Essays on hyper-preferences, polarization and information aggregation

Ali Ihsan Ozkes

ÉCOLE POLYTECHNIQUE

Département d'Économie

THÈSE

Pour l'obtention du grade de
DOCTEUR DE L'ÉCOLE POLYTECHNIQUE
Spécialité: Sciences Économiques

Essays on Hyper-preferences, Polarization and Information Aggregation

Essais sur les hyper-préférences, la polarisation
et l'aggrégation des informations

Présentée et soutenu publiquement par

Ali İhsan ÖZKES

le 1 Octobre 2014

Directeurs de Thèse: Yukio Koriyama & Remzi Sanver

Membres du Jury:

Yukio Koriyama	École Polytechnique	Directeur de Thèse
Remzi Saner	Istanbul Bilgi University	Directeur de Thèse
Nicholas Baigent	London School of Economics and Political Science	Rapporteur
Gilbert Laffond	Conservatoire National des Arts et Métiers	Rapporteur
Jean-François Laslier	École d'Économie de Paris & CNRS	Examinateur
Jean Lainé	Conservatoire National des Arts et Métiers	Examinateur

ABSTRACT

In this thesis, some important problems and properties of collective decision-making are studied. In particular, first, a stability property of preference aggregation rules is introduced and some well-known classes of rules are tested in this regard. Second, measuring preferential polarization is studied, both theoretically and empirically. Finally, strategic behavior in information aggregation situations is investigated in light of a sort of bounded rationality model, both theoretically and experimentally.

The stability notion studied in the first part of the thesis is imposed particularly on social welfare functions and requires that the outcome of these functions should be robust to reduction in preference submission that are argued to take place when individuals submit a ranking of alternatives when the outcomes are also restricted to be rankings. Given the preference profile of a society, that is a collection of rankings of alternatives, a compatible collection of rankings of rankings are extracted and the outcome of social welfare functions in these two levels are compared. It turns out that no scoring rule gives consistent results, although there might exist Condorcet-type rules.

Polarization measures studied in second part are in form of aggregation of pairwise antagonisms in a society. The public opinion polarization in the United States for the last three decades is analyzed in light of this view, by using a well-acclaimed measure of polarization introduced in the literature of income inequality. The conclusion is that no significant trend in public opinion polarization can be claimed to exist over the last several decades. Also, an adaptation of the same measure is shown to satisfy desirable properties in lieu of ordinal preference profiles when three alternatives are considered. Furthermore, a measure that is the aggregation of pairwise differences among individuals preferences is characterized by a set of axioms.

In the final part of the thesis, information aggregation situations described as in Condorcet jury model is studied in light of cognitive hierarchy approach to bounded rationality. Specifically, a laboratory experiment is run to test the theoretical predictions of the symmetric Bayesian Nash equilibrium concept. It is observed that behavior in lab is not correctly captured by this concept that assumes a strong notion of rationality and homogeneity among individuals behaviors. To better describe the findings in the experiment, a novel model of cognitive hierarchy is developed and shown to perform better

than both strong rationality approach and previous cognitive hierarchy models. This endogenous cognitive hierarchy model is compared theoretically to previous models of cognitive hierarchy and shown to improve in certain classes of games.

RÉSUMÉ

Dans cette thèse, certains problèmes importants et des propriétés de prise de décision collective sont étudiés. En particulier, d'abord, une propriété de stabilité des règles d'agrégation de préférences est introduite et certaines classes bien connues de règles sont testées à cet égard. Deuxièmement, le mesurage de la polarisation préférentielle est étudié, à la fois théorique et empirique. Enfin, le comportement stratégique dans des situations d'agrégation de l'information est étudié à la lumière d'une sorte de modèle de la rationalité limitée, à la fois théoriquement et expérimentalement.

La notion de stabilité étudié dans la première partie de la thèse est imposée en particulier sur les fonctions de bien-être sociale et exige que le résultat de ces fonctions doit être robuste à la réduction de la transmission de préférences qui sont soutenu avoir lieu lorsque les individus présentent un ordre des alternatives lorsque les résultats sont également limités à être ordres. Pour tous profils sociétaux de préférences donné, qui est une collection d'ordres des alternatives, une collection compatible d'ordre des classements est extraite et les résultats des fonctions de bien-être social dans ces deux niveaux sont comparés. Il s'avère qu'aucune règle de notation donne des résultats cohérents, bien qu'il puisse y exister des règles Condorcetien.

Mesures de polarisation qui sont étudiées en deuxième partie sont en forme d'agrégation des antagonismes par paires dans une société. La polarisation de l'opinion publique aux États-Unis pour les trois dernières décennies est analysé à la lumière de ce point de vue, en utilisant une mesure de polarisation bien acclamé qui est introduit dans la littérature de l'inégalité des revenus. La conclusion est qu'aucune tendance significative dans l'opinion publique polarisation peut être réclamé à exister au cours des dernières décennies. En outre, une adaptation de la même mesure est montrée à satisfaire des propriétés souhaitables à la place de profils de préférences ordinales lorsque trois alternatives sont considérées. En outre, une mesure qui est en effet l'agrégation des différences par paires entre les préférences des individus est caractérisée axiomatiquement.

Dans la dernière partie de la thèse, situations de l'agrégation de l'information telles que décrites comme dans le modèle du jury de Condorcet sont étudiées à la lumière d'une approche de rationalité limitée qui est connue hiérarchie cognitive. Plus précisément, une expérience de laboratoire est exécutée pour tester les prédictions théoriques de la

notion d'équilibre symétrique de Nash Bayésien. On constate que le comportement en laboratoire n'est pas correctement capturé par ce concept qui suppose une forte notion de la rationalité et de l'homogénéité entre les comportements des individus. Pour mieux décrire les résultats à l'expérience, un nouveau modèle de hiérarchie cognitive est développé et montré à faire mieux que la fois l'approche de la rationalité forte et des modèles précédentes de hiérarchie cognitive. Ce modèle de hiérarchie cognitive endogène est comparé en théorie aux modèles précédents de la hiérarchie cognitive et montré pour améliorer dans certaines catégories de jeux.

à mes parents,
Melek Nur et İhsan

ACKNOWLEDGEMENTS

I can only be happier if what I am recently told by a fellow thesis student is true: the most widely read part of a thesis is this part, acknowledgments. I should be happier because the names of those most important persons of my life I am going to remember in the following lines can only make a reader be prepared to value highly the ingredients of the thesis. I would like to extend my sincerest gratitude and appreciation to all those magnanimous souls who helped me bring into completion this dissertation.

I had the privilege of having two fantastic supervisors and I will start with one of them, Remzi Sanver. He is so special to me that I really should not dare to try to demonstrate. He was incredibly generous from the very beginning (a decade ago now, a microeconomics course in the second year of undergraduate studies). It is impossible for me to admit that I deserved any of that. Without his support, almost none of those good things that happened to and around me in the last ten years would come true. Most of what defines me and my thinking today is highly influenced by him and his presence, directly and indirectly. Wishing to be able to properly carry the honor of being his student, I thank him wholeheartedly for everything.

I learned from every second we spent together with Yukio Koriyama. His mastery in economic theory and mathematics was always enlightening, especially due to his readiness and accurate style to pass this on. I always felt extremely lucky for being supervised by Yukio. And as the years passed, we got closer and have become friends. He was always genuinely nice and kind. He even initiated me into juggling. I, hereby, apologize to Aoi if I have ever stolen from her precious time with her dad but am also sure that she will soon realize how lucky she is, if she did not yet, for having such great parents.

I will try to name several very important people in this paragraph, keeping others in mind. First, I am very grateful to Jean Lainé for not forgetting me after his supervision for my masters' thesis. It is always teaching to talk to him, and joyful. Over the years, his wisdom in not only social choice and cooperative games but life in general and human relations was generously enlightening. Uğur Özdemir is nothing less than an older brother to me. Days we didn't share ideas are as few as ignorable for almost a decade now. We shared books, beds, cars, cigarettes, pajamas, etc. as well. I was mostly

the taker and he taught me a lot, really, more than a lot. Burak Can is not different. He keeps being important and obliging even these precious days he would strongly prefer hugging his son. İpek Özkal-Sanver has always been supportive, especially when I really needed. Besides her well-designed matching theory course I have taken, I also enjoyed and learned a lot from discussing social choice problems with her. Ege Yazgan is another very important person who was always extremely indulgent and friendly. I cannot appreciate enough his kindness. I learned a lot from Jean-François Laslier, not only from his papers and books but also from our chats and discussions. His generosity in devoting time and effort for me and my questions and his part in my admittance at Ecole Polytechnique are unforgettable. Nick Baigent has been a source of energy for me throughout my postgraduate studies. At least once in a year he was sparing good enough time to hear about my work and give his invaluable comments and suggestions. What else can a student who writes a social choice thesis possibly want? I met Ton Storcken very late, in the final year of my thesis studies. But his mastery in social choice and economic theory is just so artistic and teaching that I think I learned from him priceless lessons in a very short period of time. I guess these and his kindness and friendliness would be no surprise for people who know him. Over the years we worked with Jean, I felt Gilbert Laffond's strong influence all around. And in the very final phase of my thesis work, he was actually there and helped me improve. Thanks a lot.

Fuad Aleskerov, Miguel Ballester, Jordi Brandts, Donald Campbell, Nikolaos Georgantzís, Ayça Ebru Giritligil, Emin Karagözoğlu, Özgür Kıbrıs, Bettina Klaus, Semih Koray, Hans Peters and Jack Stecher are among those esteemed scholars that alloted to me and my questions from their precious time and effort. I appreciate a lot the inspiration and stimulation I enjoyed out of these occasions.

My dear fellow thesis students Faddy Ardian, Füsun Cengiz, Emre Cenker, Cüneyt Dalgakıran, Esther Delbourg, Fatih Demirkale, Hayrullah Dindar, Onur Doğan, Vanina Forget, Arnaud Goussebaile, Margot Hovsepian, Alda Kabre, Christine Lavaur, Guilhem Lecouteux, Alexis Louaas, Antonin Macé, Deniz Nebioğlu, Özer Selçuk, Robert Somogyi, Rafael Treibich and Yusuf Varlı made it all easy and bearable. With all these names we shared thoughts, tested ideas and poured coffee on them every once in a while.

İlknur Toygar Değer, Eliane Nitiga-Madelaine, Chantal Poujouly and Lyza Racon were of incredible help throughout. They maintained the perfect environment for a thesis. Sri Srikandan is a great friend who was always there. He also helped greatly with the design and execution of an experiment.

Murat Sertel Center for Advanced Economic Studies provided great support during thesis studies. TÜBITAK BİDEB (2211) fellowship made possible to cover necessary expenditures. I am grateful in that regard to Istanbul Bilgi University as well. Finally, İSAM Library provided a very nice study environment I needed occasionally.

Families can never be over-credited. But, I feel my family delivered a really exceptional support. My mother, father, sisters and brothers.. They acted as boosters only, even in their hardest times. Thank you!

Finally, I would like to share with the readers my limitless gratitude to my wife (yes, finally), Hande, who was always there and who is next to me even while writing these lines. Keeping busy with writing an excellent thesis herself, she witnessed every single step of this thesis as well. Her name is on each and every page like a hidden watermark.

Lund, Sweden

September, 2014

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

SWF Social Welfare Function

SCF Social Choice Function

ANES American National Election Studies

CH Cognitive Hierarchy

NE Nash Equilibrium

ECH Endogenous Cognitive Hierarchy

DER Duclos et al. (2004)

Part I

INTRODUCTION

1

INTRODUCTION

On April 10, 2010, a dozen distinguished scholars from renowned institutions convened at the Division of Social Sciences of Harvard University for a symposium that was openly aimed at reaching a conclusion that would, ideally, be the social science counterpart for what great mathematician David Hilbert once has done at the International Congress of Mathematicians in Paris in 1900.[1] Over thirty problems were posed and discussed (via social media, afterwards) to eventually gather a ranking of *hardest problems in social science*. Two of those which made it in the top ten were related to collective decision-making, including Richard J. Zeckhauser's question that is formed in the following plain words:

> *A critical problem for groups, ranging from the dyad to society as a whole, is how to aggregate information possessed by different individuals so that the group can use that information to make the best decisions.*

This thesis is yet another attempt at better understanding of the collective decision-making processes, one of the major ingredients of societies of our time. An accumulation of works that are initiated out of spirit of inquiry and enlightening supervision, it consists of -mostly conceptual- studies of a multiple of subjects of investigation in collective decision-making with varying methods.

The very first of these, both chronologically and according to the current composition, deals with methods of aggregation of preferences and constitutes Part 2.[2] We take preferences as simple phenomena, representable by consistent rankings of alternatives. Given such revelation by each member of a society, methods of aggregation are employed to reach a ranking that ideally would be agreeable as the representative for the preference of the society. It is argued, in the following lines, that when the outputs of aggregation are the same kind of objects as the ones required from individuals' submissions,

1 See http://socialscience.fas.harvard.edu/hardproblems.
2 This chapter is based on a work with the same title, co-authored with Jean Lainé and Remzi Sanver.

namely rankings, since individuals might evaluate rankings as alternatives themselves, a preference over these very rankings might be claimed to be possessed by each individual. Furthermore, assuming a degree of consistency between these two levels of preference, we can obtain candidates for these *hyper-preferences*. Once that is achieved, we can now ask what would be the outcome if the current rule were to be employed in aggregation of hyper-preferences. Would the outcome be consistent with the outcome of aggregation of simple preferences? Or in other words, is it of no unfavorable consequence to get away with simple preferences instead of hyper-preferences? The chapter is formalizing these ideas and providing results as to how two most important classes of aggregation methods perform in that regard, or are they *hyper-stable* or not: scoring rules and Condorcet-type rules. It is shown that the former fails this sort of stability by nature while the latter may include hyper-stable methods.

Part 3 is on measuring polarization in (political) preferences, without any mention to reasons or consequences of it. Throughout the part, the idea that polarization can be seen as aggregation of pairwise antagonisms in a society is sustained. The first chapter in this part formulates alienation in between individuals as the distance between them and furthermore takes into consideration the effect of the size of the group of individuals with exactly the same preference on the antagonism in between. The conclusion is that a very well known class of polarization measures (introduced first in income inequality studies) can be adopted naturally to preferential polarization in order to satisfy the plausible and well established properties. The second chapter, on the other hand, is about characterizing a very simple method of such an aggregation.[3] This method is simply the summation of the occurrences of differences in preferences on pairs of alternatives and shown to be characterized by three intuitive axioms. The final chapter of part 3 comprises an empirical approach.[4] In this chapter, we investigate the trend in public opinion polarization in the United States over last couple of decades. It is shown first that the literature on the subject is remarkably inconclusive on the issue and that this relies on the fact that different approaches entail different measures. We argue that the measure of polarization mentioned above is a functional tool also for this job by talking over the implications of previously used ones. We advance on by actually employing the method to data obtained from a praised source. In doing this, we benefit from several technical procedures in order to strengthen analysis. Our final conclusion in this chapter is that no significant trend in overall polarization in public preferences

3 This chapter is based on a work with the same title, co-authored with Burak Can and Ton Storcken.
4 This chapter is based on a work co-authored with Uğur Özdemir.

can be observed in the given time period, although there might be issues on which the public got more polarized in times.

In Part 4, the final part, we turn to the study of collective decision-making from a foundational perspective; is *inclusion* even a good thing? Through a model built upon the classic works of a great Enlightenment persona, a Platonic realist -believer of objective moral truths-, Marquis de Condorcet, we revisit the question if more opinion is necessarily better in terms of making more accurate decisions. To convince the reader that this may stand a non-trivial and relevant question even if we fully agree with Condorcet, we may point to a today-well-agreed idea, which can again be illustrated by Professor Zeckhauser's formulation of his another question at the Hard Problems in Social Science Symposium at Harvard University:[5]

> *If we know that individuals are susceptible to all kinds of biases and don't always make rational decisions, how do we decide 'what's good'?*

After all, dominant in the literature of information aggregation -especially the works build on Condorcet's jury model- is assuming a sort of *perfect* rationality of individuals, either directly or due to employed game theoretic solution concepts. In Chapter 5 we investigate, first, theoretically the consequences of one particular relaxation of this assumption, which imposes heterogeneity in cognitive undertakes of individuals of the problem, the problem of collectively choosing the correct alternative with the help of voting with relying on private and imperfect informations each individual holds.[6] We develop a new approach in this lieu and investigate how composition of individuals with different cognitive hierarchies may effect the outcome in different models. We then report results from an experiment we have run with human subjects in computerized environments. The experimental observations ratify this new model which performs better than previous models with and without strict rationality requirement in describing the behaviors of the subjects. Before conclusion, the chapter points to limitations of the framework and further research questions.

Quixotic as it can be, the work in this thesis would bring incomparable jubilation to its humble author if it could succeed in contributing even an iota to its readers' understanding of collective decision-making processes. Each chapter is founded on collaborations with esteemed scholars (mentors, in fact), and all errors are completely the author's.

5 Two other relevant questions are due to Nick Bostrom: "How can humanity increase its collective wisdom?" and Gary King: "How do we understand and grapple with collective decision-making where the outcome for everyone is suboptimal (e.g., the proliferation of weapons of mass destruction)?".

6 This chapter is based on a work with the same title, co-authored with Yukio Koriyama.

Part II

HYPER-PREFERENCES

2

HYPER-STABILITY OF SOCIAL WELFARE FUNCTIONS

(joint work with Jean Lainé and Remzi Sanver)

2.1 INTRODUCTION

Many collective choice situations involve orderings of a finite set of m alternatives as resolute outcomes. Natural examples are choosing a social preference or a priority order over decisions, ranking candidates in sport or arts competitions (e.g. the Eurovision song contest) or assigning tasks to individuals. In the latter example, there are m positions to be filled by m individuals, each being assigned a specific position. Given the natural ranking $1 > ... > m$ of the positions, a social outcome is an order $f(1) \succ ... \succ f(m)$ over individuals obtained by means of a bijection f from the set of positions to the set of individuals.

The classical framework of social choice theory calls for individuals to report their preferences over social outcomes. When social outcomes are linear orders, preferences over outcomes are orders of orders, or *hyper-preferences*. However, reporting full preferences faces a problem of practical implementation: in the no-indifference case, individuals have to rank $m!$ outcomes. More generally, when outcomes are complex combinations of basic alternatives, likewise orderings or subsets, choosing from full preference profiles is hardly achievable in practice. This suggests to design procedures based on partial information about individual preferences.[1] A simple option is asking each of the individuals to report only one order. Formally, this procedure reduces to using a Social Welfare Function (SWF) α, which maps every profile of linear orders to a weak order of alternatives, completed with a tie-breaking rule.

It follows that some normative properties of SWFs cannot be investigated without retaining assumptions on how individual orders over alternatives are extended to un-

1 This is what prevails in the Eurovision song contest, where ballots are based on a partial scoring method.

derlying hyper-preferences. A typical example is given by strategy-proofness, which can be defined only conditional to the way orders over alternatives are extended to hyper-preferences. Bossert and Storcken (1992) prove impossibility results for hyper-preferences generated by the *Kemeny distance criterion*: given an order P over alternatives, the hyper-preference from P ranks an order Q above another order Q' if the Kemeny distance between P and Q is strictly lower than the one between P and Q'.[2] Bossert and Sprumont (2014) investigate strategy-proofness for hyper-preferences based on the following *betweenness criterion*: the hyper-preference from P ranks Q above Q' if the set of alternative pairs P and Q agree on contains the set of pairs P and Q' agree on.[3] Another property requiring extending orders to hyper-preferences is the Pareto property, which states that an SWF (with a tie-breaking rule) chooses at any profile over alternatives a linear order that is not unanimously less preferred than another order.

In this chapter we introduce a new property for neutral SWFs called *hyper-stability*, which also implies linking orders over alternatives to hyper-preferences. Hyper-stability is a consistency property relating two levels of choice, the one from profiles of orders over alternatives, or *basic profiles*, and the one from hyper-preference profiles, or *hyper-profiles*. Loosely speaking, an SWF is hyper-stable if its outcome at any basic profile is top-ranked at the corresponding hyper-profile. More precisely, consider an SWF α defined for any finite number of alternatives. Hence, α provides a weak order at any basic profile over m alternatives as well as at any basic profile over $m!$ alternatives. Furthermore, suppose that α is neutral, meaning that its outcomes are not sensitive to the labeling of alternatives. Thus, profiles over $m!$ alternatives can be also interpreted as hyper-preference profiles over orders of m alternatives, or in short hyper-profiles. While a basic profile clearly entails a huge loss of information about preferences over outcomes, there may nonetheless exist, in the spirit of revealed-preference theory, a class of underlying hyper-profiles (over $m!$ orders) compatible with the basic profile at which α ranks at top at least one linear extension of the weak order chosen from the basic profile. If this happens at every possible reduced profile, we say that α is hyper-stable.

As for strategy-proofness, a key-issue for hyper-stability is what is meant by a hyper-profile compatible with a basic profile. We assume here that compatibility holds when hyper-preferences are generated from orders over alternatives in accordance with the betweenness criterion. Clearly, this criterion allows to compare only a small number of orders. Therefore a basic profile generates a large class of compatible hyper-profiles.

2 The Kemeny distance between two linear orders is the number of pairs of alternatives which they disagree on.

3 See Duddy et al. (2010) for an analysis of strategy-proof SWFs based on ordinally fuzzy preferences.

Nonetheless, we prove the existence of a unanimous and hyper-stable *Condorcet SWF*.[4] However, many well-known Condorcet SWFs are not hyper-stable.

We also pay attention to the sub-class of hyper-profiles built by means of the Kemeny distance criterion. Hyper-stability relative to this sub-class is called Kemeny-stability. We show that no *scoring rule* is Kemeny-stable, hence hyper-stable, unless there are exactly three alternatives. In this case, we show that there exists a unique normalized Kemeny-stable scoring rule. Hence, our main result is that ranking by scoring is incompatible with hyper-stability, while the Condorcet criterion is not.

To the best of our knowledge, hyper-stability is a new property for SWFs, although related properties appear in several studies of collective choice. The yeast of the present study can be found in Binmore (1975), who considers a stronger notion of hyper-stability, although in a different setting. Suppose that preferences are now weak orders over three alternatives, which are aggregated to a weak order by means of a neutral SWF α. Binmore does not comment on hyper-preferences beyond writing "if a rational entity holds a certain preference preordering over a set of alternatives, then that entity must also subscribe to a certain partial preordering of the set of all preorderings" (Binmore (1975), p. 379). Moreover, weak orders are compared according to their respective top-sets. All relevant top-sets in Binmore's analysis contain at most two elements and the criterion works as follows: Given a weak order R, sets $\{x\}$, $\{y\}$, and $\{x, y\}$ are ranked in the order $\{x\}, \{x, y\}, \{y\}$ if and only if xRy. Given the 13 possible weak orders over 3 alternatives, this criterion suffices to find a family $\mathcal{T}$ of triples of weak orders on which basic preferences generate an hyper-profile.[5] Since α is neutral, it can be applied to each of these hyper-profiles, leading to a weak order R_T over each triple T in $\mathcal{T}$. Furthermore, the weak order chosen from the basic profile also induces a weak order $\widetilde{R}_T$ over each triple T in $\mathcal{T}$. Binmore shows that R_T and $\widetilde{R}_T$ coincide for all T in $\mathcal{T}$ if and only if α is either dictatorial, or anti-dictatorial or constant. There are three main differences between Binmore's approach and the present one. First, basic preferences and hyper-preferences are weak orders in Binmore's study, while we assume both are linear orders. Second, Binmore's setting defines SWFs for three alternatives only. Using neutrality

4 An SWF α is Condorcet if at any profile, it ranks alternatives as in the majority tournament whenever the latter is a linear order.

5 To see why, label alternatives as x, y and z, and consider the following weak orders R_1, R_2 and R_3 (with respective a-symmetric parts P_1, P_2 and P_3) defined by zP_1yP_1x, yP_2zP_2x and yR_3zP_3x. Denote by $\succeq_1$, $\succeq_2$ and $\succeq_3$ the respective hyper-preferences induced on $\{R_1, R_2, R_3\}$ by R_1, R_2 and R_3. Then one gets $R_1 \succ_1 R_3 \succ_1 R_2$, $R_2 \succ_2 R_3 \succ_2 R_1$ and $R_1 \sim_3 R_2 \sim_3 R_3$. It is easily seen that for each of the 13 possible weak orders, R_1, R_2 and R_3 are ranked as in $\succeq_1$, or $\succeq_2$ or $\succeq_3$. Hence, any basic profile generates an hyper-profile over the triple $\{R_1, R_2, R_3\}$.

together with a way to generate hyper-preferences, this allows to choose from hyper-profiles over triples of orders. In contrast, our setting involves a variable number of alternatives, and defines hyper-preferences as linear orders over all orders. Again, using neutrality together with a way to generate hyper-preferences, this allows to have a well-defined outcome at profiles over m alternatives and at hyper-profiles over $m!$ orders. Third, our definition of hyper-stability is clearly less demanding than Binmore's one, since it only requires that some social order chosen from basic profiles is top-ranked from hyper-profiles, imposing nothing about how this social order itself generates a social hyper-preference.

Another study related to hyper-stability can be found in Laffond and Lainé (2000), although the property is not explicitly stated there. Using the same framework as the present one, Laffond and Lainé characterize the domain of (neutral and independent) hyper-preferences such that whenever the majority tournament at a basic profile is transitive, it is a Condorcet winner of any corresponding hyper-profile.[6] This characterization result can be restated as follows in terms of hyper-stability. Call strongly Condorcet a Condorcet SWF that uniquely ranks first the Condorcet winner whenever it exists. Then there exists a class of hyper-preferences making every strongly Condorcet SWF hyper-stable.

Hyper-stability also appears, at least in watermark, in the literature of moral judgments.[7] Sen and Körner (1974) argues that morality requires to formulate judgments among preferences while rationality does not, and suggests using moral views, defined as hyper-preferences, as a way out of the Paretian liberal paradox *à la* Sen (1970).[8] If one accepts basic profiles as expressions of rationality (individuals reporting their first-best outcome) and hyper-profiles as expressions of moral judgments, hyper-stability can be interpreted as a property of moral consistency: choices made from rational preferences does not conflict with the one made from moral judgments.

Furthermore, hyper-stability also relates to a self-selectivity property. Self-selectivity is defined for a social choice function (SCF) by Koray (2000).[9] Roughly speaking, an SCF is self-selective if it chooses itself against any finite number of other social choice

6 Given a profile involving an odd number of individuals, the majority tournament is the complete and asymmetric binary relation obtained by pairwise comparisons of alternatives according to the simple majority rule. Moreover, the (necessarily unique) Condorcet winner of that profile is the alternative which defeats all other alternatives in the majority tournament.

7 One can think of hyper-preferences also as preferences of individuals over others in the society.

8 See Igersheim (2007). The reader may refer to Jeffrey (1974), McPherson (1982), and Sen (1977) for further discussion on the more general concept of a *meta-preference*.

9 A social choice function picks one alternative at every profile of preferences over alternatives. For further studies of self-selectivity, see Koray and Unel (2003) and Koray and Slinko (2008).

functions. Self-selectivity thus involves two levels of choice: choices from profiles over alternatives, and choices from profiles over choice functions. These two levels are connected by means of a consequentialist principle, which states that individuals preferring alternative x to alternative y will rank any function choosing x above any function choosing y. Koray (2000) shows that a neutral and unanimous SCF is self-selective if and only if it is dictatorial. While consequentialism allows for a canonical extension of preferences over alternatives to preferences over SCFs, this is no longer the case for SWFs. Nonetheless, self-selectivity for SWFs can be defined conditional to the definition of hyper-preferences. An individual with preference P in some basic profile P_N will prefer SWF α_1 to SWF α_2 if $\alpha_1(P_N)$ is "closer" to P than $\alpha_2(P_N)$, where closer can be in terms of the Kemeny or any other distance. More generally, once defined how a linear order generates an hyper-preference, two SWFs are compared according to the way this hyper-preference ranks their respective outcomes. Hence the consequentialist principle applies, but conditional to the way basic preferences are extended to hyper-preferences. We say that an SWF is *SW self-selective* for some preference extension if, at any basic profile it ranks itself first when compared to any finite set of SWFs. We show below that hyper-stability is a necessary condition for SW self-selectivity.

The rest of the chapter is organized as follows. Part 2 formally defines hyper-stability, and investigates its relation to self-selectivity. Hyper-stability of scoring rules is studied in Part 3. In particular, we provide examples showing that neither the Borda rule, nor the plurality and anti-plurality rules are Kemeny-stable, hence hyper-stable. Moreover, we show that no unanimous scoring rule is Kemeny-stable, and that no scoring rule is hyper-stable. Condorcet SWFs are considered in Part 4. We show that the Slater SWF, the Kemeny rule, and the Copeland SWF are not hyper-stable, whereas the transitive closure of the majority relation over alternatives is hyper-stable. The chapter ends up with comments about alternative concepts of hyper-stability, together with open questions. Finally, all proofs are postponed to an appendix.

2.2 HYPER-STABILITY

2.2.1 *Notations and definitions*

$\mathbb{N}$ denotes the set of non-zero natural numbers. We consider societies with variable numbers of individuals and of alternatives. Hence, $\mathbb{N}$ stands for the sets of potential alternatives and individuals, and each actual society involves finitely many individuals confronting finitely many alternatives. Given a finite subset X_m of $\mathbb{N}$ with cardinality m, the set of linear (resp. weak) orders over X_m is denoted by $\mathcal{L}(X_m)$ (resp. $\mathcal{R}(X_m)$). An order $P \in \mathcal{L}(X_m)$ is *a linear extension* of $R \in \mathcal{R}(X_m)$ if for any $a, b \in X_m$, $aPb \Rightarrow aRb$. The set of all linear extensions of $R \in \mathcal{R}(X_m)$ is denoted by $\Delta(R)$. Given a set N of n individuals, a weak profile over X_m is an element R_N of $\mathcal{R}(X_m)^n$, and a profile is an element P_N of $\mathcal{L}(X_m)^n$. The set of all linear extensions of the weak profile R_N is $\Delta(R_N) = \times_{i\in N}(\Delta(R_i))$.

Denoting by $\mathcal{X}_m$ the set of all subsets of $\mathbb{N}$ with cardinality m, a function $\alpha : \cup_{m,n\in\mathbb{N}} \cup_{X_m\in\mathcal{X}_m} \mathcal{L}(X_m)^n \rightarrow \cup_{m\in\mathbb{N}}\mathcal{R}(X_m)$ is a *social welfare function* (SWF) if for all $n, m \in \mathbb{N}$, for all $X_m \in \mathcal{X}_m$ and for all $P_N \in \mathcal{L}(X_m)^n$, $\alpha(P_N) \in \mathcal{R}(X_m)$. Let X_m and X'_m be two different sets in $\mathcal{X}_m$, and consider any bijection σ from X_m to X'_m. If $R \in \mathcal{L}(X_m)$, we define R^σ as the element of $\mathcal{L}(X'_m)$ such that for all $x, y \in X'_m$, $x\ R^\sigma\ y$ if and only if $\sigma^{-1}(x)\ P_i\ \sigma^{-1}(y)$. An SWF α is neutral if for all $n, m \in \mathbb{N}$, for all $X_m, X'_m \in \mathcal{X}_m$, for all bijections σ from X_m to X'_m, and for all profiles $P_N = (P_1, ., ., P_n) \in \mathcal{L}(X_m)^n$, we have $\alpha(P_N^\sigma) = [\alpha(P_N)]^\sigma$, where $P_N^\sigma = (P_1^\sigma, ..., P_n^\sigma) \in \mathcal{L}(X'_m)^n$. Given any $m \in \mathbb{N}$, define $A_m = \{1, ..., m\} \in \mathcal{X}_m$. Since neutrality states that the way to rank alternatives is non-sensitive to their labeling, we can define a neutral SWF α as a function from $\cup_{m,n\in\mathbb{N}} A_m$ to $\cup_{m\in\mathbb{N}}\mathcal{R}(X_m)$ such that for all $n, m \in \mathbb{N}$ and for all $P_N \in \mathcal{L}(A_m)^n$, $\alpha(P_N) \in \mathcal{R}(A_m)$.

Furthermore, a neutral SWF α is *unanimous* if, for any $m, n \in \mathbb{N}$, for any profile $P_N \in \mathcal{L}(A_m)^n$, for any two alternatives $a, b \in A_m$, $[a\ P_i\ b$ for all $i = 1, ..., n]$ implies that $[a\ \alpha(P_N)\ b$ and $\neg(b\ \alpha(P_N)\ a)]$. Finally, the *rational social choice correspondence attached to* α is the function $f_\alpha : \cup_{n,m\in\mathbb{N}}\mathcal{L}(A_m)^n \rightarrow 2^{A_m} \setminus \emptyset$ defined by: $\forall n, m \in \mathbb{N}$, $\forall P_N \in \mathcal{L}(A_m)^n$, $\forall a \in A_m$, $a \in f_\alpha(P_N) \iff a\ \alpha(P_N)\ b$ for all $b \in A_m$. Hence, f_α selects at each profile P_N the set of best alternatives for $\alpha(P_N)$.

2.2.2 *Preference extensions*

We turn now to the notion of hyper-preference. A *preference extension* is a function $e : \cup_{m\in\mathbb{N}}\mathcal{L}(A_m) \rightarrow \cup_{m\in\mathbb{N}}\mathcal{L}(\mathcal{L}(A_m))$ such that for all $m \in \mathbb{N}$ and all $P \in \mathcal{L}(A_m)$, $e(P) \in \mathcal{L}(\mathcal{L}(A_m))$. Hence, a preference extension maps each linear order over m alternatives to a linear order over all linear orders over alternatives. An element of $\mathcal{L}(\mathcal{L}(A_m))$ is called hyper-preference. An *extension domain* is a subset $\mathcal{E}$ of the set of all preference extensions. Given a profile $P_N = (P_1, ..., P_n) \in \mathcal{L}(A_m)^n$ together with a n-tuple $E = (e_1, ..., e_n) \in \mathcal{E}^n$, an *hyper-profile* of P_N is the element $P_N^E = (e_1(P_1), ..., e_n(P_n))$.

Given $P, Q \in \mathcal{L}(A_m)$, we define the set $A(P,Q) = \{(a,b) \in A_m \times A_m : aPb \text{ and } aQb\}$, which contains all alternative pairs P and Q agree on. We focus on the specific class of *betweenness-consistent* preference extensions.

Definition 2.2.1. *A preference extension e is betweenness-consistent if for all $m \in \mathbb{N}$ and all $P, Q, Q' \in \mathcal{L}(A_m)$, $A(P,Q) \supset A(P,Q')$ implies $Q\ e(P)\ Q'$.*

We denote by $\mathbb{B}$ the domain of betweenness-consistent preference extensions. Given $P, Q \in \mathcal{L}(A_m)$, the *Kemeny distance* between P and Q is defined by $d_K(P,Q) = |\{(a,b) \in A_m \times A_m : aPb \text{ and } bQa\}|$, that is the number of pairs of alternatives P and Q disagree on.

Definition 2.2.2. *A preference extension e is Kemeny if for all $m \in \mathbb{N}$ and all $P, Q, Q' \in \mathcal{L}(A_m)$, $d_K(P,Q) < d_K(P,Q')$ implies $Q\ e(P)\ Q'$.*

We denote by $\mathbb{K}$ the domain of Kemeny preference extensions. Pick up any $P \in \mathcal{L}(A_m)$. The Kemeny distance allows to induce from any $P \in \mathcal{L}(A_m)$ the element $\succeq_P$ $\in \mathcal{R}(\mathcal{L}(A_m))$ defined by: $\forall Q, Q' \in \mathcal{L}(A_m)$, $Q \succeq_P Q'$ iff $d_K(P,Q) \leq d_K(P,Q')$, and $Q \succ_P Q'$ iff $d_K(P,Q) < d_K(P,Q')$. In words, the weak order $\succeq_P$ induced by P ranks orders according to their respective distances to P. Given profile $P_N = (P_1, ..., P_n) \in \mathcal{L}(A_m)^n$, the Kemeny weak profile for P_N is defined by $P_N^K = (\succeq_{P_1}, ..., \succeq_{P_n})$. Thus, a preference extension e is Kemeny if for all $m \in \mathbb{N}$ and all $P \in \mathcal{L}(A_m)$, e is a linear extension of $\succeq_P$. We call Kemeny hyper-profile any linear extension of P_N^K. Clearly, every Kemeny extension is betweenness-consistent, and thus $\mathbb{K} \subset \mathbb{B}$.

The Kemeny distance criterion can be criticized by arguing that when comparing two orders, inversions in the lower tail of the ranking are less important that inversions in the upper tail. If three candidates a, b, c are to be ranked as gold, silver and bronze medal, and if they are ranked as $aPbPc$, then one may prefer order $aQcQb$ to order $bQ'aQ'c$, since reversing order for gold and silver may be seen as a more significant deviation than reversing order for silver and bronze. This calls for breaking symmetry by using weighted Kemeny distance (equivalently, this calls for some specific way to break ties in the Kemeny weak profiles). Note however that such a critic no longer holds if agendas are interpreted as task assignments. Indeed, suppose that $aQcQb$ stands for assigning task 1 to individual a, task 2 to c, and task 3 to b, a similar meaning being given to Q'. Provided that all tasks are given the same importance, Q and Q' involve only one mismatch from the viewpoint P, and nothing suggests why Q should be preferred to Q'.

The following example illustrates the construction of Kemeny hyper-profiles. Consider the following profile $P_N = (P_1, P_2, P_3)$ over 3 alternatives a, b, c:

$$P_N = \left(\begin{array}{ccc} P_1 & P_2 & P_3 \\ \hline a & c & c \\ b & b & a \\ c & a & b \end{array} \right)$$

The Kemeny weak profile P_N^K of P_N is defined by

$$P_N^K = \left(\begin{array}{ccc} \succeq_{P_1} & \succeq_{P_2} & \succeq_{P_3} \\ \hline abc & cba & cab \\ acb, bac & bca, cab & cba, acb \\ bca, cab & acb, bac & abc, bca \\ cba & abc & bac \end{array} \right)$$

where xyz stands for the linear order $xPyPz$, and where two orders belonging to the same row and column are indifferent. A Kemeny hyper-profile for P_N is any element $\dot{P}_N$ of $\Delta(P_N^K)$. For instance,

$$\dot{P}_N = \left(\begin{array}{ccc} \dot{P}_1 & \dot{P}_2 & \dot{P}_3 \\ \hline abc & cba & cab \\ bac & cab & cba \\ acb & bca & acb \\ bca & bac & abc \\ cab & acb & bca \\ cba & abc & bac \end{array} \right)$$

Contrarily to the Kemeny distance criterion, betweenness-consistency does not automatically induces a weak order over orders. For instance, $e(P_1) \in \mathbb{B}$ only if the following conditions holds: (1) $e(P_1)$ uniquely ranks P_1 first and its inverse cba last, (2) acb is ranked above cab, and (3) bac is ranked above bca. The reader will easily check that hyper-profile $\widetilde{P}_N$ below is built from a vector of betweenness-consistent preference extensions which are not Kemeny.

$$\widetilde{P}_N = \left(\begin{array}{ccc} \widetilde{P}_1 & \widetilde{P}_2 & \widetilde{P}_3 \\ \hline abc & cba & cab \\ bac & cab & cba \\ bca & bac & acb \\ acb & bca & abc \\ cab & acb & bca \\ cba & abc & bac \end{array}\right)$$

2.2.3 *Hyper-stability: definition*

We are now ready to formally define hyper-stability:

Definition 2.2.3. *A neutral social welfare function α is hyper-stable for the domain $\mathcal{E}$ of preference extensions and a number m of alternatives if for all $n \in \mathbb{N}$, for all $P_N \in \mathcal{L}(A_m)^n$, for all $E = (e_1, ..., e_n) \in \mathcal{E}^n$, we have $\Delta(\alpha(P_N)) \cap f_\alpha(P_N^E) \neq \emptyset$. Moreover, α is hyper-stable for $\mathcal{E}$ if it is hyper-stable for $\mathcal{E}$ and any $m \in \mathbb{N}$.*

A neutral SWF α is hyper-stable for domain $\mathcal{E}$ if at every finite profile P_N of linear orders over m alternatives, at least one linear extension of the weak order $\alpha(P_N)$ is ranked first by α when applied to any hyper-profile P_N^E induced from P_N by a vector of preference extensions in $\mathcal{E}$. We furthermore say that α is *Kemeny-stable* if it is hyper-stable for $\mathbb{K}$. Figure 1 below illustrates hyper-stability.

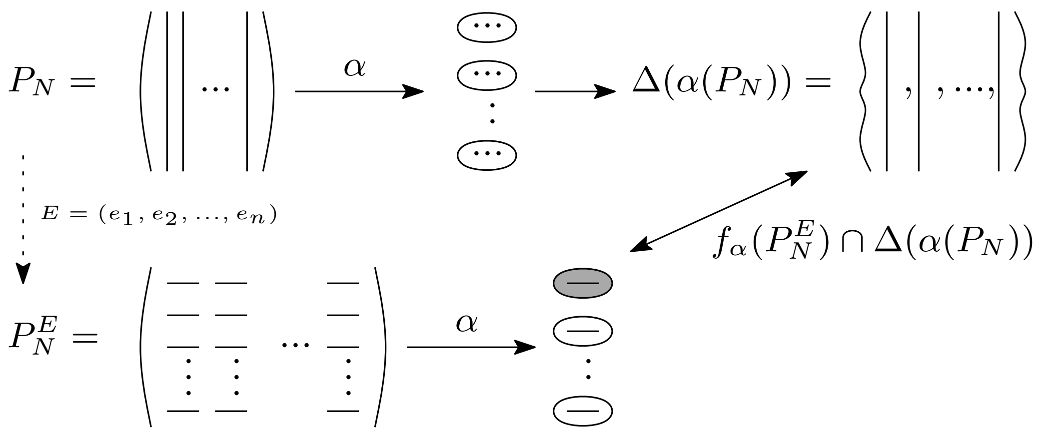


Figure 2.2.1. Hyper-stability.

A society with size n has to rank m alternatives, and has agreed on some SWF α to do so. Interpreting α as a voting rule, individual ballots are linear orders of alternatives (profile P_N), and ballots are aggregated by means of α to a weak order $\alpha(P_N)$ of alternatives. Since $\alpha(P_N)$ may involve ties, and since resolute outcomes are linear orders, the final choice results from the use of some tie-breaking rule. The set $\Delta(\alpha(P_N))$ contains all possible outcomes when some tie-breaking rule prevails. We assume that "preferences behind ballots" are induced from ballots by some n-tuple $E = (e_1, ..., e_n)$ of preference extensions. Therefore, any set of ballots P_N together with E generates an hyper-profile P_N^E over orders. Since α is neutral and defined for any number of alternatives, it can be applied to P_N^E, leading to a weak order $\alpha(P_N^E)$ over outcomes. Hyper-stability prevails for E if starting from any P_N, at least one possible final outcome from P_N is ranked first by α (or, equivalently, chosen by f_α) at the generated hyper-profile P_N^E.

2.2.4 *Hyper-stability and SW self-selectivity*

The main motivation for studying hyper-stability is that full preferences over outcomes are hardly known in practice. Another motivation stems from its close relationship with self-selectivity. Self-selectivity is defined by Koray (2000) for a neutral social choice function (SCF).[10] Suppose that the society has to choose one alternative among finitely many, as well as the SCF itself. Moreover, suppose that given individual preferences over alternatives, individuals compare SCFs by considering only their respective outcomes. According to this consequentialist principle, initial preferences over alternatives naturally extend to preferences over SCFs: consider any finite subset $\mathcal{G}$ of neutral SCFs together with a profile $P_N = (P_1, ..., P_n) \in \mathcal{L}(A_m)^n$; define for all $i = 1, ..., n$ the weak order $R(P_i)$ over $\mathcal{G}$ by: $\forall F, G \in \mathcal{G}$, $F\ R^+(P_i)\ G \Leftrightarrow F(P_N)\ P_i\ G(P_N)$, and $F\ R^\sim(P_i)\ G \Leftrightarrow F(P_N) = G(P_N)$, where $R^+(P_i)$ (resp. $R^\sim(P_i)$) is the asymmetric (resp. symmetric) part of $R(P_i)$. It follows that P_N induces a dual profile of weak orders $R_N^{\mathcal{G}} = (R(P_1), ..., R(P_n))$ over $\mathcal{G}$. *Self-selectivity* holds for an SCF F if, at any profile over alternatives, F selects itself at some linear extension of the dual profile over any finite set of SCFs. Formally, F is self-selective if for all $m, n \in \mathbb{N}$, for all $P_N \in \mathcal{L}(A_m)^n$ and for all finite subsets $\mathcal{G}$ of neutral SCFs with $F \in \mathcal{G}$, there exists a linear extension $\widetilde{P}_N^{\mathcal{G}}$ of $R_N^{\mathcal{G}}$ with $F(\widetilde{P}_N^{\mathcal{G}}) = F$. Koray

10 An SCF maps any profile of linear orders over any finite set to an element of that set. Defining neutrality along the same lines as for SWFs allows to formally define an SCF as a function $F : \cup_{m,n \in \mathbb{N}} \mathcal{L}(A_m)^n \rightarrow \cup_{m \in \mathbb{N}} A_m$ such that for all $n, m \in \mathbb{N}$ and all $P_N \in \mathcal{L}(A_m)^n$, $F(P_N) \in A_m$.

(2000) proves that, given any fixed size n of the society, a neutral and unanimous SCF is self-selective if and only if it is dictatorial.[11]

Self-selectivity for neutral SWFs can defined along the same lines: at any profile over alternatives, a self-selective SWF ranks itself first among finitely many other SWFs. However, since an SWF provides a weak order, there is no longer a natural duality between preferences over alternatives and preferences over SWFs. In order to make the consequentialist principle meaningful, we need to connect both preference levels by means of a preference extension. It follows that self-selectivity is defined conditional to some domain of preference extensions. This last point is the major difference between the SCF and the SWF settings: choosing preference extensions brings an extra degree of freedom in the analysis, which may allow to escape from Koray's impossibility result.

We formalize self-selectivity for SWFs as follows. An SWF α is called *strict* if for all $n, m \in \mathbb{N}$ and all $P_N \in \mathcal{L}(A_m)^n$, one has $\alpha(P_N) \in \mathcal{L}(A_m)$. A linearization of SWF α is a strict SWF α^* such that for all $n, m \in \mathbb{N}$, for all $a, b \in A_m$ and for all $P_N \in \mathcal{L}(A_m)$, $a\ \alpha^*(P_N)\ b$ implies $a\ \alpha(P_N)\ b$. The set of all linearizations of α is denoted by $L(\alpha)$. Pick up a profile $P_N = (P_1, ..., P_n) \in \mathcal{L}(A_m)^n$ together with a domain $\mathcal{E}$, and consider any finite subset $\mathcal{A} = \{\alpha_1, ..., \alpha_K\}$ of neutral SWFs. A strict selection of $\mathcal{A}$ is a subset $\mathcal{A}^* = \{\alpha_1^*, ..., \alpha_K^*\}$ of linearizations of $\alpha_1, ..., \alpha_K$. For all $1 \leq i \leq n$, define the weak order $\succeq_{P_i}^{\mathcal{A}^*}$over $\mathcal{A}^*$ by: $\forall 1 \leq k, k' \leq K$, $\alpha_k^* \succ_{P_i}^{\mathcal{A}^*} \alpha_{k'}^* \Leftrightarrow \alpha_k^*(P_N)\ e_i(P_i)\ \alpha_{k'}^*(P_N)$, and $\alpha_k^* \sim_{P_i}^{\mathcal{A}^*} \alpha_{k'}^* \Leftrightarrow \alpha_k^*(P_N) = \alpha_{k'}^*(P_N)$ for some $(e_1, ..., e_n) \in \mathcal{E}^n$. Thus, as for SCFs, P_N together with $E = (e_1, ..., e_n) \in \mathcal{E}^n$ induces a dual profile of weak orders $R_N^{E,\mathcal{A}^*} = (\succeq_{P_1}^{\mathcal{A}^*}, ..., \succeq_{P_n}^{\mathcal{A}^*})$ over $\mathcal{A}^*$.

Definition 2.2.4. *A neutral SWF α is SW self-selective for the domain of preference extensions E if and only if for all $m, n \in \mathbb{N}$, for all $P_N \in \mathcal{L}(A_m)^n$, for all finite subsets $\mathcal{A}$ of neutral SWFs that contain α, for all strict selections $\mathcal{A}^*$ of $\mathcal{A}$, for any $E = (e_1, ..., e_n) \in \mathcal{E}^n$, there exists a linear extension $\widetilde{P}_N^{E,\mathcal{A}^*}$ of $R_N^{E,\mathcal{A}^*}$ for which $L(\alpha) \cap \mathcal{A}^* \cap f_\alpha(\widetilde{P}_N^{E,\mathcal{A}^*}) \neq \emptyset$.*

A neutral SWF α is SW self-selective for domain $\mathcal{E}$ if the following holds: pick up any strict selection $\mathcal{A}^*$ of any finite set $\mathcal{A}$ of neutral SWFs including α, together with any profile P_N over alternatives. Every n-tuple of preference extensions in $\mathcal{E}$ generates from P_N a dual profile of weak orders over $\mathcal{A}^*$. SW selectivity holds if there exists a linear extension of this dual profile at which α ranks first at least some of its linearizations in $\mathcal{A}^*$.

11 An SCF F is dictatorial if $\exists 1 \leq i \leq n$ such that, for all $P_N \in \mathcal{L}(A_m)^n$, $F(P_N) = a \Leftrightarrow aP_ib$ for all $b \in A_m \backslash \{a\}$. Moreover, F is unanimous if for any m, for any $P_N \in \mathcal{L}(A_m)^n$, for all $a, b \in A_m$, $[aP_ib$ for all $1 \leq i \leq n] \Rightarrow b \notin F(P_N)$.

Note that, although it offers a natural adaptation of the original concept to SWFs, the formalization of SW self-selectivity looks complex for two main reasons. First, two different SWFs may have the same outcome at some profile P_N. Therefore, choosing a domain $\mathcal{E}$ is not enough to provide a dual profile of linear orders over SWFs. Second, two SWFs may produce different weak orders at P_N that admit the same linearization. Moreover, note the crucial role played by neutrality, which allows for α to be well-defined for profiles over alternatives and for dual profiles over SWFs.

Proposition 1 below states that hyper-stability is a weaker property than SW self-selectivity.

Proposition 2.2.1. *If a neutral SWF is hyper-stable for a domain $\mathcal{E}$, then it is SW self-selective for $\mathcal{E}$.*

2.3 SCORING RULES

We first study hyper-stability of scoring rules. Given a number m of alternatives, a score vector is an element $S^m = (s^{1,m}, s^{2,m}, ..., s^{m,m})$ of $\mathbb{R}_+^m$, where (1) $s^{m,m} = 0$, (2) $s^{1,m} \geq s^{2,m} \geq ... \geq s^{m,m}$, and (3) $s^{1,m} > 0$. Given a profile $P_N \in \mathcal{L}(A_m)^n$ together with a score vector S^m, the score of the alternative $x \in A_m$ in P_N is $S^m(x, P_N) = \sum_{i \in N} s^{r_i(x,P_N),m}$, where $r_i(x, P_N)$ is the rank of x in P_i. A SWF α is a *scoring rule* if there exists a sequence $\{S_\alpha^m\}_{m \geq 3} = \{S_\alpha^1, S_\alpha^2, S_\alpha^3 ...\}$ of score vectors such that, for any $m, n \in \mathbb{N}$, for any $P_N \in \mathcal{L}(A_m)^N$, for any two alternatives $x, y \in A_m$, $x\ \alpha(P_N)\ y \iff S_\alpha^m(x, P_N) \geq S_\alpha^m(y, P_N)$. Clearly, every scoring rule is neutral. We begin with the analysis of well-known scoring rules, namely the Borda rule, the plurality rule and the anti-plurality rule.

The *Borda rule* $\mathcal{B}$ is defined by: for any $m \in \mathbb{N}$, for any $k \in \{1, ..., m-1\}$, $s_\mathcal{B}^{k,m} = s_\mathcal{B}^{k+1,m} + 1$. It is easily checked that $\mathcal{B}$ is not Kemeny-stable, hence not hyper-stable for $\mathcal{E}$. Indeed, consider the following profile P_N involving 3 alternatives a, b, c and 6 individuals, where the first row indicates the number of individuals sharing the same preference order

$$P_N = \begin{pmatrix} 3 & 1 & 2 \\ \hline a & c & c \\ b & b & a \\ c & a & b \end{pmatrix}$$

Next, consider the following linear extension $\dot{P}_N$ of $P_N^\mathcal{K}$:

$$\dot{P}_N = \left(\begin{array}{ccc} 3 & 1 & 2 \\ \hline abc & cba & cab \\ bac & cab & cba \\ acb & bca & acb \\ bca & bac & abc \\ cab & acb & bca \\ cba & abc & bac \end{array} \right)$$

where abc stands for the order ranking a first, b second and c third. Then $\mathcal{B}(P_N) = \{acb\} = \Delta(\mathcal{B}(P_N))$, whereas $S^6_{\mathcal{B}}(acb, \dot{P}_N) = 16 < S^6_{\mathcal{B}}(abc, \dot{P}_N) = 19$ implies that $acb \notin f_{\mathcal{B}}(\dot{P}_N)$. Since $\Delta(\mathcal{B}(P_N)) \cap f_{\mathcal{B}}(\dot{P}_N) = \emptyset$, $\mathcal{B}$ is not Kemeny-stable.

The *plurality rule* is the scoring rule π, where, for any $m \in \mathbb{N}$, $s^{k,m}_{\pi} = 0$ for any $k = 2, ..., m$, and $s^{1,m}_{\pi} = 1$. Consider an alteration P'_N of the profile P_N above where the individual with preference cba changes to bca. Then $\pi(P'_N) = \{acb\}$, while, for any linear extension $\dot{P}'_N$ of P'^K_N, $f_{\pi}(\dot{P}'_N) = \{abc\}$. Hence, π is not Kemeny-stable.

The *anti-plurality rule* is the scoring rule λ, where, for any $m \in \mathbb{N}$, $S^{k,m}_{\lambda} = 1$ for any $1 \leq k \leq m-1$. Consider the following profile $P_N \in \mathcal{L}(A_3)^{15}$ together with its associated Kemeny weak profile P^K_N:

$$P_N = \left(\begin{array}{ccccc} 3 & 2 & 3 & 3 & 4 \\ \hline a & a & b & c & c \\ b & c & a & a & b \\ c & b & c & b & a \end{array} \right) \quad P^K_N = \left(\begin{array}{ccccc} 3 & 2 & 3 & 3 & 4 \\ \hline abc & acb & bac & cab & cba \\ acb, bac & abc, cab & abc, bca & cba, acb & cab, bca \\ bca, cab & cba, bac & acb, cba & abc, bca & bac, acb \\ cba & bca & cab & bac & abc \end{array} \right)$$

Clearly, $\lambda(P_N) = \{abc\}$. We conclude that, for all $P \in \mathcal{L}(A_6) \backslash \{abc\}$, $P \; \lambda(\dot{P}_N) \; abc$ for all $\dot{P}_N \in \Delta(P^K_N)$. Thus, $abc \notin f_{\lambda}(\dot{P}_N)$, and therefore λ is not Kemeny-stable.

We state below four negative results about Kemeny-stable scoring rules. The key-ingredient of the proofs is the following characterization of Kemeny-stable scoring rules for 3 alternatives.

Theorem 2.3.1. *A scoring rule α is Kemeny-stable for three alternatives if and only if* $s^{1,3}_{\alpha} = 2.s^{2,3}_{\alpha} > 0$ *and* $s^{1,6}_{\alpha} = \frac{4}{3}s^{2,6}_{\alpha} = \frac{4}{3}s^{3,6}_{\alpha} = 4s^{4,6}_{\alpha} = 4s^{5,6}_{\alpha} > s^{6,6}_{\alpha} = 0$.

Hence, there exists a unique pair of normalized score vectors $\{S^3_{\alpha}, S^6_{\alpha}\}$ making a scoring rule α Kemeny-stable for three alternatives.[12] Clearly, the condition stated in Theorem 1 is necessary for hyper-stability.

12 A score vector S^m is normalized if $s^{1,m} = 1$.

A scoring rule α is *non-truncated* if there exists no $m \in \mathbb{N}$ and no $k \in \{2, ..., m-1\}$ such that $s_\alpha^{k,m} = 0$: the score vector defined for some number m of alternatives gives a strictly positive score to any rank above the last one.

Theorem 2.3.2. *There is no Kemeny-stable and non-truncated scoring rule.*

A scoring rule α is *strict-at-top* if, for any $m \in \mathbb{N}$, $s_\alpha^{1,m} > s_\alpha^{2,m}$: all score vectors give a score to the top-ranked alternative strictly higher than any other score. Typical examples of strict-at-top scoring rules are the plurality and the Borda rules. Note that any convex scoring rule is also strict-at-top.[13]

Theorem 2.3.3. *There is no Kemeny-stable and strict-at-top scoring rule.*

Since a unanimous scoring rule must be strict-at-top and non-truncated, we can state the following corollary of Theorems 2 and 3.

Theorem 2.3.4. *There is no Kemeny-stable and unanimous scoring rule.*

When enlarging the Kemeny domain $\mathbb{K}$ to the domain $\mathbb{B}$ of betweenness-consistent preference extensions, we get an even stronger negative result:

Theorem 2.3.5. *No scoring rule is hyper-stable for* $\mathbb{B}$.

2.4 CONDORCET SOCIAL WELFARE FUNCTIONS

We turn now to the analysis of Condorcet SWFs. We start with some additional notations and definitions. Given a profile $P_N \in \mathcal{L}(A_m)^n$, where n is odd, the majority tournament for P_N is the complete and asymmetric binary relation $\mu(P_N)$ defined over $A_m \times A_m$ by: $\forall (x, y) \in A_m \times A_m$, $x \; \mu(P_N) \; y \Leftrightarrow |\{i \in N : xP_iy\}| > |\{i \in N : yP_ix\}|$. A SWF α is *Condorcet* if, for any $m \in \mathbb{N}$, for any $n \in 2\mathbb{N}+1$, and for any profile $P_N \in \mathcal{L}(A_m)^n$, we have $\alpha(P_N) = \mu(P_N)$ if $\mu(P_N) \in \mathcal{L}(A_m)$.

We prove below the existence of a neutral Condorcet SWF that is hyper-stable for $\mathbb{B}$. Beforehand, we show that three well-known neutral Condorcet SWFs violate Kemeny stability. The *Copeland solution* is the SWF φ defined by: $\forall m \in \mathbb{N}$, $\forall n \in 2\mathbb{N}+1$, $\forall \; P_N \in \mathcal{L}(A_m)^n$, $\forall x, y \in A_m$, $x \; \varphi(P_N) \; y \Leftrightarrow c(x, P_N) \geq c(y, P_N)$, where $c(x, P_N) = |\{z \in A_m : x \; \mu(P_N) \; z\}|$. Consider the following profile P_N together with the linear extension $\dot{P}_N$ of P_N^K:

13 A scoring rule α is convex if, for any $m \in \mathbb{N}$, the score vector $S_\alpha^m = (s_\alpha^{1,m}, ..., s_\alpha^{m,m})$ is such that $(s_\alpha^{1,m} - s_\alpha^{2,m}) \geq (s_\alpha^{2,m} - s_\alpha^{3,m}) \geq ... \geq (s_\alpha^{m-1,m} - s_\alpha^{m,m})$.

$$P_N = \left(\begin{array}{ccccc} 1 & 1 & 1 & 1 & 1 \\ \hline a & a & b & b & c \\ b & c & c & a & a \\ c & b & a & c & b \end{array}\right) \quad \dot{P}_N = \left(\begin{array}{ccccc} 1 & 1 & 1 & 1 & 1 \\ \hline abc & acb & bca & bac & cab \\ acb & cab & bac & bca & acb \\ bac & abc & cba & abc & cba \\ cab & cba & cab & acb & bca \\ bca & bac & abc & cba & abc \\ cba & bca & acb & cab & bac \end{array}\right)$$

Then, we have $\varphi(P_N) = abc$, while $c(abc, \dot{P}_N) = 3 < c(acb, \dot{P}_N) = 4$ implies that $\Delta(\varphi(P_N)) \cap f_\varphi(\dot{P}_N) = \emptyset$. Thus, φ is not Kemeny-stable.

The *Slater solution* is the social welfare correspondence[14] β defined by: $\forall m \in \mathbb{N}$, $\forall n \in 2\mathbb{N}+1$, $\forall\ P_N \in \mathcal{L}(A_m)^n$, $\forall P \in \mathcal{L}(A_m)$, $\beta(P_N) = ArgMin_{P \in \mathcal{L}(A_m)} d_K(P, \mu(P_N))$. A SWF α is *Slater-consistent* if, at any profile P_N, it always selects one linear order in $\beta(P_N)$. Consider the following profile $P_N \in \mathcal{L}(A_8)^5$:

$$P_N = \left(\begin{array}{ccccc} 1 & 1 & 1 & 1 & 1 \\ \hline b & a & d & c & d \\ c & b & a & a & b \\ d & c & b & d & c \\ a & d & c & b & a \\ a' & b' & d' & d' & c' \\ b' & c' & a' & b' & a' \\ c' & d' & b' & c' & d' \\ d' & a' & c' & a' & b' \end{array}\right)$$

Define $X = \{a, b, c, d\}$ and $Y = \{a', b', c', d'\}$ and consider the restrictions $P_N|_X$ and $P_N|_Y$ of P_N to X and Y respectively. We have that $\mu(P_N|_X)$ and $\mu(P_N|_Y)$ are isomorphic. Moreover, we observe that (1) $a\ \mu(P_N)\ b\ \mu(P_N)\ c\ \mu(P_N)\ d\ \mu(P_N)\ a$, (2) $c\ \mu(P_N)\ a$, (3) $d\ \mu(P_N)\ b$, and (4) $\forall (x,y) \in X \times Y$, $x\ \mu(P_N)\ y$. This ensures that $\beta(P_N|_X) = \{cdab\}$ and $\beta(P_N|_Y) = \{c'd'a'b'\}$. Thus, $\beta(P_N) = \{cdabc'd'a'b'\}$. Now, consider $Q = dbcad'b'c'a'$. The next table gives the Kemeny distances between each of the 5 linear orders in $P = (P_1, ..., P_5)$ and respectively, $\beta(P_N)$ and Q:

14 A *social welfare correspondence* is a mapping δ from $\underset{n,m\in\mathbb{N}}{\cup} \mathcal{L}(A_m)^n$ to $\underset{m\in\mathbb{N}}{\cup} 2^{\mathcal{R}(A_m)} \backslash \emptyset$ such that, for any $n, m \in \mathbb{N}$, for any $P_N \in \mathcal{L}(A_m)^n$, $\delta(P_N) \in 2^{\mathcal{R}(A_m)} \backslash \emptyset$, where $2^{\mathcal{R}(A_m)} \backslash \emptyset$ is the set of all non-empty subsets of weak orders over A_m.

P_i	$\beta(P_N)$	Q
P_1	$3+4$	$2+5$
P_2	$4+3$	$5+2$
P_3	$3+3$	$2+2$
P_4	$1+3$	$4+0$
P_5	$3+1$	$0+4$

.

It follows that in the Kemeny weak profile P_N^K, Q is strictly preferred to $\beta(P_N)$ by individual 3, while all other individuals are indifferent. Hence, there exists a linear extension $\dot{P}_N$ of P_N^K where Q is unanimously preferred to $\beta(P_N)$. Since the Slater solution always selects Pareto-optimal outcomes, and since $\beta(P_N)$ is a singleton, we conclude that no Slater-consistent SWF is Kemeny-stable.

The *Kemeny rule* is the Condorcet social welfare correspondence ω defined by: $\forall\ P_N = (P_1, ..., P_n) \in \mathcal{L}(A_m)^n$, $\forall P \in \mathcal{L}(A_m)$, $\omega(P_N) = ArgMin_{P \in \mathcal{L}(A_m)} \sum_{i \in N} d_K(P, P_i)$. A SWF α is *Kemeny-consistent* if, for any profile P_N, it always selects a linear order in $\omega(P_N)$. Consider the following profile $P_N \in \mathcal{L}(A_3)^9$ together with the linear extension $\dot{P}_N$ of P_N^K:

$$P_N = \begin{pmatrix} 2 & 3 & 4 \\ \hline b & c & a \\ c & a & b \\ a & b & c \end{pmatrix} \quad \dot{P}_N = \begin{pmatrix} 2 & 3 & 4 \\ \hline bca & cab & abc \\ cba & cba & acb \\ bac & acb & bac \\ cab & bca & cab \\ abc & abc & bca \\ acb & bac & cba \end{pmatrix}$$

The reader will check that $\omega(P_N) = \{abc\}$, whereas $\omega(\dot{P}_N) = \{(cab)(abc)(acb)(bca)(cba)(bac)\}$ which leads to $f_\omega(\dot{P}_N) = \{cab\}$. Hence, there is no Kemeny-stable and Kemeny-consistent SWF.

We now establish the existence of a Condorcet and unanimous SWF which is hyper-stable for $\mathbb{B}$. The *transitive closure* $\theta(P_N)$ of $\mu(P_N)$ is defined by: $\forall x, y \in A_m$, $x\ \theta(P_N)\ y$ if and only if there exist $x_1, x_2, ..., x_H \in A_m$ such that $x\ \mu(P_N)\ x_1$, $x_1\ \mu(P_N)\ x_2$, ... , $x_H\ \mu(P_N)$ y. Consider the SWF θ, which maps every profile $P_N \in \cup_{m,n} \mathcal{L}(A_m)^n$ (where n is odd) to the transitive closure $\theta(P_N)$ of $\mu(P_N)$. It is easily checked that θ is unanimous.

Theorem 2.4.1. *θ is hyper-stable for $\mathbb{B}$.*

Note that there are other neutral Condorcet SWFs that are hyper-stable for $\mathbb{B}$. Indeed, define the SWF ψ by: $\forall m, n \in \mathbb{N}$, $\forall P_N \in \mathcal{L}(A_m)^n$, $\psi(P_N) = \mu(P_N)$ if $\mu(P_N) \in \mathcal{L}(A_m)$, and otherwise, $a\ \psi(P_N)\ b$ and $b\ \psi(P_N)\ a$ for all $a, b \in A_m$. Then ψ is hyper-stable for $\mathbb{B}$. This is an immediate corollary of the Proposition 2 below. Given any $P_N \in \mathcal{L}(A_m)^n$, the *Condorcet winner* of P_N is the element $CW(P_N) \in A_m$ such that $CW(P_N)\ \mu(P_N)\ a$ for all $a \in A_m / CW(P_N)$.

Proposition 2.4.1. *Let $P_N \in \mathcal{L}(A_m)^n$ be such that $\mu(P_N) \in \mathcal{L}(A_m)$. For any $E \in \mathbb{B}^n$, either $CW(P_N^E)$ does not exist, or $CW(P_N^E) = \mu(P_N)$.*

2.5 DISCUSSION

Our main result is that no unanimous scoring rule is Kemeny-stable, hence hyper-stable for the larger domain $\mathbb{B}$ of betweenness-consistent preference extensions. However, the transitive closure of the majority relation is a unanimous Condorcet SWF which is hyper-stable for $\mathbb{B}$.

Hyper-stability does not draw a clear border between scoring rules and Condorcet SWFs. Indeed, the Kemeny SWF and several other Condorcet SWFs based on well-known tournament solutions are not Kemeny-stable. Characterizing the class of Condorcet SWFs which are hyper-stable for $\mathbb{B}$ is an open question worth being addressed. Another open problem is studying hyper-stability for non-unanimous scoring rules.

Further open questions relate to alternative concepts of hyper-stability. Consider the following property. An SWF α is *Condorcet hyper-stable* if $\forall n, m \in \mathbb{N}$, $\forall P_N \in \mathcal{L}(A_m)^n$, $\forall E \in \mathbb{B}^n$, $\alpha(P_N) \in \mathcal{L}(A_m) \Rightarrow [\alpha(P_N) = CW(P_N^E)]$. Then no Condorcet SWF is Condorcet hyper-stable. To see why, consider the $P_N \in \mathcal{L}(A_3)^5$ and its Kemeny hyper-profile $\dot{P}_N \in \Delta(P_N^K)$ shown below:

$$P_N = \begin{pmatrix} 1 & 1 & 1 \\ \hline a & a & b \\ b & c & c \\ c & b & a \end{pmatrix} \quad \dot{P}_N = \begin{pmatrix} 1 & 1 & 1 \\ \hline abc & acb & bca \\ bac & cab & cba \\ acb & abc & bac \\ bca & cba & cab \\ cab & bac & abc \\ cba & bca & acb \end{pmatrix}$$

Since $\mu(P_N) = abc$, then $\alpha(P_N) = abc$ for any Condorcet α. However, abc is defeated in $\mu(\dot{P}_N)$ by cab, hence the result. A natural question is whether any Condorcet SWF α satisfies the following weaker version of Condorcet hyper-stability: $\forall n, m \in \mathbb{N}$, $\forall P_N \in \mathcal{L}(A_m)^n$ such that $\alpha(P_N) \in \mathcal{L}(A_m)$, there exists $E \in \mathbb{B}^n$ for which $\alpha(P_N) = CW(P_N^E)$.

Remark that, in the Kemeny hyper-profile $\dot{P}_N$ above, all three individual preferences are extended through the same linear extension of the Kemeny weak order. This common linear extension can be defined as a linear order over the permutations of the set $\{1,2,3\}$ of ranks. Indeed, given two orders P and $Q = (a_1 a_2 ... a_m)$ in $\mathcal{L}(A_m)$, define $r_P(Q) = (r_P(a_1), ..., r_P(a_m))$ by $\forall h = 1, ..., m$, $r_P(a_h) = |\{b \in A_m : bPa_h\}| + 1$, that is, the rank given to a_h in P. Moreover, given $P_N = (P_1, ..., P_n) \in \mathcal{L}(A_m)^n$, we say that the hyper-profile $P_N^E = (e_1(P_1), ..., e_n(P_n))$ is *uniform* if there exists a linear order $\succ$ over the permutations of $\{1, ..., m\}$ such that, for any $i = 1, ..., n$, for any $Q, Q' \in \mathcal{L}(A_m)$, $[Q\ e^i(P_i)\ Q' \Leftrightarrow r_{P_i}(Q) \succ r_{P_i}(Q')]$. In the example above, $\succ$ is defined by: $(123) \succ (213) \succ (132) \succ (231) \succ (312) \succ (321)$. We say that an SWF α is *uniformly hyper-stable* for $\mathbb{B}$ if $\forall n, m \in \mathbb{N}$, $\forall P_N \in \mathcal{L}(A_m)^n$, $\Delta(\alpha(P_N)) \cap f_\alpha(P_N^E) \neq \emptyset$ for all uniform hyper-profiles P_N^E with $E \in \mathbb{B}^n$.

As a first step towards a complete study of uniform hyper-stability, we remark that the Borda rule $\mathcal{B}$ is not uniformly hyper-stable. Indeed, consider the following profile P_N, together with the Kemeny hyper-profile $\dot{P}_N \in \Delta(P_N^K)$:

$$P_N = \begin{pmatrix} 1 & 1 & 2 \\ \hline a & b & c \\ b & a & b \\ c & c & a \end{pmatrix} \quad \dot{P}_N = \begin{pmatrix} 1 & 1 & 2 \\ \hline abc & bac & cba \\ acb & bca & cab \\ bac & abc & bca \\ cab & cba & acb \\ bca & acb & bac \\ cba & cab & abc \end{pmatrix}$$

We get $\mathcal{B}(P_N) = bca$. Moreover, $\dot{P}_N$ is uniform (to see why, check that $\dot{P}_N$ is built in accordance with $(123) \succ (132) \succ (213) \succ (312) \succ (231) \succ (321)$). Finally, $S^6_{\mathcal{B}}(bca, \dot{P}_N) = 11 < S^6_{\mathcal{B}}(cba, \dot{P}_N) = 12$.

Finally, an alternative route worth being followed is to characterize the set of preference extensions for which a given SWF, or a given class of SWFs (e.g. scoring rules, Condorcet SWFs) is hyper-stable.

Part III

POLARIZATION

3

AN ADAPTATION OF ESTEBAN-RAY POLARIZATION MEASURE TO SOCIAL CHOICE

3.1 INTRODUCTION AND THE MODEL

Since polarization is understood to be closely related to social discord and induction of enmity, it is of interest how to understand and conceptualize polarization within a society with ordinal preferences.[1] One approach, not specifically preference based, suggests that it should be considered as the aggregation of pairwise antagonisms within society. In this note, following Esteban and Ray (1994), we suggest that antagonism felt by an individual towards another depends on the alienation between the two and also on the feeling of identification he/she enjoys by being in wherever he/she is. Alienation then can be thought of as a function of the dissimilarity between the two stances. We further postulate that identification can be seen as a function of the support of one's opinion. In a formal way, this would mean to measure polarization in a society with a function as follows.

$$\sum_j \sum_i a(I(m_i), A(d(o_i, o_j)))$$

Here the function $I(m_i)$ represents how much i is identified with his/her position, as a function of m_i, the share of the population who thinks the same with i. The function $A(d(o_i, o_j))$ represents the alienation as an increasing function of the distance of opinions of i and j. Finally, the function $a(I, A)$ represents the antagonism felt by i towards j, as a function of identification i enjoys by being wherever i is and alienation between i and j. Summing up for all directed pairs is then a measure of polarization as discussed. We construe the set of properties in Esteban and Ray (1994) and show that a subclass of the

1 See Gurer (2008) where Arrovian impossibilities are explored when preferences in a society cluster around one preference. It is furthermore conjectured that bipolar cases would resemble unipolar cases in terms of effects on impossibilities.

functions above satisfies those properties for ordinal preferences over three alternatives. To our knowledge, this is the first attempt at this question.[2]

Let A be the set of alternatives with $|A| = m$. There are $m!$ possible linear orders (or preferences) and the set of all those is denoted $\mathcal{L}(A)$. Let $\bar{P} = (P^1, P^2, ..., P^{m!})$ be an arbitrary listing of the set $\mathcal{L}(A)$. A preference profile for a society of N individuals is an element of $\mathcal{L}(A)^N$. Given $\bar{P}$ and $P_N \in \mathcal{L}(A)^N$ while letting $N_i^{\bar{P}} = |\{j \in N : P_j = P^i\}|$ we have a representing preference distribution $(\pi, \bar{P})$ where $\pi = (\pi_1, \pi_2, ..., \pi_{m!})$ is such that $\pi_i = N_i^{\bar{P}}/N$. Hence we have that $\sum_{i=1}^{m!} \pi_i = 1$. Finally, a polarization measure is a mapping $\mathcal{P} : \mathcal{D} \to \mathbb{R}_+$ where $\mathcal{D}$ denotes the set of all preference distributions $(\pi, \bar{P})$.

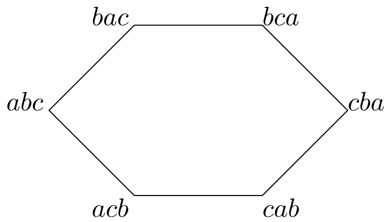


Figure 3.1.1. Kemeny graph, three alternatives.

In the following analysis, we will be employing a particular metric on preferences. Given any $P, P' \in \mathcal{L}(A)$, the Kemeny distance between P and P' is defined as $d_\kappa(P, P') = |P \backslash P'|$ where $P \backslash P'$ is the symmetric difference.[3] The graph for three alternatives is depicted in Figure 3.1.1.

Now we introduce a set of properties on preferential polarization measures following the axiomatization in Esteban and Ray (1994).[4]

Definition 3.1.1 (Property 1). *Let $\pi_i > \pi_j = \pi_k$ be the only masses of a profile such that π_j and π_k are at least as close to each other as they are to π_i, or $d_\kappa(P^j, P^k) \leq min\{d_\kappa(P^j, P^i), d_\kappa(P^k, P^i)\}$. A polarization measure is said to satisfy* Property 1 *if joining the two smaller masses at a point at least as further as the average distance to π_i increases polarization.*

Figure 3.1.2 depicts, for illustration, two situations *Property 1* apply to for the case of three alternatives. The two small masses on the right can be joined in a middle

2 Baldiga and Green (2013) analyze social choice functions in terms of assent maximizing attributes. Alcalde-Unzu and Vorsatz (2008) provide axiomatic characterizations of some consensus measures. García-Lapresta and Pérez-Román (2011) analyze properties of a class of consensus measures that are based on the distances among individual weak orders.

3 This widely known and used metric is due to Kemeny (1959) for which Bogart (1973) and Kemeny and Snell (1962), *inter alia*, provide axiomatic characterizations. Can and Storcken (2013) provides a refinement of previous characterizations. For a general treat of weighted distances between preferences, see Can (2012).

4 In fact, *Property i* here may be matched with *Axiom i* in Esteban and Ray (1994), although we keep in mind that there is no unique translation of each axiom into the current setting.

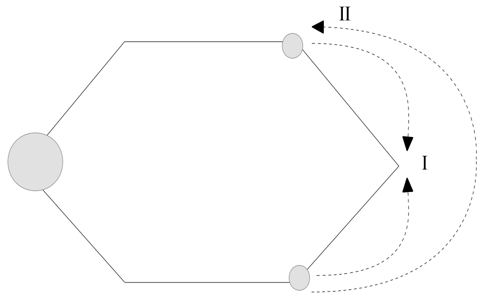


Figure 3.1.2.. Property 1.

way further away than the larger mass (as in the move 1 in the figure), or they can be imbricated at one of the positions with small densities (as in the move 2 in the figure). These types of moves do not decrease (and in case of the move 2, do not increase) the average distance, hence alienation, but increase within group support.

Let $-P^i$ denote the order where every pairwise comparison in the order P^i is reversed and π_{-i} denote the proportion of the society with the preference $-P^i$.

Definition 3.1.2 (Property 2). *A polarization measure satisfies* Property 2 *if moving a mass that is opposed mildly towards only smaller masses increases polarization. Formally, let the mass at P^i move to $P^{i'}$. If $d_\kappa(P^i, P^k) > d_\kappa(P^{i'}, P^k)$ and $\pi_k < \pi_{-k}$, then there exists $\epsilon > 0$ such that $\pi_{-i} < \epsilon$ implies that this move should increase polarization.*

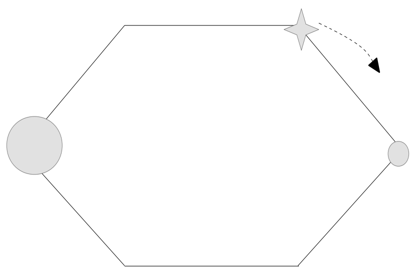


Figure 3.1.3.. Property 2.

Property 2 basically requires that polarization increases whenever an in between mass is moved towards the side which has a smaller support. This idea is explained in Figure

3.1.3. Required is that the move should be towards only smaller poles and is not carried by a strongly opposed mass.

Definition 3.1.3 (Property 3). *A polarization measure satisfies* Property 3 *if for any profile of preferences with three consecutive masses, dissolution of the middle mass equally into two sides increases polarization.*

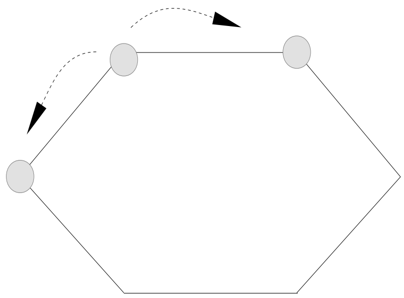


Figure 3.1.4.. Property 3.

Property 3 is an exact counterpart of Axiom 3 in Esteban and Ray (1994) and the intuition is clear. The following homotheticity property requires that the size of the society does not matter. Let λP_N denote a profile where $N_i^{\lambda \bar{P}} = \lambda N_i^{\bar{P}}$ for all $i \leq m!$.

Definition 3.1.4 (Homotheticity). *Let P_N and $P'_{\tilde{N}}$ be two profiles with possibly distinct sizes; N and $\tilde{N}$. A polarization measure $\mathcal{P}$ satisfies* homotheticity *if $\mathcal{P}(P_N) \geq \mathcal{P}(P'_{\tilde{N}})$ implies $\mathcal{P}(\lambda P_N) \geq \mathcal{P}(\lambda P'_{\tilde{N}})$ for all $\lambda > 0$.*

Finally, *bipolarity*, as a regularity property, requires that full concentration at one order is the least polarized case and that bipolar society where one half is at the exact opposite of the other is the most polarized.

Definition 3.1.5 (Bipolarity). *Suppose P_N is such that $\pi_Q = \frac{1}{2} = \pi_{-Q}$ for some $Q \in \mathcal{L}(A)$ and that P'_N is such that $\pi'_{Q'} = 1$ for some $Q' \in \mathcal{L}(A)$. A polarization measure $\mathcal{P}$ is said to satisfy* bipolarity *if $\mathcal{P}(P_N) \geq \mathcal{P}(\dot{P}_N) \geq \mathcal{P}(P'_N)$ for all $N \in \mathbb{N}$ and $\dot{P}_N \in \mathcal{L}(A)^N$.*

3.2 RESULTS

Proposition 3.2.1. *The polarization measure defined as*

$$\mathcal{P}^*_\alpha(\pi, \bar{P}) = \sum_{j=1}^{m!}\sum_{i=1}^{m!} \pi_i^{1+\alpha}\pi_j(d_{\mathcal{K}}(P^i, P^j))$$

satisfies properties 1-3 for any $\alpha \in (0, \alpha^*)$ with $\alpha \simeq 1.6$ and $m = 3$.

Note that this function reduces to Gini Index with Kemeny distance when $\alpha = 0$ and the higher α is the higher the sensitivity to identification. In what follows, let's fix $\bar{P} = ((abc) = P^1, (bac) = P^2, (acb) = P^3, (bca) = P^4, (cab) = P^5, (cba) = P^6)$ and denote the initial profile by P_N and the profile after a described move by P'_N.

Proof. Property 1: Without loss of generality, three cases may apply.
(i) $\pi_1 > \pi_4 = \pi_6$, and move $\pi_4 \to P^6$: We have $\mathcal{P}(P'_N) - \mathcal{P}(P_N) = \pi_4(\pi_1^{1+\alpha} - \pi_4^{1+\alpha}) + \pi_4^{1+\alpha}(6\pi_1(2^\alpha - 1))$.
(ii) $\pi_1 > \pi_2 = \pi_4$, and move $\pi_2 \to P^6$: We have $\mathcal{P}(P'_N) - \mathcal{P}(P_N) = \pi_4(\pi_1^{1+\alpha} - \pi_4^{1+\alpha}) + \pi_4^{1+\alpha}(4\pi_1(2^\alpha - 1))$.
(iii) $\pi_1 > \pi_4 = \pi_5$, and move (iii.a) $\pi_4 \to P^5$ or (iii.b) $\{\pi_4, \pi_5\} \to P^6$: (iii.a) induces that $\mathcal{P}(P'_N) - \mathcal{P}(P_N) = 4\pi_1\pi_4^{1+\alpha}(2^\alpha - 1)$ whereas (iii.b) induces $\mathcal{P}(P'_N) - \mathcal{P}(P_N) = 2\pi_4\pi_1^{1+\alpha} + 2\pi_1\pi_4^{1+\alpha}(3 \cdot 2^\alpha - 2)$. In all cases, we have $\mathcal{P}(P'_N) > \mathcal{P}(P_N)$ if $\alpha > 0$.
Property 2: Let π_4 move to P^6 to induce P'_N. We have $\mathcal{P}(P'_N) - \mathcal{P}(P_N) > \pi_4(\pi_1^{1+\alpha} + \pi_2^{1+\alpha} - \pi_3^{1+\alpha} - \pi_5^{1+\alpha} - \pi_6^{1+\alpha}) + \pi_4^{1+\alpha}(\pi_1 + \pi_2 - \pi_3 - \pi_5 - \pi_6)$ which is positive if the move is only towards smaller poles, or formally both $\pi_1 > \pi_6$ and $\pi_2 > \pi_5$, for any support π_3 of opposition to P^4 smaller than ϵ^* where $\epsilon^* = min\{\pi_1 + \pi_2 - \pi_5 - \pi_6, (\pi_1^{1+\alpha} + \pi_2^{1+\alpha} - \pi_5^{1+\alpha} - \pi_6^{1+\alpha})^{\frac{1}{1+\alpha}}\}$ whenever $\alpha > 0$.
Property 3: The question reduces to that of the Axiom 3 in Esteban and Ray (1994), and the proof that α is bounded above approximately by 1.6, which pertains to the established (in the same text) fact that there exists $\alpha^* > 0$ such that $max_{z \geq 0}[(1+\alpha)(z - \frac{z^\alpha}{2} - z^{1+\alpha}) - \frac{1}{2}] < 0$ if and only if $\alpha < \alpha^*$, can be found in the last paragraph of the proof of the Theorem 1 in that paper, p. 837. □

Furthermore, it is quick to observe that the measure $\mathcal{P}^*$ is *anonymous* and *neutral* in the sense that it treats individuals and preferences equally. A polarization measure $\mathcal{P}$ is said to be anonymous if for any profile P_N, for any permutation $\sigma : N \to N$ of individuals, we have that $\mathcal{P}(P_N) = \mathcal{P}(P_{\sigma N})$, where $P_{\sigma N} = (P_{\sigma(i)})_{i \in N}$. Similarly, a polarization measure $\mathcal{P}$

is said to be neutral if for any profile P_N, for any permutation $\delta : A \to A$ of alternatives, we have that $\mathcal{P}(P_N) = \mathcal{P}(\delta P_N)$ where $\delta P_N = (\delta P_i)_{i \in N}$ with $aP_ib \iff \delta(a)\delta P_i \delta(b)$.

Proposition 3.2.2. *The polarization measure $\mathcal{P}^*_\alpha$ is anonymous, neutral and homothetic for any $\alpha \geq 0$ and $m \in \mathbb{N}$.*

Proof. Take any $N \in \mathbb{N}$ and let $\sigma : N \to N$ be an arbitrary permutation. Since $N_i^P = N_i^{\sigma P}$, we have $\pi_i = \pi'_i$ for all $i \in \mathbb{N}_{m!}$ where $\pi'_i = N_i^{\sigma P}/N$, which demonstrates anonymity. For neutrality it suffices to observe that $d_\kappa(P, P') = d_\kappa(\delta P, \delta P')$ for any $P, P' \in \mathcal{L}(A)$, under any permutation $\delta : A \to A$.[5] The homotheticity follows the fact that the summation is taken over the supports as percentages. □

Finally, if we take α as 1, which reflects a simple antagonism function where alienation is represented as the distance and identification as just the relative societal support of one's preference[6], the measure satisfies the *bipolarity* condition. Let us denote by # *support profile* a profile in the support of which there are only # different preference orderings.

Proposition 3.2.3. *The measure $\mathcal{P}^*_1$ satisfies bipolarity, for $m = 3$.*

Proof. The measure is zero if we have full concentration at one order, and strictly positive in all other cases. For any two support profile, increasing distance in between increases polarization clearly. Once they are at exact opposites, making them equal in density will increase the measure.

Lemma 3.2.1. *The bipolar case is more polarized than any three support profile under $\mathcal{P}^*_1$.*

Proof. Suppose first that the three masses are not equidistant to each other. Two cases are possible. (*Case i*) *One is further away from the two.* Let, without loss of generality, the support be $\{\pi_1, \pi_2, \pi_5\}$. If $\pi_5 > \pi_2$, moving π_1 to P^2 increases the value if $\pi_5^2\pi_1 + \pi_1^2(\pi_5 - \pi_2) + 6\pi_1\pi_2\pi_5 > \pi_2^2\pi_1$ which is true. If $\pi_2 > \pi_5$, moving π_1 to P^5 increases the value if $\pi_2^2\pi_1 + \pi_1^2(\pi_2 - \pi_5) + 3\pi_1\pi_2\pi_5 > \pi_5^2\pi_1$ which is true. (*Case ii*) *One is at unit distance to both.* Let, without loss of generality, the support be $\{\pi_1, \pi_2, \pi_3\}$ and that $\pi_2 > \pi_3$. Moving π_1 to P^3 increases the value if $\pi_2^2\pi_1 + \pi_1^2(\pi_2 - \pi_3) + 4\pi_1\pi_2\pi_3 > \pi_3^2\pi_1$ which is true.

Now suppose the three masses are equidistant to each other. Let, without loss of generality, the support be $\{\pi_1, \pi_4, \pi_5\}$ and we move π_1 to P^5. For this to increase polarization

5 A characterization of Kemeny distance incorporating neutrality can be found in Bogart (1973).
6 Formally, $T(I, a) = a \cdot I$ where $I(\pi_i) = \pi_i$ and $a(d(P^i, P^j)) = d_{\mathcal{K}}(P^i, P^j)$.

it is enough to have π_4 to be the greatest of the three. If they are all equal size $\frac{1}{3}$, we have $\mathcal{P}_1^*(P_N) = 3(\frac{1}{3})^2(2\frac{1}{3}+2\frac{1}{3}) = \frac{4}{9} < 2(\frac{1}{2})^2(3\frac{1}{2}) = \frac{3}{4}$ where the latter is of the bipolar case. □

The following lemma shows that from any five or six support profile, we can reduce the domain to four masses and increase polarization.

Lemma 3.2.2. *For any six (five) support profile, there exists a move that reduces the support of the profile to five (four) and increases the value of P_1^*.*

Proof. Let masses at each (or all but one) location be positive. Without loss of generality, either (*i*) $\pi_1 > \pi_4 + \pi_5 + \pi_6$ or (*ii*) $\pi_5 > \pi_4 + \pi_2 + \pi_1$ is true. Suppose the first is true and move the mass π_2 to P^4.[7] The difference $\mathcal{P}(P'_N) - \mathcal{P}(P_N) = \pi_2[\pi_1^2 + \pi_3^2 - (\pi_4^2 + \pi_5^2 + \pi_6^2)] + \pi_2^2[\pi_1 + \pi_3 - (\pi_4 + \pi_5 + \pi_6)]$ is clearly positive. This is regardless of π_3 being non-zero or not (in other words, the support being five or six masses) hence the lemma demonstrated. □

Corollary 1. *For any six support profile there exists a consecutive pair of moves that induces a four support profile with higher value under $\mathcal{P}_1^*$.*[8]

The final lemma below concludes the proof of the proposition by showing that from any (out of three possible) four support profile we can reduce to a two support profile with higher measure.

Lemma 3.2.3. *The bipolar profile is more polarized than any P_N with four masses under $\mathcal{P}_1^*$.*

Proof. Let $\Sigma \subseteq \mathcal{L}(A)$ be the orders that have nonzero support in P_N. Without loss of generality, we have three cases.

(a) $\Sigma = \{P^1, P^3, P^4, P^6\}$. We have three distinct cases; *either (i)* $\pi_4 \geq \pi_3$ and $\pi_1 \geq \pi_6$ *, or (ii)* $\pi_4 < \pi_3$ and $\pi_1 < \pi_6$ *, or (iii)* $\pi_4 > \pi_3$ and $\pi_1 < \pi_6$ (not distinctively $\pi_4 < \pi_3$ and $\pi_1 > \pi_6$ is also possible.). For the first two cases, moving π_4 to P^6 and π_3 to P^1 increases polarization if $(\pi_4 - \pi_3)(\pi_1 - \pi_6) \geq 0$, which is true. For the third case, moving π_1 to P^4 and π_6 to P^3 increases polarization if $(\pi_4 - \pi_3)(\pi_6 - \pi_1) \geq 0$, which is true.

(b) $\Sigma = \{P^1, P^4, P^5, P^6\}$. Moving π_6 to P^1 induces $\mathcal{P}(P'_N) - \mathcal{P}(P_N) > 3\pi_1\pi_6(\pi_1 - \pi_6) > 0$ if $\pi_1 \geq \pi_6$. If otherwise, moving π_4 to P^2 and π_5 to P^3 induces $\mathcal{P}(P'_N) - \mathcal{P}(P_N) = (\pi_4 + \pi_5)(\pi_6^2 - \pi_5^2) + (\pi_4^2 + \pi_5^2)(\pi_6 - \pi_1) > 0$ which leaves us with the exact situation as before and hence moving π_1 to P^6 increases polarization.

7 If the second is true, moving π_6 to P^4 will do the same.
8 Proof left to the reader, since it only incorporates application of the move in the Lemma 2.4 consecutively.

(c) $\Sigma = \{P^1, P^2, P^4, P^6\}$. Moving π_2 to P^3 induces $\mathcal{P}(P'_N) - \mathcal{P}(P_N) = 2\pi_2\pi_4(\pi_2 + \pi_4)$ which is positive and leaves us with the case (a). $\square$

Thus established is the proof of sthe proposition. $\square$

4

MEASURING POLARIZATION IN PREFERENCES

(joint work with Burak Can and Ton Storcken)

4.1 INTRODUCTION

Higher polarization in ideologies or preferences over policies is generally considered as a bad feature in politics mainly due to representational concerns. It is argued to cause policy gridlock (Jones (2001)), decrease turnout if it is only in elite level (Hetherington (2008)) and increase economic inequality (McCarty et al. (2003)).

Due to the disagreements in measurement, we see disparity in the results of polarization analyses. For instance, the increase in polarization in the U.S. politics is somewhat unequivocal for the elite level although the literature on public polarization is inconclusive. For a review in line with this conclusion, see Hetherington (2009). This paper introduces yet another approach to the measurement of polarization. However, the major component of our contribution is in that of the subject of measurement. Although there are quite a number of articles analyzing the measurement of polarization for distributions that can be represented on a line[1], this paper is among the very first attempts for analyzing polarization measures for ordinal preference profiles.[2]

Some of the related concepts that could be found analyzed in the social choice literature could be listed as *consensus* (Herrera-Viedma et al., 2011), *assent* (Baldiga and Green, 2013) and *cohesiveness* (Alcalde-Unzu and Vorsatz, 2013). According to numerous authors, as formulated in Bosch (2006), consensus can be formulated such that it can be measured with mappings that assign to any profile of preferences a value in unit interval, which has the following two properties necessarily: first, the value given to a

1 See, *inter alia*, Esteban and Ray (1994) and Montalvo and Reynal-Querol (2005).
2 For a measure of ordinal preference polarization which adopts the methodology of Esteban and Ray (1994) with the use of a metric *à la* Kemeny (1959), see previous chapter.

profile is highest, namely 1, if and only if all individuals agree on how to rank alternatives and second, the same value given to any two profiles if the only difference in between them is the names of either the alternatives or individuals. Alcalde-Unzu and Vorsatz (2008) have introduced some axiomatic characterizations in this vein. García-Lapresta and Pérez-Román (2011) analyze properties of a class of consensus measures that are based on the distances among individual weak orders.

Baldiga and Green (2013) define conflict between two individuals as the disagreement in their top choices. They then use an aggregate-assent maximizing approach to the selection of the choice rule, where the assent between preferences is the probability that these preferences would be conflictual on a random feasible set.

Finally, Alcalde-Unzu and Vorsatz (2013) denote the level of similarity among preferences in a profile as cohesiveness and characterize a class of cohesiveness measures with a set of plausible axioms. This class of functions falls within the above definition of consensus.

In what follows we argue, first, that polarization is not necessarily the opposite of consensus and hence calls for a particular treatment. The least polarized case naturally coincides with a fully consensual state, which is easily defined as a unanimous preference profile. However, there is no unique way of framing the most polarized situation. This would entail a normative approach, which we embrace in this paper as follows. Since we investigate polarization in preferences that are represented as linear orders, we restrict the most polarized situations to societies which are divided equally into two completely opposite linear orders.

Second, we impose that the polarization level should not depend on the number of individuals in a society but stay the same if the supporting individuals of each preference is multiplied by equal terms. Furthermore, we require a form of equal treatment of marginal changes in the composition of preferences. More precisely, if a single individual changes her preference to conform with the majority view on a single issue, then the change in polarization should not depend on the size of this majority. Finally, we impose neutrality towards alternatives.

In this paper, we show that interpreting polarization as an aggregation of antagonisms in a society is the only way of measuring polarization with the properties above. In this context antagonisms are taken as disagreements over pairwise comparisons of alternatives.

We proceed as follows. In the next section we introduce basic notations and formal definitions regarding the axiomatic model. Section 3 provides our main results and proofs thereof. We conclude in Section 4 by pointing to a possible direction for further research. We show the logical independence of axioms in the appendix.

4.2 MODEL

4.2.1 *Preliminaries*

Let A be a finite and nonempty set of m alternatives. For any finite and nonempty set of individuals N, and for any individual i in N, let $p(i)$ denote the preference of i in terms of a linear order, i.e., a complete, antisymmetric and transitive binary relation on A. Furthermore, p indicates a *profile*, a combination of such individual preferences and $\mathbb{L}$ the set of all preferences on A. So, p is an element of $\mathbb{L}^N$.

We denote by $\bar{A}$ the set of all subsets of A with cardinality 2. For a given profile p in $\mathbb{L}^N$ and different alternatives a and b in A let $n_{ab}(p)$ denote the number of individuals who prefer a to b, i.e., $n_{ab}(p) = \#\{i \in N : (a,b) \in p(i)\}$. Let $d_{ab}(p) = |n_{ab}(p) - n_{ba}(p)|$ denote the absolute difference between the number of voters preferring a to b and those preferring b to a at profile p.

For a preference R, let R^N denote the unanimous profile where all individuals have preference R. Let $-R = \{(y,x) : (x,y) \in R\}$ be the preference where all pairs in R are reversed. If π denotes a permutation on A, then the permuted preference of R is $\pi R = \{(\pi(a), \pi(b)) : (a,b) \in R\}$ which naturally defines the permuted profile πp in a coordinate-wise manner, i.e., $(\pi p)(i) = \pi(p(i))$.

For two profiles p and q of two disjoint sets of individuals, say N_1 and N_2 respectively, let (p,q) denote the profile, say r, such that $r(i) = p(i)$ if i is in N_1 and $r(i) = q(i)$ if i is in N_2. Similarly define $p^2 = (p', p'')$, where $p' \in \mathbb{L}^{N'}$ and $p'' \in \mathbb{L}^{N''}$, to be a replication of p if there are bijections $\sigma' : N \leftrightarrow N'$ and $\sigma'' : N \leftrightarrow N''$ such that $p(i) = p'(\sigma'(i))$ and $p(i) = p''(\sigma''(i))$ for all $i \in N$. This naturally extends to define $p^3, p^4, ..$ accordingly.

Let p and q be two profiles in $\mathbb{L}^N$. We say that p and q form an *elementary change from ab to ba* whenever there is an individual i in N who ranks a and b consecutively in p and furthermore $q(i) = \Big(p(i) \cup \{(b,a)\}\Big) \setminus \{(a,b)\}$ and for all j in $N \setminus \{i\}$, $p(j) = q(j)$. This means that $q(i)$ can be obtained from $p(i)$ by only reversing the ordered pair (a,b).

Finally, a *polarization measure* Ψ assigns to any profile p in $\mathbb{L}^N$ a real number $\Psi(p)$, where N is any finite and nonempty set of individuals. Next we discuss a few normatively appealing conditions on the polarization measures.

4.2.2 *Conditions on Polarization Measures*

We first impose a regularity condition on polarization measures to normalize between 0 and 1. The former value is attained by profiles wherein each individual has the same preferences, i.e., a unanimous profile. In this regard, we see the maximal consensus as a case of minimal polarization. However, we furthermore restrict the maximally polarized case. The profiles (with even number of individuals) where half of the individuals have a preference R and the rest have $-R$, for some $R \in \mathbb{L}$ are considered to be the maximally polarized profiles.

Regularity : $\Psi(p) \in [0,1]$ for all $p \in \mathbb{L}^N$ and furthermore $\Psi(R^N) = 0$ and $\Psi(R^{N_1}, (-R)^{N_2}) = 1$ for all preferences R and all finite and nonempty sets N_1 and N_2 of individuals such that N_1 and N_2 are disjoint and equal in size, i.e., $\#N_1 = \#N_2$.

Neutrality is a standard property in social choice. In this context, it requires that a renaming of the alternatives does not change the polarization level.

Neutrality : $\Psi(p) = \Psi(\pi p)$ for all permutations π on A and all profiles p.

The following condition requires that when societies are replicated by some positive integer, the polarization is unchanged. Note that this also implies *anonymity*, i.e., renaming the individuals does not change the polarization level. Formally:

Replication invariance : $\Psi(p^k) = \Psi(p)$ for all positive integers k, and all profiles p.

Finally, we introduce our final condition which we call *support independence*. This condition requires that selementary changes in favor of an alternative that has a majoritarian support against another lead to identical changes in polarization across profiles. For instance, if a majority of individuals agree that a is better than b in each of the two profiles, then an increase in the support of a over b should lead to the same amount of change in the polarization for both of these profiles.

Support independence : $\Psi(p) - \Psi(q) = \Psi(\widehat{p}) - \Psi(\widehat{q})$ for any two elementary changes $p, q \in \mathbb{L}^N$ and $\widehat{p}, \widehat{q} \in \mathbb{L}^N$ both from ba to ab for some alternatives a and b with $n_{ab}(p) \geq n/2$ and $n_{ab}(\widehat{p}) \geq n/2$.

4.3 RESULT

Assume for simplicity that the issue in hand is a binary choice, that there are only two alternatives. If the absolute difference between numbers of individuals preferring a to b and b to a, i.e., d_{ab}, is 0, then the polarization should intuitively be maximal. If this number is equal to n, then we have that everyone prefers a over b or vice versa, a full agreement. Hence polarization should be minimal. Therefore, the polarization can be related to $n - d_{ab}$. If we normalize by dividing by n, then we have a bound on the polarization (between 0 and 1) therefore regularity is also satisfied. For profiles on more than two alternatives, we iterate this process over all pairs of distinct alternatives. Thereafter we normalize this value with respect to the number of such pairs and the number of individuals. Hence we obtain the following polarization measure:

$$\Psi^*(p) = \sum_{\{a,b\} \in \bar{A}} \frac{n - d_{ab}(p)}{n \cdot \binom{m}{2}}.$$

It is easy to verify that Ψ^* satisfies the conditions introduced in Section 4.2.2. In the sequel, we will show that it is indeed the only measure that satisfies these conditions. Before, we discuss some features regarding elementary changes that are instrumental in what follows.

Let p and q form an elementary change from ab to ba, so that $n_{ab}(p) - 1 = n_{ab}(q)$ and $n_{ba}(p) + 1 = n_{ba}(q)$. This change can be of one of the following three;

(i) a minority decrement if $n_{ab} \leq n/2$,

(ii) a majority decrement if $n_{ab} \geq n/2$ and

(iii) a swing if $n_{ab}(p) > n/2$ and $n_{ab}(q) < n/2$.[3]

The first two changes are straightforward. For the third, consider the case where 4 individuals prefer a to b and 3 prefer b to a. An elementary change, in this case, from ab to ba makes the former minority a majority.

Remark 1. *Note that if p and q form an elementary change from ab to ba that is a minority decrement, then q and p form an elementary change from ba to ab that is a majority decrement. This duality allows us to construct the forthcoming lemmas by focusing on either of the two first elementary changes.*

The following Lemma shows that all minority decrements yield an equal change in polarization regardless of what alternatives are involved. By Remark 1, the result also holds for majority decrements. Let Ψ satisfy the four conditions; regularity, neutrality, replication invariance and support independence.

3 Hence $n_{ab}(p) = n_{ba}(q)$.

Lemma 4.3.1. *Let p and q be a minority elementary change from ab to ba and let $\widehat{p}$ and $\widehat{q}$ be a minority elementary change from xy to yx. We have*

$$\Psi(p) - \Psi(q) = \Psi(\widehat{p}) - \Psi(\widehat{q}).$$

Proof. Let $p^{\{a,b\}}$ be the profile obtained from p by shifting a and b to the two top positions for each individual while leaving preference between a and b as well as those between alternative in $A \backslash \{a,b\}$ unchanged. That is for all individuals i in N let $p^{\{a,b\}}(i) = p(i)|_{\{a,b\}^2} \cup (\{a,b\} \times A \backslash \{a,b\}) \cup\ p(i)|_{(A \backslash \{a,b\})^2}$. Similarly define $q^{\{a,b\}}$. Then by support independence we have

$$\Psi(p) - \Psi(q) = \Psi(p^{\{a,b\}}) - \Psi(q^{\{a,b\}}).$$

Considering the permutation π on A such that $\pi(a) = x$, $\pi(x) = a$, $\pi(b) = y$, $\pi(y) = b$ and $\pi(z) = z$ for all $z \in A \backslash \{a,b,x,y\}$ neutrality implies

$$\Psi(p^{\{a,b\}}) - \Psi(q^{\{a,b\}}) = \Psi(\pi p^{\{a,b\}}) - \Psi(\pi q^{\{a,b\}}).$$

Note that $\pi p^{\{a,b\}}$ and $\pi q^{\{a,b\}}$ are preferences at which the alternatives x and y are in the two top position for every individual. Furthermore they form a minority elementary change from xy to yx. Therefore support independence implies

$$\Psi(\pi p^{\{a,b\}}) - \Psi(\pi q^{\{a,b\}}) = \Psi(\widehat{p}) - \Psi(\widehat{q}).$$

So, all in all

$$\Psi(p) - \Psi(q) = \Psi(\widehat{p}) - \Psi(\widehat{q}).$$

□

Next we prove that all minority elementary changes yield a decrease of polarization by $\alpha = \frac{2}{n \cdot \binom{m}{2}}$. By Remark 1, then, all elementary changes in majority increase polarization by that same amount α.

Lemma 4.3.2. *Let p and q be a minority elementary change from ab to ba. Then*

$$\Psi(p) - \Psi(q) = \frac{2}{n \cdot \binom{m}{2}}.$$

Proof. By Lemma 4.3.1 it is sufficient to prove that at some minority elementary change, polarization decreases by $\alpha = \frac{2}{n \cdot \binom{m}{2}}$. Replication invariance implies that we may assume that the set of individuals is even, that is $n = 2 \cdot k$. Consider any two set of individuals $\#N_1 = \#N_2 = k$ with $N_1 \cap N_2 = \emptyset$ and a combined set of individuals $N = N_1 \cup N_2$. Given any preference R, consider the following two profiles $(R^{N_1}, (-R)^{N_2})$ and R^N. Note that there is a path of $k \cdot \binom{m}{2}$ elementary changes from the former to the latter. By regularity $\Psi(R^{N_1}, (-R)^{N_2}) = 1$ $\Psi(R^N) = 0$. By Lemma 4.3.1, each step in this path cause the same change in polarization, say α. Note that the amount of swaps from $(-R)$ to R is $\binom{m}{2}$. The number of individuals requiring this many swaps is $n/2$. Therefore each elementary change should decrease the polarization by $2/n\binom{m}{2}$. □

We have shown that each minority (or majority) elementary change causes the same amount of decrease (or increase) in the polarization. Next we show that swing elementary changes does not affect the polarization level.

Lemma 4.3.3. *Let p and q be a swing elementary change from ab to ba. Then*

$$\Psi(p) = \Psi(q).$$

Proof. Consider the profiles p^2, (p, q) and q^2. Both p^2 and (p, q) as well as q^2 and (p, q) form minority elementary changes. The former pair from ab to ba the latter pair from ba to ab. So, $\Psi(p^2) - \Psi(p, q) = \alpha = \Psi(q^2) - \Psi(p, q)$. Hence, $\Psi(p^2) = \Psi(q^2)$. Therefore by replication invariance we have $\Psi(p) = \Psi(q)$. □

Now we can state our main theorem.

Theorem 4.3.1. *A polarization measure Ψ satisfies regularity, neutrality, replication invariance and support independence if and only if $\Psi = \Psi^*$.*

Proof. Assume Ψ satisfies the conditions. Take any preference R and consider the profile R^N. By regularity, $\Psi(R^N) = \Psi^*(R^N) = 0$. Any profile p in L^N can be acquired by a sequence of elementary changes beginning from R^N by minority decrements, majority decrements or swings. By Lemmas 4.3.2 and 4.3.3, the increase (or decrease) induced by each of the elementary changes should be the same. Hence for any p in L^N, we conclude $\Psi(p) = \Psi^*(p)$. □

4.4 CONCLUSION

In this paper, we have modeled polarization as an aggregation of antagonisms per issues within a profile. The polarization measure we introduce simply check for each issue, i.e., pairwise comparison of alternatives, and compares the strength of a majority versus minority. These pairwise comparisons on issues are then aggregated and normalized to a real number between 0 and 1. The measure is very intuitive and is characterized by a few plausible conditions.

There are many directions for future research. The relation between the extent of polarization and the social aggregation outcomes would be a natural route of inquiry. Gurer (2008) studies the Arrovian impossibilities when the preferences in the society cluster, in some sense, around a preference, where it is also conjectured that in a bipolar society the sum of the distances from the two opposite clusters, around which the society is polarized, will be decisive concerning whether we end up with possibilities. The analysis is dependent on a metric-based approach to alienation between preferences. Thus, the relevance of polarization measures based on pairwise comparisons of alternatives to social aggregation outcomes is an open and immediate question one might ask.

Note that the current analysis treats pairs of alternatives impartially, i.e., every issue is of equal importance for polarization. Of course, in many real life situations we may have differing weights on issues. Another question for future research would be analyzing richer domains of preferences, e.g., weak orders, or restricted ones, e.g., single-peaked domains which are politically relevant and interesting.

5

MEASURING U.S. PUBLIC OPINION POLARIZATION

(joint work with Uğur Özdemir)

> *"I believe we can seize this future together because we are not as divided as our politics suggests. We're not as cynical as the pundits believe. We are greater than the sum of our individual ambitions and we remain more than a collection of red states and blue states. We are, and forever will be, the United States of America."*
> *Barack Obama*
> *Victory Speech on November, 7, 2012*

5.1 INTRODUCTION

There is a never ending public and academic debate on the trend of public polarization in the U.S. for the last couple of decades. Although politicians were never exempt from it[1], Obama's victory speech is particularly relevant for he reveals his corner on the issue by taking side with those who claim that the public is not polarized but the elites are. In the academic dimension, there are two major camps on the issue: those who forefront an increase in the polarization and those who hold that if anything happened it is not polarization but sorting. [2]

In this chapter we argue that the measures used in the literature are not theoretically connected to a notion of polarization and most of the disagreement arises from this very fact. We adapt an axiomatically derived measure due to Esteban and Ray (1994) ,which originates in the measurement of income inequality but is nevertheless conceptually

1 In fact, as in the example of the influential "*culture war speech*" by Patrick Buchanan in 1992 Republican National Convention, politicians have embraced the issue deeply.

2 *Sorting* refers to the situation where party affiliations of individuals are getting more aligned with ideologies; more and more liberals vote for democrats and vice versa. Today these camps are mainly represented by Alan Abromowitz and colleagues and Morris Fiorina and colleagues respectively.

suitable for measuring preferential polarization. Before going into detail as to why we preferred this method and how we proceed let us first clarify a bit more the subject matter.

Aside from the conceptual discussion of *what polarization is*, we should first make clear *polarization of whom and what* we will be analyzing. As it is prevalent in the literature, either *public* (or *mass*) or *elite* polarization is to be measured. In another dimension we have to choose between *partisan* (marked with the distance between party affiliates) and *preferential* (or attitudinal, opinion) polarization. Hence, we might summarize the issues with the following matrix.

Elite Partisan	Elite Preferential
Public Partisan	Public Preferential

This chapter is an attempt at providing insights into the public preferential polarization in the United States since 1980s.[3] For that purpose, we introduce the use of a method for measuring polarization of the public opinion in the form of responses to survey questions while discussing the conceptualization and assumptions that lie behind our methods.

Citizens do not have a House-like institution with devices such as roll call votes to directly reveal their political ideologies intensively in a dynamic manner. Also, it is not always the best way to extract ideologies from the votes of the public for offices (presidential, house or senate) simply due to representation issues, as argued in Fiorina and Abrams (2008). Hence it comes to public opinion surveys where large numbers of individuals are asked about many issues that are thought to be salient and representative regarding ideological preferences.

Aside from the measure, there are two further methodological apparatuses we employ in the chapter. First, we use a scaling due to Aldrich and McKelvey (1977) to correct for differential-item functioning that arises when respondents interpret issue scales like the standard seven-point liberal-conservative scale differently and distort their placements of the stimuli and themselves in estimating ideal points of the individuals. Second, we use the scaling due to Poole (1998) in estimating the latent ideological dimension of respondents.

We delve into the American National Election Survey data for years between 1982-2008 to implement the framework developed in the chapter. Our findings suggest a significant increase in the polarization in the latent ideological dimension during the

3 The methodology followed here is readily applicable for elite preferential polarization as well. We do this in an ongoing project.

1990s. Although there is an upward trend during the 2000s, we are unable to classify this as significant. The trends in the separate issue areas show some significant changes but do not present a clear pattern.

The organization of the chapter is as follows. In Section 2 we explore the previous literature on the subject and expose some measures of polarization utilized before. Section 3 is devoted to introduction of measures while Section 4 is devoted to the methodology and data. Section 5 comments on the results. We conclude afterwards in section 6 while addressing some limitations and further research interests.

5.2 LITERATURE REVIEW

Although the concept of polarization has a clear intuitive meaning, it is not as clear when it comes to formally defining and measuring it. DiMaggio et al. (1996) conceptualize polarization process as the motion in the opinions toward the poles of a distribution. A large body of research, in light of this view, hence searches for clusters around poles, or simply bi-modality, in distributions representing public preferences. In fact what we see in Fiorina et al. (2005), Fiorina et al. (2008), Fiorina and Abrams (2008), Fiorina and Abrams (2010) along side DiMaggio et al. (1996) (updated in Evans (2003)) is this kind of an approach: commenting on the variation in distribution of preferences over years in form of a decrease in the center and increase in the extreme ends or comparing variances, kurtosis or simply the weights of extreme category responses.[4] As strongly emphasized in Fiorina and Levendusky (2006), no polarization is observed by these works for the last couple of decades[5], instead, voters aligned better: correlation between policy views and partisan identification increased.[6]

But, let us look at what Downey and Huffman (2001) noted, illustrated in Figure 5.2.1 below. Here we end up in a trimodal distribution from a normal-like distribution when half-way masses on both sides move in equal weights to middle and extreme points. In this case kurtosis and variance stay silent.[7]

Next, consider the following Table 5.2.1 reproduced from Fiorina et al. (2008), which shows the change from 1984 to 2004 in percentages of respondents that chose each point

4 In Abramowitz and Saunders (2006), Abramowitz (2006) and Jacobson (2006) also we see utilization of standard deviation variation as an indicator of polarization.

5 Fiorina and Abrams (2008), however, acknowledges that the activists (in fact, campaign-active partisans) are polarized since 1970s.

6 See also, *inter alia*, Davis and Robinson (1997), Baker (2005), Layman and Carsey (2002) and Levendusky (2009).

7 The changes in variances are insignificant due to Downey and Huffman (2001) who provide also evidence on how prevalent trimodal distributions, for example, in General Social Survey data.

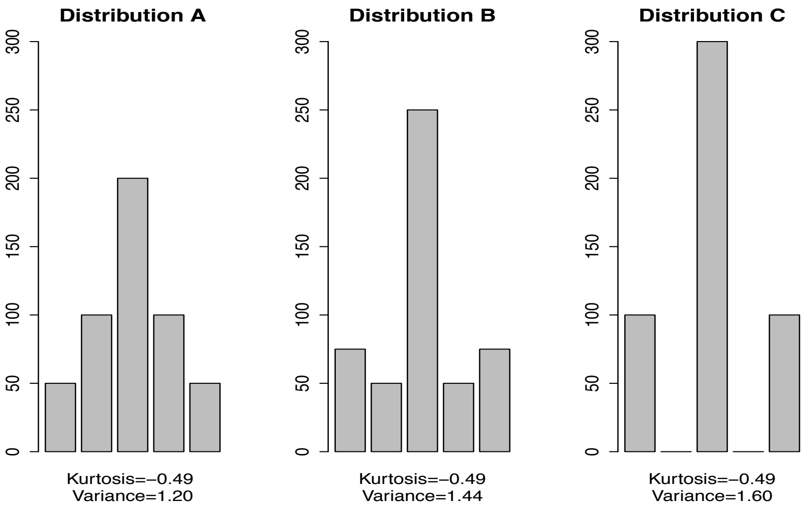


Figure 5.2.1.. Kurtosis and variance cannot tell the difference.

on the scales for six items, for illustration of another example of an approach in the same vein.

	Left Shift						
Item 1	6	2	3	0(-9)	0	-2	-2
Item 2	5	4	5	-3(-5)	-3	-3	-2
	Right Shift						
Item 3	0	-2	-5	-5(-7)	-1	6	8
Item 4	-5	-4	-3	-5(-4)	8	4	2
	Polarization						
Item 5	2	1	0	-2(-7)	0	1	3
	No Change						
Item 6	1		-1		3		-1

Table 5.2.1.. Percentage changes from 1984 to 2004 in positions for six items in the American National Election Studies. Numbers in parentheses are changes when "Don't Know"s are treated as moderates.

The items 1-5 has seven points and the item 6 has 4 points in the scale, going from extremely liberal to extremely conservative, from left to right. For example, the entry "-1" next to "Item 3" stands for the fact that the percentage of respondents who chose the first conservative position from center decreased by one, from 1984 to 2004, while "2" in the row means the percentage of respondents that chose the second liberal position from center for "Item 1" increased by 2. The authors argue that polarization can only be claimed to be observed for the fifth item. In items 1 and 2 they see a shift towards left and in items 3 and 4 towards right that leads exclusion of these from the polarized in their view. In item 6, they see no change and the only case they see polarization is item 5 where a movement from center to extremes are observed. However if the moves are from the high supported side to the low so that the side with low approaches to the side with high in size, we would be inclined to see polarization.[8] Consider the following case, in Figure 5.2.2, where a right to left shift induces a distribution with two equal masses. The latter situation seems clearly more compelling to be tagged as more polarized.

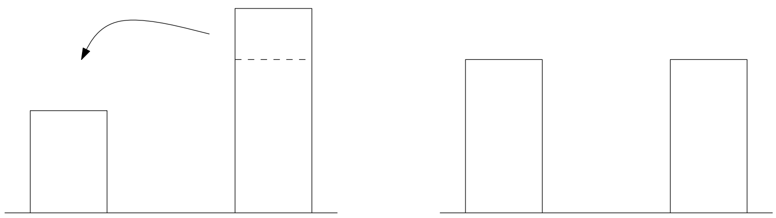


Figure 5.2.2.. A left shift.

Although not as severe as Hunter (1992) depicts, some prominent authors claimed an increasing trend of polarization among American public for recent decades by employing another set of measures. Leaving behind phenomena such as partisan, geographical or religion-based polarization -that are argued sometimes wrongly in place of public preferential polarization- we move forward to cast an eye on two prevalent methods that can be found especially in Alan Abramowitz and his colleagues' works.

The first of these methods includes a comparison of some sort of correlation between issues. For a better exposition see Figure 5.2.3 where each respondent i, j and k picks positions on seven different issues, from extremely liberal to extremely conservative, in the same manner. To keep everything comparable with Abramowitz and Saunders (2008), where we gathered the following method, we depict one of the issues (the issue number 1) as a four-scale issue.[9]

Respondent i has four extremely liberal positions, two centrist positions and one extremely conservative position. Respondent j has four slightly liberal positions, one liberal position, one centrist position and one slightly conservative position. Respondent k has five extremely conservative positions and two extremely liberal positions. The method proposes to label respondents with 0 or 1 as absolute value difference between liberal and conservative positions with low, respondents with 2 or 3 as absolute value

8 Not that we suggest increase in polarization for that particular data, but the point is that this method might lead to inappropriate conclusions.

9 The two positions with bold markers in between the extreme points are to denote interim positions for the four-scaled issue.

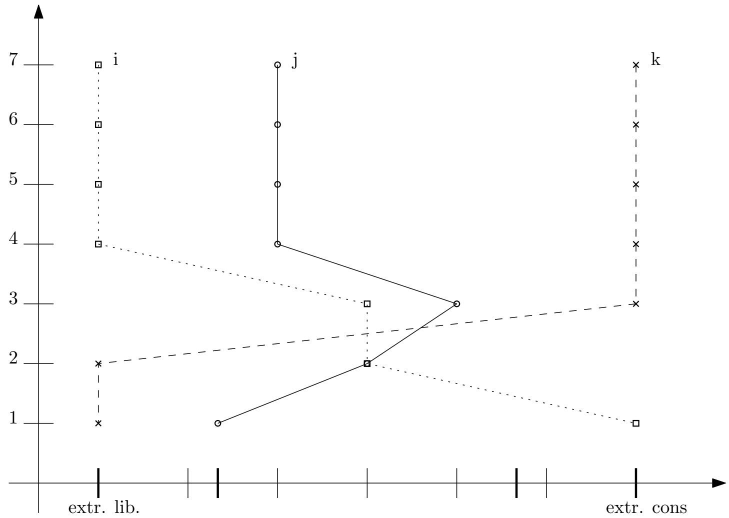


Figure 5.2.3.. Clustering effect.

difference with moderate and respondents with 4 or beyond with high in polarizedness scale. It advances on by comparing the proportions of highly polarized respondents among years. So in a year with only *j* type respondents, we will have only highly polarized profile hence high polarization while in a year with any combinations of *i* and *k* (even when half are of *i* type and half are of *k*) we will have a moderate profile.[10]

One other prevalent indicator used in some papers such as Fiorina and Abrams (2010), Abramowitz (2010), Abramowitz (2006), Jacobson (2006), Abramowitz and Saunders (2005) and Abramowitz and Saunders (2006) is the change in the ratio of the weights on the extremist responses to the weights on the centrist. However, let's consider the situation similar to the distribution A in Figure 5.2.1 with the only difference that the middle mass is now 100 instead of 200. If we similarly dissolve the masses in between extremes and center into both sides equally, the ratio of the extreme masses to the middle mass doesn't change whereas this change, we argue, should increase polarization.

For another review of methods in the literature, we refer the reader to Hill and Tausanovitch (2014) where provided an analysis of measures used previously, related especially to the application to the data from American National Election Studies.[11] Other interesting reviews may include, among others, Hetherington (2009), Prior (2013), Fiorina and Abrams (2008), Nivola and Brady (2006) and McCarty et al. (2006).

10 These methods are also applied to some omnibus scales that are prone to some rescaling issues. We provide, in the Appendix B, examples of rescalings found in the aforementioned literature.
11 The authors come up with a list of ten different measures employed.

We now turn to a novel approach to the measurement of polarization, borrowed from the economics literature on inequality measurement.

5.3 MEASURING PREFERENTIAL POLARIZATION

Polarization is usually seen as an intensifying disagreement in preferences (or ideologies). Naturally, assessment of it depends highly on the context. In American politics, polarization is generally seen as *a separation of politics into liberal and conservative camps*, remarked in McCarty et al. (2006). At the congress and elites level, academic literature reflects more or less a consensus on that it has increased.[12] On the contrary, the literature is rather polarized on if the polarization of public opinion has increased (or took place, if taken as a process) or not.

Obviously, most of the debate[13] until today rests on the fact that there does not exist a unique definition of polarization, especially for the case of public preferences. Following the work of Joan Esteban and Debraj Ray that is initiated as a study on income polarization, we argue that the (public preferential) polarization can be seen as the aggregation of pairwise antagonisms within society. The antagonism between two individuals can be seen as a function of the distance between positions (*alienation*) and increasing with the support of the held positions (*identification*). Esteban and Ray (1994) characterized first a class of functions of this feature with some plausible axioms. This class of measures, basically, consists of those that sum up all pairwise alienations, weighted in a certain way with the population shares of the positions of the components.

To illustrate, suppose we have N individuals and $y \in \mathbb{R}^N$ is the vector of positions. For each such vector y let there be $m(y) \in \mathbb{N}$ positions with nonzero support. Let $M(y) = (M_1(y), ..., M_{m(y)}(y)) \in \mathbb{R}^{m(y)}$ be the vector of those positions. Furthermore, let $\phi(y) = (\phi_1(y), ..., \phi_{m(y)}(y)) \in \mathbb{R}^{m(y)}_{++}$ be the vector of frequencies for each reported position, hence $\sum_{j\in\{1,...,m(y)\}} \phi_i(y) = 1$, for all y. The class of functions given by

$$\mathcal{P}(y, \alpha) = K \sum_{i=1}^{m(y)} \sum_{j=1}^{m(y)} [\phi_j(y)(\phi_i(y))^{1+\alpha}|M_i(y) - M_j(y)|] \tag{5.1}$$

12 Reader might be referred to, *inter alia*, Hetherington (2009), Theriault (2008), McCarty et al. (2006), Layman and Carsey (2002) and Layman et al. (2010) and for the rare contrary argument to Harbridge (2009) and Krehbiel and Peskowitz (2012).

13 We remind the reader that *Fiorina vs Abramowitz* is only an example of the debate, there are others for both sides. The recent *The Monkey Cage* blog at the Washington Post website can give a hint about the ongoing saliency of the discussion today. http://www.washingtonpost.com/blogs/monkey-cage/

where $\alpha \in [1, \alpha^*]$ with α^* being approximately 1.6 is the only class within the description above that satisfies the aforementioned axioms.[14] We think this measure is conceptually better suited for measuring public polarization for the cases where the distribution can be, or originally is, described with clusters on a line. To illustrate, we discuss some properties along with examples we introduced before.

The function has a value zero whenever there is full consensus in the society. It is hence the case also when we have all of the respondents have the same type (i, j or k) in the clustering example above. Naturally, the case where half is type i and half is type k is more polarized than the consensus case for any plausible reduction of dimensions. For left (or right) shift discussion of Figure 5.2.2 above, consider the case where p portion of the society hold a leftist position and the rest $1 - p = p + \epsilon$ hold a rightist position. The polarization, if the distance is d and $\alpha = 1$, in this case is: $p^2(p+\epsilon)d + p(p+\epsilon)^2 d = dp(p+\epsilon)[2p+\epsilon]$. When ϵ is equally shared by both positions however, the new measure is: $d(p+\frac{\epsilon}{2})^2[2p+\epsilon]$. The difference, $\varepsilon^2/4$, is always positive. Finally, the kurtosis-variance example goes as follows: The distribution A has value 0.264 while distribution C has almost double, 0.448. Thus, we established that the measure gives the desired comparisons in the discussions above.

This is not the first work to propose the use of this measure for polarization in politics. Clark (2009) employs it to measure ideological polarization on the Supreme Court while Oosterwaal and Torenvlied (2010) compares the trend in polarization in the Netherlands measured by $\mathcal{P}$. However, as it will be elaborated further below, we will be dealing with data generated from large surveys that describe the distributions of preferences with (estimated) density functions. This very fact brings the need of a measure applicable to such distributions and the version of the measure to be utilized in this chapter, thus, is due to Duclos et al. (2004) where the authors provide a natural extension. The class of functions, (denoted by DER from now on)

$$\mathcal{P}_\alpha(f) \equiv \int \int f(x)^{1+\alpha} f(y)|x-y|dxdy \tag{5.2}$$

with $\alpha \in [0.25, 1]$, defined on all continuous densities in $\mathbb{R}_+$ is shown to be characterized by a set of axioms in the same vein as before. A detailed analysis is provided in the Appendix C.

14 Reader is advised to refer to the original paper by Esteban and Ray (1994) for an exhaustive analysis. The parameter α stands for the weight given to concentrations compared to distances.

5.4 DATA AND THE ESTIMATION METHODOLOGY

We use the ANES data for the election years between 1984 and 2008 for our empirical analysis. There are seven questions we are studying:[15]

1. Liberal/conservative, with scale 1 (extremely liberal) to 7 (extremely conservative)
2. Government aid to blacks, with scale 1 (help) to 7 (no help)
3. Defense spending, with scale 1 (greatly decrease) to 7 (greatly increase)
4. Jobs and living standards, with scale 1 (provide) to 7 (let go)
5. Government services and spending, with scale 1 (few provision) to 7 (more provision)
6. Health insurance, with scale 1 (govt plan) to 7 (private plan)
7. Abortion, with scale 1 (never permit) to 4 (always permit)

These are chosen following the earlier literature Abramowitz and Stone (2006); Abramowitz and Saunders (2008); Abrams and Fiorina (2012); Fiorina and Abrams (2008); Fiorina et al. (2005).

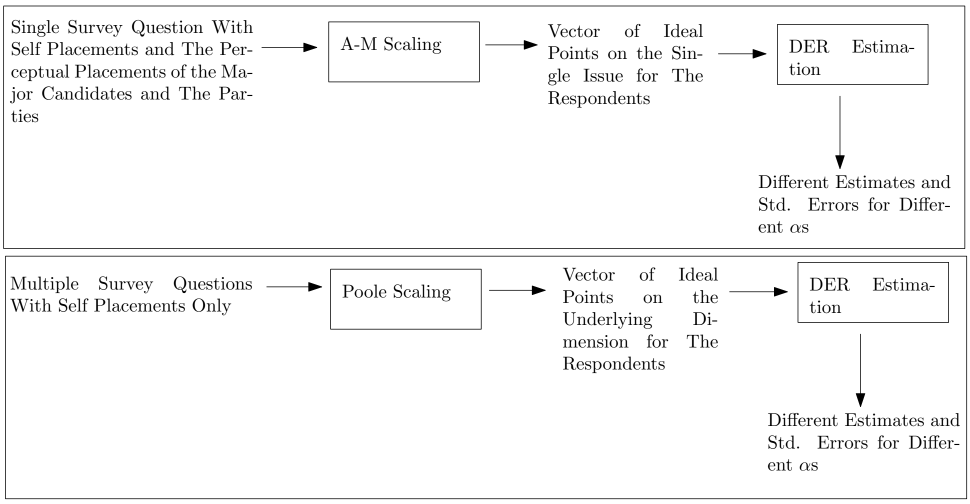


Figure 5.4.1.. The Outline of the Empirical Strategy.

Figure 5.4.1 summarizes our empirical approach. Basically there are two different procedures we follow. In the first one, we use the Aldrich-McKelvey scaling in order to

15 The exact wording of these questions are given in Appendix D.

correct for the differential item functioning bias and measure polarization in separate issues using these ideal points. Note that A-M technique produces ideal points on a continuous scale hence we are able to estimate the distribution confidently with DER. In the second one, we initially derive a single policy dimension from the underlying seven issues using the Poole's scaling and estimate the polarization using this "aggregate" distribution.

In what follows we are going to elaborate these different steps involved in our analysis.

5.4.1 *Estimation of the DER Measure*

The estimator for the function in (5.2) *à la* Duclos et al. (2004) is given by:

$$\pi_\alpha(f) = \frac{\sum_{i=1}^{n} w_i f(y_i)^\alpha a(y_i)}{\sum_{i=1}^{n} w_i} \tag{5.3}$$

where w_is are the weights given to the positions, $f(.)$ is the density function of the ideal point distribution, y_is are empirical quantile for percentiles between $(i-1)/n$ and i/n, $\alpha \in [0.25, 1]$ and

$$a(y_i) = \mu + y_i \left(\frac{2\sum_{j=1}^{i} w_j - w_i}{\sum_{i=1}^{N} w_i} - 1 \right) - \left(\frac{2\sum_{j=1}^{i-1} w_j y_j + w_i y_i}{\sum_{i=1}^{N} w_i} \right). \tag{5.4}$$

If the weights are taken to be equal, the function π_α reduces to

$$\pi_\alpha(f) = \frac{1}{n} \sum_{i=1}^{n} f(y_i)^\alpha a(y_i) \tag{5.5}$$

where

$$a(y_i) = \mu - y_i + \frac{2}{n}[y_i(r(i) - 1) - \sum_{j=1}^{i-1} y_j]. \tag{5.6}$$

The following normalization leads to the final version of the function which is used to estimate polarization:

$$P_\alpha(f) = \frac{1}{2\mu^{1-\alpha}} \pi_\alpha(f). \tag{5.7}$$

Hence, in order to find the polarization index, we need to have the empirical distribution of individual ideal points. We find this distribution using the Gaussian kernel density estimation:[16]

$$\hat{f}(y) \equiv n^{-1}\sum_{i=1}^{n} K_h(y - y_i) \tag{5.8}$$

$$K_h(z) \equiv h^{-1}K(z/h) \tag{5.9}$$

$$K(u) = (2\pi)^{-0.5}\exp^{-0.5u^2} \tag{5.10}$$

where h is the bandwidth for kernel function and it is chosen so as to minimize the mean squared error.

The meaning of α needs some further elaboration since it will be critical in interpreting our results. As it was discussed earlier, there are two forces which determine the polarization level: alienation and identification. An individual located at point x feels alienation vis-a-vis another located at y, and this alienation increases with the distance between these individuals, $|x - y|$. However, for this alienation to be translated into polarization, the individual must - to a greater or a lesser degree- identify with the rest of the society. An individual located at ideal point x experience a sense of identification and this is given by $f(x)$.

The weight of the density function on the measure increases with α as seen in the functional forms in (5.5) and (5.7), which would mean that the weight assigned to identification increases. In terms of what we see from the empirical distributions, the shadow of peaks and multi-modality kicks in as α gets bigger. Note that there is always a "trade-off" between the height of the distribution and the width of the distribution since the area underneath any density function is constant and equal to 1. The former is related to identification and the latter with increasing distances, hence alienation. The interplay between these two forces is the most critical dynamic behind the DER measure, and α parametrizes the relative weights of these two forces.

Finally, the standard errors are computed by bootstrapping and all analyses are carried out with the statistical computing environment R.

5.4.2 *Aldrich-McKelvey Scaling*

A-M Scaling Aldrich and McKelvey (1977) is a solution offered for the differential item functioning problem in estimating the positions of political stimuli and survey respon-

16 We have tried other kernels for robustness, but results do not change.

dents along a latent policy dimension from issue scale data. This problem arises in all cases where the respondents interpret issue scales differently and distort their placements of the stimuli and themselves. The method treats placements as linear distortions of the "true" positions of the stimuli. By estimating each respondents perceptual distortion parameters, it is possible to recover the locations of the stimuli as well as the respondents.

Let z_{ij} be the perceived location of the political stimulus j (party or candidate) by individual i. The A-M model assumes that the individuals reports a noisy linear transformation of the true location of stimulus, z_j:

$$z_{ij} = \alpha + \beta_i z_j - u_{ij}$$

where α_i is the shift term, β_i is the stretch term, and u_{ij} satisfies the usual Gauss-Markov assumptions for error terms.[17] So in cases where you not only have the self-placements of the voters on the issue scale but also their perceptual placements of different political stimuli such as political parties and candidates, A-M scaling offers a way to correct for the bias due to interpersonal differences in interpreting the scale.

Aldrich and McKelvey estimate these parameters using a Lagrange multiplier. We use the maximum-likelihood based method for A-M scaling available in the `R` package `basicspace` Poole et al. (2012). In fact, very recently, a Bayesian implementation has been introduced in Armstrong et al. (2014). There are two main advantages of this Bayesian approach: it can handle missing data and it allows for idiosyncratic error variances. The latter is an important improvement since the assumption that respondents are assumed to have an unequal likelihood of reporting an incorrect ordering of the stimuli is quite unrealistic given the variation in respondents' political sophistication. We will be replicating our results using this Bayesian methodology in further research.

5.4.3 *Poole's Scaling*

This is a scaling procedure, introduced in Poole (1998), for estimating the latent (unobservable) dimensions underlying a set of manifest (observable) variables. We will be using this technique to derive the underlain ideological positions of the respondents that are, we believe, observable through their self-placements on different issues. In contrast to more prevalent dimension reduction techniques such as factor analysis which work with a covariance matrix computed from the data matrix, it analyzes the data matrix di-

17 Zero in expectation, homoscedasticity and independence.

rectly without any intermediary transformations of the original data.[18] This allows for the method to handle missing observations which is an important problem especially with the survey data. Poole (1998) provides Monte Carlo experimentation to show that the procedure performs well, in deed, in reproducing the missing elements of a matrix even when the level of missing data is high.

5.5 OBSERVATIONS ON THE RESULTS

The plots which summarize our findings are given in the Appendix A, Figures C.1.1-C.1.4. These are the polarization estimates with %95 confidence intervals around them. Our main finding is depicted in Figure C.1.1a. There is an upward trend in polarization for the primary ideology dimension after 1992. The increase is significant[19] between years 1992 and 1996 and between 1996 and 2000. Although the trend continues up until year 2008, these increases are not significant. We also observe a decrease in polarization from year 1988 to 1992. Hence our results suggest that 1990s was the decade in which public polarization increased significantly. One other observation is that the polarization differences get more evident as α gets smaller, or as the relative weight of alienation increases. This is due to a "release" in the distribution which increases alienation, as can be seen in Figure C.1.5a.

We are not going discuss the results for each and every issue-year pair. Instead we will make some particular observations which we think are interesting for these issues.

The polarization on the blacks issue also increases from 2000 to 2004 and from 2004 to 2008, but this time for $\alpha = 1$ and $\alpha = 0.5$. The reason why the polarization also increases for $\alpha = 0.5$ is that the multi-modality is not as severe and peaks are not as high in 2008 and the distances are greater in 2004 compared to the job issue as depicted in Figure C.1.6a. One rather interesting observation about this alienation-identification framework is that the polarization on the blacks issue *increases* significantly for $\alpha = 0.25$ and *decreases* significantly for $\alpha = 1$ from 1984 to 1988. This is a nice illustration how the interaction between alienation and identification can yield significantly different results under different values of α.

The polarization on the defense spending issue has an upward trend from 1992 until 2000 for $\alpha = 0.25$ and $\alpha = 0.5$. This is in accord with the fact that American public was quite divided on this issue along the partisan lines starting with the First Gulf War

18 In effect this is a method to obtain an Eckart-Young lower-rank approximation matrix of a matrix with missing entries.

19 Unless otherwise noted, we refer to a significance at the 0.05 level.

in 1990. This divide continued with other military operations US participated such as the Bosnian and Kosovo wars. It is also interesting to note that the polarization in this issue decreases significantly from 2000 to 2004. This can be attributed to the increased external threat awareness and nationalistic sentiments which led to a convergence of opinions on this issue after the September 11 terrorist attacks.

The polarization on the job creation issue increases from 2004 to 2008 only for $\alpha = 1$, i.e, when the weight of identification is highest. As it can be seen from the Figure C.1.5b, the multimodal (peaked) structure of the distribution derives this result. This observation is quite consistent with the fact that the public became more polarized about the extent to which government should intervene an economy in recession starting with the 2000s. A similar pattern is visible for the government services issue which supports this explanation.

The polarization on the abortion issue makes a significant peak in year 2000. The fact that, together with gay/lesbian rights, it was particularly an important issue in the 2000 presidential election concurs this observation. Although the debate on abortion is still alive, it never gained that momentum it had in between Al Gore and George W. Bush.

5.6 CONCLUSION

In this chapter we have suggested a methodology to measure public polarization and implemented it for the election years in the U.S. from 1984 to 2008. Pointing out the need for a theoretically supported measure in the literature, we adapted a measure *à la* Duclos et al. (2004) for this purpose. With its intuitive axiomatic support, DER is a reliable measure to help us understand the dynamics of preferential polarization. Rather than using some summary statistic, it uses the distribution of ideal points in the society as a whole. By parametrizing the relative weight it puts on alienation versus identification, it provides a nice micro-behavioral foundation. In fact, Identifying how the interaction between alienation and identification lead to changes in polarization can be an important step towards having a more nuanced understanding of the nature of polarization in different issues. The two other tools we employed were the A-M scaling Aldrich and McKelvey (1977) to correct for the differential item-functioning and Poole's scaling Poole (1998) to derive the latent policy dimension.

Our empirical findings show that although the US polarization measured on the underlying latent ideology dimension has an upward trend after 1992, this is only significant during the 1990s. There is not a common pattern when we do an issue-wise analysis. Although there are significant changes between some time periods, issues do not seem to have a shared pattern. This discrepancy between the aggregated ideology dimension and its components is probably due to the fact that individuals are changing their opinions on different issues but becoming more consistent about their choices on different issues overall. This certainly needs a more careful analysis of the particular issues and time periods.

5.6.1 *Limitations of the Framework and Future Work*

We study polarization on a unidimensional policy space. However, there is a robust finding in the spatial voting literature that a two dimensional policy space is needed in order to capture the ideological spectrum Benoit and Laver (2006). This suggests that developing a polarization measure for multidimensional spaces will be a significant contribution.

We used the same question (issues) to measure polarization every year. Even if this approach is convenient in the sense that it makes comparisons more meaningful, it implicitly assumes that the salient issues do not change from one year to another. This is certainly a restrictive assumption. We could try to identify different issues for different years depending on their saliency. One way to do this would be to run a logit model using the vote choice of respondents as the dependent variable and choose those issues which seem to have significant effect.

We might as well consider giving more weight to those individuals who have higher engagement in the political issues. If public polarization is the sum of all binary antagonisms, then it can be argued that the more "activist" the individuals are the more their effect should be on the level of polarization. As we discussed earlier, the DER measure is general enough to handle these weights. ANES includes questions that can be uses to determine the activism level of the respondents.[20]

Using alternative dimensions for alienation and identification seems to be another promising direction for future study. One might argue, for instance, that people identify themselves with their income group but the alienation takes place on the liberal-conservative scale.

We can employ the estimated polarization measure as a dependent variable and investigate the causes of polarization and use it as an explanatory variable in order to understand its consequences. The latter is particularly important in that it is related to the fundamental question of why we care about polarization in the first place. After all, the consequences of polarization "are not entirely clear and may include some benefi-

20 See Schofield et al. (2011).

cial as well as detrimental consequences." as discussed in Epstein and Graham (2007). There have been some work on the consequences of *elite* polarization on voter turnout Crepaz (1990), on foreign policy McCormick and Wittkopf (1990) and on the judicial system Binder (2000) but the effects of *public* polarization has not been studied.[21] The absence of a well-defined continuous measure was probably one of the reasons why this has been the case.

21 One exception is Abramowitz and Stone (2006). This chapter however uses proxy measures for polarization, like the ratio of the extreme voters.

Part IV

INFORMATION AGGREGATION

6

THE CONDORCET JURY THEOREM UNDER COGNITIVE HIERARCHIES

(joint work with Yukio Koriyama)

6.1 INTRODUCTION

Since Condorcet's earlier work in 1785, mathematical support has been provided for the idea of increasing accuracy of collective decisions by including more individuals in the process. In his seminal *Essai*, Condorcet considered nonstrategic individuals voting to make a decision on a binary issue where each alternative is commonly preferred to the other one in one of the two states of the world. Each individual receives independently an imperfectly informative private signal about the true state of the world. Then under majority rule, the probability of reaching a correct decision monotonically increases with the size of the electorate and converges to certainty in the limit.

Although allowing strategic behavior may revoke assumptions of this basic model this property survives in various circumstances of collective decision making (e.g. Austen-Smith and Banks (1996) and Feddersen and Pesendorfer (1997)). In most models, probability of making a right decision increases and converges to one as the group size increases, even when strategic players may vote against their signals. This asymptotic efficiency is often coined as the *Condorcet property*.

In the literature of the binary-state, binary-issue decision making, often the signals are also assumed to be binary. However, we find that this assumption does not necessarily span most of the situations considered to be captured with these models.

On the other hand, assuming complete rationality of the voters may be too much requiring when the strategy space and the number of players are large. This intuition finds support from experimental evidence and information aggregation situations are not, to say the least, exempt from it. Especially when a major aspect of the strategic models comprises the presumption that voters take into account their probability of being pivotal (Downs (1957)). As shown by Esponda and Vespa (2013) in their experimental study, hypothetical thinking in direction to extract information from others' strategies required in strategic voting models are too strong as an assumption and does not find support in lab. Battaglini et al. (2008) report from another experimental study an increase in irrational non-equilibrium play as the size of electorate increases.[1]

Models of nonequilibrium strategic thinking have been proposed to explain structural deviations from equilibrium thinking in a variety of games. A sizable part of bounded rationality literature is devoted to the models of cognitive hierarchy, starting with Nagel (1995) and Stahl and Wilson (1995), which allow heterogeneity among the individuals in levels of strategic thinking. The construction of this model dictates a foundational level of cognitive hierarchy, level-0, which represents a strategically naïve initial approach to game. Then a level-k player (hereafter Lk, where $k \geq 1$) is assumed to best respond to others with a cognitive hierarchy of level $k-1$. The construction of levels resonate with *rationalizability*, as in Bernheim (1984), due to the fact that the decisions made by a level-k survive k rounds of iterated deletion of strictly dominated strategies in two-person games.

The closely related *Poisson-CH* model is introduced in Camerer et al. (2004). They allow heterogeneity in beliefs on others' levels in a particular way. A level-k type best responds to a mixture of lower levels, which is estimated by consistent truncations up to level $k-1$ from a Poisson distribution, for each $k > 0$. The relevant Poisson distribution is either obtained from maximum likelihood estimations applied to data or calibrated from previous estimates.

The set of level-1 strategies in this *Poisson-CH* model (hereafter *CH1*) is exactly the same as that of *L1*. For higher levels, *Lk* and *CHk* differ. For example, strategies in *CHk* are not rationalizable in general.[2] Common in these models is the assumption that level-k players do not assign any probability to the levels higher than k. This assumption captures the idea that there is a hierarchy in cognitive limits among players.

Another assumption shared by these two models is that of *overconfidence*. Both models presumes that no individual assigns a positive possibility to opposing another player with the same level of cognitive hierarchy. In this paper, we propose a new model, the model of *endogenous cognitive hierarchy* (*ECH*), without imposing the overconfidence assumption. ECH builds on the Poisson-CH model by allowing individuals to best respond to others who may belong to the same level of cognitive hierarchy. While it may

1 As Camerer (2003), chapter 7, stresses the effect of group size on behavior in strategic interactions is a persistent phenomenon, especially towards coordination.
2 Crawford et al. (2013) provides a fine review of these models and applications.

be reasonable to assume ignorance on higher levels due to cognitive limits, we show in this paper that exclusion of the same level opponents leads to significant consequences both theoretically and empirically.

There are three reasons why overconfidence assumption should not be taken as granted. First, we show that in a large class of games exclusion of self awareness, or imposing overconfidence, leads to a significant difference in the asymptotic behavior of the players. In these games, the prescribed strategies are extreme (distance from the symmetric Nash strategy diverges away) in both level-k and CH. We argue that such divergence is not coherent with the idea of cognitive hierarchy. Second, in our experiments, we find that ECH can explain the observed behavior much better than level-k or CH. Third, a large number of subjects, i.e., 97%, answered in the questionnaire that they do not exhibit such overconfidence, as shown in Figure 6.1.1. In fact, most of them think that most of others use the same reasoning, as shown in Figure D.1.1 in the Appendix.

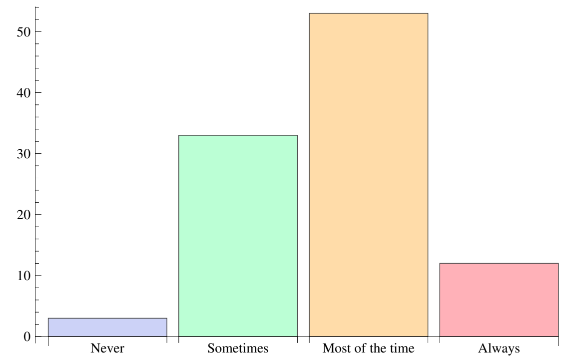


Figure 6.1.1. Frequencies of responses to the Question 2 in a post-experimental questionnaire: "When you made decisions, did you think that the other participants in your group used exactly the same reasoning as you did?".

We present results from a laboratory experiment designed to explore the effect of the size of electorate on accuracy of decisions under majority rule. We deviate from previous literature by setting signals to induce differing posterior probabilities on the true state of the world, contrary to binary signals.

Related Literature

Gerling et al. (2005) provides an extensive survey on the studies of collective decision making modeled similar to Condorcet jury model. Palfrey (2013) presents an ideal survey on experiments in political economy, and particularly in strategic voting. Costinot and Kartik (2007) study voting rules and show that the optimal voting rule is the same when players are sincere, playing according to Nash equilibrium, to level-k, or belonging to a mixture of these. Koriyama and Szentes (2009) provides a general analysis of optimal group size under costly information acquisition and Bhattacharya et al. (2013) tests experimentally the theoretical predictions about individual behavior and group decision under costly information acquisition. They find poor support for the comparative statics predictions of the theory.

The paper proceeds as follows. We introduce the endogenous cognitive hierarchy model formally in the following section where we discuss its properties while relating to previous models. We furthermore compare the implications of models in a stylized setting of information aggregation. In Section 3 we introduce our experimental design that carries novelties due to our modeling concerns and signal setup. Section 4 provides the results of the experiment and compares how different models fit to data. We conclude by summarizing our findings and future research possibilities after Section 5, where we further investigate the implications of both ECH and CH models. Appendices are devoted to proofs and detailed experimental data.

6.2 THE MODEL

Let (N, S, u) be a symmetric normal-form game where $N = \{1, \ldots, n\}$ is the set of players, $S \subset \mathbb{R}$ is a convex set of pure strategies, and $u : S^n \to \mathbb{R}^n$ is the payoff function.

In the cognitive hierarchy models, each player forms a belief on the frequency with respect to the cognitive hierarchy levels of the other players. Let $g_k(h)$ denote the belief of a level-k player on the frequency that the other player belongs to level-h.

In the standard level$-k$ model *à la* Nagel (1995), first, a naïve, nonstrategic behavior is specified. This constitutes the initial level of cognitive hierarchy, the *level*-0, or $L0$. Second, Lk for $k \in \mathbb{N}_+$ believes that all the other players belong to one level below:

$$g_k(h) = \begin{cases} 1 & \text{if } h = k-1, \\ 0 & \textit{otherwise}. \end{cases}$$

In the cognitive hierarchy model introduced in Camerer et al. (2004), every level$-k$, or CHk, best responds to a mixture of lower levels in a consistent manner. To be more specific, let $f = (f_0, f_1, \ldots)$ be a distribution over $\mathbb{N}$ that represents the composition of cognitive hierarchy levels. Then a CHk has a distributional belief that is specified by $g_k = (g_k(0), g_k(1), \ldots, g_k(k-1))$ where

$$g_k(h) = \frac{f_h}{\sum_{m=0}^{k-1} f_m} \text{ for } h = 0, \ldots, k-1,$$

for all $k \in \mathbb{N}_+$. Given a global distribution of congnitive hierarchy levels, consistency requires each level to embrace a belief on distribution of other players' levels that is a truncation up to one level below. Thus, these two models share the following two assumptions.

Assumption 1. (COGNITIVE LIMIT) $g_k(h) = 0$ *for all* $h > k$.

Assumption 2. (OVERCONFIDENCE) $g_k(k) = 0$ *for all* $k > 0$.

In what follows, we introduce the *endogenous cognitive hierarchy model* of games formally, in which overconfidence is dropped while cognitive limit assumption is preserved.

6.2.1 *Endogenous Cognitive Hierarchy Model*

Fix an integer $K > 0$ that prescribes the highest level considered in the model.[3] Then, a sequence of truncated distributions $g = (g_1, \ldots, g_K)$ is uniquely defined from f. As in the cognitive hierarchy model, we focus on the symmetric equilibirum, i.e. the mixed startegy assigned to each level describes the distribution of pure strategies played in the level. A sequence of mixed strategies $\sigma = (\sigma_0, \ldots, \sigma_K)$ where $\sigma_k \in \Delta(S)$ for all $k \in \{1, ..., K\}$ is an ECH equilibrium when there exists a global distribution f such that for each k, σ_k is a best response assuming that the other players' levels are drawn from the truncated distribution.

Definition 6.2.1. *A sequence of symmetric strategies $\sigma = (\sigma_0, \ldots, \sigma_K)$ is called **endogenous cognitive hierarchy equilibrium** when there exists a distribution f over $\mathbb{N}$ under which*

$$\mathit{supp}(\sigma_k) \subset \arg\max_{s_i \in S} \mathbb{E}_{s_{-i}} \left[u(s_i, s_{-i}) \mid g_k, \sigma \right], \ \forall k \in \mathbb{N}_+,$$

where g_k is the truncated distribution induced by f such that

$$g_k(h) = \frac{f_h}{\sum_{m=0}^{k} f_m} \textit{ for } h = 0, \ldots, k,$$

3 We assume $f_i > 0$ for all $i \leq K$. For the truncated distribution to be well-defined, it is sufficient to assume $f_0 > 0$, but we restrict ourselves to the cases where all levels are present with a positive probability.

and the expectation over s_{-i} is drawn from a distribution

$$\gamma_k(\sigma) := \sum_{m=0}^{k} g_k(m)\,\sigma_m,$$

for each player $j \neq i$.

A standard assumption is that the levels follow a Poisson distribution:

$$f^{\tau}(k) = \tau^k \frac{e^{-\tau}}{k!}.$$

This assumption is adopted in the literature and a detailed argument can be found in Camerer et al. (2004). We discuss the implications and limitations of the assumption later in following sections.

Next, we move to the analysis of voting situations modeled as a Condorcet jury model.

6.2.2 *A Condorcet Jury Model*

We consider a binary-state, binary decision making in the group of n players. The true state of the world takes one of the two values, $\omega \in \{-1, 1\}$, with a common prior of equal probabilities. The utility is a function of the realized state and the decision

$$u(d,\omega) = \begin{cases} 0 & \text{if } d \neq \omega, \\ q & \text{if } d = \omega = 1, \\ 1-q & \text{if } d = \omega = -1, \end{cases}$$

for each individual.[4] Voter i receives a private signal that is distributed with a Normal distribution with known common error around the true state, $s_i \sim \mathcal{N}(\omega, \sigma)$. Then, i submits a vote v_i, upon receiving a signal, which is in the form of added bias, i.e., $v_i(s_i) = s_i + b_i$, so an individual i's strategy, in fact, is to pick a bias, $b_i \in \mathbb{R}$. The decision is reached by the unbiased rule, that is by the sign function on the summation of votes: $\mu(v) = \operatorname{sgn}(\sum_{i \in N} v_i)$. The expected utility of $v = (v_1, \ldots, v_i, \ldots, v_n)$ can be written as

$$\mathbb{E}_{(s_i)_{i\in N},\omega}[u(\mu((v_i(s_i))_{i\in N}, \omega))].$$

4 The assumption of symmetric prior is without loss of generality since we allow the utility loss of the two types of errors to be heterogeneous.

We obtain the best response function

$$\beta(b_{-i}) = -\sum_{j\in N\setminus\{i\}} b_j + \frac{\sigma^2}{2}\ln(q/(1-q)),$$

where b_{-i} denotes the biases of all voters except i.[5]

6.2.2.1 *Group accuracy under cognitive hierarchies*

In this section we consider the group accuracy under cognitive hierarchy models. To do this let us fix $K = 2$, that is the highest level in the group to be 2.[6] We first show that if the composition of levels stays the same, or in other words if the Poisson parameter that moves with the average level in the composition is fixed, increasing the jury size will increase the probability of correct decision. In other words, Condorcet property continues to hold.

Proposition 6.2.1. *Let τ be fixed. The probability of correct decision in ECH model converges to 1 as $n \to \infty$ while it goes to $1/2$ in CH.*

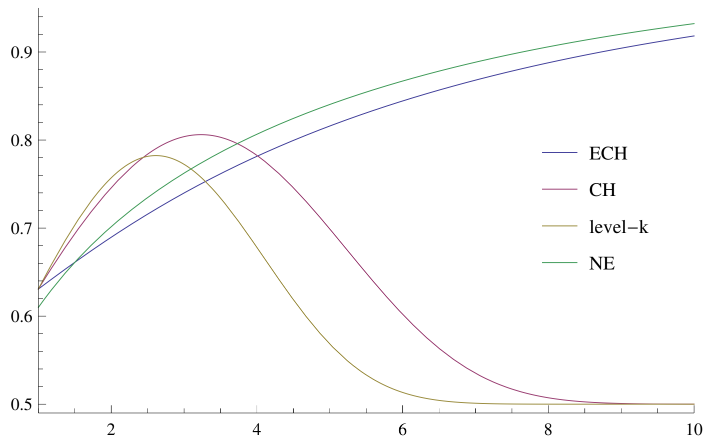


Figure 6.2.1.. The probability of correct group decision under different models. In CH, ECH and level$-k$ models $\sigma = 2$, $q = 3/4$, $\tau = 3$ and $b_0 = -1$ are taken, while for NE only $\sigma = 2$ and $q = 3/4$ are relevant.

5 The algebra leading to this can be found in Appendix D.2.
6 To specify an upper bound for levels is necessary for the analysis to be complete and in picking this particular value we rely on, first, the previous literature on cognitive hierarchy applications and second, ex-post, this assumption proves a good fit for our experimental results.

It is also shown that this is not the case in CH model where the accuracy is as good only as random voting, in the limit. Figure 6.2.1 shows the comparison of models in an exemplary modification. Proof of the proposition can be found in Appendix D.2.

6.2.3 *Cognitive hierarchy as a model of complexity induced by group size*

The idea of group bounded rationality may be argued to be represented best when we have decreasing sophistication as the complexity of the problem increases. In our context, we can think of the complexity as positively correlated with the size of the jury. And as shown in what follows, under certain assumptions, this leads to moving away from the Condorcet property.

Proposition 6.2.2. *Let $\tau = n^{-1}$. Then the probability of correct decision converges to 1 if $b_0 \in (-2, 2)$, collapses to $1/2$ if $|b_0| > 2$ and converges to $3/4$ if $b_0 \in \{-2, 2\}$.*

In fact, we conjecture that the speed of the fall in τ is negatively correlated with the bound on the size of the level-0 bias to preserve Condorcet property. In our calculations we see that if $\tau = n^{-\gamma}$, the interval $(-a, a)$ where a b_0 within secures Condorcet property is shrinking as $\gamma > 1$ is increasing. Proof of the proposition can be found in Appendix D.2.

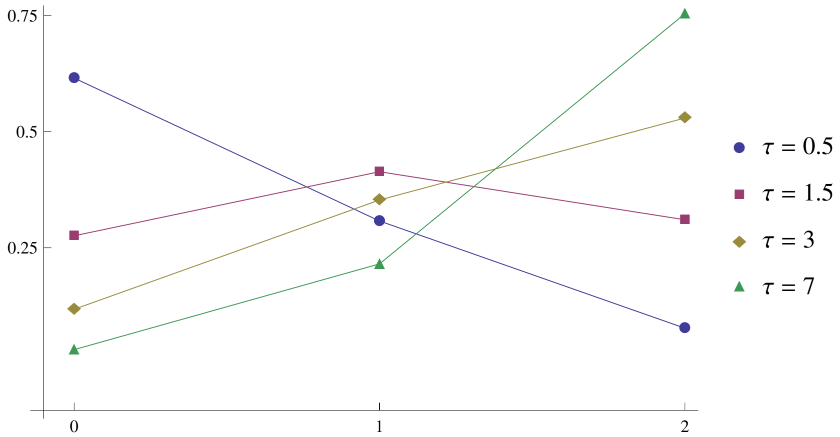


Figure 6.2.2. The truncated distributions for several Poisson parameters. The corresponding means are 0.46 ($\tau = 0.5$), 1.03 ($\tau = 1.5$), 1.41 ($\tau = 3$) and 1.72 ($\tau = 7$).

The conclusions of this section are based on the Poisson distribution assumption, for which truncated distributions for several parameters are shown in Figure 6.2.2. A foundation for this assumption given in Camerer et al. (2004) is as follows. If it is assumed, due to working memory limits, that as k increases fewer and fewer players do the next step of thinking beyond k, we may constrain $f(k)/f(k-1)$ to decline with k. If, furthermore, this decline is proportional to $1/k$, we obtain Poisson distribution.

However, since what matters is the change in the composition of levels, the relevant conclusions would not be affected by the particular choice of distribution. The reasoning is that in the case cognitive aspects of the current situation is not related at all to the complexity induced by the group size, having larger groups would comprise higher presence of naïve thinking only in absolute terms, not in proportional terms, and hence the bias can be addressed accurately in ECH due to presence of also higher levels but not in CH. In the case complexity induced by higher group size translates into more naïve thinking in proportional terms, large enough level-0 bias may worsen the outcome in ECH model as well.

6.3 experimental design

All our -computerized- experimental sessions were run at the Experimental Economics Laboratory of Ecole Polytechnique in November and December 2013.[7] In total we had 180 actual participants and 9 sessions. Each session with 20 subjects lasted about one hour. Earnings were expressed in points and exchanged for cash to be paid right after each session. Participants earned an average of about 21 Euros, including 5 Euros of show-up fee. Complete instructions and details can be found in an online appendix.[8] The instructions pertaining to whole experiment were read aloud in the beginning of each session. Before each phase (treatment) during the experiment, the changes from the previous phase are read aloud and an information sheet including the relevant details of the game is distributed. These sheets are collected while distributing the sheets for the following phase.

We employed a within-subjects design where each subject played all four phases consecutively in a session. Following a direct-response method, in each phase there were 15 periods of play, which makes 60 periods in total that are played by each participant.[9] Since the question of our research relates to the strategic aspect of group decision, our

7 We utilized a z-Tree (Fischbacher (2007)) program and a website for registrations, both developed by Sri Srikandan.

8 The online appendix can be found at http://sites.google.com/site/ozkesali.

9 In third phase, two groups of nine randomly chosen members are formed before each period. Having 20 participants in total, hence, two randomly chosen subjects waited during each period. The same method is applied in fourth phase as well, this time one random subject waited at each period. Subjects are told both orally and through info sheets that in the case the random lottery incentive mechanism picks a period where a subject has been waiting, the payoff from that phase for this subject will be taken as 500 points.

experiment was presented to subjects as an abstract group decision-making task where natural language is used to avoid any reference to voting or election of any sort. In the beginning of each period, computer randomly formed groups of subjects, where the common size of the groups is determined by the current phase. Then the subjects are shown a box full of hundred cards, all colorless (gray in z-Tree). This is also when the unknown true color of the box for each group is determined randomly by the computer. Subjects were informed that the color of the box would be either blue or yellow, with equal probability. Furthermore, it was common knowledge that the blue box contained 60 blue and 40 yellow cards whereas the yellow box contained 60 yellow and 40 blue cards. After confirming to see the next page, they are shown only 10 randomly drawn cards with random locations in the box on the screen, this time with their true colors. These draws were independent among all subjects but were coming from the same box for subjects in same groups.

On the very screen subjects observed their 10 randomly drawn cards, they are required to vote for either blue or yellow. The decision for the group is reached by majority rule, which was conclusive all the time since we only had odd number of subjects in groups and abstention was not allowed. Once everyone in a group voted by clicking the appropriate button, on a following screen, subjects are shown the true color of the box, the number of the votes for blue, the number of votes for yellow and the earning for that period. Once everyone confirmed, the next period started.

	Phase 1	Phase 2	Phase 3	Phase 4
Sessions	$n = 5$	$n = 5$	$n = 9$	$n = 19$
A,B	500 : 500	800 : 300	800 : 300	800 : 300
1-7	500 : 500	900 : 200	900 : 200	900 : 200

Table 6.3.1.. Experimental design. Each phase consists of 15 periods.

The payoffs were symmetric at the first phase of each session. The size of the groups for each period was fixed to be 5 in this phase and each subject earned 500 points in any correct group decision (i.e., blue decision when true color of the box is blue and yellow decision when true color of the box is yellow). In case of incorrect decision no points earned.

The following three phases for each treatment differed only in the size of the groups (5,9 or 19) where a biased payoff scheme is fixed. For two sessions (Sessions A and B), the correct group decision when true color of the box is blue earned each subject 800 points whereas the correct group decision when true color of the box is yellow earned 300 points. For seven sessions (Sessions 1 to 7), the correct group decision when true color of the box is blue earned each subject 900 points whereas the correct group decision when true color of the box is yellow earned 200 points. The summary of these can be found in Table 6.3.1. We implemented a random lottery incentive system where final payoffs at each phase is determined by the payoffs from a randomly drawn period. In the beginning of each session, during general instructions being read aloud and as part of instructions, subjects played two forced trial periods.

Each session concluded after a short questionnaire, which can be found in the online appendix.

6.3.1 *Equilibrium Predictions*

In what follows we provide the theoretical predictions given the parameter values used in our experiment. In doing so, we consider only (noisy) symmetric Nash equilibrium for now. To compare with the data, we first estimate individual strategies that are restricted to be in form of cutoff strateies. Specifically, each player is assumed to have a cutoff value, so that she votes for blue in case of observing higher number of blue cards than this cutoff and for yellow otherwise. Since we have a direct-response method, these cutoff strategies must be estimated from observed behaviors. Thus, given 12 plays in a phase after excluding the first three periods, we estimate by maximum likelihood method a cutoff strategy for each player for each phase, assuming behaviors to follow a logit probabilistic function with individual-specific error parameters that are common across phases.[10] Table 6.3.2 provides the cutoff strategy of the unique symmetric Nash equilibrium for each phase.[11] We don't have experimental observations with unbiased prior and group sizes 9 and 19.

Session	Payoff	$n = 5$	$n = 9$	$n = 19$	#
$1 - 7$	900 : 200	4.43	4.67	4.84	140
A, B	800 : 300	4.66	4.84	4.93	40
A, B & $1 - 7$	500 : 500	5	–	–	180

Table 6.3.2.. The average symmetric equilibria cutoff strategies. Rightmost column is the number of observations.

The equilibrium strategies get closer to 5, the unbiased play, as the size of the groups increases. This feature also holds as the payoff bias decreases, for which if the group

10 The scale parameter values are also endogenously determined by the maximum likelihood estimations and are used throughout the analysis of the experimental data as individual's logistic error values.

11 These values are averages for each phase. The values calculated by using each session's average errors can be found in the Table D.1.1.

size is 5 we also have unbiased case. Table 6.3.3 below shows the predicted accuracy of group decisions for each phase.

Session	Payoff	$n = 5$	$n = 9$	$n = 19$
$1 - 7$	$900 : 200$	0.833	0.913	0.980
A, B	$800 : 500$	0.868	0.935	0.987
A, B & $1 - 7$	$500 : 500$	0.879	–	–

Table 6.3.3.. The average predicted accuracy of group decisions given equilibrium strategies.

As discussed before, the group accuracy in each case increases as the group size increases. Also, as the prior bias decreases we see that the group accuracy increases, for which if the group size is 5 we also have unbiased case. We will investigate the experimental data in light of these comparisons at both the individual and group level. Table 6.3.3 gives the global averages, session specific averages can be found in detail in Table D.1.1.

6.4 EXPERIMENTAL RESULTS

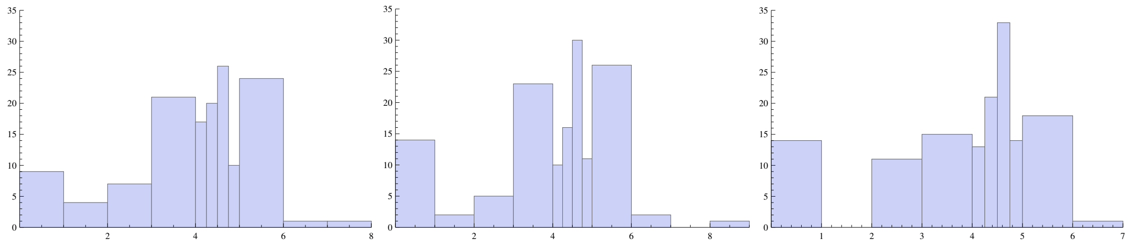
In this section we present and analyze our experimental results by investigating behaviors of our subjects at both individual level and group level.

6.4.1 *Individual behavior*

In the whole experiment, only five percent of the subjects are observed to behave that seems to be different than a cutoff strategy. Figure 6.4.1 below shows the histograms for distributions of estimated cutoff values from those sessions we take into consideration throughout the analysis, namely Sessions 1 to 7.

The frequencies of the interval $[4, 5)$ for group sizes 5,9 and 19 are 73%, 67% and 81%, respectively. On the other hand, the frequencies of the interval $[0, 1)$ are 9%, 14% and 14%. The frequencies of cutoffs that are higher than or equal to 5 are 26%, 29% and 19%. We want to emphasize at this point that these two latter significant frequencies constitute the major subject of our analysis. Table 6.4.1 below provides the averages for each phase.

We see that as the payoff bias increases the average biases towards the favored alternative (namely, blue) increases, which resonates with Table 6.3.2. This observation extends also to the comparisons including the unbiased phase.



(a) $n = 5$
Average $= 4.06$

(b) $n = 9$
Average $= 3.99$

(c) $n = 19$
Average $= 3.92$

Figure 6.4.1.. The global histograms of the cutoff strategies for each phase within Sessions $1-7$. Hence we have, in total, 140 observations for each case where the payoffs are $900 : 200$. Note that the intervals $[4, 5)$, which are the most voluminous, are divided into four equal subintervals.

Session	Payoff	$n = 5$	$n = 9$	$n = 19$
$1-7$	$900 : 200$	4.06	3.99	3.92
A, B	$800 : 300$	4.26	4.46	4.34
A, B & $1-7$	$500 : 500$	4.85	n/a	n/a

Table 6.4.1.. The averages for estimated cutoff strategies.

However, the claim that the bias in strategies towards the favored alternative should decrease as groups get larger due to pivotality approach argument, which is suggested by our theoretical predictions as well, occurs to be rejected for our data. Consider the first row of Table 6.4.1, which shows the averages for the cutoff estimations for each biased payoff phase in Sessions 1 to 7.[12] The average cutoff values in data appears to be slightly decreasing with the size of the group in the case of highly biased payoffs. Furthermore, although the comparisons in between group sizes 5 and 9, and 5 and 19 conform with theoretical predictions, the comparison between the sizes 9 and 19 differs highly. Finally, relying on the comparison within columns -especially the first one- of Table 6.4.1, we can say that our subjects appear to react in a way predicted by the theory to the biases in payoffs.

6.4.2 *Group decision accuracy*

Table 6.4.2 below shows the averages for observed accuracies of group decisions.[13]

12 For session-specific phase averages for cutoff estimations, see Table D.1.2.
13 For session-specific phase averages for group decision accuracies, see Table D.1.3.

Session	Payoff	$n = 5$	$n = 9$	$n = 19$
$1 - 7$	$900 : 200$	0.776	0.833	0.774
A, B	$800 : 300$	0.813	0.833	0.958
A, B & $1 - 7$	$500 : 500$	0.806	n/a	n/a

Table 6.4.2.. The averages for observed frequencies of correct group decisions.

We see some apparent discrepancies with the theoretical predictions. First thing to note is that in all cases groups do worse than theory's predictions. Furthermore, comparing first rows of Tables 6.3.3 and 6.4.2, we see that the Condorcet property cannot be confirmed with our data. Namely, as the size of the group increases we should see an increase in the accuracy of group decisions. Although significantly different from the absolute values of theoretical predictions, we see increase in accuracy, as expected, as moving from group size 5 to 9 or 5 to 19. More important is the fact that the observed accuracy of group decision in the case of 19 players is the lowest of the three. However, if we run a difference in proportions test for the Sessions $1 - 7$, the differences don't seem to be significant.

Guarnaschelli et al. (2000) also report decreasing accuracy with larger juries under majority rule and binary signals. They conclude that group accuracy might not be as robust as to accommodate small changes in individual behavior. However, in our experimental data, since we see discrepancies from theoretical predictions in individual behaviors as well we cannot relate inaccuracy observations fully to vulnerability.

The fall in the group accuracy when moving from the $800 : 300$ to unbiased can be explained by the facts that the unbiased phases were the very first phases of each session and if we look within session values it disappears.[14]

6.4.3 *Cognitive hierarchy models*

6.4.3.1 *Level-k approach*

Since a Lk player is reacting against her belief of $n - 1$ players playing a $k - 1$ level strategy, the need for playing in an opposite manner that is necessary to correct for the bias of the Lk-1 is amplified. So, if $L0$ strategy is to play a cutoff rule that is biased towards an alternative, the $L1$ can do best only by playing in an opposite way and in

14 Specifically, 0.885 is the average of the frequencies of correct decisions in the first phases of the two sessions (A and B).

higher magnitude.[15] Thus, a $L1$ voter will use strategy 10, under her belief that level-0 play is 0. The same argument applies and $L2$ plays 0. This oscillation continues, of course, to the infinity.

Level-0	Level-1	Level-2
0	10	0

Table 6.4.3.. Strategies under level$-k$ approach.

We will fix the strategy 0 to represent the level-0 play throughout our analysis of cognitive hierarchy models since this appears as a natural candidate: the bias towards voting for blue created by payoff scheme may tempt the easy or unsophisticated thinking of the problem. To confirm a level$-k$ thinking approach then, we should see only 0s and 10s in our observations, a nonexistent phenomenon, so we do not attempt at explaining our experimental data with such model. Battaglini et al. (2010) are the first to observe the anomaly of level$-k$ thinking model in Condorcet jury models.

6.4.3.2 *Poisson-CH model*

The Poisson-CH model postulates that a CHk player is reacting against her belief of $n-1$ players coming from a truncated Poisson distribution of levels up to her own level. The $CH1$ play is based on beliefs that are exactly the same with the level$-k$ approach, hence it is to have cutoff strategy 10 if $CH0$ is to play the strategy 0.

To calculate predictions for Poisson-CH model, we need to specify the Poisson parameter, τ, and the upper bound for how many levels exist. For the latter, relying on the previous experimental findings and the fact that it is enough to see some sort of convergence, we restrict attention to the case where we have only up to 2 levels of cognitive hierarchy. For the former we find the best fitting parameter within the interval $[0,3]$.

Table 6.4.4 shows the plays of second level voters, predicted by the Poisson-CH model for sessions 1 to 7, as well as the log likelihood values if we are to explain data by the relevant specification. The log-likelihood values are observed to monotonically increase within $\tau \in [0,3]$, so we provide in Table 6.4.4 the boundary case, i.e., $\tau = 3$.

15 Not that if level-0 play is assumed to be to vote for blue all the time, a $L1$ voter will see that she is never pivotal and will be indifferent. However, since we base our analysis on cutoff strategies that take the form of logistic functions, we always have nonzero probability for being pivotal.

	$n = 5$		$n = 9$		$n = 19$	
Session	$CH2$	LL	$CH2$	LL	$CH2$	LL
1	3.176	-89.275	3.156	-107.466	3.101	-111.062
2	2.440	-73.222	2.330	-76.270	2.049	-83.437
3	3.040	-80.131	3.006	-89.057	2.916	-94.087
4	2.774	-78.152	2.712	-74.629	2.560	-78.758
5	3.173	-153.911	3.152	-93.234	3.097	-97.726
6	2.834	-69.145	2.779	-74.191	2.643	-79.065
7	3.040	-74.943	3.006	-84.370	2.917	-88.523

Table 6.4.4.. The $CH2$ strategies and the log-likelihood values when level-0 is 0, $CH1$ is 10 and the Poisson parameter is 3 under the assumption that there are only up to two levels of cognitive hierarchy existing in the group.

6.4.4 *Endogenous Cognitive Hierarchy model*

For the analysis of ECH model, we keep the restrictions applied to the Poisson-CH. Hence, the Poisson parameter is fixed to lie in the interval $[0, 3]$ and the level distribution is restricted to include only up to level-2.

Table 6.4.5 shows the best fitting ECH models for the Phase 2, under aforementioned restrictions. As seen in the second column, most of the sessions have the feature of being better a fit with highest possible τ, namely 3. Sessions 2 and 4 are exceptions, where we see lower τ values that makes the model fit better, which is partly due to the relatively high frequencies of cutoff estimates that are lower than 4. In particular, we have five subjects with cutoffs lower than 3 in both Sessions 2 and 4, and two subjects with approximately 0 in the Session 2 while it is three in Session 4.

Session 1 also has two subjects with cutoff value 0. However there is only one more subject with a cutoff value lower than 3. The other sessions have at most one subject with a cutoff value of 0.

Letting τ be larger than 3, but still below 10, we have the following further observations for Phase 2. Sessions 1,3 and 7 can be explained best with the truncated Poisson ECH model with a Poisson parameter lower than 10. These are 5 for Sessions 1 and 3 with log-likelihoods -26.12 and -25.45, respectively. For Session 7, $\tau = 7$ gives the lowest LL value which is -32.69. Sessions 5 and 6, hits the boundary, i.e., 10, in terms of best fitting τ, which explains the comparison between ECH and NE in Table 6.4.8.

When we move to the phase 3, namely the case $n = 9$, we have a similar picture. Table 6.4.6 provides the model specifications and log-likelihood values for phase 3. First thing to note is that only the model specification for Session 4 fits best the data when τ is

Session	τ^*	Level 0	Level 1	Level 2	LL
1	3	0	4.705	4.535	-26.291
2	2.2	0	4.636	4.425	-39.647
3	3	0	4.687	4.508	-25.739
4	2	0	4.726	4.519	-32.101
5	3	0	4.704	4.534	-23.688
6	3	0	4.652	4.464	-28.340
7	3	0	4.688	4.508	-33.976

Table 6.4.5.. ECH model specifications when for the case $n = 5$ with log-likelihood values.

lower than the bound we imposed, namely 3. This occurs at around the same value as in the previous phase, namely 2.1. This is not observed to be the case for Session 2.

Session	τ^*	Level 0	Level 1	Level 2	LL
1	3	0	5.123	4.761	-33.818
2	3	0	5.047	4.690	-33.68
3	3	0	5.111	4.756	-25.238
4	2.1	0	5.175	4.792	-35.984
5	3	0	5.123	4.761	-33.237
6	3	0	5.090	4.731	-38.772
7	3	0	5.111	4.756	-41.256

Table 6.4.6.. ECH model specifications when for the case $n = 9$ with log-likelihood values.

Now consider the conjecture that the best fitting τ will be decreasing as the group size increases. Clearly, data from Session 2 contradict already with this and Session 4 does not represent a clear difference. As noted above, the Sessions 1 and 3 had 5 as the best fitting τ value. While it is not lower than 5 for Session 3, for Session 1 we have 4.5 now, complying with the argument.

The best fitting Poisson parameter for Session 5 is not lower in Phase 3 than its value in Phase 2. In fact, this phase continues to be better explained by NE. See Table 6.4.8.

On the other hand, significantly lower τ values specify the best fitting ECH models for Sessions 6 and 7. These values are 4 and 6, respectively, in Phase 3, whereas these were 10 and 7 in Phase 2.

Moving to the final phase, we do not see anymore Poisson parameter values of best fitting ECH models that are lower than our first upper bound of 3. Furthermore, none of the sessions have the complying feature, comparing to Phase 3 values. Table 6.4.7 is devoted to Phase 4.

Session	τ^*	Level 0	Level 1	Level 2	*LL*
1	3	0	5.4657	4.927	-29.06
2	3	0	5.4524	4.893	-37.529
3	3	0	5.4634	4.9	-47.63
4	3	0	5.4586	4.9	-38.449
5	3	0	5.4657	4.927	-41.454
6	3	0	5.4597	4.9	-38.114
7	3	0	5.4634	4.9	-36.027

Table 6.4.7.. ECH model specifications when for the case $n = 19$ with log-likelihood values.

6.4.5 *Which model fits the best?*

In this section we compare the performances of models in explaining our experimental data. First, note that in all sessions ECH fits better than CH. In some sessions and phases we have NE fitting better than ECH, although not plenty. But it is quick to observe that there exists high enough τ that could result in ECH dominating NE. This is due to the fact that as τ increases, the level-2 becomes dominant and at the limit only level-2 remains which is basically the symmetric Nash play.

	$n = 5$			$n = 9$			$n = 19$		
Session	CH	ECH	NE	CH	ECH	NE	CH	ECH	NE
1	-89.27	-26.29	-53.31	-107.46	-33.82	-58.30	-111.06	-29.06	-34.42
2	-73.22	-39.64	-46.52	-76.27	-33.68	-45.40	-83.43	-37.53	-51.35
3	-80.13	-25.73	-36.74	-89.05	-25.24	-31.07	-94.08	-47.63	-35.85
4	-78.15	-32.10	-53.91	-74.62	-35.98	-72.28	-78.75	-38.45	-70.04
5	-153.91	-23.68	-21.32	-93.23	-33.24	-24.06	-97.72	-41.45	-47.55
6	-69.14	-28.34	-25.10	-74.19	-38.77	-51.13	-79.06	-38.11	-40.14
7	-74.94	-33.97	-41.85	-84.37	-41.26	-49.07	-88.52	-36.03	-54.26

Table 6.4.8.. Comparison of log-likelihoods of the models. Here, τ is restricted to $[0, 3]$ and the best fitting one is taken, for both of ECH and CH.

As discussed above also, in Phase 2, data belonging to Sessions 5 and 6 can be explained better by NE, if $\tau \leq 10$, although the difference is not large. Only Session 5 carries this comparison to Phase 3, and none to Phase 4. However, Phase 4 of Session 3 appears to be explained better by NE. Thus, we have ECH dominating the two other models in 18 cases out of 21, under current restrictions on τ.

6.4.5.1 *Group Accuracy under ECH*

Phase	1	2	3	4	5	6	7	ave.
n=5	0.824	0.730	0.824	0.752	0.844	0.822	0.832	0.804
n=9	0.904	0.852	0.903	0.839	0.918	0.887	0.908	0.887
n=19	0.979	0.946	0.977	0.957	0.984	0.969	0.978	0.970

Table 6.4.9.. The predicted group accuracy for each phase in each session under best fitting ECH models.

Table 6.4.9 shows the group accuracy predictions of the best fitting ECH models. Comparing with the first row of Table 6.4.2, the last column of Table 6.4.9 does not appear to be a good predictor for the observed group accuracy. This holds true for the session-wise comparisons as well. But it should be noted that the differences in observations failed significance tests. Nonetheless, ECH gives predictions that are closer than NE predictions to data for any session. The average values for the latter is already presented in Table 6.3.3.

Furthermore, as discussed in Section 2.3, the Poisson distribution assumption plays a significant role here. As can be seen also in Figure 6.2.2, for all high enough values of τ, for instance, we have level-1 having higher share in the distribution compared to level-0 which clearly restricts the estimations. Removing any such assumption on distribution and estimating using maximum likelihood methods from data might give better results.[16] It should be noted that current analysis avoids -in a way- overfitting that would emerge under such approach.

6.5 DISCUSSION ON ECH AND CH MODELS

In certain classes of games, ECH behaviors differ significantly from those of CH.

6.5.1 *Perfect substitution*

Suppose that the game has perfect strategic substitutability. The best reply of a player is to correct the sum of the biases caused by the other players' strategies:

$$BR_i\left(s_{-i}\right) = -\sum_{j \neq i} s_j.$$

16 We obtained some better fitting *Poisson-free* models of ECH for some sessions, these can be provided upon request.

Proposition 6.5.1. *Suppose that the game has perfect strategic substitutability. Let b_k be the bias from symmetric Nash strategy. Then, for any $b_0 \neq 0$, the ratio b_k/b_0 grows in the order of n^k for Lk and CHk while it grows in the order of n^0 for ECHk.*

Proof of the proposition can be found in Appendix D.2.

6.5.2 *Cournot competition*

Consider a standard Cournot competition with linear demand and constant marginal cost of production which represents a game with strategic substitutability. We have that the best response function is a contraction if $n = 2$, an expansion if $n > 3$. In particular, let the best reply function be $q' = \frac{1}{2}(1 - Q_{-i})$. Since the profit function is quadratic, we have the best response function under uncertainty as a function of the expectation:

$$q' = \frac{1}{2}\left(1 - \mathbb{E}\left[Q_{-i}\right]\right).$$

The Nash equilibrium production level is $q = \frac{1}{n+1}$.

$\hookrightarrow$ *Level-k Approach*

Let q_0 denote $L0$ play. Then $L1$'s best response is to correct for the bias created by $n-1$ others:

$$q_1 = \frac{1}{2}\left(1 - (n-1)\, q_0\right).$$

Similarly, $L2$'s best response is:

$$\begin{aligned} q_2 &= \frac{1}{2}\left(1 - (n-1)\frac{1}{2}\left(1 - (n-1)\, q_0\right)\right), \\ &= \frac{1}{2}\left(1 - \frac{1}{2}(n-1)\right) + \frac{(n-1)^2}{4} q_0. \end{aligned}$$

Hence $L2$ will do best by reacting in an opposite and amplified way compared to $L1$. So $L2$'s reaction will be destined to be biased at least as much as and in the direction of $L0$. Naturally, $L3$ will be tending towards hitting the other corner. Iterating along in k, a bang-bang type oscillation in strategy space is required.

$\hookrightarrow$ *CH Model*

By definition, $CH1$ is the same as $L1$:

$$q_1 = \frac{1}{2}\left(1-(n-1)\,q_0\right).$$

For $CH2$, we have

$$\begin{aligned} q_2 &= \frac{1}{2}\left(1-(n-1)\left(\frac{f_0}{f_0+f_1}q_0+\frac{f_1}{f_0+f_1}q_1\right)\right), \\ &= \frac{1}{2}\left(1-\frac{(n-1)f_1}{2(f_0+f_1)}\right)-(n-1)\frac{f_0-(n-1)f_1}{2(f_0+f_1)}q_0, \end{aligned}$$

where f_0 and f_1 represent the weights of $CH0$ and $CH1$ under global distribution f, respectively.

↪ *ECH Model*

An $ECH1$ player's belief assigns weighted frequencies to $ECH0$ and $ECH1$ that are the same with a $CH2$ belief above. Hence $ECH1$ best responds as

$$q_1 = \frac{1}{2}\left(1-(n-1)\left(\frac{f_0}{f_0+f_1}q_0+\frac{f_1}{f_0+f_1}q_1\right)\right),$$

which can be rearranged as

$$\left(2+(n-1)\,\frac{f_1}{f_0+f_1}\right)q_1 = \left(1-(n-1)\,\frac{f_0}{f_0+f_1}q_0\right).$$

Similarly found is $ECH2$'s best response:

$$q_2 = \frac{1}{2}\left(1-(n-1)\left(g_2(0)q_0+g_2(1)q_1+g_2(2)q_2\right)\right).$$

That is,

$$q_2 = \frac{1-(n-1)\left(g_2(0)q_0+g_2(1)q_1\right)}{2+(n-1)\,g_2(2)}.$$

We have symmetric Nash equilibria where $q^* = 0.0909$. Suppose now that f follows Poisson distribution with coefficient $\tau = 3$ and let $n = 10$. Tables 6.5.1 and 6.5.2 give the model specifications when $q_0 = 0.05$ and $q_0 = 0.2$ respectively.

	q_0	q_1	q_2
level$-k$	0.05	0.275	0
CH	0.05	0.275	0
ECH	0.05	0.101	0.092

Table 6.5.1.. Model specifications for Cournot competition when $q_0 = 0.05$, $\tau = 3$ and $n = 10$.

	q_0	q_1	q_2
level$-k$	0.2	0	0.5
CH	0.2	0	0.275
ECH	0.2	0.063	0.087

Table 6.5.2.. Model specifications for Cournot competition when $q_0 = 0.2$, $\tau = 3$ and $n = 10$.

6.5.3 *Keynesian Beauty contest*

On the other hand, in the games where the best reply function is a contraction mapping, behavior under ECH do not differ much from those under CH. Consider a standard Keynesian beauty contest game, a game with strategic complementarity, where n players simultaneously submit a number between 0 and 100, and the winner is the one who chooses a number closest to the $0 < p < 1$ times average of all submissions. The best response function is a contraction mapping defined as $BR_i(x_{-i}) = \frac{p}{n-p}\sum_{j\neq i} x_j$. Suppose that the utility function is quadratic as a function of the distance from the winning number. Then the best reply under uncertainty is the expected winning number:

$$BR_i(x_{-i}) = \frac{p}{n-p}\mathbb{E}\left[\sum_{j\neq i} x_j\right].$$

hence, the unique NE is $x = 0$.

$\hookrightarrow$ *Level$-k$ Approach*
$L1$'s best response is:

$$b_1 = p\frac{n-1}{n-p}b_0.$$

$L2$'s best response is:

$$b_2 = p\frac{n-1}{n-p}b_1 = \left(p\frac{n-1}{n-p}\right)^2 b_0.$$

The strategies converge to Nash equilibrium levels.

$\hookrightarrow$ *CH Model*
By definition, $CH1$ is the same as $L1$:

$$b_1 = p\frac{n-1}{n-p}b_0..$$

For $CH2$:

$$\begin{aligned} b_2 &= p\frac{n-1}{n-p}\left(\frac{f_0}{f_0+f_1}b_0 + \frac{f_1}{f_0+f_1}b_1\right), \\ &= p\frac{n-1}{n-p}\left(\frac{f_0}{f_0+f_1} + \frac{f_1}{f_0+f_1}p\frac{n-1}{n-p}\right)b_0. \end{aligned}$$

$\hookrightarrow$ *ECH Model*
$ECH1$'s best response is:

$$b_1 = p\frac{n-1}{n-p}\left(\frac{f_0}{f_0+f_1}b_0 + \frac{f_1}{f_0+f_1}b_1\right).$$

Hence,

$$\left(1 - p\frac{n-1}{n-p}\frac{f_1}{f_0+f_1}\right)b_1 = p\frac{n-1}{n-p}\frac{f_0}{f_0+f_1}b_0.$$

$ECH2$'s best response is:

$$b_2 = p\frac{n-1}{n-p}\left(g_2(0)b_0 + g_2(1)b_1 + g_2(2)b_2\right).$$

Hence,

$$b_2 = \frac{p\frac{n-1}{n-p}\left(g_2(0)b_0 + g_2(1)b_1\right)}{1 - p\frac{n-1}{n-p}g_2(2)}.$$

The unique Nash equilibrium strategy when $p < 1$ is 0. Now suppose that f follows Poisson distribution with coefficient $\tau = 3$ and let $n = 10$. Tables 6.5.3 and 6.5.4 give the model specifications when $p = 2/3$ and $p = 1/2$ respectively.

	b_1/b_0	b_2/b_0
level$-k$	0.643	0.413
CH	0.643	0.470
ECH	0.310	0.221

Table 6.5.3.. Model specifications for the Keynesian beauty contest when $p = 2/3$, $\tau = 3$ and $n = 10$.

	b_1/b_0	b_2/b_0
level$-k$	0.474	0.224
CH	0.474	0.287
ECH	0.184	0.115

Table 6.5.4.. Model specifications for the Keynesian beauty contest when $p = 1/2$, $\tau = 3$ and $n = 10$.

6.6 CONCLUSION

We have delivered results from an experiment designed to test the effect of group size on decision accuracy. The experimental results suggest systematic heterogeneity among individual behaviors. We provide a cognitive hierarchy model that advances upon the current ones by allowing self-level consideration by individuals to account for this systematic deviation from symmetric equilibrium thinking. This model is theoretically shown to be able to accommodate our observations and performs better than previous models of cognitive hierarchy as well as symmetric Bayesian Nash equilibrium in explaining individual behavior.

The modeling consists of specification of distribution for cognitive hierarchy levels. In this paper we employed Poisson distribution for that regard. This specification hints that the bounded rationality of a group that is related to the sophistication induced by the size of the group resonates with the best fitting Poisson parameter. In other words, we would expect to see that as group size increases, since the problem in hand would get more demanding, nonstrategic play that is seen to appear in a systematic way should be amplified and hence best fitting Poisson parameters would be lower. And if this is severe enough, we could capture the decrease in group decision accuracy. However, we don't see this in our data. Consequently, the model falls short of explaining lower group decision accuracy in larger groups. It should be noted, on the other hand, that the particular restriction of Poisson distribution assumption plays a role here and dropping that would lead to get closer to the group bounded rationality argument above, since we already see increase in tendency towards lower levels.

In current experimental design learning and ordering effects were not targets of investigation. Battaglini et al. (2010) and Bhattacharya et al. (2013) reports no observation of significant learning effects. Furthermore, we excluded first three periods of all treatments in our analysis and implemented a random matching design so that at each period subjects knew that their group is randomly formed with possibly different members.

BIBLIOGRAPHY

Abramowitz, A. and Saunders, K. (2005). Why cant we all just get along? the reality of a polarized america. In *The Forum*, volume 3, pages 1–22. bepress.

Abramowitz, A. I. (2006). Disconnected, or joined at the hip? *Red and Blue Nation?: Characteristics and Causes of America's Polarized Politics*, 1:72–85.

Abramowitz, A. I. (2010). *The disappearing center: Engaged citizens, polarization, and American democracy*. Yale University Press.

Abramowitz, A. I. and Saunders, K. L. (2006). Exploring the bases of partisanship in the american electorate: Social identity vs. ideology. *Political Research Quarterly*, 59(2):175–187.

Abramowitz, A. I. and Saunders, K. L. (2008). Is polarization a myth? *The Journal of Politics*, 70(02):542–555.

Abramowitz, A. I. and Stone, W. J. (2006). The bush effect: Polarization, turnout, and activism in the 2004 presidential election. *Presidential Studies Quarterly*, 36(2):141–154.

Abrams, S. J. and Fiorina, M. P. (2012). the big sort that wasn't: A skeptical reexamination. *PS: Political Science & Politics*, 45(02):203–210.

Alcalde-Unzu, J. and Vorsatz, M. (2008). The measurement of consensus: an axiomatic analysis. *Documento de Trabajo*, 2008:28.

Alcalde-Unzu, J. and Vorsatz, M. (2013). Measuring the cohesiveness of preferences: an axiomatic analysis. *Social Choice and Welfare*, 41(4):965–988.

Aldrich, J. H. and McKelvey, R. D. (1977). A method of scaling with applications to the 1968 and 1972 presidential elections. *The American Political Science Review*, pages 111–130.

Alesina, A., Baqir, R., and Easterly, W. (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics*, 114(4):1243–1284.

Armstrong, D., Bakker, R., Carroll, R., Hare, C., and Poole, K. T. (2014). Using bayesian aldrich-mckelvey scaling to study citizens ideological preferences and perceptions. *Working Paper*.

Austen-Smith, D. and Banks, J. S. (1996). Information aggregation, rationality, and the condorcet jury theorem. *American Political Science Review*, pages 34–45.

Baker, W. E. (2005). *America's crisis of values: Reality and perception*. Princeton University Press.

Baldiga, K. A. and Green, J. R. (2013). Assent-maximizing social choice. *Social Choice and Welfare*, 40(2):439–460.

Battaglini, M., Morton, R. B., and Palfrey, T. R. (2008). Information aggregation and strategic abstention in large laboratory elections. *The American Economic Review*, pages 194–200.

Battaglini, M., Morton, R. B., and Palfrey, T. R. (2010). The swing voter's curse in the laboratory. *The Review of Economic Studies*, 77(1):61–89.

Benoit, K. and Laver, M. (2006). *Party policy in modern democracies*. Routledge.

Bernheim, B. D. (1984). Rationalizable strategic behavior. *Econometrica: Journal of the Econometric Society*, pages 1007–1028.

Bhattacharya, S., Duffy, J., and Kim, S. (2013). Information acquisition and voting mechanisms: Theory and evidence. Technical report, Working paper, University of Pittsburgh.

Binder, S. A. (2000). Going nowhere: A gridlocked congress? *The Brookings Review*, pages 16–19.

Binmore, K. (1975). An example in group preference. *Journal of Economic Theory*, 10(3):377–385.

Bogart, K. (1973). Preference structures i: Distances between transitive preference relations. *Journal of Mathematical Sociology*, 3(1):49–67.

Bosch, R. (2006). Characterizations on voting rules and consensus measures. *Ph.D.Dissertation,Tilburg University, The Netherlands*.

Bossert, W. and Sprumont, Y. (2014). Strategy-proof preference aggregation: Possibilities and characterizations. *Games and Economic Behavior*, 85:109–126.

Bossert, W. and Storcken, T. (1992). Strategy-proofness of social welfare functions: the use of the kemeny distance between preference orderings. *Social Choice and Welfare*, 9(4):345–360.

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Camerer, C. F., Ho, T.-H., and Chong, J.-K. (2004). A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, pages 861–898.

Can, B. (2012). Weighted distances between preferences. Technical report, Maastricht: METEOR, Maastricht Research School of Economics of Technology and Organization.

Can, B. and Storcken, T. (2013). A re-characterization of the kemeny distance. *mimeo*.

Clark, T. S. (2009). Measuring ideological polarization on the united states supreme court. *Political Research Quarterly*, 62(1):146–157.

Costinot, A. and Kartik, N. (2007). On optimal voting rules under homogeneous preferences. Technical report, working paper.

Crawford, V. P., Costa-Gomes, M. A., and Iriberri, N. (2013). Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature*, 51(1):5–62.

Crepaz, M. M. (1990). The impact of party polarization and postmaterialism on voter turnout. *European Journal of Political Research*, 18(2):183–205.

Davis, N. J. and Robinson, R. V. (1997). A war for americas soul? the american religious landscape. *Cultural wars in American politics: Critical reviews of a popular myth*, pages 39–61.

DiMaggio, P., Evans, J., and Bryson, B. (1996). Have american's social attitudes become more polarized? *American journal of Sociology*, pages 690–755.

Downey, D. J. and Huffman, M. L. (2001). Attitudinal polarization and trimodal distributions: measurement problems and theoretical implications. *Social Science Quarterly*, 82(3):494–505.

Downs, A. (1957). *An economic theory of democracy*. New York: Harper Collins Publishers.

Duclos, J.-Y., Esteban, J., and Ray, D. (2004). Polarization: concepts, measurement, estimation. *Econometrica*, 72(6):1737–1772.

Duddy, C., Perote-Peña, J., and Piggins, A. (2010). Manipulating an aggregation rule under ordinally fuzzy preferences. *Social Choice and Welfare*, 34(3):411–428.

Epstein, D. and Graham, J. D. (2007). *Polarized politics and policy consequences*. Rand Corporation.

Esponda, I. and Vespa, E. (2013). Hypothetical thinking and information extraction: Strategic voting in the laboratory. Technical report, Mimeo, NYU Stern.

Esteban, J.-M. and Ray, D. (1994). On the measurement of polarization. *Econometrica*, 62:819–819.

Evans, J. H. (2003). Have americans' attitudes become more polarized?an update*. *Social Science Quarterly*, 84(1):71–90.

Feddersen, T. and Pesendorfer, W. (1997). Voting behavior and information aggregation in elections with private information. *Econometrica: Journal of the Econometric Society*, pages 1029–1058.

Fiorina, M. P. and Abrams, S. A. (2010). Where's the polarization? *Controversies in voting behavior, edited by Niemi, Richard G and Weisberg, Herbert F and Kimball, David C*, pages 309–318.

Fiorina, M. P., Abrams, S. A., and Pope, J. C. (2008). Polarization in the american public: Misconceptions and misreadings. *The Journal of Politics*, 70(02):556–560.

Fiorina, M. P. and Abrams, S. J. (2008). Political polarization in the american public. *Annu. Rev. Polit. Sci.*, 11:563–588.

Fiorina, M. P., Abrams, S. J., and Pope, J. (2005). *Culture war?* Pearson Longman New York.

Fiorina, M. P. and Levendusky, M. S. (2006). Disconnected: The political class versus the people. *Red and Blue Nation?: Characteristics and Causes of America's Polarized Politics*, 1:49–72.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178.

García-Lapresta, J. L. and Pérez-Román, D. (2011). Measuring consensus in weak orders. In *Consensual Processes*, pages 213–234. Springer.

Gerling, K., Grüner, H. P., Kiel, A., and Schulte, E. (2005). Information acquisition and decision making in committees: A survey. *European Journal of Political Economy*, 21(3):563–597.

Guarnaschelli, S., McKelvey, R. D., and Palfrey, T. R. (2000). An experimental study of jury decision rules. *American Political Science Review*, pages 407–423.

Gurer, E. (2008). Do impossibility results survive in historically standard domains? Master's thesis, Bilkent University.

Harbridge, L. M. (2009). *Bipartisanship in a polarized Congress*. Stanford University.

Herrera-Viedma, E., García-Lapresta, J. L., Kacprzyk, J., Fedrizzi, M., Nurmi, H., and Zadrożny, S. (2011). *Consensual processes*, volume 267. Springer.

Hetherington, M. J. (2008). Turned off or turned on? how polarization affects political engagement. *Red and blue nation*, 2:1–33.

Hetherington, M. J. (2009). Review article: Putting polarization in perspective. *British Journal of Political Science*, 39(02):413–448.

Hill, S. J. and Tausanovitch, C. (2014). Non-polarization in the american public. *Working Paper*.

Hunter, J. D. (1992). *Culture wars: The struggle to control the family, art, education, law, and politics in America*. Basic Books.

Igersheim, H. (2007). Du paradoxe libéral-parétien à un concept de métaclassement des préférences. *Recherches économiques de Louvain*, 73(2):173–192.

Jacobson, G. C. (2006). Comment. *Red and Blue Nation?: Characteristics and Causes of America's Polarized Politics*, 1:85–95.

Jeffrey, R. C. (1974). Preference among preferences. *The Journal of Philosophy*, pages 377–391.

Jones, D. R. (2001). Party polarization and legislative gridlock. *Political Research Quarterly*, 54(1):125–141.

Kemeny, J. and Snell, L. (1962). Preference ranking: an axiomatic approach. *Mathematical models in the social sciences*, pages 9–23.

Kemeny, J. G. (1959). Mathematics without numbers. *Daedalus*, 88(4):577–591.

Koray, S. (2000). Self-selective social choice functions verify arrow and gibbard-satterthwaite theorems. *Econometrica*, 68(4):981–996.

Koray, S. and Slinko, A. (2008). Self-selective social choice functions. *Social Choice and Welfare*, 31(1):129–149.

Koray, S. and Unel, B. (2003). Characterization of self-selective social choice functions on the tops-only domain. *Social Choice and Welfare*, 20(3):495–507.

Koriyama, Y. and Szentes, B. (2009). A resurrection of the condorcet jury theorem. *Theoretical Economics*, 4(2):227–252.

Krehbiel, K. and Peskowitz, Z. (2012). Legislative organization and ideal-point bias. Technical report, Stanford University.

Laffond, G. and Lainé, J. (2000). Majority voting on orders. *Theory and Decision*, 49(3):249–287.

Layman, G. C. and Carsey, T. M. (2002). Party polarization and" conflict extension" in the american electorate. *American Journal of Political Science*, pages 786–802.

Layman, G. C., Carsey, T. M., Green, J. C., Herrera, R., and Cooperman, R. (2010). Party polarization, party commitment and conflict extension among american party activists. *American Political Science Review*, 104(2):324–346.

Levendusky, M. S. (2009). *The partisan sort: How liberals became Democrats and conservatives became Republicans*. University of Chicago Press.

McCarty, N., Poole, K. T., and Rosenthal, H. (2003). Political polarization and income inequality. *Unpublished paper*.

McCarty, N. M., Poole, K. T., and Rosenthal, H. (2006). *Polarized America: The dance of ideology and unequal riches*. mit Press Cambridge.

McCormick, J. M. and Wittkopf, E. R. (1990). Bipartisanship, partisanship, and ideology in congressional-executive foreign policy relations, 1947–1988. *The Journal of Politics*, 52(04):1077–1100.

McPherson, M. S. (1982). Mill's moral theory and the problem of preference change. *Ethics*, pages 252–273.

Montalvo, J. G. and Reynal-Querol, M. (2005). Ethnic polarization, potential conflict, and civil wars. *American economic review*, pages 796–816.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review*, pages 1313–1326.

Nivola, P. S. and Brady, D. W. (2006). *Red and Blue Nation?: Consequences and Correction of America's Polarized Politics*, volume 2. Brookings Institution Press.

Oosterwaal, A. and Torenvlied, R. (2010). Politics divided from society? three explanations for trends in societal and political polarisation in the netherlands. *West European Politics*, 33(2):258–279.

Palfrey, T. R. (2013). Experiments in political economy. *Handbook of Experimental Economics*, 2.

Poole, K., Rosenthal, H., Lewis, J., Lo, J., and Carroll, R. (2012). basicspace: A package to recover a basic space from issue scales. *R package version 0.05. URL: http://cran. r-project. org/web/packages/basicspace/index. html*.

Poole, K. T. (1998). Recovering a basic space from a set of issue scales. *American Journal of Political Science*, pages 954–993.

Prior, M. (2013). Media and political polarization. *Annual Review of Political Science*, 16:101–127.

Schofield, N., Claassen, C., Ozdemir, U., and Zakharov, A. (2011). Estimating the effects of activists in two-party and multi-party systems: comparing the united states and israel. *Social choice and welfare*, 36(3-4):483–518.

Sen, A. (1970). The impossibility of a paretian liberal. *The journal of political economy*, pages 152–157.

Sen, A. and Körner, S. (1974). Choice, ordering and morality. *Practical reason*.

Sen, A. K. (1977). Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy & Public Affairs*, pages 317–344.

Stahl, D. O. and Wilson, P. W. (1995). On players models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10(1):218–254.

Theriault, S. M. (2008). *Party polarization in congress*. Cambridge University Press.

A

APPENDIX TO CHAPTER 2

A.1 PROOF OF PROPOSITION 2.2.1

Let α be a neutral SFW that is SW self-selective for some domain $\mathcal{E}$. Pick up any profile $P_N \in \mathcal{L}(A_m)^n$ where $\Delta(\alpha(P_N)) = \{Q_1, ..., Q_H\}$, together with any $E = (e_1, ..., e_n) \in \mathcal{E}^n$. Consider the set of SWFs $\mathcal{A} = \{\alpha_1, \alpha_2, ..., \alpha_H, \rho_1, ..., \rho_{m!-H}\}$ such that:
- $\alpha_1(P_N) = Q_1, ..., \alpha_H(P_N) = Q_H$
- $\forall k \neq k' \in \{1, ..., m! - H\}$, $\alpha_k(P_N) \neq \alpha_{k'}(P_N)$
- $\cup_{1 \leq k \leq m!-H} \rho_k(P_N) = \mathcal{L}(A_m) \mid_{\Delta(\alpha(P_N))}$

Since all elements of $\mathcal{A}$ are strict SWFs, then $\mathcal{A}$ is a strict selection. Moreover, all elements of $\mathcal{A}$ having different outcomes from P_N, then $R_N^{E,\mathcal{A}}$ is a profile of linear orders over $\mathcal{A}$. Furthermore, $[\cup_{1 \leq h \leq H} \alpha_h(P_N)] \cup [\cup_{1 \leq k \leq m!-1} \rho_k(P_N)] = \mathcal{L}(A_m)$ implies that $R_N^{E,\mathcal{A}}$ is a profile over all $m!$ linear orders, so that $L(\alpha) \cap \mathcal{A} = L(\alpha)$. It follows from definition that $R_N^{E,\mathcal{A}}$ is isomorphic to P_N^E. Since α is SW self-selective for E, then $L(\alpha) \cap f_\alpha(\widetilde{P}_N^{E,\mathcal{A}}) \neq \emptyset$. Finally, since $R_N^{E,\mathcal{A}}$ is isomorphic to P_N^E, the neutrality of α ensures that $\Delta(\alpha(P_N)) \cap f_\alpha(P_N^E) \neq \emptyset$ and the conclusion follows.

A.2 PROOF OF THEOREM 2.3.1

We first establish three necessary conditions for Kemeny-stability.

Proposition A.2.1. *A scoring rule α is Kemeny-stable only if $s_\alpha^{2,3} > 0$ and $s_\alpha^{1,6} > s_\alpha^{2,6} = s_\alpha^{3,6} > s_\alpha^{4,6} = s_\alpha^{5,6} > 0$.*

Proposition A.2.2. *A scoring rule α is Kemeny-stable only if $s_\alpha^{1,6} = s_\alpha^{2,6} + s_\alpha^{4,6}$.*

Proposition A.2.3. *A scoring rule α is Kemeny-stable only if $s_\alpha^{1,3} = 2s_\alpha^{2,3}$.*

A.2.1 *Proof of Proposition A.2.1*

The proof is organized in six 6 intermediate lemmas:

Lemma A.2.1. *If α is a Kemeny-stable scoring rule, then $s_\alpha^{2,3} > 0$.*

Proof: Suppose that $s_\alpha^{2,3} = 0$, and consider the following profile $P_N \in \mathcal{L}(A_3)^{n_1+n_2+n_3+n_4}$, where $n_1 > n_2 > n_3 + n_4$, together with the following linear extension $\dot{P}_N$ of P_N^K:

$$P_N = \left(\begin{array}{cccc} n_1 & n_2 & n_3 & n_4 \\ \hline a & b & c & c \\ c & c & a & b \\ b & a & b & a \end{array} \right) \quad \dot{P}_N = \left(\begin{array}{cccc} n_1 & n_2 & n_3 & n_4 \\ \hline acb & bca & cab & cba \\ cab & bac & acb & cab \\ abc & cba & cba & bca \\ bac & cab & bca & acb \\ cba & abc & abc & bac \\ bca & acb & bac & abc \end{array} \right)$$

One gets that $\Delta(\alpha(P_N)) = \{abc\}$. Kemeny-stability requires that $S_\alpha^6(abc, \dot{P}_N) = n_1 s_\alpha^{3,6} + (n_2 + n_3) s_\alpha^{5,6} \geq S_\alpha^6(cab, \dot{P}_N) = (n_1 + n_4) s_\alpha^{2,6} + n_2 s_\alpha^{4,6} + n_3 s_\alpha^{1,6}$, hence that $n_1(s_\alpha^{3,6} - s_\alpha^{2,6}) + n_2(s_\alpha^{5,6} - s_\alpha^{4,6}) + n_3(s_\alpha^{5,6} - s_\alpha^{1,6}) \geq n_4 s_\alpha^{2,6}$, which is clearly impossible $\square$

Lemma A.2.2. *If α is a Kemeny-stable scoring rule, then $s_\alpha^{2,6} = s_\alpha^{3,6}$ and $s_\alpha^{4,6} = s_\alpha^{5,6}$.*

Proof: Suppose first that $s_\alpha^{1,3} > 2 s_\alpha^{2,3}$, and consider $P_N \in \mathcal{L}(A_3)^4$, and $\dot{P}_N \in \Delta(P_N^K)$:

$$P_N = \left(\begin{array}{cccc} 1 & 1 & 1 & 1 \\ \hline a & a & b & c \\ b & c & c & b \\ c & b & a & a \end{array} \right) \quad \dot{P}_N = \left(\begin{array}{cccc} 1 & 1 & 1 & 1 \\ \hline abc & acb & bca & cba \\ bac & cab & cba & bca \\ acb & abc & bac & cab \\ bca & cba & cab & bac \\ cab & bac & abc & acb \\ cba & bca & acb & abc \end{array} \right)$$

Since $S_\alpha^3(a, P_N) = 2 s_\alpha^{1,3}$, and $S_\alpha^3(b, P_N) = S_\alpha^3(c, P_N) = s_\alpha^{1,3} + 2 s_\alpha^{2,3}$, then $\Delta(\alpha(P_N)) = \{abc, acb\}$. Moreover, we have (1) $S_\alpha^6(abc, \dot{P}_N) = S_\alpha^3(acb, \dot{P}_N) = s_\alpha^{1,6} + s_\alpha^{3,6} + s_\alpha^{5,6}$, and (2) $S_\alpha^6(bca, \dot{P}_N) = s_\alpha^{1,6} + s_\alpha^{2,6} + s_\alpha^{4,6}$. Kemeny stability implies from (1) and (2) that $s_\alpha^{3,6} + s_\alpha^{5,6} \geq s_\alpha^{2,6} + s_\alpha^{4,6}$ (3), which in turn leads to $s_\alpha^{2,6} = s_\alpha^{3,6}$ and $s_\alpha^{4,6} = s_\alpha^{5,6}$.

Suppose now that $s_\alpha^{1,3} < 2 s_\alpha^{2,3}$, and consider profile P_N above and the hyper-profile $\dot{P}'_N \in \Delta(P_N^K)$ obtained from $\dot{P}_N$ by switching in each order alternatives respectively ranked (1) second and third, and (2) fourth and fifth. We get $\Delta(\alpha(P_N)) = \{bca, cba\}$, and we reach the same conclusion as above by a symmetric argument. Finally, suppose that $s_\alpha^{1,3} = 2s_\alpha^{2,3}$, and consider the profile $P_N \in \mathcal{L}(A_3)^{4Z-1}$ below, where $Z > 1$, together with the Kemeny weak profile P_N^K:

$$P_N = \begin{pmatrix} Z & Z & Z-1 & Z \\ \hline a & b & c & c \\ b & a & b & a \\ c & c & a & b \end{pmatrix} \quad P_N^K = \begin{pmatrix} Z & Z & Z-1 & Z \\ \hline abc & bac & cba & cab \\ acb, bac & abc, bca & bca, cab & cba, acb \\ bca, cab & cba, acb & acb, bac & abc, bca \\ cba & cab & abc & bac \end{pmatrix}$$

Then $\alpha(P_N) = abc$. Moreover, there exists $\dot{P}_N \in \Delta(P_N^K)$ such that $S_\alpha^6(abc, \dot{P}_N) = Z(s_\alpha^{1,6} + s_\alpha^{3,6} + s_\alpha^{5,6})$ and $S_\alpha^6(cab, \dot{P}_N) = Zs_\alpha^{1,6} + (Z-1)s_\alpha^{2,6} + Zs_\alpha^{4,6}$. Kemeny stability requires that $s_\alpha^{3,6} + s_\alpha^{5,6} \geq \frac{Z-1}{Z} s_\alpha^{2,6} + s_\alpha^{4,6}$ for all $Z > 1$. Thus, $s_\alpha^{2,6} + s_\alpha^{4,6} \leq s_\alpha^{3,6} + s_\alpha^{5,6}$, and hence $s_\alpha^{2,6} = s_\alpha^{3,6}$ and $s_\alpha^{4,6} = s_\alpha^{5,6}$ □

We assume in the sequel that α is such that $s_\alpha^{2,6} = s_\alpha^{3,6}$ and $s_\alpha^{4,6} = s_\alpha^{5,6}$ (property $(*)$). Clearly, $(*)$ implies that given any profile P_N over 3 alternatives, given any Kemeny-stable SWF α, one has $\alpha(\dot{P}_N) = \alpha(\widetilde{P_N})$ for any two $\forall \dot{P}_N, \widetilde{P_N} \in \Delta(P_N^K)$.

Lemma A.2.3. *If α is a Kemeny-stable scoring rule, then $[s_\alpha^{1,6} = s_\alpha^{2,6}] \Rightarrow [s_\alpha^{4,6} = s_\alpha^{5,6} > 0]$.*

Proof: Consider the following $P_N \in \mathcal{L}(A_3)^{3Z+W}$ below, where $Z, W \geq 1$ are chosen such that $W < \frac{s_\alpha^{2,3}}{s_\alpha^{1,3}} Z$:

$$P_N = \begin{pmatrix} Z & Z & Z & W \\ \hline a & b & c & a \\ b & a & b & c \\ c & c & a & b \end{pmatrix}$$

Then $\alpha(P_N) = bac$. Furthermore, using $(*)$ together with Kemeny stability and $s_\alpha^{1,6} = s_\alpha^{2,6}$, one must have $S_\alpha^6(bac, \dot{P}_N) = 2Zs_\alpha^{1,6} + (Z+W)s_\alpha^{5,6} \geq S_\alpha^6(abc, \dot{P}_N) = (2Z+W)s_\alpha^{1,6}$. Thus, $s_\alpha^{1,6} \leq \frac{Z+W}{W} s_\alpha^{5,6}$. Finally, since $s_\alpha^{1,6} > 0$, then $s_\alpha^{5,6} > 0$ □

Lemma A.2.4. *If α is a Kemeny-stable scoring rule, then $[s_\alpha^{1,6} = s_\alpha^{2,6}] \Rightarrow [2s_\alpha^{1,3} = 3s_\alpha^{2,3}]$.*

Proof: Define the two profiles $P_N \in \mathcal{L}(A_3)^5$ and $P'_N \in \mathcal{L}(A_3)^{3Z+1}$, where $Z > 1$, as follows:

$$P_N = \begin{pmatrix} 2 & 1 & 1 & 1 \\ \hline a & a & c & b \\ b & c & b & c \\ c & b & a & a \end{pmatrix} \quad P'_N = \begin{pmatrix} 2Z & 1 & Z \\ \hline a & c & c \\ b & a & b \\ c & b & a \end{pmatrix}$$

Suppose first that $2s_\alpha^{1,3} > 3s_\alpha^{2,3}$. It follows from $2s_\alpha^{1,3} > 3s_\alpha^{2,3}$ that $\alpha(P_N) = abc$. Using $(*)$, we have $s_\alpha^{1,6} = s_\alpha^{2,6} = s_\alpha^{3,6} \geq s_\alpha^{4,6} = s_\alpha^{5,6}$. Hence, $\forall \dot{P}_N \in \Delta(P_N^K)$, $S_\alpha^6(abc, \dot{P}_N) = 3s_\alpha^{1,6} + s_\alpha^{5,6}$, and $S_\alpha^6(bac, \dot{P}_N) = 3s_\alpha^{1,6} + 2s_\alpha^{5,6}$. Since Kemeny-stability requires $S_\alpha^6(abc, \dot{P}_N) \geq S_\alpha^6(bac, \dot{P}_N)$, then we get $s_\alpha^{5,6} = 0$, in contradiction with Lemma 3.

Similarly, suppose that $2s_\alpha^{1,3} < 3s_\alpha^{2,3}$. From $0 < 2s_\alpha^{1,3} < 3s_\alpha^{2,3}$, we get that $\alpha(P'_N) = bac$ for Z large enough. Moreover, $\forall \dot{P}_N \in \Delta(P_N'^K)$, $S_\alpha^6(bac, \dot{P}_N) = Z(2s_\alpha^{1,6} + s_\alpha^{5,6}) < S_\alpha^6(acb, \dot{P}_N) = Z(2s_\alpha^{1,6} + s_\alpha^{5,6}) + s_\alpha^{1,6}$, in contradiction with Kemeny stability $\square$

Lemma A.2.5. *If α is a Kemeny-stable scoring rule, then $s_\alpha^{1,6} > s_\alpha^{2,6}$.*

Proof: Suppose that $s_\alpha^{1,6} = s_\alpha^{2,6}$. From Lemma 3 and 4 together with $(*)$, we have $s_\alpha^{1,6} = s_\alpha^{2,6} = s_\alpha^{3,6}$, $2s_\alpha^{1,3} = 3s_\alpha^{2,3}$, and $s_\alpha^{4,6} = s_\alpha^{5,6} > 0$. Then, consider the following profile $P_N \in \mathcal{L}(A_3)^4$:

$$P_N = \left(\begin{array}{cccc} 2 & 1 & 1 & 1 \\ \hline a & b & c & c \\ b & a & b & a \\ c & c & a & b \end{array} \right)$$

Since $S_\alpha^3(a, P_N) = 2s_\alpha^{1,3} + 2s_\alpha^{2,3}$, $S_\alpha^3(b, P_N) = s_\alpha^{1,3} + 3s_\alpha^{2,3}$, and $S_\alpha^3(c, P_N) = 2s_\alpha^{1,3}$, then, using Lemma 1 and Lemma 4, $\alpha(P_N) = abc$. From Kemeny-stability, we have that for any $\dot{P}_N \in \Delta(P_N^K)$, $S_\alpha^6(abc, \dot{P}_N) = 3s_\alpha^{1,6} + s_\alpha^{5,6} \geq S_\alpha^6(acb, \dot{P}_N) = 3s_\alpha^{1,6} + 2s_\alpha^{5,6}$. But this implies that $s_\alpha^{5,6} = 0$, in contradiction with Lemma 3 $\square$

Lemma A.2.6. *If α is a Kemeny-stable scoring rule, then $s_\alpha^{3,6} > s_\alpha^{4,6}$.*

Proof: Suppose that $s_\alpha^{3,6} = s_\alpha^{4,6}$. It follows from Lemma 2 together with Lemma 5 that $s_\alpha^{1,6} > s_\alpha^{2,6} = s_\alpha^{3,6} = s_\alpha^{4,6} = s_\alpha^{5,6} \geq s_\alpha^{6,6} = 0$. Using Lemma 1, we get the following possible cases:

Case 1: $s_\alpha^{1,3} = s_\alpha^{2,3} > 0$

Consider the 4 following profiles:

$$P_N = \left(\begin{array}{ccccc} 3 & 2 & 3 & 3 & 4 \\ \hline a & a & b & c & c \\ b & c & a & a & b \\ c & b & c & b & a \end{array} \right) P'_N = \left(\begin{array}{ccc} 1 & 3 & 1 \\ \hline a & b & c \\ b & a & a \\ c & c & b \end{array} \right) P''_N = \left(\begin{array}{cccc} 1 & 1 & 1 & 1 \\ \hline a & a & b & c \\ b & c & a & b \\ c & b & c & a \end{array} \right)$$

$$P'''_N = \left(\begin{array}{ccc} 2 & 2 & 1 \\ \hline a & b & c \\ c & a & a \\ b & c & b \end{array} \right)$$

If $s_\alpha^{2,6} > 0$, then $\alpha(P_N) = abc$. Since $S_\alpha^6(abc, \dot{P}_N)$ $= 3s_\alpha^{1,6} + 8s_\alpha^{2,6} < S_\alpha^6(cba, \dot{P}_N) = 4s_\alpha^{1,6} + 8s_\alpha^{2,6}$ for all $\dot{P}_N \in \Delta(P_N^K)$, then α is not Kemeny-stable. If $s_\alpha^{2,6} = 0$, then $\alpha(P'_N) = abc$. Since $f^\alpha(\dot{P}_N) = \{bac\}$ for all $\dot{P}_N \in \Delta(P_N'^K)$, then α is not Kemeny-stable.

Case 2: $s_\alpha^{1,3} > s_\alpha^{2,3} > 0$

If $s_\alpha^{2,6} > 0$, then $\alpha(P''_N) = abc$. Since $S_\alpha^6(abc, \dot{P}_N)$ $= s_\alpha^{1,6} + 2s_\alpha^{2,6} < S_\alpha^6(bac, \dot{P}_N) = s_\alpha^{1,6} + 3s_\alpha^{2,6}$ for all $\dot{P}_N \in \Delta(P_N''^K)$, then α is not Kemeny-stable. Finally, if $s_\alpha^{2,6} = 0$, then $\alpha(P'''_N) = abc$. Since $S_\alpha^6(abc, \dot{P}_N) = 0 < S_\alpha^6(acb, \dot{P}_N) = 2s_\alpha^{1,6}$ for all $\dot{P}_N \in \Delta(P_N'''^K)$, then α is not Kemeny-stable. Thus, Kemeny-stability requires that $s_\alpha^{3,6} > s_\alpha^{4,6}$ □

By combining the six lemmas above, we get that any Kemeny-stable scoring rule α must satisfy (1) $s_\alpha^{1,6} > s_\alpha^{2,6} = s_\alpha^{3,6} > s_\alpha^{4,6} = s_\alpha^{5,6} \geq 0$, and (2) $s_\alpha^{1,3} \geq s_\alpha^{2,3} > 0$, hence Proposition 3.

A.2.2 *Proof of Proposition A.2.2*

Suppose first that $s_\alpha^{1,3} > 2s_\alpha^{2,3}$, and consider the two profiles $P_N, P'_N \in \mathcal{L}(A_3)^4$ below:

$$P_N = \begin{pmatrix} 1 & 1 & 1 & 1 \\ \hline a & a & b & c \\ b & c & c & b \\ c & b & a & a \end{pmatrix} \quad P'_N = \begin{pmatrix} 1 & 2 & 1 \\ \hline a & a & b \\ b & c & c \\ c & b & a \end{pmatrix}$$

Since $S_\alpha^3(a, P_N) = 2s_\alpha^{1,3}$ and $S_\alpha^3(b, P_N) = S_\alpha^3(c, P_N) = s_\alpha^{1,3} + 2s_\alpha^{2,3}$, then $s_\alpha^{1,3} > 2s_\alpha^{2,3} \Rightarrow$ $\Delta(\alpha(P_N) = \{abc, acb\}$. Using Proposition 3, we get that for any $\dot{P}_N \in \Delta(P_N^K)$, $S_\alpha^6(abc, \dot{P}_N) =$ $s_\alpha^{1,6} + s_\alpha^{2,6} + s_\alpha^{5,6}$, while $S_\alpha^6(bac, \dot{P}_N) = 2s_\alpha^{2,6} + 2s_\alpha^{5,6}$. Therefore, Kemeny-stability requires $s_\alpha^{1,6} \geq s_\alpha^{2,6} + s_\alpha^{5,6}$. Similarly, since $S_\alpha^3(a, P'_N) = 3s_\alpha^{1,3}$, $S_\alpha^3(b, P'_N) = s_\alpha^{1,3} + s_\alpha^{2,3}$ and $S_\alpha^3(c, P'_N) =$ $3s_\alpha^{2,3}$, then $s_\alpha^{1,3} > 2s_\alpha^{2,3} \Rightarrow \alpha(P'_N) = abc$. For any $\dot{P}'_N \in \Delta(P_N'^K)$, $S_\alpha^6(abc, \dot{P}'_N) = s_\alpha^{1,6} + 2s_\alpha^{2,6} +$ $s_\alpha^{5,6}$, while $S_\alpha^6(acb, \dot{P}'_N) = 2s_\alpha^{1,6} + s_\alpha^{2,6}$. Thus, Kemeny-stability requires $s_\alpha^{1,6} \leq s_\alpha^{2,6} + s_\alpha^{5,6}$. Therefore, if $s_\alpha^{1,3} > 2s_\alpha^{2,3}$, then $s_\alpha^{1,6} = s_\alpha^{2,6} + s_\alpha^{5,6}$.

Now suppose that $s_\alpha^{1,3} < 2s_\alpha^{2,3}$, and consider profiles $\widetilde{P}_N \in \mathcal{L}(A_3)^{5Z+1}$, where $Z > 1$, and $\overline{P}_N \in \mathcal{L}(A_3)^4$ below:

$$\widetilde{P}_N = \begin{pmatrix} 2Z & Z+1 & Z & Z \\ \hline a & c & c & b \\ b & b & a & c \\ c & a & b & a \end{pmatrix} \quad \overline{P}_N = \begin{pmatrix} 1 & 1 & 1 & 1 \\ \hline a & a & b & c \\ b & c & c & b \\ c & b & a & a \end{pmatrix}$$

Since $S^3_\alpha(a,\widetilde{P}_N) = 2Zs^{1,3}_\alpha + Zs^{2,3}_\alpha$, $S^3_\alpha(b,\widetilde{P}_N) = Zs^{1,3}_\alpha + (3Z+1)s^{2,3}_\alpha$ and $S^3_\alpha(c,\widetilde{P}_N) = (2Z+1)s^{1,3}_\alpha + Zs^{2,3}_\alpha$, then if Z is chosen large enough, $s^{1,3}_\alpha < 2s^{2,3}_\alpha \Rightarrow \alpha(P_N) = bca$. Moreover, using again Proposition 3, Kemeny-stability implies that for any $Z > 1$ and any $\dot{P}_N \in \Delta(\widetilde{P}^K_N)$, $S^6_\alpha(bca,\dot{P}_N) \geq S^6_\alpha(abc,\dot{P}_N)$. Thus, $Zs^{1,6}_\alpha + (Z+1)s^{2,6}_\alpha + 3Zs^{5,6}_\alpha \geq 2Z(s^{1,6}_\alpha + s^{5,6}_\alpha)$, and therefore $s^{1,6}_\alpha \leq (1+\frac{1}{Z})s^{2,6}_\alpha + s^{5,6}_\alpha$ for all $Z > 1$, leading to $s^{1,6}_\alpha \leq s^{2,6}_\alpha + s^{5,6}_\alpha$. Similarly, we get $\Delta(\alpha(\overline{P}_N)) = \{bca, cba\}$, while for any $\dot{P}_N \in \Delta(\overline{P}^K_N)$, $S^6_\alpha(bca,\dot{P}_N) = S^6_\alpha(cba,\dot{P}_N) = s^{1,6}_\alpha + s^{2,6}_\alpha + s^{5,6}_\alpha$, while $S^6_\alpha(bac,\dot{P}_N) = 2s^{2,6}_\alpha + 2s^{5,6}_\alpha$. Thus, Kemeny-stability implies $s^{1,6}_\alpha \geq s^{2,6}_\alpha + s^{5,6}_\alpha$. Therefore, if $s^{1,3}_\alpha < 2s^{2,3}_\alpha$, then $s^{1,6}_\alpha = s^{2,6}_\alpha + s^{5,6}_\alpha$.

Finally, suppose that $s^{1,3}_\alpha = 2s^{2,3}_\alpha$ and consider $Q_N \in \mathcal{L}(A_3)^{4Z+3}$ and $Q'_N \in \mathcal{L}(A_3)^{10Z+1}$, where $Z > 1$:

$$Q_N = \begin{pmatrix} Z & Z+1 & Z+1 & Z+2 \\ \hline a & b & b & c \\ c & c & a & a \\ b & a & c & b \end{pmatrix} \quad Q'_N = \begin{pmatrix} 3Z & 3Z+1 & 2Z & 2Z \\ \hline a & b & b & c \\ c & c & a & a \\ b & a & c & b \end{pmatrix}$$

Since $s^{1,3}_\alpha = 2s^{2,3}_\alpha$, then $\alpha(Q_N) = cba$. From Proposition 3, one has for any $\dot{P}_N \in \Delta(Q^K_N)$ that $S^6_\alpha(cba,\dot{P}_N) = (2Z+3)s^{2,6}_\alpha + (2Z+1)s^{5,6}_\alpha$ and $S^6_\alpha(acb,\dot{P}_N) = Zs^{1,6}_\alpha + (Z+2)s^{2,6}_\alpha + (Z+1)s^{5,6}_\alpha$. Kemeny-stability implies $s^{1,6}_\alpha \leq (1+\frac{1}{z})s^{2,6}_\alpha + s^{5,6}_\alpha$, and thus $s^{1,6}_\alpha \leq s^{2,6}_\alpha + s^{5,6}_\alpha$. Furthermore, we have $\alpha(Q'_N) = bca$, while Kemeny stability implies that for any $\dot{P}_N \in \Delta(Q'^K_N)$ that $S^6_\alpha(bca,\dot{P}_N) \geq S^6_\alpha(bac,\dot{P}_N)$. Hence, $(3Z+1)s^{1,6}_\alpha + 2Zs^{2,6}_\alpha + 2Zs^{5,6}_\alpha \geq 2Zs^{1,6}_\alpha + (3Z+1)s^{2,6}_\alpha + 3Zs^{5,6}_\alpha$, leading to $s^{1,6}_\alpha \geq s^{2,6}_\alpha + \frac{Z}{Z+1}s^{5,6}_\alpha$ for all $Z > 1$, Therefore, if $s^{1,3}_\alpha = 2s^{2,3}_\alpha$, then $s^{1,6}_\alpha = s^{2,6}_\alpha + s^{5,6}_\alpha$, and the proof is complete.

A.2.3 *Proof of Proposition A.2.3*

Consider the following profiles $P_N \in \mathcal{L}(A_3)^{2Z+2}$ and $P'_N \in \mathcal{L}(A_3)^{56Z+1}$, where $Z > 1$:

$$P_N = \begin{pmatrix} Z+1 & 1 & Z \\ \hline a & a & c \\ b & c & b \\ c & b & a \end{pmatrix} \quad P'_N = \begin{pmatrix} 11Z & 28Z & 17Z & 1 \\ \hline a & a & b & c \\ b & c & c & b \\ c & b & a & a \end{pmatrix}$$

Suppose that $s^{1,3}_\alpha < 2s^{2,3}_\alpha$. Then $\alpha(P_N) = bac$ for Z large enough. Moreover, from Proposition 3, $S^6_\alpha(bac,\dot{P}_N) = (Z+1)(s^{2,6}_\alpha + s^{5,6}_\alpha) < S^6_\alpha(acb,\dot{P}_N) = s^{1,6}_\alpha + (Z+1)s^{2,6}_\alpha + Zs^{5,6}_\alpha$ for all $\dot{P}_N \in \Delta(P^K_N)$, in contradiction with Kemeny-stability.

Suppose that $s^{1,3}_\alpha > 2s^{2,3}_\alpha$. Then $\alpha(P'_N) = abc$ for Z large enough. Using again Proposition 3, $S^6_\alpha(abc,\dot{P}_N) = 11Zs^{1,6}_\alpha + 28Zs^{2,6}_\alpha + 17Zs^{5,6}_\alpha$ while $S^6_\alpha(acb,\dot{P}_N) = 28Zs^{1,6}_\alpha + 11Zs^{2,6}_\alpha +$

$s_\alpha^{5,6}$ for all $\dot{P}_N \in \Delta(P_N'^K)$. Since $s_\alpha^{5,6} > 0$ from Proposition 3, we get by using Proposition 4, $S_\alpha^6(abc, \dot{P}_N) = 39Zs_\alpha^{2,6} + 28Zs_\alpha^{5,6} < S_\alpha^6(abc, \dot{P}_N) = 39Zs_\alpha^{2,6} + 28Zs_\alpha^{5,6} + s_\alpha^{5,6}$, in contradiction with Kemeny-stability.

A.2.4 *End of proof of Theorem 2.3.1*

(Necessary part) Using Propositions 3,4 and 5, it suffices to prove that if α is Kemeny-stable, then $s_\alpha^{2,6} = 3s_\alpha^{5,6}$. Consider the following profiles $P_N \in \mathcal{L}(A_3)^{3Z+1}$ and $P'_N \in \mathcal{L}(A_3)^{3Z-4}$, where $Z > 2$:

$$P_N = \begin{pmatrix} 2Z+1 & Z \\ \hline a & b \\ b & c \\ c & a \end{pmatrix} \quad P'_N = \begin{pmatrix} Z-1 & Z-1 & Z-2 \\ \hline a & b & c \\ c & a & b \\ b & c & a \end{pmatrix}$$

Suppose that $s_\alpha^{2,6} > 3s_\alpha^{5,6}$. Since $s_\alpha^{1,3} = 2s_\alpha^{2,3}$ from Proposition 5, then $\alpha(P_N) = abc$. For any $\dot{P}_N \in \Delta(P_N^K)$, we get from Proposition 3 together with Proposition 4 that $S_\alpha^6(abc, \dot{P}_N) = (2Z+1)s_\alpha^{1,6} + Zs_\alpha^{5,6} = (2Z+1)s_\alpha^{2,6} + (3Z+1)s_\alpha^{5,6}$, while $S_\alpha^6(bac, \dot{P}_N) = (3Z+1)s_\alpha^{2,6}$. But since $s_\alpha^{2,6} > 3s_\alpha^{5,6}$, we get $S_\alpha^6(bac, \dot{P}_N) > S_\alpha^6(abc, \dot{P}_N)$ for all $Z > 2$, in contradiction with Kemeny-stability.

Suppose that $s_\alpha^{2,6} < 3s_\alpha^{5,6}$. Using again $s_\alpha^{1,3} = 2s_\alpha^{2,3}$ from Proposition 5, we get $\alpha(P'_N) = abc$. For any $\dot{P}_N \in \Delta(P_N'^K)$, we get from Proposition 3 together with Proposition 4 that $S_\alpha^6(abc, \dot{P}_N) = (2Z-2)s_\alpha^{2,6}$, while $S_\alpha^6(cba, \dot{P}_N) = (Z-2)s_\alpha^{1,6} + (2Z-2)s_\alpha^{5,6} = (Z-2)s_\alpha^{2,6} + (3Z-4)s_\alpha^{5,6}$. Thus, $S_\alpha^6(bac, \dot{P}_N) > S_\alpha^6(abc, \dot{P}_N)$ for Z large enough, in contradiction with Kemeny-stability. Hence one must have $s_\alpha^{2,6} = 3s_\alpha^{5,6}$, which proves the necessary part.

(Sufficiency part). Consider any $n \in \mathbb{N}$ together with any profile $P_N \in \mathcal{L}(A_3)^n$ having the form

$$P_N = \begin{pmatrix} n_1 & n_2 & n_3 & n_4 & n_5 & n_6 \\ \hline a & a & b & b & c & c \\ b & c & a & c & a & b \\ c & b & c & a & b & a \end{pmatrix}$$

with $\sum_{h=1}^6 n_h = n$. Pick up any scoring rule α fulfilling the conditions $(*)$ $s_\alpha^{1,3} = 2s_\alpha^{2,3} > 0$, and $(**)$ $s_\alpha^{1,6} = \frac{4}{3}s_\alpha^{2,6} = \frac{4}{3}s_\alpha^{3,6} = 4s_\alpha^{4,6} = 4s_\alpha^{5,6} > s_\alpha^{6,6} = 0$. We get that:

- $S_\alpha^3(a, P_N) = (2n_1 + 2n_2 + n_3 + n_5)s_\alpha^{2,3}$
- $S_\alpha^3(b, P_N) = (2n_3 + 2n_4 + n_1 + n_6)s_\alpha^{2,3}$
- $S_\alpha^3(c, P_N) = (2n_5 + 2n_6 + n_2 + n_4)s_\alpha^{2,3}$

Moreover, suppose without loss of generality that $s_\alpha^{5,6} = 1$ and $abc \in \Delta(\alpha(P_N))$. It follows that:

- $n_1 + 2n_2 + n_5 \geq n_3 + 2n_4 + n_6$ *(1)*
- $2n_1 + n_2 + n_3 \geq n_4 + n_5 + 2n_6$ *(2)*
- $n_1 + 2n_3 + n_4 \geq n_2 + 2n_5 + n_6$ *(3)*

Now, pick up any $\dot{P}_N \in \Delta(P_N^K)$. Then we get from $(**)$ that:

- $S_\alpha^6(abc, \dot{P}_N) = 4n_1 + 3(n_2 + n_3) + (n_4 + n_5)$
- $S_\alpha^6(acb, \dot{P}_N) = 4n_2 + 3(n_1 + n_5) + (n_3 + n_6)$
- $S_\alpha^6(bac, \dot{P}_N) = 4n_3 + 3(n_1 + n_4) + (n_2 + n_6)$
- $S_\alpha^6(bca, \dot{P}_N) = 4n_4 + 3(n_3 + n_6) + (n_1 + n_5)$
- $S_\alpha^6(cab, \dot{P}_N) = 4n_5 + 3(n_2 + n_6) + (n_1 + n_4)$
- $S_\alpha^6(cba, \dot{P}_N) = 4n_6 + 3(n_4 + n_5) + (n_2 + n_3)$

Then one easily checks that *(3)* $\Rightarrow S_\alpha^6(abc, \dot{P}_N) \geq S_\alpha^6(acb, \dot{P}_N)$, *(1)* $\Rightarrow S_\alpha^6(abc, \dot{P}_N) \geq S_\alpha^6(bac, \dot{P}_N)$, *(1)+(2)* $\Rightarrow S_\alpha^6(abc, \dot{P}_N) \geq S_\alpha^6(bca, \dot{P}_N)$, *(2)+(3)* $\Rightarrow S_\alpha^6(abc, \dot{P}_N) \geq S_\alpha^6(cab, \dot{P}_N)$, and *(1)+(2)+(3)* $\Rightarrow S_\alpha^6(abc, \dot{P}_N) \geq S_\alpha^6(cba, \dot{P}_N)$. Hence $abc \in \Delta(\alpha(P_N)) \cap f_\alpha(\dot{P}_N)$, and the proof is complete.

A.3 PROOF OF THEOREM 2.3.2

Let α be a non-truncated and Kemeny-stable scoring rule. Consider profile $P_N \in \mathcal{L}(A_6)^{A+B+C+1}$, where $A > B > C > 1$:

$$P_N = \left(\begin{array}{cccc} A & B & C & 1 \\ \hline a & b & c & f \\ c & c & a & e \\ b & a & b & d \\ d & d & d & c \\ e & e & f & b \\ f & f & e & a \end{array} \right)$$

Using Theorem 1, and normalizing S_α^6 by setting $s_\alpha^{1,6} = 1$, we get $S_\alpha^6(a, P_N) = A + \frac{3}{4}(B + C)$, $S_\alpha^6(b, P_N) = \frac{3}{4}(A + C) + B + \frac{1}{4}$, $S_\alpha^6(c, P_N) = \frac{3}{4}(A + B) + C + \frac{1}{4}$, $S_\alpha^6(d, P_N) = \frac{1}{4}(A + B + C) + \frac{3}{4}$, $S_\alpha^6(e, P_N) = \frac{1}{4}(A + B) + \frac{3}{4}$, and $S_\alpha^6(f, P_N) = \frac{1}{4}C + 1$. Obviously, A,B and C can be chosen to ensure that $\alpha(P_N) = abcdef$. Consider the following Kemeny hyper-profile $\dot{P}_N \in \Delta(P_N^K)$

$$\dot{P}_N = \begin{pmatrix} A & B & C & 1 \\ \hline acbdef & bcadef & cabdfe & fedcba \\ cabdef & cbadef & cabdef & ... \\ abcdef & bacdef & cbadfe & ... \\ ... & bcdaef & cadbfe & ... \\ ... & bcaedf & cabfde & ... \\ ... & bcadfe & acbdfe & ... \\ ... & cabdef & ... & ... \\ ... & abcdef & ... & ... \\ ... & ... & ... & ... \end{pmatrix}$$

We get that $S^{6!}_{\alpha}(cabdef, \dot{P}_N) = (A + C)s^{2,6!}_{\alpha} + Bs^{7,6!}_{\alpha} + s^{z,6!}_{\alpha}$, where $z < 6!$, whereas $S^{6!}_{\alpha}(abcdef, \dot{P}_N) = As^{3,6!}_{\alpha} + Bs^{8,6!}_{\alpha} + Cs^{w,6!}_{\alpha}$, where $w > 6$. Finally, Kemeny-stability implies that $s^{1,6!}_{\alpha} = ... = s^{8,6!}_{\alpha}$, and $s^{z,6!}_{\alpha} = 0$, which contradicts that α is non-truncated.

A.4 PROOF OF THEOREM 2.3.3

The proof is similar to the one above. Consider profile $P_N \in \mathcal{L}(A_6)^9$ below:

$$P_N = \begin{pmatrix} 4 & 3 & 1 & 1 \\ \hline a & b & c & c \\ c & c & a & b \\ b & a & b & a \\ d & d & d & d \\ e & e & e & e \\ f & f & f & f \end{pmatrix}$$

We get $S^6_{\alpha}(a, P_N) = 4s^{1,6}_{\alpha} + s^{2,6}_{\alpha} + 4s^{3,6}_{\alpha}$, $S^6_{\alpha}(b, P_N) = 3s^{1,6}_{\alpha} + s^{2,6}_{\alpha} + 5s^{3,6}_{\alpha}$, $S^6_{\alpha}(c, P_N) = 2s^{1,6}_{\alpha} + 7s^{2,6}_{\alpha}$, $S^6_{\alpha}(d, P_N) = 9s^{4,6}_{\alpha}$, $S^6_{\alpha}(e, P_N) = 9s^{5,6}_{\alpha}$, and $S^6_{\alpha}(f, P_N) = 0$. If α is Kemeny-stable, it follows from Theorem 1 that $s^{2,6}_{\alpha} = s^{3,6}_{\alpha}$, which implies that $\Delta(\alpha(P_N)) \subseteq \{P \in \mathcal{L}(A_6) : P = (abc \to Q)$, where $Q \in \mathcal{L}(\{d, e, f\})\}$. Consider the following Kemeny hyper-profile $\dot{P}_N \in \Delta(P^K_N)$

$$\dot{P}_N = \begin{pmatrix} 4 & 3 & 1 & 1 \\ \hline acbdef & bcadef & cabdef & cbadef \\ cabdef & cbadef & acbdef & cabdef \\ abcdef & bacdef & cbadef & bcadef \\ ... & bcdaef & cadbef & cbdaef \\ ... & bcaedf & cabedf & cbaedf \\ ... & bcadfe & cabdfe & cbadfe \\ ... & cabdef & abcdef & abcdef \\ ... & abcdef & ... & ... \\ ... & ... & ... & ... \end{pmatrix}$$

We get that $S_\alpha^{6!}(cabdef, \dot{P}_N) = 4s_\alpha^{2,6!} + 3s_\alpha^{7,6!} + s_\alpha^{1,6!} + s_\alpha^{2,6!}$, whereas $S_\alpha^{6!}(abcdef, \dot{P}_N) = 4s_\alpha^{3,6!} + 3s_\alpha^{8,6!} + s_\alpha^{7,6!} + s_\alpha^{7,6!}$. Using $s_\alpha^{2,6!} \geq s_\alpha^{3,6!}$ and $s_\alpha^{7,6!} \geq s_\alpha^{8,6!}$ together with the strict-at-top property, we have that $S_\alpha^{6!}(cabdef, \dot{P}_N) > S_\alpha^{6!}(abcdef, \dot{P}_N)$. The conclusion follows from the fact that $abcdef$ maximizes $S_\alpha^{6!}(P, \dot{P}_N)$ in $\Delta(\alpha(P_N))$.

A.5 PROOF OF THEOREM 2.3.5

Consider profile $P_N \in \mathcal{L}(A_3)^5$ below:

$$P_N = \begin{pmatrix} 2 & 1 & 1 & 1 \\ \hline a & b & c & c \\ b & a & b & a \\ c & c & a & a \end{pmatrix}$$

Pick up a scoring rule α hyper-stable for $\mathbb{B}$. Since $\mathbb{K} \subset \mathbb{B}$, then α is Kemeny-stable. It follows from Theorem 1 that score vectors must be such that $(*)$ $s_\alpha^{1,3} = 2s_\alpha^{2,3} > 0$, and $(**)$ $s_\alpha^{1,6} = \frac{4}{3}s_\alpha^{2,6} = \frac{4}{3}s_\alpha^{3,6} = 4s_\alpha^{4,6} = 4s_\alpha^{5,6} > s_\alpha^{6,6} = 0$. It follows that $\alpha(P_N) = abc$. It is straightforward to check that the following hyper-profile P_N^E is built from a 5-tuple $E = (e_1, e_1, e_3, e_4, e_5)$ of betweenness-consistent preference extensions:

$$P_N^E = \begin{pmatrix} 2 & 1 & 1 & 1 \\ \hline abc & bac & cba & cab \\ bac & bca & bca & acb \\ acb & abc & bac & cba \\ bca & acb & cab & bca \\ cab & cba & acb & abc \\ cba & cab & abc & bac \end{pmatrix}$$

Note that all extensions in E but e_4 are Kemeny. We get $S^6_\alpha(abc, P^E_N) = 12s^{5,6}_\alpha < S^6_\alpha(bac, P^E_N) = 13s^{5,6}_\alpha$, which contradicts hyper-stability for $\mathbb{B}$.

A.6 PROOF OF THEOREM 2.4.1

Given any $Q \in \mathcal{L}(A_m)$, we write $Q = (Q_1 \to Q_2 \to ... \to Q_H)$, where, for any $1 \leq h \leq H$, $Q_h \in \mathcal{L}(B_h)$ is a segment of Q, and where $\{B_1, B_2, ..., B_H\}$ is a partition of A_m into non-empty sets. Lemma 7 is a useful intermediate step towards the proof.

Lemma A.6.1. *Let $Q, Q' \in \mathcal{L}(A_m)$ be respectively defined by $Q = (Q_1 \to x \to Q_2 \to y \to Q_3)$ and $Q' = (Q_1 \to y \to Q_2 \to x \to Q_3)$, where $Q_h \in \mathcal{L}(B_h)$, for $1 \leq h \leq 3$. Then, for any $P_N \in \mathcal{L}(A_m)^n$ with n is odd, and any $E = (e_1, ..., e_n) \in \mathbb{B}^n$, $[x\ \mu(P_N)\ y] \Rightarrow [Q\ \mu(P^E_N)\ Q']$.*

Proof: Define $B = \{x, y\} \cup B_2$, where $Q_2 \in \mathcal{L}(B_2)$. Pick up any $P_i \in \mathcal{L}(A_m)$ where xP_iy, and consider the restriction $P_i|_B$ of P_i to B. We can write $P_i|_B = (V_1 \to x \to V_2 \to y \to V_3)$, where V_1, V_2, and V_3 are segments of $P_i|_B$, with $V_h \in \mathcal{L}(B_{2h})$, $1 \leq h \leq 3$, and $\{B_{21}, B_{22}, B_{23}\}$ being a partition of B_2. Then $A(P_i|_B, Q|_B) = \{x, y\} \cup [\{x\} \times (B_{22} \cup B_{23})] \cup [(B_{21} \cup B_{22}) \times \{y\}] \cup A(P_i|_{B_2}, Q_2)$, while $A(P_i|_B, Q'|_B) = A(P_i|_B, Q|_B) \backslash \{x, y\}$. Hence, $A(P_i|_B, Q'|_B) \subset A(P_i|_B, Q|_B)$. Since Q and Q' have the same segment Q_1 at top and the same segment Q_3 at bottom, then $A(P_i, Q') \subset A(P_i, Q)$. From betweenness-consistency of e_i, we get $Q\ e_i(P_i)\ Q'$. Finally, $x\ \mu(P_N)\ y$ implies that $|\{i : xP_iy\}| > \frac{n}{2}$, hence that $|\{i : Q\ e_i(P_i)\ Q'\}| > \frac{n}{2}$ and the conclusion follows $\square$

Given $P_N \in \mathcal{L}(A_m)^n$, the top-cycle for P_N is the subset $T(B, P_N)$ of A_m containing all maximal elements for $\theta(P_N)$. The transitive closure partition of A_m is the ordered set $S(\theta, P_N) = (S_1, S_2, ..., S_J)$ of indifference classes for $\theta(P_N)$, where $\forall j \leq j' \in \{1, ..., J\}$, $\forall (x, x') \in S_j \times S_{j'}$, $x\ \theta(P_N)\ x'$ and $\neg(x'\ \theta(P_N)\ x)$ if $j < j'$. By definition of θ, one has $\Delta(\theta(P_N)) = \{Q \in \mathcal{L}(A_m) : Q = (Q_1 \to Q_2 \to ... \to Q_J)$ where, for each $j = 1, ..., J$, $Q_j \in \mathcal{L}(S_j)\}$. The proof of Theorem 6 is complete if we show that for any $E = (e_1, ..., e_n) \in \mathbb{B}^n$, $\Delta(\theta(P_N)) \cap T(\mathcal{L}(A_m), P^E_N) \neq \emptyset$.

Pick up any $P \in \mathcal{L}(A_m) \backslash \Delta(\theta(P_N))$ and any $E = (e_1, ..., e_n) \in \mathbb{B}^n$. Define $B(P) = \{x \in A_m : x \in S_j$ for some j and $\forall y \in S_{j'} \backslash \{x\}$, $xPy \Rightarrow j' > j\}$, and $B = A_m \backslash B(P)$. Consider order $Q(P) \in \Delta(\theta(P_N))$ such that:

- $Q(P) \mid_{B(P)} = P \mid_{B(P)}$
- $xPy \Rightarrow x\ Q(P)\ y$ for all $x, y \in B \cap S_j$ for some $j \in \{1, ..., J\}$.

Write $P \mid_B = b_1 b_2 ... b_T$, where $T = |B|$. There exists a permutation σ of $\{1, ..., T\}$ such that $Q(P) \mid_B = b_{\sigma(1)} b_{\sigma(2)} ... b_{\sigma(T)}$. Then, there is a finite sequence $\{\omega_h\}_{1 \leq h \leq H}$ of transpositions of A_m, where $H \leq T$, such that ω_1 swaps b_1 and $b_{\sigma(1)}$ in $P \mid_B$, leading to $P^1 \mid_B = b_{\omega_1(1)} b_{\omega_1(2)} ... b_{\omega_1(T)}$, ω_2 swaps $b_{\omega_1(2)}$ and $b_{\sigma(2)}$ in $P^1 \mid_B$, leading to $P^2 \mid_B = b_{\omega_2 \circ \omega_1(1)} b_{\omega_2 \circ \omega_1(2)} ... b_{\omega_2 \circ \omega_1(T)}$, ..., ω_H swaps $b_{\sigma(T)}$ and $b_{\omega_{T-1} \circ ... \circ \omega_1(T)}$ in $P^{T-1} \mid_B$, leading to $P^T \mid_B = Q(P) \mid_B$. Since $b_{\sigma(1)}\ \mu(P_N)\ b_{\sigma(2)}\ \mu(P_N) ... \mu(P_N)\ b_{\sigma(T)}$, then Lemma 7 ensures that for all $1 \leq h \leq H$, either $P^{h+1} \mid_B = P^h \mid_B$ or $(P^{h+1} \mid_B)\ \mu(P_N^E \mid_B)\ (P^h \mid_B)$. Hence $(Q(P) \mid_B)\ \theta(P_N^E \mid_B)\ (P \mid_B)$, and thus $Q(P)\ \theta(P_N^E)\ P$. This proves that for any order P not in $\Delta(\theta(P_N))$, there exists $Q \in \Delta(\theta(P_N))$ such that $Q\ \theta(P_N^E)\ P$.

Finally, since $T(\mathcal{L}(A_m), P_N^E \mid_{\Delta(\theta(P_N))}) \neq \emptyset$, there exists $Q \in \Delta(\theta(P_N))$ such that $Q\ \theta(P_N^E)$ Q' for all $Q' \in \Delta(\theta(P_N)) \backslash \{Q\}$. Thus, there exists $Q \in \Delta(\theta(P_N))$ such that $Q\ \theta(P_N^E)\ Q'$ for all $Q' \in \mathcal{L}(A_m) \backslash \{Q\}$. Thus $\Delta(\theta(P_N)) \cap T(\mathcal{L}(A_m), P_N^E)) \neq \emptyset$ and the proof is complete.

A.7 PROOF OF PROPOSITION 2.4.1

Let $P_N = (P_1 ..., P_i, ..., P_n) \in \mathcal{L}(A_m)^n$. Choose any $E = (e_1, ..., e_n) \in \mathbb{B}^n$. Suppose without loss of generality that $\mu(P_N) = a_1 a_2 ... a_m$. Moreover, suppose that $CW(P_N^E) = Q = b_1 b_2 ... b_m$, with $b_1 \neq a_1 = b_h$ for some $2 \leq h \leq m$. Now define $Q' = a_1 b_2 ... b_{h-1} b_h b_{h+1} ... b_m \in \mathcal{L}(A_m)$. It follows from Lemma 7 above that $[1\ \mu(P_N)\ b_1] \Rightarrow [Q'\ \mu(P_N^E)\ Q']$, which contradicts $CW(P_N^E) = Q$. Thus, $b_1 = a_1$. We conclude by iterating the same argument for $b_2, ..., b_m$.

B

APPENDIX TO CHAPTER 4

B.1 APPENDIX: LOGICAL INDEPENDENCE OF AXIOMS

Regularity: The following measure satisfies replication invariance, neutrality and support independence but not regularity:

$$\Psi'(p) = \sum_{\{a,b\} \in \bar{A}} \frac{d_{ab}(p)}{n}.$$

To show it satisfies the first three axioms is rather straightforward. To see violation of regularity it would suffice to consider a unanimous profile.

Support independence: Consider the function $\bar{\Psi}$ which assigns 1 to all but unanimous profiles, to which it assigns 0. It is quick to observe that his function fails only the support independence axiom.

Neutrality: For any set of alternatives A, let $x, y \in A$ be a predefined choice of pairs. The following measure satisfies replication invariance, regularity and support independence but not neutrality:

$$\dot{\Psi}(p) = \frac{n - d_{xy}(p)}{n}$$

Replication Invariance: Let $m = 2, n = 3$. We first construct a function $\bar{K}$ such that $\bar{K}(p) = 0$ for all unanimous profiles, and $\bar{K}(p) = 1$ for all other profiles. Consider the measure below which for $n = 3$ and $m = 2$ equals $\bar{K}(p)$ and in all other cases equals $\Psi^*(p)$:

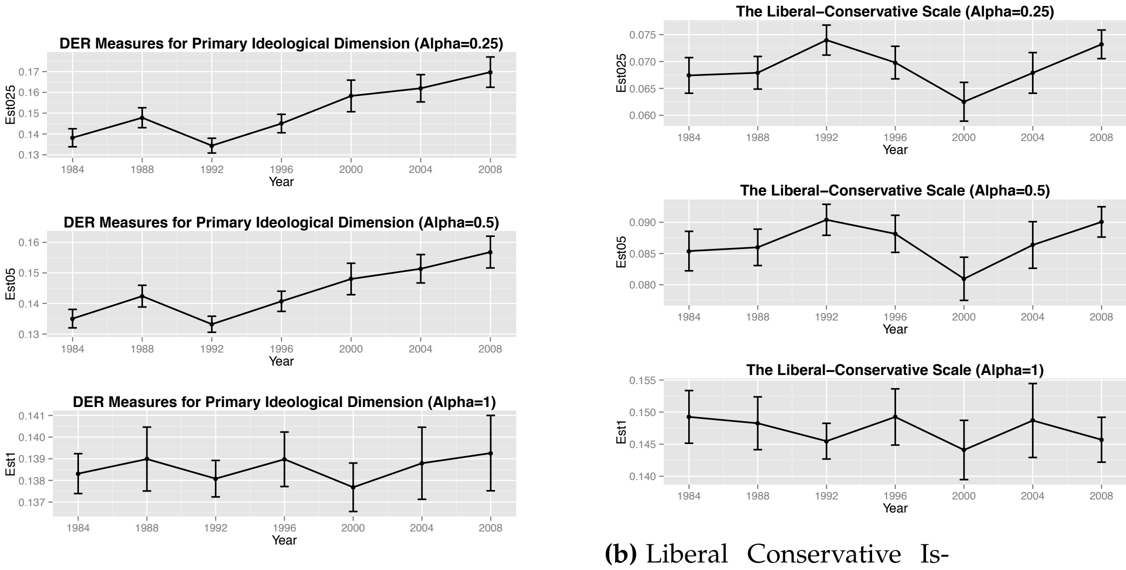
$$\hat{\Psi}(p) = \begin{cases} \bar{K}(p) & \text{if } m = 2 \text{ and } n = 3 \\ \Psi^*(p) & \text{otherwise.} \end{cases}$$

This measure satisfies neutrality, support independence, regularity but not replication invariance.

C

APPENDIX TO CHAPTER 5

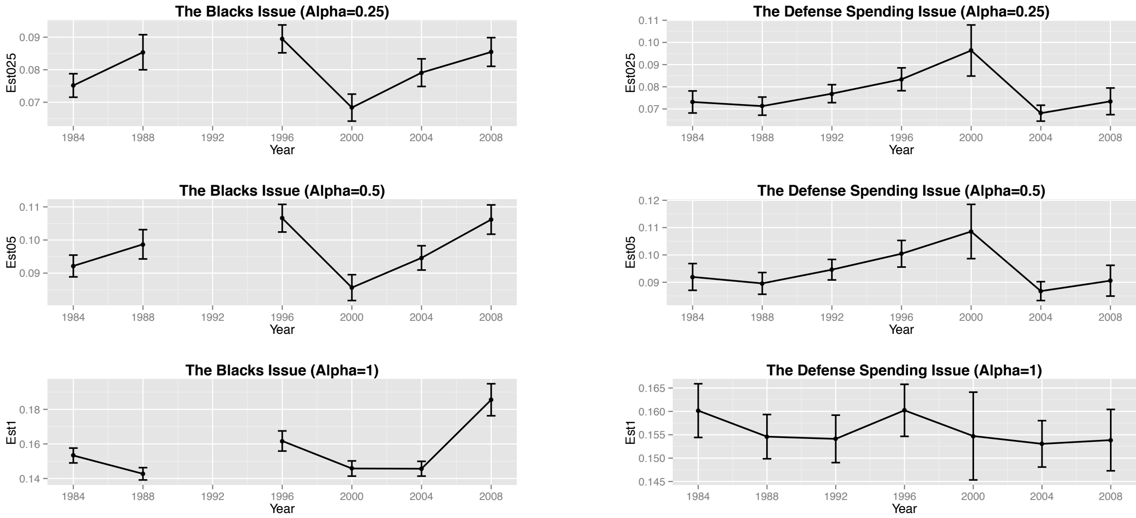
C.1 RESULTS



(a) Latent Ideology

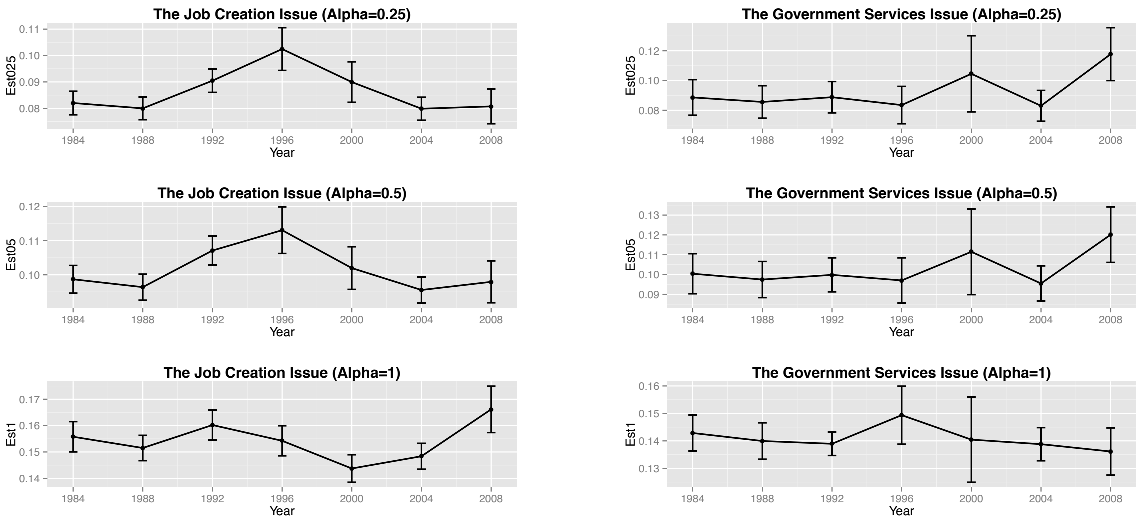
(b) Liberal Conservative Issue

Figure C.1.1.. Ideology.



(a) Aid to Blacks Issue **(b)** Defense Spending Issue

Figure C.1.2.. Issues.



(a) Jobs Creation Issue **(b)** Govt Services Spending

Figure C.1.3.. Issues.

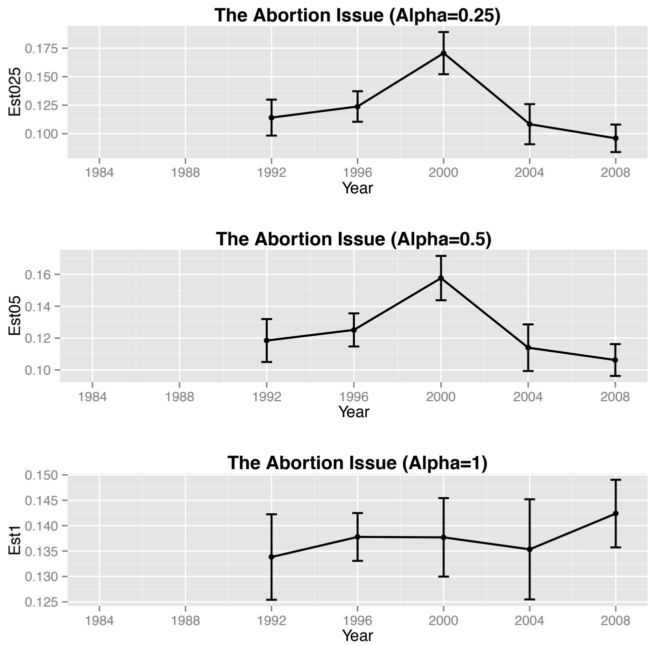
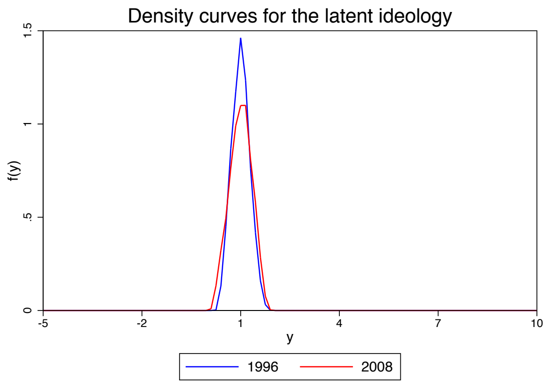
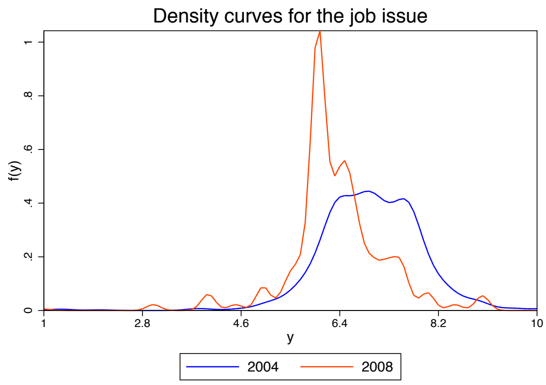


Figure C.1.4.. Abortion Issue

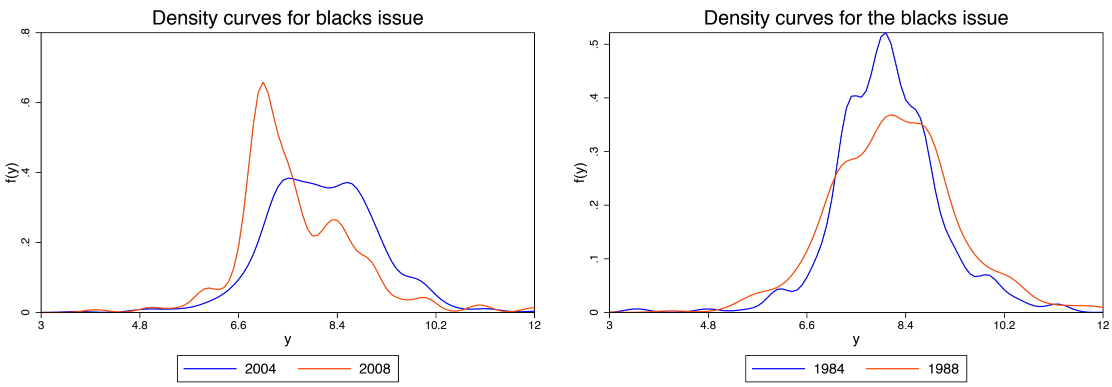


(a) Latent ideology 1996-2008



(b) Jobs issue 2004-2008

Figure C.1.5.. Latent ideology and jobs issue.



(a) 2004-2008

(b) 1984-1988

Figure C.1.6.. Blacks issue comparisons.

C.2 RESCALING AND SEVERAL OTHER MEASURES

The following analysis of polarization with a rescaling method is provided to make sample of the approaches in aggregating issue-wise data.

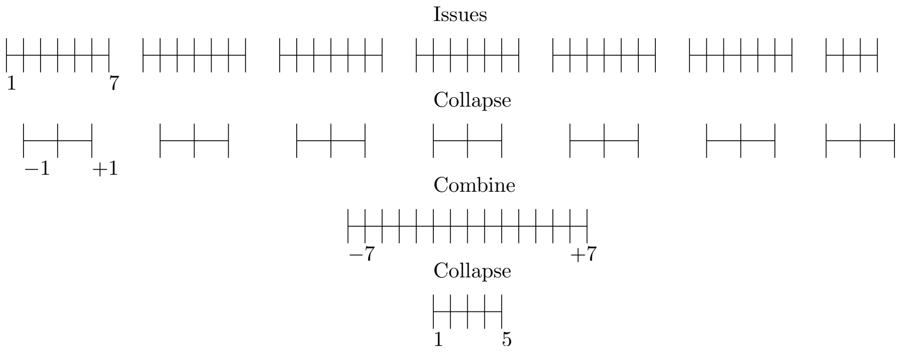


Figure C.2.1.. The rescaling method employed in (Abramowitz, 2006).

Abramowitz (2006) employs the scaling method depicted in Figure C.2.1 in the following way.[1] First, the responses from the survey questions are collapsed into three categories (liberal, moderate and conservative) with groupings, respectively, $1-3$, 4 and $5-7$ for questions with seven scales. For questions with four scale, the groupings were 1-2, 3 and 4. Second this new scale then combined into a fifteen point liberal-conservative scale, i.e. a respondent with liberal position on all issues got -7. Then, this fifteen point scale collapsed again into a 5 point scale: $(-7,-5)$ as consistent liberals, $(-4,-2)$ as moderate liberals, $(-1,+1)$ as inconsistents etc. Finally, on this 5 point scale, the percentage at the central position, position 3, and positions 1 and 5 is compared over years. A decrease at position 3, for instance, indicated an increase in polarization.

Abramowitz and Saunders (2008) also use the average correlation between responses to different issues and see the values 0.20 in 1980s, 0.26 in 1990s and 0.32 in 2004 as a proof of increase in ideological thinking, hence in polarization.

Another measure in the same paper is introduced by first arguing that the ones whose opinions matter first are the ones that are well informed and politically engaged. Political engagement is measured with interest (in political campaign), knowledge (an aggregate of 10 questions including accurately placing candidates on abortion and liberal/conservative scales) and participation. Then comparing the weight of individuals with high stances (on an omnibus scale where responses to 16 questions are collapsed, combined and recoded), for example, among the low and high knowledge individuals, it is shown that polarization is more among the high knowledge individuals.

1 A very similar approach can be found in Abramowitz and Saunders (2008), among others.

Finally, the authors also create an overall index of political engagement by combining the three measures of interest, knowledge and participation to eventually show that (although the means are the same) the standard deviation of high engagement group being the double of the low engagement group on the omnibus scale is a strong indicator of the high polarization among the high engagement group. Furthermore, they compare the number of respondents on the extreme (left and right) points on the scale for each group.

In another very influential work, Alesina et al. (1999), we see the median distance to median (MDM) as a measure for polarization where authors argue that polarization in the form of preference variations among ethnic groups is the underlying reason for public good underprovision.

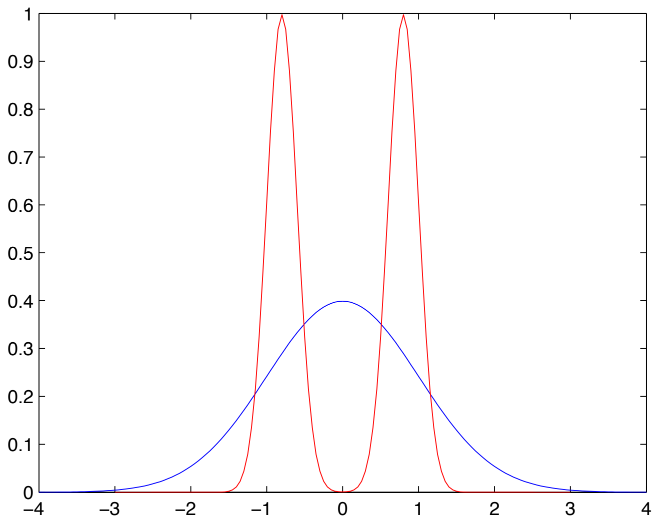


Figure C.2.2.. A case for median distance to median.

Figure C.2.2 presents two distributions, the uni-modal distribution is the standard Normal distribution and the bimodal distribution is a mixture of two Normal distributions with means -0.8 and 0.8 and a common variance of 0.2. Hence MDM for the bimodal distribution is smaller than the unimodal distribution, leading to conclude under the analysis of Alesina et al. (1999) that it is less polarized. But this would contradict an intuitive approach that considers heavy concentrations as an ingredient of polarization. In fact MDM is just a proxy for the variance of the distribution and its implications regarding to polarization is no more that the implications of the variance, that of which we demonstrated in the above text as a bad measure of polarization.

C.3 AXIOMATIC ANALYSIS OF THE DER MEASURE

The overarching idea of the DER measure is to aggregate pairwise antagonisms in the society. For the sake of illustration, suppose we have a distribution of preferences that can be described by a function f defined on $\mathbb{R}_+$. Assuming that the distance between any two point on $\mathbb{R}_+$ represents the alienation between two individuals with preferences represented with those points, the antagonism, as argued in Duclos et al. (2004), can be seen as a function of the alienation as well as the identification which would be a function of the density at the point. So the *effective antagonism of x towards y under f* can be represented by a function $T(f(x), |x-y|)$. It is assumed that T is increasing in $|x-y|$ and that $T(\cdot, 0) = T(0, \cdot)$. The polarization then would be (proportional to) the sum of all effective antagonisms:

$$P(f) = \int\int T(f(x), |x-y|) f(x) f(y) dx dy.$$

Within this large class of functions, the analysis advances on searching for subclasses that satisfy certain plausible axioms. Before getting into the presentations of axioms, we need to define one central item of the analysis, namely, a *squeeze*. A squeeze is a sort of mean-preserving reduction in the spread of a distribution. More specifically, a *λ-squeeze*, used in what follows, of f is a transformation such that:

$$f^{\lambda}(x) \equiv \frac{1}{\lambda} f\left(\frac{x - [1-\lambda]\mu}{\lambda}\right).$$

Any λ-squeeze collapses the density inwards towards the global mean and the following properties can be proved. *(i)* For each $\lambda \in (0,1)$, f^{λ} is a density function. *(ii)* For each $\lambda \in (0,1)$, f^{λ} shares the same mean with f. *(iii)* $0 < \lambda < \lambda' < 1$ implies that f^{λ} second-order stochastically dominates $f^{\lambda'}$. Finally, *(iv)* as $\lambda \downarrow 0$, f^{λ} converges weakly to the degenerate measure granting all weight to μ.

Given this, a measure in the above class is shown in Duclos et al. (2004) to satisfy the following four axioms, DER1-4, if and only if it is proportional to

$$\mathcal{P}_{\alpha}(f) \equiv \int\int f(x)^{1+\alpha} f(y) |x-y| dy dx$$

for some $\alpha \in [0.25, 1]$.

DER1 If a distribution is composed of a single basic density, then a squeeze of that density cannot increase polarization.

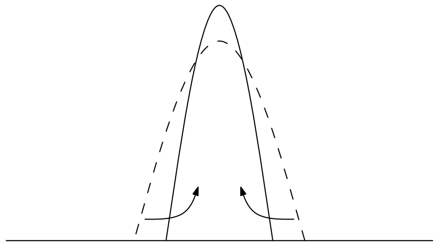


Figure C.3.1.. A squeeze should not increase polarization.

DER2 If a symmetric distribution is composed of three basic densities with the same root and mutually disjoint supports, then a symmetric squeeze of the side densities cannot reduce polarization.

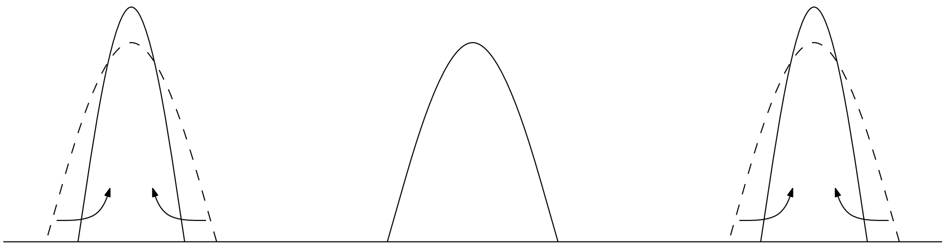


Figure C.3.2.. A symmetric double squeeze should not decrease polarization.

DER3 Consider a symmetric distribution composed of four basic densities with the same root and mutually disjoint supports, as in Figure 4. Slide the two middle densities to the side as shown (keeping all supports disjoint).Then polarization must go up.

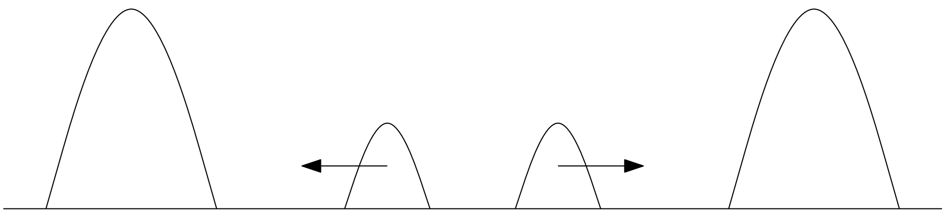


Figure C.3.3.. A symmetric outward slide should increase polarization.

DER4 If $P(F) \geq P(G)$ and $p > 0$, then $P(pF) \geq P(pG)$, where pF and pG represent (identical) population scalings of F and G, respectively.

C.4 DATA

The wording of the questions we used in the empirical analysis change slightly in different years. We quote here some representative version. The questions are the same in cases where respondents locate themselves and the political stimuli (parties and candidates).

1. **Liberal-Conservative Scale**

   We hear a lot of talk these days about liberals and conservatives. Here is a seven-point scale on which the political views that people might hold are arranged from extremely liberal to extremely conservative.

   Q. Where would you place yourself on this scale, or haven't you thought much about this?

   1. extremely liberal
   2. liberal
   3. slightly liberal
   4. moderate, middle of road
   5. slightly conservative
   6. conservative
   7. extremely conservative
   8. Don't know
   9. NA
   0. Haven't thought much

2. **Aid to Blacks**

   Some people feel that the government in Washington should make every effort to improve the social and economic position of blacks. Others feel that the government should not make any special effort to help blacks because they should help themselves.

   Q. Where would you place yourself on this scale, or haven't you thought much about this?

   1. government should help blacks
   2., 3., 4., 5., 6.,
   7. blacks should help themselves
   8. Don't know

9. NA
0. Haven't thought much

3. **Defense Spending**

   Some people believe that we should spend much less money for defense. Others feel that defense spending should be greatly increased.

   Q. Where would you place yourself on this scale, or haven't you thought much about this?

   1. greatly decrease defense spending
   2., 3., 4., 5., 6.,
   7. greatly increase defense spending
   8. Don't know
   9. NA
   0. Haven't thought much

4. **Jobs and Living Standards**

   Some people feel the government in Washington should see to it that every person has a job and a good standard of living. others think the government should just let each person get ahead on his own.

   Q. Where would you place yourself on this scale or haven't you thought much about this?

   1. government see to a job and good standard of living
   2., 3., 4., 5., 6.,
   7. government let each person get ahead on own
   8. Don't know
   9. NA
   0. Haven't thought much

5. **Health Insurance**

   There is much concern about the rapidrise in medical and hospital costs. Some people feel there should be a government insurance plan which would cover all medical and hospital expenses for everyone. Others feelthat all medical expenses should be paid by individuals, and through private insurance plans like blue cross or other company-paid plans.

   Q. Where would you place yourself on this scale, or haven't you thought much about this?

1. government insurance plan
2., 3., 4., 5., 6.,
7. private insurance plan
8. Don't know
9. NA
0. Haven't thought much

6. **Abortion**

   There has been some discussion about abortion during recent years.

   Q. Which one of the opinions on this page best agrees with your view? You can just tell me the number of the opinion you choose.

   1. By law, abortion should never be permitted.
   2. the law should permit abortion only in case of rape, incest or when the woman's life is in danger.
   3. The law should permit abortion for reasons other than rape, incest, or danger to the woman's life, but only after the need for the abortion has been clearly established.
   4. By law, a woman should always be able to obtain an abortion as a matter of personal choice.
   7. Other
   8. Don't know
   9. NA

D

APPENDIX TO CHAPTER 6

D.1 PREDICTIONS AND EXPERIMENTAL RESULTS

	Phase 1		Phase 2		Phase 3		Phase 4		
Session	σ	w	σ	w	σ	w	σ	w	e
A	5	0.888	4.67	0.868	4.84	0.935	4.93	0.987	0.164
B	5	0.887	4.66	0.868	4.83	0.935	4.93	0.987	0.185
1	5	0.886	4.48	0.849	4.70	0.922	4.87	0.984	0.233
2	5	0.856	4.33	0.802	4.61	0.890	4.81	0.970	0.554
3	5	0.882	4.46	0.841	4.68	0.918	4.85	0.983	0.297
4	5	0.871	4.40	0.823	4.65	0.907	4.83	0.978	0.420
5	5	0.886	4.48	0.849	4.70	0.922	4.87	0.984	0.229
6	5	0.874	4.42	0.827	4.66	0.910	4.83	0.979	0.396
7	5	0.882	4.46	0.841	4.68	0.918	4.85	0.983	0.297
$1-7$ ave.			4.43	0.833	4.67	0.913	4.84	0.980	0.346
A, B ave.			4.66	0.868	4.84	0.935	4.93	0.987	0.175
All ave.		0.879							0.308

Table D.1.1.. Nash equilibrium strategies, σ, and predicted group accuracies, w, including the unbiased phase and A and B sessions. Rightmost column gives session averages for logistic error value estimations that are used in calculating strategies.

Phase	A	B	1	2	3	4	5	6	7
1	5.01	4.78	4.97	4.52	4.74	5.20	5.13	4.46	4.85
2	4.27	4.24	3.91	4.00	4.00	3.74	4.53	4.19	4.05
3	4.64	4.29	4.12	3.86	4.25	3.23	4.39	4.05	4.01
4	4.47	4.21	4.29	3.54	4.06	3.29	4.17	4.08	4.00

Table D.1.2.. Phase averages for cutoff estimations in each session.

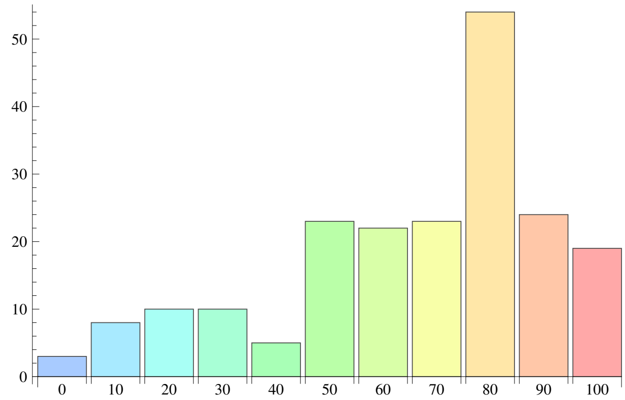


Figure D.1.1.. Frequencies of responses to the Question 2a: "If answered "Yes" in Question 2, what is the percentage of the other participants using the same reasoning, according to your estimation?".

Phase	A	B	1	2	3	4	5	6	7
1	0.938	0.938	0.833	0.708	0.896	0.708	0.854	0.750	0.625
2	0.833	0.792	0.792	0.667	0.854	0.792	0.792	0.792	0.771
3	0.833	0.833	0.792	0.708	0.917	0.833	0.875	0.917	0.792
4	1.000	0.917	0.917	0.500	0.917	0.667	0.833	0.750	0.833

Table D.1.3.. Phase averages for accurate group decision frequencies in each session.

D.2 PROOFS

The expected utility

$$\mathbb{E}_{(s_i)_{i\in N},\omega}[u(\mu((v_i(s_i))_{i\in N},\omega))]$$

can be written as

$$\frac{q}{2}\Pr\left[\sum_{1\leq j\leq n} s_j + \sum_{1\leq j\leq n} b_j > 0|\omega = 1\right] + \frac{1-q}{2}\Pr\left[\sum_{1\leq j\leq n} s_j + \sum_{1\leq j\leq n} b_j < 0|\omega = -1\right].$$

Since the sum of signals are also distributed normally, i.e. with $\mathcal{N}(\omega n, \sigma\sqrt{n})$, we have the following first order condition

$$q\phi\left(\frac{-\sum_{j\neq i} b_j - b_i - n}{\sigma\sqrt{n}}\right) - (1-q)\phi\left(\frac{-\sum_{j\neq i} b_j - b_i - (-n)}{\sigma\sqrt{n}}\right) = 0,$$

where ϕ is the normal density function which gives the best response function

$$\beta(b_{-i}) = -\sum_{j\in N\setminus\{i\}} b_j + \frac{\sigma^2}{2}\ln(q/(1-q)), \tag{$*$}$$

where b_{-i} denotes the biases of all voters except i.

An $ECH1$ player has a belief that the frequency of $ECH0$ players is $g_1(0)$ and the frequency of $ECH1$ is $1 - g_1(0)$ where

$$g_1(0) = \frac{f^\tau(0)}{f^\tau(0) + f^\tau(1)} = \frac{1}{1+\tau}.$$

Denoting the $ECH1$ players' strategy by b_1^{ECH} and $ECH0$ strategy by b_0, we have

$$b_1^{ECH} = -\frac{n-1}{1+\tau n} b_0 + \frac{\sigma^2}{2}\ln(q/(1-q)),$$

using ($*$). So we have that, as $n \to \infty$,

$$b_1^{ECH}(b_0) \to -\frac{1}{\tau} b_0 + \frac{\sigma^2}{2}\ln(q/(1-q)).$$

Moving to second level, consider first the CH model. In what follows, we write b_1^{CH} in place of $\beta(b_0)$ in ($*$) above.

$$\begin{aligned} b_2^{CH} &= -(n-1)[b_0 g_1(0) + b_1^{CH} g_1(1)] + \frac{\sigma^2}{2}\ln(q/(1-q)), \\ &= -(n-1)\left(\frac{1-(n-1)\tau}{1+\tau}\right) b_0 + \left(1 - \frac{(n-1)\tau}{1+\tau}\right)\frac{\sigma^2}{2}\ln(q/(1-q)). \end{aligned}$$

For $ECH2$, note first that

$$g_2(0) = \frac{f^\tau(0)}{f^\tau(0) + f^\tau(1) + f^\tau(2)} = \frac{2}{2+2\tau+\tau^2},$$

similarly, that

$$g_2(1) = \frac{2\tau}{2+2\tau+\tau^2} \text{ and } g_2(2) = \frac{\tau^2}{2+2\tau+\tau^2}.$$

So an $ECH2$ best responds, according to this belief as follows

$$
\begin{aligned}
b_i &= -(n-1)[b_0 g_2(0) + b_1^{ECH} g_2(1) + b_2 g_2(2)] + \frac{\sigma^2}{2}\ln(q/(1-q)),\\
&= \frac{-(n-1)}{2+2\tau+\tau^2}\left(2 - \frac{2\tau(n-1)}{1+\tau n}\right) b_0 - \frac{\tau^2(n-1)}{2+2\tau+\tau^2} b_2 + \left(1 - \frac{2\tau(n-1)}{2+2\tau+\tau^2}\right)\frac{\sigma^2}{2}\ln(q/(1-q)).
\end{aligned}
$$

For equilibrium, we have,

$$
\begin{aligned}
b_2^*\left(1 + \frac{\tau^2(n-1)}{2+2\tau+\tau^2}\right) &= \frac{-(n-1)}{2+2\tau+\tau^2}\left(2 - \frac{2\tau(n-1)}{1+\tau n}\right) b_0 + \left(1 - \frac{2\tau(n-1)}{2+2\tau+\tau^2}\right)\frac{\sigma^2}{2}\ln(q/(1-q)),\\
b_2^* &= -\frac{2(n-1)(\tau+1)}{(2+2\tau+\tau^2 n)(1+\tau n)} b_0 + \left(\frac{2+4\tau+\tau^2-2\tau n}{2+2\tau+\tau^2 n}\right)\frac{\sigma^2}{2}\ln(q/(1-q)).
\end{aligned}
$$

Where we have the first coefficient vanishing as $n \to \infty$, while the second one converges to $-\frac{2}{\tau}$:

$$
b_2^{ECH}(b_0) \to -\frac{\sigma^2}{\tau}\ln\left(\frac{q}{1-q}\right).
$$

D.2.1 *Proof of Proposition 1*

Proof. CH

Given b_0 and τ, we have $(b_0, b_1^{CH}(b_0), b_2^{CH}(b_0))$ with frequencies $p_\tau = \frac{1}{2+2\tau+\tau^2}(2, 2\tau, \tau^2)$ as above. The probability of the correct decision is $(\Pr[d=1|\omega=1] + \Pr[d=-1|\omega=-1])/2$ where

$$
\Pr[d=1|\omega=1] = \Pr\left[\sum_{1\le j\le n} s_j + \Sigma(b_0,\tau,\sigma,n,q) > 0|\omega=1\right],
$$

if $\Sigma(b_0, \tau, \sigma, n, q)$ denotes the summation of biases under *CH*. Algebra shows that

$$(2+2\tau+\tau^2)\Sigma(b_0,\tau,\sigma,n,q) = \left(2 - 2(n-1)\tau - \tau^2(n-1)\frac{1-(n-1)\tau}{1+\tau}\right) b_0 + \left(\sigma^2\tau + \frac{\tau^2\sigma^2}{2}\left(1 - \frac{(n-1)\tau}{1+\tau}\right)\right)\ln\left(\frac{q}{1-q}\right).$$

We see that $\partial\Sigma/\partial n$ is a linear function of and has the same sign with b_0. Furthermore, as $n \to \infty$,

$$\Sigma \to \text{sgn}(b_0)\infty$$

and

$$\Pr\left[\sum_{1\le j\le n} s_j + \Sigma(b_0,\tau,\sigma,n,q) > 0|\omega = 1\right] = \frac{1}{2}\text{Erfc}\left(\frac{-n-\Sigma(b_0,\tau,\sigma,n,q)}{\sqrt{2n}\sigma}\right)$$

where

$$\frac{-n-\Sigma(b_0,\tau,\sigma,n,q)}{\sqrt{2n}\sigma} \to \text{sgn}(-b_0)\infty.$$

Similarly,

$$\Pr\left[\sum_{1\le j\le n} s_j + \Sigma(b_0,\tau,\sigma,n,q) < 0|\omega = -1\right] = \frac{1}{2}\text{Erfc}\left(\frac{-n+\Sigma(b_0,\tau,\sigma,n,q)}{\sqrt{2n}\sigma}\right)$$

where

$$\frac{-n+\Sigma(b_0,\tau,\sigma,n,q)}{\sqrt{2n}\sigma} \to \text{sgn}(b_0)\infty,$$

as $n \to \infty$. Without loss of generality, assume $b_0 > 0$. Then the former probability converges to 1 whereas the latter to 0. Thus, the probability of a correct decision converges to $1/2$.

ECH

Given b_0 and τ, we have $(b_0, b_1^{ECH}(b_0), b_2^{ECH}(b_0))$ with frequencies $p_\tau = \frac{1}{2+2\tau+\tau^2}(2, 2\tau, \tau^2)$ since we assume from now on that the level-2 has the correct belief. The probability of the correct decision is $(\Pr[d = 1|\omega = 1] + \Pr[d = -1|\omega = -1])/2$ where

$$\Pr[d = 1|\omega = 1] = \Pr\left[\sum_{1\le j\le n} s_j > \Sigma(b_0,\tau,\sigma,n,q)|\omega = 1\right],$$

if $\Sigma(b_0, \tau, \sigma, n, q)$ denotes minus the summation of biases under *ECH*. Algebra shows that

$$\left(-\frac{n}{2+2\tau+\tau^2}\right)^{-1}\Sigma(b_0,\tau,\sigma,n,q) = \left(2-\frac{(n-1)}{1+\tau n}2\tau-\frac{2(n-1)(1+\tau)}{(2+2\tau+\tau^2 n)(1+\tau n)}\tau^2\right)b_0 + \left(2\tau+\frac{2+4\tau+\tau^2-2\tau n}{2+2\tau+\tau^2 n}\tau^2\right)\frac{\sigma^2}{2}\ln\left(\frac{q}{1-q}\right),$$

or, simply,

$$\Sigma(b_0,\tau,\sigma,n,q) = -\frac{2n(1+\tau)}{(1+n\tau)(2+2\tau+n\tau^2)}b_0 - \frac{n\tau(2+\tau)}{2+2\tau+n\tau^2}\frac{\sigma^2}{2}\ln\left(\frac{q}{1-q}\right).$$

As $n \to \infty$, we have

$$\Sigma(b_0,\tau,\sigma,n,q) \to -\frac{(2+\tau)\sigma^2}{2\tau}\ln\left(\frac{q}{1-q}\right).$$

Hence the probability

$$\Pr\left[\sum_{1\le j\le n} s_j < \Sigma(b_0,\tau,\sigma,n,q)|\omega=-1\right] = \frac{1}{2}\mathrm{Erfc}\left(\frac{-n-\Sigma(b_0,\tau,\sigma,n,q)}{\sqrt{2n}\sigma}\right),$$

as well as

$$\Pr\left[\sum_{1\le j\le n} s_j > \Sigma(b_0,\tau,\sigma,n,q)|\omega=1\right] = \frac{1}{2}\mathrm{Erfc}\left(\frac{-n+\Sigma(b_0,\tau,\sigma,n,q)}{\sqrt{2n}\sigma}\right),$$

are approaching 1, which means that probability of correct decision is converging to certainty under both states of the world. Note that here we have $\mathrm{Erfc}(\frac{x}{\sqrt{2}}) = 1-2\Phi(x)$, where Φ is the cumulative distribution function for a standard normal variable. □

D.2.2 *Proof of Proposition 2*

Proof. In this case we have

$$\Sigma(b_0,\tau,\sigma,n,q)|_{\tau=1/n} = -\frac{n(1+n)}{3+2n}b_0 + \frac{1+2n}{6+4n}\sigma^2\ln\left(\frac{q}{1-q}\right),$$

which we denote by $\Sigma(b_0, 1/n, \sigma, n, q)$.

So, first

$$\Pr\left[\sum_{1\le j\le n} s_j > \Sigma(b_0, 1/n, \sigma, n, q) | \omega = 1\right] = \frac{1}{2}\mathrm{Erfc}\left(\frac{-n + \Sigma(b_0, 1/n, \sigma, n, q)}{\sqrt{2n}\sigma}\right),$$

where

$$\frac{-n + \Sigma(b_0, 1/n, \sigma, n, q)}{\sqrt{2n}\sigma} \to \mathrm{sgn}(-1 - \frac{b_0}{2})\infty.$$

And also

$$\Pr\left[\sum_{1\le j\le n} s_j < \Sigma(b_0, 1/n, \sigma, n, q) | \omega = -1\right] = \frac{1}{2}\mathrm{Erfc}\left(\frac{-n - \Sigma(b_0, 1/n, \sigma, n, q)}{\sqrt{2n}\sigma}\right),$$

where

$$\frac{-n - \Sigma(b_0, 1/n, \sigma, n, q)}{\sqrt{2n}\sigma} \to \mathrm{sgn}(-2 + b_0)\infty.$$

We know that $\mathrm{Erfc}(-\infty) = 2$, $\mathrm{Erfc}(\infty) = 0$ and $\mathrm{Erfc}(0) = 1$, hence probability of correct decision is 1 in the limit if $b \in (-2, 2)$, 0 if $|b| > 2$ and $3/4$ if $b \in \{-2, 2\}$. □

D.2.3 *Proof of Proposition 3*

Proof. Denote the level-0 strategy as b_0 for all models.

Level-k approach

For higher levels, we have

$$\begin{aligned} b_1(n) &= -(n-1)b_0, \\ b_2(n) &= -(n-1)b_1 = -(n-1)[-(n-1)b_0] = [-(n-1)]^2 b_0, \\ &\vdots \\ b_k(n) &= [-(n-1)]^k b_0. \end{aligned}$$

Hence we have that $b_k(n)/b_0(n) \in O(n^k)$.

Cognitive Hierarchy

The proof is done by induction. Observe first that, as in level$-k$, $b_1(n) = -(n-1)b_0$, hence $b_1(n)/b_0(n) \in O(n)$. Now suppose for $t \in \mathbb{N}$, we have that $b_t(n)/b_0(n) \in O(n^t)$. For $t+1$, it holds that:

$$b_{t+1}(n) = -(n-1)\left(\frac{1}{\sum_{j=0}^{t} f(j)} \sum_{j=0}^{t} f(j) b_j(n)\right),$$
$$b_{t+1}(n)/b_0(n) = -(n-1)\left(\frac{1}{\sum_{j=0}^{t-1} f(j)} \sum_{j=0}^{t-1} f(j) b_j(n)/b_0(n)\right) - (n-1)f(t)b_t(n)/b_0(n).$$

Hence, it can be shown that the first term in the summation is dominated by the latter, which belongs to $O(n^{t+1})$.

Endogenous Cognitive Hierarchy

Oberve first that we have the following:

$$b_1(n) = -(n-1)\left(\frac{f(0)}{f(0)+f(1)} b_0 + \frac{f(1)}{f(0)+f(1)} b_1(n)\right),$$

which can be simplified as;

$$b_1(n) = \frac{-(n-1)f(0)b_0}{f(0)+f(1)-(n-1)f(1)}.$$

In fact, can be shown also is that

$$b_k(n) = \frac{-(n-1)f(0)b_0(n) - (n-1)\sum_{j=1}^{k-1} f(j)b_j(n)}{\sum_{j=0}^{k} f(j) - (n-1)f(k)}.$$

So we have $b_k(n)/b_0 \to C$ for some constant C as $n \to \infty$. $\square$