{r_{l-1}+1}, \ldots, P_{r_l}\}$ for some $r_0 = 0 < r_1 < r_2 < \cdots < r_{s-1} < r = r_s \in \mathbb{N}$. For $l = 1, \ldots, s$, let $G^l = \bigcap_{k=r_{l-1}+1,\ldots,r_l} G_k$. By Lemma 11, part 2, we have $G^l = \bigcap_{k=r_{l-1}+1,\ldots,r_l} G_k \in C^\star(\mathcal{R})(Q_l)$. Seeing that $\bigcap_{l=1,\ldots,s} G^l = \bigcap_{l=1,\ldots,s} \left(\bigcap_{k=r_{l-1}+1,\ldots,r_l} G_k \right) = \bigcap_{k=1,\ldots,r} G_k = \emptyset$, $C^\star(\mathcal{R})$ violates (IPC), a contradiction.

C.4.4 Proof of Proposition 5

Proof of part 1

Let $\widetilde{\mathcal{C}} \in \mathbb{C}$. We show that there exists some almost critical $\mathcal{C}_{\widetilde{\mathcal{C}}} \in \mathbb{C}$ such that $\widetilde{\mathcal{C}} \subseteq \mathcal{C}_{\widetilde{\mathcal{C}}}$. Let $\mathcal{C} \in \mathbb{F}$ be some arbitrary almost critical collection. Let $P \in \widetilde{\mathcal{C}}, P \notin \mathcal{C}$ (if no such P exists, we have $\widetilde{\mathcal{C}} \subseteq \mathcal{C}$; thus, letting $\mathcal{C}_{\widetilde{\mathcal{C}}} = \mathcal{C}$ completes the proof) and $Q \in \mathcal{C}$. By Lemma 14, part 2, there exists some $\mathcal{C}' \in \mathbb{C}$ such that $|\mathcal{C}'| \geq 2$ and $P \in \mathcal{C}' \Vdash^\star \{Q\}$. Let $Q' \in \mathcal{C}' \backslash \{Q\} \neq \emptyset$. By Lemma 14, part 1, there exists some $\mathcal{C}'' \in \mathbb{C}$ such that $Q \in \mathcal{C}'' \Vdash^\star \{Q'\}$. Note that $\mathcal{C} \sqcup \{P\} \subseteq \mathcal{C}'' \sqcup (\mathcal{C}' \backslash \{Q'\}) \sqcup (\mathcal{C} \backslash \{Q\})$. Using Lemma 9, part 4, $\mathcal{C}'' \sqcup (\mathcal{C}' \backslash \{Q'\}) \sqcup (\mathcal{C} \backslash \{Q\}) \Vdash^\star \mathcal{C}' \sqcup (\mathcal{C} \backslash \{Q\}) \Vdash^\star \mathcal{C}$. As $\mathcal{C}$ is almost critical, so is $\mathcal{C}'' \sqcup (\mathcal{C}' \backslash \{Q'\}) \sqcup (\mathcal{C} \backslash \{Q\})$. Thus, there exists an almost critical collection containing $\mathcal{C} \sqcup \{Q\}$. In the same fashion, if $P' \in \widetilde{\mathcal{C}}, P' \notin \mathcal{C} \sqcup \{P\}$, we can find an almost critical collection containing $\mathcal{C} \sqcup \{P\} \sqcup \{P'\}$. Consequently, in finitely many steps, we can construct some almost critical collection $\mathcal{C}_{\widetilde{\mathcal{C}}}$ such that $\widetilde{\mathcal{C}} \subseteq \mathcal{C}_{\widetilde{\mathcal{C}}}$.

Let $\mathcal{P}_{2^n}$ be the collection that contains every property from $\mathcal{P}$ exactly 2^n times. As a consequence of what we just showed, there is some almost critical $\mathcal{C}_{\mathcal{P}_{2^n}} \supseteq \mathcal{P}_{2^n}$. If $\mathcal{R}$ is trivial, let $X^n \ni \boldsymbol{x} = (x_1, \ldots, x_n) \mapsto f_{i^\star}(\boldsymbol{x}) = x_{i^\star}$ for some $i^\star \in \bigcap_{G \in \mathcal{R}(P),\, P \in \mathcal{P}} G$. Then $f_{i^\star}$ respects rights and is onto and monotone independent. Conversely, suppose $\mathcal{R}$ is consistent with voting by properties. Note that, for every $P \in \mathcal{P}$, $\mathcal{R}(P) \subseteq 2^N$. Thus, $|\mathcal{R}(P)| \leq 2^n$. By Theorem 7, $\mathcal{R}$ satisfies (IPAC). When evaluated over $\mathcal{C}_{\mathcal{P}_{2^n}}$, this yields that $\bigcap_{G \in \mathcal{R}(P),\, P \in \mathcal{P}} G \neq \emptyset$.

Proof of part 2

As every critical family is almost critical, we only need to show the converse. We establish an auxiliary result first. By the definition of $\vdash$, on a median space $(X, \mathcal{P})$, we have that for all $\mathcal{F} \in \mathbb{F}, P \in \mathcal{P}$: $\mathcal{F} \vdash P \iff (\exists Q \in \mathcal{P} : \mathcal{F} = \{Q\} \vdash P) \iff (\exists Q \in \mathcal{P} : \mathcal{F} = \{Q\}$ and $\{Q, P^c\}$ is critical$) \iff (\exists Q \in \mathcal{P} : \mathcal{F} = \{Q\}$ and $Q \subseteq P)$. Consequently, for all $\mathcal{C} \in \mathbb{C}, P \in \mathcal{P}$: $\mathcal{C} \Vdash \{P\} \iff (\exists Q \in \mathcal{P} : \mathcal{C} = \{Q\} \Vdash P) \iff (\exists Q \in \mathcal{P} : \mathcal{C} = \{Q\}$ and $Q \subseteq P)$. By transitivity of $\subseteq$, the same statements carry over to $\Vdash^\star$.

Let $\mathcal{C} \in \mathbb{C}$ be almost critical. That is, there exists some critical $\mathcal{G} \in \mathbb{F}$ such that $\mathcal{C} \Vdash^\star \mathcal{G}$. As $(X, \mathcal{P})$ is median, $\mathcal{G} = \{P_1, P_2\}$ for some $P_1, P_2 \in \mathcal{P}$. By Lemma 8, part 4, there exist $\mathcal{C}_1, \mathcal{C}_2 \in \mathbb{C}, \mathcal{C} = \mathcal{C}_1 \sqcup \mathcal{C}_2$ such that $\mathcal{C}_1 \Vdash^\star \{P_1\}$ and $\mathcal{C}_2 \Vdash^\star \{P_2\}$. By what we just showed, there exist $Q_1, Q_2 \in \mathcal{P}$ such that $\mathcal{C}_1 = \{Q_1\}, \mathcal{C}_2 = \{Q_2\}$ and $Q_1 \subseteq P_1, Q_2 \subseteq P_2$. We have $Q_1 \subseteq P_1 \subseteq P_2^c \subseteq Q_2^c$. Thus, $\{Q_1, (Q_2^c)^c\} = \{Q_1, Q_2\} = \mathcal{C}$ is critical.

As the almost critical collections are exactly the critical families (all of length two), it is without loss of generality to choose $\mathcal{R}(P) \ni G_P \neq N$ for both properties $P \in \mathcal{G}$ when checking (IPAC) over critical $\mathcal{G}$ (otherwise, the intersection is trivially non-empty). Thus, (IPAC) is equivalent to (IPC). Equivalence of (a) and (b) follows (from Theorems 6 and 7). To prove equivalence of (c) and (d), note that on median spaces all entailments are unconditional (hence direct). Thus, two properties are dependent if and only if they are directly dependent. It follows that weak independence and independence are equivalent concepts on median spaces. Lastly, to show equivalence of (a) and (d), we note that, as all critical fragments have length two on median spaces, $\mathcal{R}$ is weakly independent if and only if it satisfies (IPC) (i.e., is consistent).

C.4.5 Proof of Fact 5

For all $P, Q \in \mathcal{P}$ define $P \equiv Q \iff (P \trianglerighteq^\star Q$ and $Q \trianglerighteq^\star P)$. Note that $\equiv$ is an equivalence relation on $\mathcal{P}$.

We show that $\equiv$ induces exactly two distinct and non-empty equivalence classes. As $(X, \mathcal{P})$ is not totally blocked, there must be at least two equivalence classes. Let $[\equiv]_1, [\equiv]_2$ be two such classes and consider some $P \in [\equiv]_1, Q \in [\equiv]_2$. We have $P \not\equiv Q$. Thus, by semi-blockedness of $(X, \mathcal{P})$, $Q \equiv P^c$; i.e., $P^c \in [\equiv]_2$. Now, for every $\widehat{P} \in \mathcal{P} \backslash \{P, P^c\}$, $\widehat{P} \equiv P$ or $\widehat{P} \equiv P^c$. Thus, there are at most two distinct and non-empty equivalence classes.

As $[\equiv]_1 \neq [\equiv]_2$, we must have $[\equiv]_1 \not\trianglerighteq^\star [\equiv]_2$ or $[\equiv]_2 \not\trianglerighteq^\star [\equiv]_1$. W.l.o.g., assume the former. Define $\mathcal{P}^- = [\equiv]_1$ and $\mathcal{P}^+ = [\equiv]_2$. Then 2. holds by construction. To verify 1., suppose for a contradiction that $Q \equiv Q^c$ for some $Q \in \mathcal{P}$. Then, for all $P, P' \in \mathcal{P}$, $P \equiv Q$ or $P \equiv Q^c$ and $P' \equiv Q$ or $P' \equiv Q^c$. As $Q \equiv Q^c$, $P \equiv P'$ and $(X, \mathcal{P})$ is totally blocked.

To verify 3., we still need to show that $\mathcal{P}^+ \trianglerighteq^\star \mathcal{P}^-$. We first show that, when $|X| > 2$, there exists some critical family $\mathcal{G}$ such that $|\mathcal{G}| \geq 3$. Indeed, suppose for a contradiction that $|\mathcal{G}| = 2$ for all critical $\mathcal{G} \subseteq \mathcal{P}$. Then we have for all $\widehat{P}, \widehat{Q} \in \mathcal{P}$: $\widehat{P} \trianglerighteq^\star \widehat{Q} \iff \widehat{P} \subseteq \widehat{Q}$; i.e., $\widehat{P} \equiv \widehat{Q} \iff \widehat{P} = \widehat{Q}$. Consequently, $|\mathcal{P}| = 2$. It follows that, when $|X| > 2$, there are $x \neq y \in X$ such that, for all $P \in \mathcal{P}$, $x \in P \iff y \in P$, in contradiction to $(X, \mathcal{P})$ being a property space. Thus, there must exist some critical $\mathcal{G} \in \mathbb{F}$ with $|\mathcal{G}| \geq 3$. Let $\mathcal{G} \subseteq \mathcal{P}$ be critical, $|\mathcal{G}| \geq 3$. From part 1 and the fact that $\mathcal{P}^- \not\trianglerighteq^\star \mathcal{P}^+$, we have $|\mathcal{G} \cap \mathcal{P}^-| \leq 1$. Thus, $|\mathcal{G} \cap \mathcal{P}^+| \geq 2$ and there exist $\{P, Q\} \subseteq \mathcal{G} \cap \mathcal{P}^+$. We have $P \trianglerighteq^\star Q^c$. By part 1, $Q^c \in \mathcal{P}^-$. Hence for all $\widehat{P} \in \mathcal{P}^+, \widehat{Q} \in \mathcal{P}^-$: $\widehat{P} \trianglerighteq^\star P \trianglerighteq^\star Q^c \trianglerighteq^\star \widehat{Q}$.

Let $\widetilde{\mathcal{C}} = \{P_1, \ldots, P_r\}$ be almost critical and suppose there exist $k', k'' \in \{1, \ldots, r\}, k' \neq k''$ such that $P_{k'}, P_{k''} \in \mathcal{P}^-$. By Lemma 10, $P_{k''} \in \widetilde{\mathcal{C}} \backslash \{P_{k'}\} \Vdash^\star \{P_{k'}^c\}$. Thus, by Lemma 14, part 1 and part 1 from above, $\mathcal{P}^- \ni P_{k''} \trianglerighteq^\star P_{k'}^c \in \mathcal{P}^+$ contradicting $\mathcal{P}^- \not\trianglerighteq^\star \mathcal{P}^+$. Thus, every almost critical $\widetilde{\mathcal{C}}$ can contain at most one element from $\mathcal{P}^-$.

Let $\mathcal{C}$ be some multiset over $\mathcal{P}^+$ and $P \in \mathcal{P}^-$. We show that there exists some almost critical collection that contains $\mathcal{C} \sqcup \{P\}$. By part 1, $P^c \in \mathcal{P}^+$. Additionally, $\{P, P^c\}$ is critical; a fortiori, almost critical. Let $Q \in \mathcal{C}$. By Lemma 14, part 1 and part 3 above, there exists some $\mathcal{C}' \in \mathbb{C}, |\mathcal{C}'| \geq 2$ such that $Q \in \mathcal{C}' \Vdash^\star \{P^c\}$. Thus, $\{P, Q\} \subseteq \{P\} \sqcup \mathcal{C}'$ and $\{P\} \sqcup \mathcal{C}'$ is almost critical (seeing that $\{P\} \sqcup \mathcal{C}' \Vdash^\star \{P, P^c\}$). As $|(\{P\} \sqcup \mathcal{C}') \backslash \{P, Q\}| \geq 1$ and $(\{P\} \sqcup \mathcal{C}') \backslash \{P, Q\} \subseteq \mathcal{P}^+$, we can use Lemma 14, part 2 repeatedly until we have constructed an almost critical family containing $\mathcal{C} \sqcup \{P\}$ after finitely many steps.

C.4.6 Proof of Proposition 6

Suppose some monotone independent and onto $f : X^n \to X$ respects $\mathcal{R}$. By Fact 2 and the discussion in the main text, $f = F_{\mathcal{R}'}$ for some consistent exhaustive $\mathcal{R}'$ that extends $\mathcal{R}$. Thus, for all $i \in N$, $\mathcal{R}'$ satisfies (MR-i). That is, for all $i \in N$, there exist $P_i \in \mathcal{P}, G_i \subseteq N$ such that $i \in G_i \in \mathcal{R}'(P_i)$ and $(N \backslash G_i) \cup \{i\} \in \mathcal{R}'(P_i^c)$.

Consider any $P \in \mathcal{P}^-$. For all $Q \in \mathcal{P}^-$, Fact 5 establishes that there exists some almost critical $\mathcal{C} \in \mathbb{C}$ such that $\{P^c, Q^c, Q\} \subseteq \mathcal{C}$. Thus, by Lemma 10, $\{Q^c, Q\} \subseteq \mathcal{C} \backslash \{P^c\} \Vdash^\star \{P\}$. Using this with $Q = P_i$ as well as Lemma 12, part 2, we have for all $i \in N$: $\{i\} = G_i \cap ((N \backslash G_i) \cup \{i\}) \in \mathcal{R}'(P)$ (note that $N \in \bigcap_{P \in \mathcal{P}} \mathcal{R}(P)$). As $\mathcal{R}'$ is monotone, we have $\mathcal{R}'(P) = 2^N \backslash \{\emptyset\}$ and, consequently, $\mathcal{R}'(P^c) = \{N\}$. As $P \in \mathcal{P}^-$ was arbitrary, $f = F_{\mathcal{R}'}$ is a unanimity rule with default $\bigcap \mathcal{P}^-$.

D Appendix to Part V

D.1 Relation of Equation (16.2) and the General Case of Exchangeable Errors.

How general is equation (16.2)? Instead of equation (16.2) we may simply assume that $\left(e_i^k\right)_{i\in\mathbb{N}}$ are exchangeable for all $k = 1, \ldots, K$ (with homogeneous positive pairwise correlation $\rho \geq 0$) and independent (across k). By De Finetti's Theorem, there exist (latent) random variables $\tilde{c}^k$ taking on values in $[0,1]$ such that for all $k = 1, \ldots, K$:

1. $\left(\tilde{e}_i^k\right)_{i\in\mathbb{N}}$ are independent (for given k) when conditioned on $\tilde{c}^k$.

2. $\mathbb{E}[\tilde{e}_i^k | \tilde{c}^k] = \tilde{c}^k$

3. $\tilde{c}^k = \lim\limits_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \tilde{e}_i^k$

If the c^k are themselves binary r.v.s (say taking on values $a, b \in [0,1]$), then specifying $\mathbb{P}(\tilde{c}^k = a)$ pins down $\mathbb{P}(\tilde{e}_i^k = 0 | \tilde{c}^k = a)$ and $\mathbb{P}(\tilde{e}_i^k = 0 | \tilde{c}^k = b)$ (where, w.l.o.g., $\mathbb{P}(\tilde{e}_i^k = 0 | \tilde{c}^k = a) \geq \mathbb{P}(\tilde{e}_i^k = 0 | \tilde{c}^k = b)$)[155] and hence the joint distribution of $\left(\tilde{e}_i^k\right)_{i\in\mathbb{N}}$. This is the case since we must have that

1. $1 - p = \mathbb{E}[\tilde{e}_i^k] = \mathbb{E}[\mathbb{E}[\tilde{e}_i^k | \tilde{c}^k]] = \mathbb{P}(\tilde{c}^k = a)(1 - \mathbb{P}(\tilde{e}_i^k = 0 | \tilde{c}^k = a)) + (1 - \mathbb{P}(\tilde{c}^k = a))(1 - \mathbb{P}(\tilde{e}_i^k = 0 | \tilde{c}^k = b))$ and

2. $p(1-p)\rho = CoV(\tilde{e}_i^k, \tilde{e}_j^k) = \mathbb{E}[\tilde{e}_i^k \tilde{e}_j^k] - \mathbb{E}[\tilde{e}_i^k]^2 = \mathbb{P}(\tilde{c}^k = a)(1 - \mathbb{P}(\tilde{e}_i^k = 0 | \tilde{c}^k = a))^2 + (1 - \mathbb{P}(\tilde{c}^k = a))(1 - \mathbb{P}(\tilde{e}_i^k = 0 | \tilde{c}^k = b))^2 - (1-p)^2$;

a system of two equations in two unknown with a unique solution. Letting $p_a := \mathbb{P}(\tilde{e}_i^k = 0 | \tilde{c}^k = a)$ and $p_b := \mathbb{P}(\tilde{e}_i^k = 0 | \tilde{c}^k = b)$, this system reduces to

$$p = \mathbb{P}(\tilde{c}^k = a)p_a + (1 - \mathbb{P}(\tilde{c}^k = a))p_b \tag{D.1}$$

$$p(1-p)\rho = \mathbb{P}(\tilde{c}^k = a)(1 - p_a)^2 + (1 - \mathbb{P}(\tilde{c}^k = a))(1 - p_b)^2. \tag{D.2}$$

Solving (D.1) for p_a delivers $p_a = \frac{p}{\mathbb{P}(\tilde{c}^k = a)} - \frac{1 - \mathbb{P}(\tilde{c}^k = a)}{\mathbb{P}(\tilde{c}^k = a)} p_b$. Plugging into (D.2) and solving for p_b, we have

$$p_b = p \pm \sqrt{\frac{\mathbb{P}(\tilde{c}^k = a)}{1 - \mathbb{P}(\tilde{c}^k = a)}(1-p)p\rho}.$$

155 Thus, $c^k = a$ is the "good" state in which latent variable c^k delivers valuable information in the sense of increasing individual competence above the average level p.

Thus, using (D.1) again,

$$p_a = p \mp \sqrt{\frac{1-\mathbb{P}(\tilde{c}^k = a)}{\mathbb{P}(\tilde{c}^k = a)}(1-p)p\rho}.$$

The unique solution such that $p_a \geq p_b$ is $p_a = p + \sqrt{\frac{1-\mathbb{P}(\tilde{c}^k=a)}{\mathbb{P}(\tilde{c}^k=a)}(1-p)p\rho}$ and $p_b = p - \sqrt{\frac{\mathbb{P}(\tilde{c}^k=a)}{1-\mathbb{P}(\tilde{c}^k=a)}(1-p)p\rho}$. For $\mathbb{P}(\tilde{c}^k = a) = p$, we obtain $p_a = p + \sqrt{\rho}(1-p)$ and $p_b = p - \sqrt{\rho}p$ (cf. Equation (16.4)).

D.2 Auxiliary Results

Lemma 15. *For all $\alpha, q, q' \in [0,1]$, all $k = 1, \ldots, K$: $\hat{f}_x(q_1, \ldots, q_{k-1}, \alpha q + (1-\alpha)q', q_{k+1}, \ldots, q_K) = \alpha \hat{f}_x(q_1, \ldots, q_{k-1}, q, q_{k+1}, \ldots, q_K) + (1-\alpha)\hat{f}_x(q_1, \ldots, q_{k-1}, q', q_{k+1}, \ldots, q_K)$.*

Proof. We show the case $k = 1$. The cases $k = 2, \ldots, K$ are analogous. We have $\hat{f}_x(\alpha q + (1-\alpha)q', q_2, \ldots, q_K) =$

$$\begin{aligned}
&\sum_{\substack{v\in\{0,1\}^K:\\ F(v)=F(x)\\ v_1=x_1}} (\alpha q + (1-\alpha)q') \prod_{k=2}^{K} q_k^{1-|x_k-v_k|}(1-q_k)^{|x_k-v_k|} \\
&+ \sum_{\substack{v\in\{0,1\}^K:\\ F(v)=F(x)\\ v_1\neq x_1}} (\alpha(1-q) + (1-\alpha)(1-q')) \prod_{k=2}^{K} q_k^{1-|x_k-v_k|}(1-q_k)^{|x_k-v_k|} \\
=&\alpha \sum_{\substack{v\in\{0,1\}^K:\\ F(v)=F(x)}} q^{1-|x_1-v_1|}(1-q)^{|x_1-v_1|} \prod_{k=2}^{K} q_k^{1-|x_k-v_k|}(1-q_k)^{|x_k-v_k|} \\
&+ (1-\alpha) \sum_{\substack{v\in\{0,1\}^K:\\ F(v)=F(x)}} q'^{1-|x_1-v_1|}(1-q')^{|x_1-v_1|} \prod_{k=2}^{K} q_k^{1-|x_k-v_k|}(1-q_k)^{|x_k-v_k|} \\
=&\alpha \hat{f}_x(q, q_2, \ldots, q_K) + (1-\alpha)\hat{f}_x(q', q_2, \ldots, q_K)
\end{aligned}$$

□

D.3 Proofs for the Main Text

Proof of Proposition 8

Let the agenda be given by $F : X \to \{0,1\}$ and consider any $x \in X$.

1. Suppose that x induces increasing competence and consider any $p \in (0.5, 1)$. As $g_n(p) > 0.5$ is increasing in n and $f_x(\cdot)$ is increasing on $(0.5, 1)$, we have for any two odd $n' > n$: $f_x(g_{n'}(p)) > f_x(g_n(p))$.

2. Suppose x is truth-conducive and consider any $p \in (0.5, 1)$. As $f_x(p) > 0.5$, $g_n(f_x(p))$ is increasing in n. Else suppose there exists some $p' \in (0.5, 1)$ such that $f_x(p') \leq 0.5$. If $f_x(p') = 0.5$, then $g_n(f_x(p')) = 0.5$ for all odd n. Else if $f_x(p') < 0.5$, then $g_n(f_x(p'))$ is decreasing in n.

Proof of Proposition 9

Using Lemma 15, we have $P^x_{\rho,n}(p) =$

$$\begin{aligned}
&\sum_{y\in\{0,1\}^K} p^{K-||y||_1}(1-p)^{||y||_1}\hat{f}_x(g_n((1-y_1)p_H + y_1p_L), \dots, g_n((1-y_K)p_H + y_Kp_L))\\
&= \sum_{z\in\{0,1\}^{K-1}} p \cdot p^{K-1-||z||_1}(1-p)^{||z||_1}\hat{f}_x(g_n(p_H), g_2((1-y_2)p_H + y_2p_L), \dots)\\
&+ \sum_{z\in\{0,1\}^{K-1}} (1-p) \cdot p^{K-1-||z||_1}(1-p)^{||z||_1}\hat{f}_x(g_n(p_L), g_2((1-y_2)p_H + y_2p_L), \dots)\\
&= \sum_{z\in\{0,1\}^{K-1}} p^{K-1-||z||_1}(1-p)^{||z||_1}\hat{f}_x(pg_n(p_H) + (1-p)g_n(p_L), g_2((1-y_2)p_H + y_2p_L), \dots)\\
&= \sum_{w\in\{0,1\}^{K-2}} p \cdot p^{K-2-||w||_1}(1-p)^{||w||_1}\hat{f}_x(pg_n(p_H) + (1-p)g_n(p_L), g_n(p_H), \dots)\\
&+ \sum_{w\in\{0,1\}^{K-2}} (1-p) \cdot p^{K-2-||w||_1}(1-p)^{||w||_1}\hat{f}_x(pg_n(p_L) + (1-p)g_n(p_L), g_n(p_L), \dots)\\
&= \dots\\
&= \hat{f}(pg_n(p_H) + (1-p)g_n(p_L), \dots, pg_n(p_H) + (1-p)g_n(p_L))\\
&= f(pg_n(p_H) + (1-p)g_n(p_L))
\end{aligned}$$

Bibliography

Aizerman, M. A. and Malishevski, A. V. (1981). General theory of best variants choice: Some aspects. *IEEE Transaction on Automatic Control*, 26(5):1030–1040.

Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2013). Discounting behaviour and the magnitude effect: Evidence from a field experiment in Denmark. *Economica*, 80(320):670–697.

Andreoni, J. and Sprenger, C. (2012). Estimating time preferences from convex budgets. *American Economic Review*, 102(7):3333–3356.

Arrow, K. J. (1959). Rational choice functions and orderings. *Economica*, 26(102):121–127.

Arrow, K. J. (1963). *Social Choice and Individual Values: 2nd Edition.* Yale University Press, New Haven.

Augenblick, N., Niederle, M., and Sprenger, C. (2015). Working over time: Dynamic inconsistency in real effort tasks. *Quarterly Journal of Economics*, 130(3):1067–1115.

Augenblick, N. and Rabin, M. (2019). An experiment on time preference and misprediction in unpleasant tasks. *Review of Economic Studies*, 86(3):941–975.

Austen-Smith, D. and Banks, J. S. (1996). Information aggregation, rationality, and the Condorcet Jury Theorem. *American Political Science Review*, 90(1):34–45.

Ballard, I. C., Aydogan, G., Kim, B., and McClure, S. M. (2018). Causal evidence for the dependence of the magnitude effect on dorsolateral prefrontal cortex. *Nature: Scientific Reports*, 8(16545).

Ballard, I. C., Kim, B., Liatsis, A., Aydogan, G., Cohen, J. D., and McClure, S. M. (2017). More is meaningful: The magnitude effect in intertemporal choice depends on self-control. *Psychological Science*, 28(10):1443–1454.

Baumeister, R. F., Bratslavsky, E., Muraven, M., and Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, 74(5):1252–1265.

Baumeister, R. F., Hofmann, W., Summerville, A., Reiss, P. T., and Vohs, K. D. (2020). Everyday thoughts in time: Experience sampling studies of mental time travel. *Personality and Social Psychology Bulletin*, 46(12):1631–1648.

Baumeister, R. F., Vohs, K. D., and Oettingen, G. (2016). Pragmatic prospection: How and why people think about the future. *Review of General Psychology*, 20(1):3–16.

Baumeister, R. F., Vohs, K. D., and Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science*, 16(6):351–355.

Benhabib, J. and Bisin, A. (2005). Modeling internal commitment mechanisms and self-control: A neuroeconomics approach to consumption-saving decisions. *Games and Economic Behavior*, 52(2):460–492.

Benhabib, J., Bisin, A., and Schotter, A. (2010). Present-bias, quasi-hyperbolic discounting, and fixed costs. *Games and Economic Behavior*, 69(2):205–223.

Bermúdez, J. P., Murray, S., Chartrand, L., and Barbosa, S. (2023). What's inside is all that counts? The contours of everyday thinking about self-control. *Review of Philosophy and Psychology*, 14(1):33–55.

Bernholz, P. (1974). Is a Paretian liberal really impossible? *Public Choice*, 20(1):99–107.

Blau, J. H. (1975). Liberal values and independence. *The Review of Economic Studies*, 42(3):395–401.

Boros, E., Elbassioni, K., Gurvich, V., and Makino, K. (2010). On effectivity functions of game forms. *Games and Economic Behavior*, 68(2):512–531.

Boureau, Y.-L., Sokol-Hessner, P., and Daw, N. D. (2015). Deciding how to decide: Self-control and meta-decision making. *Trends in Cognitive Sciences*, 19(11):700–710.

Bovens, L. and Rabinowicz, W. (2006). Democratic answers to complex questions – an epistemic perspective. *Synthese*, 150(1):131–153.

Bozbay, I. (2019). Truth-tracking judgment aggregation over interconnected issues. *Social Choice and Welfare*, 53(2):337–370.

Bozbay, I., Dietrich, F., and Peters, H. (2014). Judgment aggregation in search for the truth. *Games and Economic Behavior*, 87:571–590.

Condorcet, N. M. d. (2014). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Cambridge University Press, Cambridge.

Coughlin, P. J. (1986). Rights and the private Pareto principle. *Economica*, 53(211):303–320.

de Clippel, G. and Eliaz, K. (2015). Premise-based versus outcome-based information aggregation. *Games and Economic Behavior*, 89:34–42.

Deb, R. (1994). Waiver, effectivity and rights as game forms. *Economica*, 61(242):167–178.

Deb, R. (2004). Rights as alternative game forms. *Social Choice and Welfare*, 22(1):83–111.

Deb, R., Pattanaik, P. K., and Razzolini, L. (1997). Game forms, rights, and the efficiency of social outcomes. *Journal of Economic Theory*, 72(1):74–95.

Dekel, E., Lipman, B. L., and Rustichini, A. (2009). Temptation-driven preferences. *The Review of Economic Studies*, 76(3):937–971.

Dietrich, F. (2007). A generalised model of judgment aggregation. *Social Choice and Welfare*, 28(4):529–565.

Dietrich, F. and List, C. (2007). Arrow's theorem in judgment aggregation. *Social Choice and Welfare*, 29(1):19–33.

Dietrich, F. and List, C. (2008). A liberal paradox for judgment aggregation. *Social Choice and Welfare*, 31(1):59–78.

Dietrich, F. and Spiekermann, K. (2013). Epistemic democracy with defensible premises. *Economics and Philosophy*, 29(1):87–120.

Dietrich, F. and Spiekermann, K. (2019). Jury Theorems. In Fricker, M., Graham, P. J., Henderson, P., and Pedersen, N. J., editors, *The Routledge Handbook of Social Epistemology*, pages 386–396. Routledge, New York and London.

Dietrich, F. and Spiekermann, K. (2021). Jury Theorems. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2021 edition.

Dokow, E. and Holzman, R. (2010). Aggregation of binary evaluations. *Journal of Economic Theory*, 145(2):495–511.

Duckworth, A. L., Gendler, T. S., and Gross, J. J. (2014). Self-control in school-age children. *Educational Psychologist*, 49(3):199–217.

Duckworth, A. L., Gendler, T. S., and Gross, J. J. (2016). Situational strategies for self-control. *Perspectives on Psychological Science*, 11(1):35–55.

Figner, B., Knoch, D., Johnson, E. J., Krosch, A. R., Lisanby, S. H., Fehr, E., and Weber, E. U. (2010). Lateral prefrontal cortex and self-control in intertemporal choice. *Nature Neuroscience*, 13(5):538–539.

Fine, B. (1975). Individual liberalism in a Paretian society. *Journal of Political Economy*, 83(6):1277–1281.

Fishburn, P. C. and Rubinstein, A. (1986). Aggregation of equivalence relations. *Journal of Classification*, 3(1):61–65.

Fleurbaey, M. and Van Hees, M. (2000). On rights in game forms. *Synthese*, 123(3):295–326.

Frederick, S., Loewenstein, G., and O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40(2):351–401.

Fudenberg, D. and Levine, D. K. (2006). A dual-self model of impulse control. *The American Economic Review*, 96(5):1449–1476.

Fudenberg, D. and Levine, D. K. (2012). Timing and self-control. *Econometrica*, 80(1):1–42.

Gaertner, W., Pattanaik, P. K., and Suzumura, K. (1992). Individual rights revisited. *Economica*, 59(234):161–177.

Gärdenfors, P. (1981). Rights, games and social choice. *Noûs*, 15(3):341–356.

Geddis, A. (2006). A dual track democracy? The symbolic role of the Māori seats in New Zealand's electoral system. *Election Law Journal*, 5(4):347–371.

Gibbard, A. (1974). A Pareto-consistent libertarian claim. *Journal of Economic Theory*, 7(4):388–410.

Gibbard, A. (2014). Social choice and the Arrow conditions. *Economics and Philosophy*, 30(3):269–284.

Gilboa, I. (2012). *Rational choice.* MIT Press, Cambridge (MA).

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *The American Psychologist*, 54(7):493–503.

Gollwitzer, P. M., Gawrilow, C., and Oettingen, G. (2010). The power of planning: Self-control by effective goal striving. In Hassin, R. R., Ochsner, K., and Trope, Y., editors, *Self Control in Society, Mind, and Brain*, pages 279–296. Oxford University Press, Oxford.

Gollwitzer, P. M. and Oettingen, G. (2011). Planning promotes goal striving. In Vohs, K. D. and Baumeister, R. F., editors, *Handbook of Self-Regulation: Research, Theory, and Applications*, pages 162–185. Guilford Press, New York and London.

Gollwitzer, P. M. and Sheeran, P. (2006). Implementation intentions and goal achievement. *Advances in Experimental Social Psychology*, 38:69–119.

Green, L., Myerson, J., and McFadden, E. (1997). Rate of temporal discounting decreases with amount of reward. *Memory & Cognition*, 25(5):715–723.

Grofman, B., Owen, G., and Feld, S. L. (1983). Thirteen theorems in search of the truth. *Theory and decision*, 15(3):261–278.

Gul, F. and Pesendorfer, W. (2001). Temptation and self-control. *Econometrica*, 69(6):1403–1435.

Gul, F. and Pesendorfer, W. (2004). Self-control and the theory of consumption. *Econometrica*, 72(1):119–158.

Gul, F. and Pesendorfer, W. (2005). The revealed preference theory of changing tastes. *Review of Economic Studies*, 72:429–448.

Gul, F. and Pesendorfer, W. (2006). The simple theory of temptation and self-control. *Working Paper*.

Halevy, Y. (2015). Time consistency: Stationarity and time invariance. *Econometrica*, 83(1):335–352.

Hartmann, S., Pigozzi, G., and Sprenger, J. (2010). Reliable methods of judgement aggregation. *Journal of Logic and Computation*, 20(2):603–617.

Hartmann, S. and Sprenger, J. (2012). Judgment aggregation and the problem of tracking the truth. *Synthese*, 187(1):209–221.

Hennecke, M. and Bürgler, S. (2020). Many roads lead to Rome: Self-regulatory strategies and their effects on self-control. *Social and Personality Psychology Compass*, 14(6).

Herzberg, F. (2017). Respect for experts vs. respect for unanimity: The liberal paradox in probabilistic opinion pooling. *Economics Letters*, 151:44–47.

Hsiaw, A. (2013). Goal-setting and self-control. *Journal of Economic Theory*, 148(2):601–626.

Inzlicht, M., Werner, K. M., Briskin, J. L., and Roberts, B. W. (2021). Integrating models of self-regulation. *Annual Review of Psychology*, 72:319–345.

Janssen, M. C. W. (2016). Microfoundations. In *The New Palgrave Dictionary of Economics*, pages 1–8, London. Palgrave Macmillan UK.

Kelly, J. S. (1976). Rights exercising and a Pareto-consistent libertarian claim. *Journal of Economic Theory*, 13(1):138–153.

Kirby, K. N. (1997). Bidding on the future: Evidence against normative discounting of delayed rewards. *Journal of Experimental Psychology: General*, 126(1):54–70.

Koch, A. K. and Nafziger, J. (2011). Self-regulation through goal setting. *Scandinavian Journal of Economics*, 113(1):212–227.

Kornhauser, L. A. and Sager, L. G. (1986). Unpacking the court. *The Yale Law Journal*, 96(1):82–117.

Kornhauser, L. A. and Sager, L. G. (1993). The one and the many: Adjudication in collegial courts. *California Law Review*, 81(1):1–59.

Kretz, C. (2021). Consistent rights on property spaces. *Journal of Economic Theory*, 197:105323.

Ladha, K. K. (1993). Condorcet's Jury Theorem in light of de Finetti's theorem. *Social Choice and Welfare*, 10(1):69–85.

List, C. (2005). The probability of inconsistencies in complex collective decisions. *Social Choice and Welfare*, 24(1):3–32.

List, C. and Goodin, R. E. (2001). Epistemic democracy: Generalizing the Condorcet Jury Theorem. *Journal of Political Philosophy*, 9(3):277–306.

List, C. and Pettit, P. (2002). Aggregating sets of judgments: An impossibility result. *Economics and Philosophy*, 18(1):89–110.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1:19–46.

Ludwig, R. M., Srivastava, S., and Berkman, E. T. (2018). Planfulness: A process-focused construct of individual differences in goal achievement. *Collabra: Psychology*, 4(1):28.

Lynch, J. G., Netemeyer, R. G., Spiller, S. A., and Zammit, A. (2010). A generalizable scale of propensity to plan: The long and the short of planning for time and for money. *Journal of Consumer Research*, 37(1):108–128.

Mas-Colell, A. and Sonnenschein, H. (1972). General possibility theorems for group decisions. *The Review of Economic Studies*, 39(2):185–192.

Masatlioglu, Y., Nakajima, D., and Ozdenoren, E. (2020). Willpower and compromise effect. *Theoretical Economics*, 15(1):279–317.

Masicampo, E. J. and Baumeister, R. F. (2011). Consider it done! Plan making can eliminate the cognitive effects of unfulfilled goals. *Journal of Personality and Social Psychology*, 101(4):667–683.

McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G., and Cohen, J. D. (2007). Time discounting for primary rewards. *Journal of Neuroscience*, 27(21):5796–5804.

Meier, S. and Sprenger, C. (2010). Present-biased preferences and credit card borrowing. *American Economic Journal: Applied Economics*, 2(1):193–210.

Meyer, A. G. (2015). The impacts of elicitation mechanism and reward size on estimated rates of time preference. *Journal of Behavioral and Experimental Economics*, 58:132–148.

Mischel, W. and Patterson, C. J. (1976). Substantive and structural elements of effective plans for self-control. *Journal of Personality and Social Psychology*, 34(5):942.

Moulin, H. (1983). *The Strategy of Social Choice*. North-Holland, New York.

Moulin, H. (1985). Choice functions over a finite set: A summary. *Social Choice and Welfare*, 2(2):147–160.

Nehring, K. (2006a). Oligarchies in judgment aggregation. *Working Paper*.

Nehring, K. (2006b). Self-control through second-order preferences. *Working Paper*.

Nehring, K. and Puppe, C. (2002). Strategy-proof social choice on single-peaked domains. *Working Paper*.

Nehring, K. and Puppe, C. (2007). The structure of strategy-proof social choice - Part I: General characterization and possibility results on median spaces. *Journal of Economic Theory*, 135(1):269–305.

Nehring, K. and Puppe, C. (2010). Abstract arrowian aggregation. *Journal of Economic Theory*, 145(2):467–494.

Noor, J. (2011). Intertemporal choice and the magnitude effect. *Games and Economic Behavior*, 72(1):255–270.

Noor, J. and Takeoka, N. (2010). Uphill self-control. *Theoretical Economics*, 5(2):127–158.

Noor, J. and Takeoka, N. (2015). Menu-dependent self-control. *Journal of Mathematical Economics*, 61:1–20.

Nozick, R. (1974). *Anarchy, State, and Utopia*. Basic Books, New York.

Peleg, B. (1998). Effectivity functions, game forms, games, and rights. *Social Choice and Welfare*, 15(1):67–80.

Peleg, B. (2002). Game-theoretic analysis of voting in committees. In Arrow, K. J., Sen, A. K., and Suzumura, K., editors, *Handbook of Social Choice and Welfare*, volume 1, chapter 8, pages 395–423. Elsevier.

Pettit, P. (2001). Deliberative democracy and the discursive dilemma. *Philosophical Issues*, 11:268–299.

Plott, C. R. (1973). Path independence, rationality, and social choice. *Econometrica*, 41(6):1075–1091.

Schuessler, A. A. (2000). Expressive voting. *Rationality and Society*, 12(1):87–119.

Sen, A. (1969). Quasi-transitivity, rational choice and collective decisions. *The Review of Economic Studies*, 36(3):381–393.

Sen, A. K. (1970). The impossibility of a Paretian liberal. *Journal of Political Economy*, 78(1):152–157.

Sen, A. K. (1976). Liberty, unanimity and rights. *Economica*, 43(171):217–245.

Shelah, S. (2005). On the Arrow property. *Working Paper*.

Sjåstad, H. and Baumeister, R. F. (2018). The future and the will: Planning requires self-control, and ego depletion leads to planning aversion. *Journal of Experimental Social Psychology*, 76:127–141.

Stovall, J. E. (2010). Multiple temptations. *Econometrica*, 78(1):349–376.

Stovall, J. E. (2018). Temptation with uncertain normative preference. *Theoretical Economics*, 13(1):145–174.

Strotz, R. H. (1955). Myopia and inconsistency in dynamic utility maximization. *The Review of Economic Studies*, 23(3):165–180.

Sugden, R. (1985). Liberty, preference, and choice. *Economics and Philosophy*, 1(2):213–229.

Sun, C. and Potters, J. (2022). Magnitude effect in intertemporal allocation tasks. *Experimental Economics*, 25(2):593–623.

Sutter, M., Kocher, M. G., and Glätzle-Rützler, D. (2013). Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *The American Economic Review*, 103(1):510–531.

Takeoka, N. (2008). Temptation, certainty effect, and diminishing self-control. *Working Paper*.

Taylor, S. E., Pham, L. B., Rivkin, I. D., and Armor, D. A. (1998). Harnessing the imagination: Mental simulation, self-regulation, and coping. *American Psychologist*, 53(4):429–439.

Thaler, R. (1981). Some empirical evidence on dynamic inconsistency. *Economics Letters*, 8(3):201–207.

Thaler, R. H. and Shefrin, H. M. (1981). An economic theory of self-control. *Source: Journal of Political Economy*, 89(2):392–406.

Wakker, P. P. (1989). Additive representations. In *Additive Representations of Preferences: A New Foundation of Decision Analysis*, pages 41–77. Springer Netherlands, Dordrecht.