Opinions and Preferences as Socially Distributed Attitudes

Ignacio Maria Ojea Quintana

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2019

Abstract

Opinions and Preferences as Socially Distributed Attitudes

Ignacio Maria Ojea Quintana

The dissertation focuses on how to best represent the consensus and attitude dynamic of a group given the attitudes of its individuals. This is done in the Bayesian epistemology framework using pooling with imprecise probabilities, and in utility theory extending Harsanyi's aggregation theorem to characterize other directed attitudes like spite and altruism. The final part of the dissertation considers attitudes within social networks and provides explanations and simulation models for online segregation and tribalism as well as the spread of rumors through contagion. The dissertation hopes to contribute to foundational issues like that of epistemic consensus, but also to new emerging phenomena in social epistemology.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my gratitude, first of all, to my advisor and committee members: Philip Kitcher, Achille Varzi, Jessica Collins, Kevin Zollman, and Eleonora Cresto. They not only reviewed this work, but also offered guidance throughout its writing and the whole program. Eleonora first introduced me to formal epistemology back in Argentina, and without her support I would have never applied to Columbia. She also provided invaluable feedback and advice during the program. Jessica and Achille encouraged me at several points during my time in the program, not only through their classes but also helping me navigate what was a foreign culture and academic system. Kevin has been very influential to me since I met him during my visit to Carnegie Mellon University. From him I learned new techniques and approaches, most saliently simulation models and network epistemology. Yet my interests in social epistemology, as well as the Pragmatist tradition, were mostly inspired by some of Philip's ideas. Its been a privilege to have him as an advisor, his support and conversations made me grow immensely.

There have been others that, while not on my committee, took interest in my philosophical education and development. Issac Levi's ideas deeply affected my work. I was very lucky to have the chance to meet him regularly at his apartment to discuss ideas, which also allowed me to learn how to fiercely defend my views. Haim Gaifman advised me during my first years at Columbia, while I was still pursuing a career as a logician. I learned very much from him, I happy to have had him as an advisor, and that he was supportive even when I decided to move to epistemology. Rohit Parikh taught me a fair amount of game theory, as well as epistemic logic. Our meetings were very helpful and he open many gates for me. Jeff Helzner was extremely helpful to me during our overlapping time at Columbia. I met Greg Wheeler in Munich in the summer of 2015. he was crucial in my development as a programmer, and has been helping to guide my research and professional development ever since.

I have learned a lot from friends and colleagues, in particular those who were part of the Formal Philosophy Group at Columbia: Rush Stewart, Robby Finley, Michael Nielsen, Yang Liu, and Arthur Heller. That group developed into perhaps the most significant source of learning in my graduate education. Rush deserves a mention since two of our joint papers have been published (a version of Chapter 2 in the *Journal of Philosophical Logic* and a version of Chapter 3 in *Erkentnnis*). One very valuable lesson was to learn how fruitful and enjoyable collaborative work can be, even if it remains less popular in philosophy than in other disciplines. Thanks also to other fellow graduate students at Columbia for conversations, advice, and friendship: Adam Blazej, Mariana Noe, Jorge Morales, Porter Williams, Sebastien Rivat, and others. Aristotle said it best: "Without friends, no one would want to live, even if he had all other goods."

To conclude, I would like to express my deepest gratitude to my family. My father, Rodolfo Ojea, and my mother, Graciela Imaz provided support and affection at every instance. My sister Celina, my brother Tomas as well as my nieces and nephews (Mora, Vicente, Matias, Malena, and Joaquin) were a source of bliss during the most challenging times. I regard such a supportive affective network fundamental for a good life.

CHAPTER 1

Introduction

1. The Basic Outline of the Project

The present dissertation consists in a collection of papers on a variety of related topics. Unlike more traditional Ph.D Thesis where a single theme or argument is developed exhaustively in several chapters, I here present six essays that address different questions, make different arguments, and use different methodologies. The purpose of this introduction is to bring together all these works in a way that the particular set of philosophical issues that motivate them are clear, their methodologies are justified, and such that a history of my intellectual path reveals where my future research is directed.

There are several schematic ways of partitioning the material. If papers are organized by theme, then four of them are about social epistemology and two about social preferences. If they are divided by methodology, four involve attitude aggregation theory and two attitude distribution in social networks. If they are divided by developmental stage, two of them are already published, two have been submitted, and two are first presented in this dissertation. Table 1 below summarizes this by chronological order.

A guiding question through all of the papers is how to best conceptualize the relation between individual attitudes (opinions or preferences) and social ones, and how to deal with some of the philosophical issues that result from such relation. Sometimes this is done by presenting an account of group consensus in the presence of disagreement, even of the most radical kind. Sometimes by reflecting on how individual preferences are affected by those of friends and foes, or how preferential treatment of others at the individual level may lead to larger social outcomes like segregation.

Paper	***Theme***	***Methodology***	***Stage***
Probabilistic Opinion Pooling with Imprecise Probabilities	Social Epistemology	Opinion Aggregation	Published
Learning and Pooling, Pooling and Learning	Social Epistemology	Opinion Aggregation	Published
Radical Pooling and Imprecise Probabilities	Social Epistemology	Opinion Aggregation	Under Review
Harsanyi meets Smith's Altruism and Spite	Social Preferences	Preference Aggregation	Under Review
Schelling Segregation in Online Social Networks	Social Preferences	Preference Distribution	First Presented
Social Centrality in the Contagion of Rumors	Social Epistemology	Opinion Distribution	First Presented

TABLE 1. Papers by Category

Finally, by studying how an individual's power or centrality in a social arrangement affects the diffusion of rumors or misinformation.

Through the years, the concern for the relation between the individual and the social made me identify different philosophical problems and struggle to find the appropriate conceptual tools to address them. The best way of giving an introduction to this dissertation is *genealogical*, by telling the story of how it was originated and the conceptual transformations that led to the final chapters. This corresponds to the "Stage" partition in the table above. The next section will summarize the discussion around opinion pooling, and the contribution made by the first two papers. Section 3 will describe my pragmatic turn, how my concerns shifted from foundational issues to more pressing irritating ones, and how that turn taints the four final papers. The fourth section briefly presents the papers on radical pooling and preference aggregation. The fifth section begins by justifying the shift in methodology: from attitude aggregation to attitude distribution in social networks, from proofs to simulations and data analysis. Then, it describes the two final projects that use these new methods. The sixth and final section is conclusive and suggests future directions for my research.

2. Opinion Pooling

A particular line of research in the area of Social Epistemology that received much attention lately is the epistemology of collective agents. One of the central questions in this research project is how to aggregate or pool the opinions of the individuals to get the opinion of the group or collective. This question is being prominently addressed by Philip Pettit and Christian List in the Philosophical literature, but it was also the focus of attention in statistical studies during the late 80's and early 90's. This research area is a natural extension of the work of Kenneth Arrow in preference aggregation in the fifties and subsequent work of Amartya Sen on social choice functions, continued by the development of different voting methods; but now the objects to aggregate are judgments rather than preferences. List and Pettit (2002, 2011), Dietrich and List (2011) are helpful survey works.

The problem of opinion aggregation is considered in the literature to be the problem of determining a sensible formula for representing the opinions of a group, given the opinion of the individuals. Representations of group opinion are important in a number of contexts, from scientific advisory panels (on climate change, for example), to joint efforts in scientific inquiry, to decision making in various kinds of unions. In a Bayesian setting, group consensus is particularly important from a theoretical standpoint. The received view is that probabilistic opinions are subjective de Finetti (1964); Ramsey (1990); Savage (1954). Forms of intersubjective agreement have been sought to replace the surrendered notion of objectivity Genest and Zidek (1986); Nau (2002). Probabilistic opinion pooling is one proposal for finding such consensus. It is widely assumed that, for a group of Bayesians, a representation of group opinion should take the form of a (single) probability distribution. For example, Dietrich and List (2014a, 2017) review and generalize some important work in pooling while making novel contributions, but do not consider treating imprecise probabilities (IP).

Since there seem to be advantages in considering imprecise probabilities (in a number of contexts), its natural to ask whether the aggregation problem can be approached in these terms.

The main contribution in this stage was to call into question the assumption that group opinion should be represented by a single probability distribution when precision holds at the level of the individuals. This is done in two different papers.

In the first paper I follow two lines of argumentation. On the one hand, we use a strategy initiated by Seidenfeld et al. (1989, 2010); Walley (1991) that uses limitative results concerning aggregation as a springboard into IP. Even in cases in which individual probabilities are precise, demanding that the output of an aggregation method be a single probability function is overly restrictive. Indeed, as it is shown in several results, representations of group opinion in terms of sets of probability functions satisfies a number of the central pooling axioms that are not jointly satisfied by any of the standard, precise pooling recipes. On the other hand, following some ideas by Levi (1985); Seidenfeld et al. (1989), we argue that IP allows for a plausible philosophical account of rational consensus as common ground at the outcome of inquiry . This co-authored paper (with Rush Stewart) is published in the *Journal of Philosophical Logic*.

The second paper is focused on group *learning*. Among the pooling axioms that serve as required desiderata for any aggregation function there are what are called *Bayesian* axioms. These axioms encode the following intuition: the opinion of a collective agent should be the same if we first decide to aggregate the opinions of the members and then provide the collective agent with some information (and hence provoking a belief updating) and if we first provide each of the members with the same information and then we decide to aggregate the individual opinions into a collective opinion. There are several ways in which this updating could be done: Probabilistic Conditionalization, Conditionalization using likelihood functions, Jeffrey updating and Imaging are among the most well known. With my co-author, Rush Stewart, were successful in showing which

of these principles are satisfied in our approach, and we argue that our account does better than the other available pooling methods. The results of this investigation were published in *Erkenntnis.*

3. A Pragmatic Turn

There is an unspoken rule at Columbia according to which no formal or social epistemologist can graduate without paying respects to the university's pragmatist tradition. This is not the result of the tyrannical enforcing of a social norm. Rather, this is a consequence of the sort of philosophical conversations that are carried out within the community. Through those exchanges, intellectual styles and sensitivities are inherited and reinterpreted. Here I will briefly describe how my work and ideas were transformed by this tradition.

One core pragmatist stand is the rejection of foundational approaches in philosophy and the adoption of a problem solving methodology. This idea was originally articulated by Peirce's Belief-Doubt Model in "The Fixation of Belief." This epistemological model presents a dynamic that starts when the peaceful state of belief is interrupted by an irritating doubt. Resolving that heartfelt doubt is the sole purpose of inquiry, which will bring us back to a new peaceful state of belief. Let us develop the point in more detail.

Peirce characterized belief as a state of mind which is "calm and satisfactory state which we do not wish to avoid" [CP 5.372] and also a habit, a rule for action or disposition to act under particular circumstances: "Belief does not make us act at once, but puts us into such a condition that we shall behave in some certain way, when the occasion arises" [CP 5.373]. The Peircean view in this text is that the fundamental relation between mental and non-mental elements is not that of a correspondence between representation and external objects, but that of practical success in carrying out (in action) the guides and plans contained in the mental elements (beliefs); plans and guides oriented to satisfying the needs and desires of the agent.

As with belief, Peirce's characterization of doubt has an emotional and a practical component. While the state of belief is desirable, the state of doubt "is an uneasy and dissatisfied state from which we struggle to free ourselves and pass into the state of belief" [CP 5.372]. Peirce rejects the Cartesian idea that philosophy should begin with hyperbolic doubt, and that it should pursue some foundational certainty. Proper doubt is not simply an intellectual or rhetorical exercise, but it is genuinely *irritating:* "Let us not pretend to doubt in philosophy what we do not doubt in our hearts." [CP 5.265] The practical component of doubt resides precisely in the fact that it is irritating, and therefore motivates inquiry; "with the doubt, therefore, the struggle begins, and with the cessation of doubt it ends. Hence, the sole object of inquiry is the settlement of opinion." [CP 5.375]

Inquiry, in particular philosophical inquiry, must be rooted in a pressing problem and not in armchair speculations. Furthermore, research should be fundamentally oriented in resolving the crucial issues at hand rather than challenging all established beliefs in order to get a new foundation. The revision of our belief systems, technologies and dispositions should be proportional to the problem that needs resolution. Anything can be put into question, but questioning everything at once is putting the cart before the horse. Our habits, our moral and belief systems, help us navigate the world, and only those routes that do not take us to safe port must be challenged.

The first two papers on pooling were motivated by the need of providing a new formal account of consensus that was not present in the literature and that, I argued, offered certain axiomatic advantages. They were not driven by a pressing, irritating, problem that needed resolution through inquiry. In contrast, the next four papers were motivated by some contemporary social and political issues and events. The most salient of them is the 2016 United States' presidential election, but this is just the tip of the iceberg. It became clear to me in the years before the election that there was an increase in political and ideological polarization, that negative emotions and discourse were gaining track in public discourse, and that social media and online networks were disrupting the

way in which organize inquiry and politics. These issues constituted for me a fertile ground for social epistemology that required rethinking the relation between individual and social attitudes. This was my pragmatic turn, placing the pressing problem before the foundational speculation.

The next section will describe an intermediate pragmatist stage. In thinking about irreconcilable ideological differences, and connecting that problem with the original papers on consensus, I developed an account of *radical* pooling. Concerned by the raise in the expression of negative emotions in the media, I tried to develop an account of other directed attitudes that not only factors benevolent or altruistic tendencies but also those that are spiteful. The section after reveals an even more radical pragmatic turn. In order to conceptualize problems like social segregation and the spread of rumors and unreliable information, I recognized the need to go beyond the literature in aggregation and incorporated networks and statistical methods. This will all become more clear shortly.

There is, someone may object, something profoundly *antiphilosophical* in the pragmatist stand. After all, much of the research presented in this dissertation could be characterized as *formal sociology*. Pragmatism is not defended, but assumed. Perice's belief-doubt model, the pragmatist conceptions of truth and progress, are here not put into question or even reinterpreted in a novel way. Furthermore, formal methodologies adopted like deductive proofs, simulations, and regression analysis would be foreign to many people in philosophy departments. There is certainly truth in this objection, but only because it fails to recognize some bit of pragmatist wisdom. Some philosophical questions, in particular those that demand a foundation, are sometimes overcome without ever being fully answered. They are deemed irrelevant in the face of the new ideas and technologies developed in problem solving. The reader should then follow me and evaluate the next four papers. After all, the proof of the pudding is in the eating.

4. Radical Pooling and Other Directed Attitudes

Section 2 introduced the question of opinion pooling and the contributions made to that topic. As explained, group opinion was conceptualized as consensus as common ground and a formal representation was given in terms of imprecise probabilities. This showed to have some formal advantages. First, it satisfies a number of the central pooling axioms that are not jointly satisfied by any of the standard, precise pooling recipes. Second, it outperforms precise pooling under certain accounts of probabilistic learning. Here I argue that there is a further reason to endorse the view. Consensus as common ground at the outset of inquiry, formalized by pooling imprecise probabilities, is a helpful model to attain group opinion when individuals disagree at the most fundamental level.

Radical pooling is the term coined for the problem of how to aggregate credences when there is a disagreement among agents about which are the relevant sample or event spaces. This question has been neglected in the pooling literature, but it has been treated differently in other branches of the discipline. In philosophy of science, this is related to the issue of theory change treated by Kitcher (1978). In belief revision, this amount to structural changes explored by Bradley (2017) in the probabilistic case, and by Cresto (2008) in an AGM-style. Within game theory, this resebles the literature on unawareness by Fagin and Halpern (1987), Halpern (2001), Dekel et al. (1998), and Modica and Rustichini (1999). None of these approaches, nevertheless, focus on *consensus*.

The solution to the problem of radical pooling advanced in the paper is once again based on the notion of consensus as common ground at the outset of inquiry first introduced by Isaac Levi (1985) and formalized using imprecise probabilities. This is not claimed to be the *only* solution, but a sound one on different grounds. On the one hand, the essay uses the Priestley and Lavoisier debate in Chemistry as a guiding example of the conceptual incommesurability that can be modeled as a disagreement about what are the relevant sample and event spaces. The solution advanced shows

that taking a common ground of those spaces gives a more refined set of possibilities from which to start inquiry. At the outset, disagreements about what variables are relevant for explanation should begin by giving a fair treatment to all the possibilities. During inquiry, some of those possibilities may be discarded. On the other hand, the paper argues that imposing formal desiderata like marginalization, rigidity, or minimum divergence on how to extend probabilities to larger algebras is satisfied by using imprecise probabilities.

The upshot of the paper is that even the case when individual agents disagree on their conceptual frameworks, suspending judgment on personal world views and trying to find a common ground that is a starting point for collective inquiry is a feasible and healthy alternative.

The question of opinion pooling can be thought as a natural extension of the work of Kenneth Arrow in preference aggregation in the fifties and subsequent work of Amartya Sen on social choice functions. In this line, "Harsanyi meets Smith's Altruism and Spite" returns to the issue of preference aggregation and modifies Harsanyi's Impartial Observer Theorem to characterize positive and negative preferences towards others.

Harsanyi (1955, 1977) showed that if welfare preferences satisfy a Pareto (or indifference) condition with respect to individual preferences, then the welfare utility function can be linearly decomposed by the individual utility functions. Although Harsanyi (1977) provided an interpretation of the result in terms of Smith's Impartial Observer, much of the subsequent literature vindicated it as utilitarian: the social or welfare utility is nothing more than the (linear) aggregation of individual utilities. The paper suggests a return to its Smithian roots.

Smith (2002) famous theory of moral sentiments distinguishes between three types of other-directed attitudes: social, unsocial, and selfish passions. Following some ideas in Kitcher (2010), the paper modifies Harsanyi's original result to characterize altruistic, spiteful, and selfish preferences. One agent's preferences are altruistic (spiteful) towards another's if, all other things being equal, the

first would rather chose a social outcome that (dis)favors the second. This can be formalized using Pareto conditions to obtain the result that an agents other-directed utilities can be decomposed linearly between the positive private utilities of those they care about and the negative private utilities of those they despise.

Negative, unsocial, emotions have been neglected by the literature on preferences and utilities; and much of the focus was placed in the economics of altruism. This may be because, as Smith acutely notes, unsocial passions are naturally disagreeable. But they need to be acknowledged and dealt with. In particular, they need to be distinguished from *self-interest*. *Das Adam Smith Problem* is an argument that arose among German scholars during the second half of the nineteenth century concerning the compatibility of the conceptions of human nature advanced in Smith's *Theory of Moral Sentiments* [TSM] and his *Wealth of Nations*. Roughly, the question is how to reconcile the emphasis on sympathy and benevolence in Smith's first book, with the emphasis on self-interest in Smith's second. In other words, how to reconcile in Smith the relation between the kind of interpersonal relations that are at the base of our morality, and the relations that are constitutive of the economic market. But without the recognition of negative emotions towards others, Smith's account of human nature in TMS is simply incomplete. *Das Adam Smith Problem* is made *worse* here, because it requires reconciling all three dimensions: altruism, self-interest, but also spite.

Both papers summarized in this section are currently under review.

5. From Aggregation to Networks

The final stage of the dissertation completes the pragmatic turn in my research. For the purposes of engaging with contemporary problems like online segregation and disinformation I decided to incorporate new techniques and formalisms. In a nutshell, this last stage of the dissertation is a move from *group* epistemology to *network* epistemology.

As I see it, the listed contributions in opinion pooling are within group epistemology. The guiding question is how to aggregate individual opinions to obtain the consensus view, which is usually interpreted as the opinion of the group. Similarly, the paper on Harsanyi fundamentally builds on the literature in preference aggregation, which seeks to represent the welfare utility of the social group. Moving to network epistemology involves an ontological and methodological shift. As I understand them, networks fall between individuals and groups.

Methodological individualism was first presented by Max Weber as the principle that social and economic phenomena must be explained in terms of individual actions, which in turn must be explained by reference to their intentional states. In economics this principle was famously defended by Kenneth Arrow, and we can find contemporary defendants in Jon Elster in political theory. In contrast, in part of the social sciences and social philosophy groups and collectives are usually taken as units of analysis that cannot be reduced to their individual components. Although groups supervene on individuals, as defended by List, C. and Pettit, P. (2011), they constitute an explanatory layer of their own. The past few years saw a great development in understanding the ontology of groups, recognizing their varieties and the relationships among their individual components; for treatment of the issue see Uzquiano (2018). In my view, networks are ontologically lighter than groups, but heavier than mere individuals. The basic units of analysis are both individuals and relations between individuals. The assume no agency over an above individual agency, but relations between agents that cannot be reduced to individual intentional states.

This intermediate methodological stance between individualism and group ontology is not defended in the dissertation, and is left for future research. Rather, I do want to mention that the transition involved incorporating new technologies: agent based modeling, statistical methods, and network theory. The fruits of this labor are summarized below.

The first essay presents a Schelling-like model for Online Social Networks where agents form and cut bonds with others for the purpose of satisfying homophily and heterophily preferences.

Like in Schelling's original model, segregation emerges as a macro phenomenon even when individual agents have mild homophily preferences. Yet, experiments show that heterophily has more effect than homophily in unequal populations. The results are demonstrated via simulations using *Netlogo.* Since the model is designed to represent the dynamics of friending/unfriending and following/unfollowing in social media like Twitter, Facebook or Instagram, it might provide some insight on how online communities self-organize according to tribes.

Schelling's model as been generalized and modified in multiple ways. For example, there may be more than two agent-types, like in Wilensky (1997). Hatna and Benenson (2015) allow for agents to have a preference for diversity. Gandica et al. (2016) generalize by exploring different topologies, allowing for superposition. Paolillo and Lorenz (2018) consider segregation across two dimensions, which lead to cross-contagion effects. Rogers and J. McKane (2011) present a unified framework for segregation with an even more general topology. Examples abound, yet most of them focus on *geographical* segregation and therefore assume a fixed topological space representing a geography which agents inhabit and explore. Even if the topology is generalized by networks, the Schelling dynamics implemented in those studies does not modify the network in any way. Here, that is precisely what we do.

The model hopes to resemble the kind of dynamic a user of social media would find familiar. Agents are part of a social network in which they are connected with other agents. They also have some agency with respect to whom they connect: they can befriend or follow other agents, and they can also sever their links. More precisely, directed networks in which agents decide whether follow others are akin to Twitter or Instagram. Undirected networks in which connections are reciprocal resemble Facebook's friending and unfriending options. This identifications are purely motivational, and the results can be stated in abstract terms. At each stage of the dynamic, each individual can assess their neighbors' (friends or following) tribe. Much like in Schelling, agents will be satisfied if a sufficient amount of their neighbors are of their type. If they are not satisfied,

they will (randomly) sever some of their links and (randomly) seek for new friends to follow. This dynamic on the network stabilizes when all agents are satisfied, or when cycles emerge. How *segregated* or *tribal* a network is depends on how much agents of one tribe are connected with agents of the other.

Much like racial segregation, internet tribalism is a widely recognized phenomenon. It has been long reported by the media (Cragg, 2011), diagnosed by NGO's dealing with political polarization (More in Common (2018)), and heavily studied by data scientists and programmers for the purposes of identifying tribe membership (Bryden et al., 2013; Gloor et al., 2018). In highly polarized networks, members of different tribes do not communicate enough, which might lead to misunderstandings, information bubbles, empathy gaps, etc. Here we are not interested in recognizing tribes, much less manipulating them. Rather, we want to provide an explanation of the sort that Schelling provided for geographic segregation. Abstracting from racism and other forms of in-group bias, Schelling showed that segregation can emerge from very weak conditions. In a similar vein, we intend to show that online tribalism can emerge from a dynamic with mild micro preferences, even when we abstract the well established human tribal tendencies.

The second essay is a study of the contagion of rumors in a variety of social networks. This is done by constructing a computational model using *Netlogo*, and assessing the statistical relation between the centrality (or social power) of the initial *influencer* and the contagion success rate. The study informs us about three epistemological problems around rumors.

To begin, Clifford (1879) originally identified several problems with *credulity*, the habit of endorsing believes without sufficient justification. Not only this constitute an epistemic vice, but one that may infect society at large. Credulous individuals are willing to spread information without sufficient evidence. Such a cultural norm reduces the epistemic standards of others, fostering a credulous *character*. By lowering the standards and promoting epistemic vices, the whole social network becomes vulnerable to manipulation and *fake* information: "The credulous man is father

to the liar and the cheat." The results show that rumors or misinformation can very easily infect much of the network, vindicating some of Clifford's worries.

Second, it is shown that the centrality of the initial influencer has a statistically significant effect on the contagion success. Surprisingly, some centrality measures like *PageRank* are *negatively* correlated. But more broadly, both centrality and network properties affect contagion significantly. This may lead to the recognition of a new form of *epistemic injustice*, a concept developed by Fricker (2007) and much discussed lately in feminist epistemology.

Finally, and relatedly, in conversation with some ideas in Mößner and Kitcher (2017), the essay argues that it may be too soon to conceive of the Internet and Online Social Networks as an epistemically democratizing force.

The two papers summarized in this section were concluded only recently, and might be a bit rough on the edges. I decided to include them in the dissertation because they reflect my intellectual development, and because they introduce some of the techniques that I am hoping to refine in the future.

6. Wrap Up

This introduction was meant to bring together the six essays that constitute the dissertation. I decided to do that genealogically by explaining how my ideas, motivations, and methods evolved in the course of the program. As the title suggests, the guiding issue through all the papers is that of social opinions and preferences. The first stage essays are concerned with the notion of consensus, and this is treated axiomatically in probabilistic pooling. After a pragmatic turn, my interested shifted to more concrete problems in social epistemology. This resulted in an essay on radical pooling, concerned with how to bridge disagreement in conceptual frameworks, and an essay characterizing altruism and spite using techniques from preference aggregation. The third and final stage papers shift from aggregation to networks, and they have an even more pragmatist vein. The

first one, motivated by understanding segregation in online social networks, presents a Schelling-like dynamic that can illuminate how easily tribalism can emerge. The second one provides some insights on the relation between the spread of rumors via contagion and the centrality of those who want to propagate them.

I sincerely think that a network approach to social epistemology can be very fruitful, and I am pursuing this research path in my future. Networks offer a clear and well studied formal model for social relations, and one that can be supplemented with empirical data from our online interactions. Technologies like the Internet and Social Media had a serious impact in the way we organize our knowledge and information, and with that some problems emerged. My hope for now is to develop the conceptual problems to deal with them.

To conclude, some essays were not included in this dissertation, and some others were left for future development. More importantly, all of the ideas introduced here are perfectible, and I hope to continue to improve my thinking through philosophical engagement.

CHAPTER 2

Probabilistic Opinion Pooling with Imprecise Probabilities

1. Introduction

The problem of opinion aggregation is "the problem of determining a sensible formula for representing the opinions of a group" (Genest et al., 1986). Representations of group opinion are important in a number of contexts, from scientific advisory panels (on climate change, for example), to joint efforts in scientific inquiry, to decision making in various kinds of groups. In a Bayesian setting, group consensus is particularly important from a theoretical standpoint. The received view is that probabilistic opinions are *subjective* (Ramsey, 1990; Savage, 1954; de Finetti, 1964). Forms of intersubjective agreement have been sought to replace the surrendered notion of objectivity (Genest and Zidek, 1986; Nau, 2002). Probabilistic opinion pooling is one proposal for finding such consensus. It is widely assumed that, for a group of Bayesians, a representation of group opinion should take the form of a (single) probability distribution. The central position of this essay is that, in certain philosophically interesting and important cases, such an assumption is not always appropriate.

At the end of their review article on pooling, Dietrich and List mention other approaches that "redefine the aggregation problem itself" (2014b, p. 20). According to them, one such approach is the aggregation of imprecise probabilities.[1] Of the few accounts of aggregating probabilities that

A version of this chapter is published as a paper, coauthored with Rush Stewart, under the same title in the *Journal of Philosophical Logic*.

[1]Here I use *IP* as a general term, abstracting from the important distinction Isaac Levi makes between what he calls *imprecise* and *indeterminate* probability, or what Walley calls the *Bayesian sensitivity analysis* and *direct* interpretations, respectively. Roughly speaking, according to the first interpretation, while an agent is normatively committed to or descriptively in a state of numerically precise judgments of credal probability, these precise judgments

deal with imprecision, many tend to focus on cases in which the individual opinions are already imprecise. And such accounts do not proceed by generalizing the pooling framework, axioms, etc. (Moral and Del Sagrado, 1998; Nau, 2002). A general account of probabilistic consensus should cover cases in which probabilities are imprecise at the level of the individual (a topic to which I return towards the end of the essay). However, our aim is to call into question the assumption that group opinion should be represented by a single probability distribution when precision holds at the level of the individuals. In this effort, I extend a line of argument that uses limitative results concerning aggregation—results demonstrating the impossibility of jointly satisfying a set of formal pooling criteria for precise aggregation methods—as a springboard into IP (Walley, 1982; Seidenfeld et al., 1989). That is, the limitations of precise pooling motivate IP in the sense that certain IP models *do* satisfy desiderata for "group" opinion that precise models do not.

After presenting the basic mathematical framework for probabilistic opinion pooling, I review some of the central limitative results (Section 2). One contribution of the present essay is generalizing the pooling framework, framing pooling with imprecise probabilities in the mathematical language common in research on probability aggregation with precise probabilities (Sections 4 and 5). The particular IP model that I primarily focus on in this dissertation, as a proof of concept, is presented in Section 4 (in a sense that will be made clear and precise, our case for considering IP models of pooling does *not* rise and fall with this particular format). Even in cases in which individual probabilities are precise, demanding that the output of an aggregation method be a single probability function is overly restrictive. As I show, representations of group opinion in terms of sets of probability functions have some very nice features. On the one hand, IP allows for a plausible philosophical account of rational consensus (Section 3). On the other hand, the construction I study satisfies a number of the central pooling axioms that are not jointly satisfied

may not be precisely elicited or introspected. On the second interpretation, imprecision is a feature of the credal state itself and is not attributable to imperfect elicitation or introspection. It is possible, of course, for a credal state to be imprecise in both senses, that is, an indeterminate credal state could be incompletely elicited.

by any of the standard, precise pooling recipes on pain of triviality (Sections 5 and 6). I close by considering some potential objections (Section 7).

2. Pooling

A general framework for aggregating the probabilistic opinions of a group to form a collective opinion is that of *pooling*. Formally, a pooling method for a group of n individuals is a function

$$F\colon \mathbb{P}^n \to \mathbb{P}$$

mapping profiles of probability functions for the n agents (or simply the n distributions under consideration), $(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$, to *single* probability functions intended to represent group opinion, $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$. The probabilities are assigned to events, which we represent as subsets of a sample space, Ω. We assume that Ω is countable. The *agenda*, or the set of events under consideration, is assumed to be an algebra $\mathscr{A}$ of events over Ω, that is, a set of subsets of Ω closed under complementation and finite unions (in the general case, closure under countable unions yields a σ-algebra).[2] A function $\boldsymbol{p} : \Omega \to [0, 1]$ is a *probability mass function* (pmf) *iff* $\sum_{\omega\in\Omega} \boldsymbol{p}(\omega) = 1$. Abusing notation, we can define a probability *measure*, $\boldsymbol{p}$, on general events for a given pmf by $\boldsymbol{p}(E) = \sum_{\omega\in E} \boldsymbol{p}(\omega)$. Pooling can be formulated in terms of pmfs, and we will appeal to pmfs in discussing geometric pooling functions and the external Bayesianity constraint below.

Various interpretations of pooling are proposed in the literature. Wagner, for example, offers the following (2009, pp. 336-337):

[2] For completeness, I include the probability axioms. A *probability function* is a mapping $\boldsymbol{p} : \mathscr{A} \to \mathbb{R}$ that satisfies the following conditions:

(i) $\boldsymbol{p}(A) \geqslant 0$ for any $A \in \mathscr{A}$;
(ii) $\boldsymbol{p}(\Omega) = 1$;
(iii) $\boldsymbol{p}(A \cup B) = \boldsymbol{p}(A) + \boldsymbol{p}(B)$ for any $A, B \in \mathscr{A}$ such that $A \cap B = \varnothing$.

If, in addition, $\mathscr{A}$ is a σ-algebra and $\boldsymbol{p}$ satisfies the following condition, $\boldsymbol{p}$ is called *countably additive*:

(iv) If $\{A_n\}_{n=1}^{\infty} \subseteq \mathscr{A}$ is a collection of pairwise disjoint events, then $\boldsymbol{p}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \boldsymbol{p}(A_n)$.

In this essay, I assume countable additivity for convenience, not because I take it to be rationally mandatory.

(1) A rough summary of the current probabilities of the n individuals;

(2) a "compromise" adopted by the individuals for the purpose of group decision making;

(3) a rational consensus to which the individuals revise their probabilities after discussion;

(4) the opinion a decision maker external to the group adopts upon being informed of the n expert opinions in the group;

(5) the opinion an individual in the group adopts upon being informed of the $n-1$ opinions of his "epistemic peers" in the group.

These five interpretations do not exhaust the possibilities. Our target interpretation is rational consensus, adopted either *for the sake of the argument* (a compromise) in order to perform some task in group inference or decision making (2) or genuinely by individual group members (3, 5). However, the account I consider could also be used by a decision maker external to the group.

2.1. Criteria for Pooling Functions. What properties should a pooling function have? Let's review some of the most popular properties discussed in this connection. It is important to consider, for each property, the extent to which it is normatively compelling for a particular interpretation and use of pooling functions. Surveys of the material presented here include Simon French's (1985), Genest and Zidek's (1986), and Dietrich and List's (2014b).

McConway (1981) and Lerher and Wagner (1981) introduce a convenient property of pooling functions called the *strong setwise function property* and *strong label neutrality* by the respective authors. The property has it that the individual probabilities for an event—and not the entire distributions of each individual—are all that is required to determine the collective probability of that event.

Strong Setwise Function Property. There exists a function $G : [0,1]^n \rightarrow [0,1]$ such that, for every event A, $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) = G(\boldsymbol{p}_1(A), ..., \boldsymbol{p}_n(A))$.

What case can be made for the strong setwise function property (SSFP) as a pooling *norm*? SSFP can be seen as a probabilistic analogue of the independence of irrelevant alternatives constraint in the social choice literature. Consider two profiles $(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ and $(\boldsymbol{p}'_1, ..., \boldsymbol{p}'_n)$. Suppose that, for some event A, $\boldsymbol{p}_i(A) = \boldsymbol{p}'_i(A)$ for $i = 1, ..., n$, but the two profiles differ on some other event (so for pooling probabilities for A, "irrelevant" parts of the probability functions differ). It can happen that $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) \neq F(\boldsymbol{p}'_1, ..., \boldsymbol{p}'_n)(A)$ despite the fact that $(\boldsymbol{p}_1(A), ...\boldsymbol{p}_n(A)) = (\boldsymbol{p}'_1(A), ..., \boldsymbol{p}'_n(A))$. That is, the "consensus" probabilities for A differ for the two profiles despite no change in individual opinions concerning A. So, the pooled probability for A is not a function merely of the individual probabilities for A. For such an F, no function G exists because such a function would have to map one profile of values in $[0, 1]^n$ to two distinct outputs in $[0, 1]$. Admittedly, such a case for the normative status of SSFP is incomplete.

Many of the axioms proposed in the literature on pooling require that some property of the individual probability functions be preserved under pooling. When the algebra contains at least three events, one such preservation property follows immediately from SSFP, as McConway observes (1981, Theorem 3.2).

Zero Preservation Property. For any event A, if $\boldsymbol{p}_i(A) = 0$ for $i = 1, ..., n$, then $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) = 0$.

As Genest and Zidek remark, the zero preservation property (ZPP) is one in a class of constraints requiring that the pool preserves initial shared agreements. The normative status of this sort of preservation axiom has been called into question in the literature (Genest and Wagner, 1987). Of course, ZPP is forced upon those endorsing SSFP. For conceptions of *consensus* on which common ground is sought, that is, a non-question begging position of agreement, ZPP is more compelling. I return to *consensus as shared agreement* or *common ground* below in Section 3.

McConway's Theorem 3.2 shows more. Taken together, the *marginalization property* (MP) and the zero preservation property (ZPP) are equivalent to SSFP. McConway's formal setup differs somewhat from the one presented here. He is concerned with classes of pooling functions that take into account all $\sigma-$algebras on Ω. We, however, are considering pooling functions for a fixed algebra (which seems to be the more common approach). The formal properties of concern to McConway must be modified accordingly. A pooling function satisfies MP if marginalization and pooling commute. We adopt the modification of MP proposed by Genest and Zidek (1986, p. 118). Let $\mathscr{A}'$ be a subalgebra of $\mathscr{A}$.[3] Suppose that $\boldsymbol{p}$ is a distribution over $(\Omega, \mathscr{A})$. The *marginal* distribution $\boldsymbol{p}\restriction_{\mathscr{A}'}$ given by $\boldsymbol{p}$ over $(\Omega, \mathscr{A}')$ is the restriction of $\boldsymbol{p}$ to $\mathscr{A}'$ such that $\boldsymbol{p}(A) = \boldsymbol{p}\restriction_{\mathscr{A}'}(A)$ for all $A \in \mathscr{A}'$. $[\boldsymbol{p}\restriction_{\mathscr{A}'}]$ is a Carathéodory extension of $\boldsymbol{p}\restriction_{\mathscr{A}'}$ to $\mathscr{A}$.

Marginalization Property. Let $\mathscr{A}'$ be a sub-σ-algebra of $\mathscr{A}$. For any $A \in \mathscr{A}'$, $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) = F([\boldsymbol{p}_1\restriction_{\mathscr{A}'}], ..., [\boldsymbol{p}_n\restriction_{\mathscr{A}'}])(A)$.

Below, I will state an analogue of another of McConway's results. That result says that MP is equivalent to the *weak setwise function property* (WSFP) (1981, Theorem 3.1). Instead of $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A)$ depending just on the $\boldsymbol{p}_i(A), i = 1, ...n$, those pooling functions merely satisfying WSFP depend on both $\boldsymbol{p}_i(A)$ *and* the event, A. The difference is that a profile in $[0, 1]^n$ may be mapped to more than one output, so long as the associated event differs.

Weak Setwise Function Property. There exists a function $G : \mathscr{A} \times [0, 1]^n \to [0, 1]$ such that, for any event $A \in \mathscr{A}$, $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) = G(A, \boldsymbol{p}_1(A), ..., \boldsymbol{p}_n(A))$ for each profile in the domain of F.

Probabilistic independence is another natural candidate property for preservation under pooling. In the precise setting, there are a number of equivalent formulations of probabilistic independence. For example, two events, A and B, are said to be *stochastically independent* according to $\boldsymbol{p}$

[3] $\mathscr{A}'$ is a boolean subalgebra of $\mathscr{A}$ if $\mathscr{A}' \subseteq \mathscr{A}$ and $\mathscr{A}'$, with the distinguished elements and operations of $\mathscr{A}$, is a boolean algebra. That is, the operations must be the restrictions of the operations of the whole algebra; being a subset that is a boolean algebra is not sufficient for being a subalgebra of $\mathscr{A}$ (Halmos, 1963).

if $\boldsymbol{p}(A \cap B) = \boldsymbol{p}(A)\boldsymbol{p}(B)$. Dividing both sides by $\boldsymbol{p}(B)$, provided $\boldsymbol{p}(B) > 0$, yields $\frac{\boldsymbol{p}(A \cap B)}{\boldsymbol{p}(B)} = \boldsymbol{p}(A)$ when A and B are independent. But the lefthand side of the equation is a standard definition of the probability of A *conditional* on B: $\boldsymbol{p}(A|B) = \frac{\boldsymbol{p}(A \cap B)}{\boldsymbol{p}(B)}$, when $\boldsymbol{p}(B) > 0$. This observation allows us to state another standard formulation of probabilistic independence. A and B are independent according to $\boldsymbol{p}$ if $\boldsymbol{p}(A|B) = \boldsymbol{p}(A)$. The *conditionalization* of $\boldsymbol{p}$ with respect to an event B, $\boldsymbol{p}^B$, is given by setting $\boldsymbol{p}^B(A) = \boldsymbol{p}(A|B)$ for all A. I will return to stochastic independence below, but it will be convenient for us to adopt the definition in terms of conditional probabilities.

Probabilistic Independence Preservation. If $\boldsymbol{p}_i(A|B) = \boldsymbol{p}_i(A)$ for $i = 1, ..., n$, then $F^B(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) = F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A)$.

This axiom says that two events that are probabilistically independent according to every individual probability function are independent according to the pool.

Another preservation axiom is *unanimity preservation*, which requires that, if all of the functions being pooled are identical, then the output of the pooling function is that probability function. So if all the individual opinions are the same, the group opinion is identical to that common distribution.

Unanimity Preservation. For every opinion profile $(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) \in \mathbb{P}^n$, if all $\boldsymbol{p}_i$ are identical, then $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = \boldsymbol{p}_i$.

Other sorts of pooling axioms, like MP above, demand that some operation or other commutes with pooling. A very interesting example of such an operation is a type of Bayesian updating. Standard Bayesian conditionalization goes *via* Bayes' theorem:

$$\boldsymbol{p}^B(A) = \boldsymbol{p}(A|B) = \frac{\boldsymbol{p}(A)\boldsymbol{p}(B|A)}{\boldsymbol{p}(B)}, \text{ when } \boldsymbol{p}(B) > 0.$$

By the law of total probability, the denominator, $\boldsymbol{p}(B)$, can be rewritten. Where $\{C_j : j = 1, 2, ...\}$ is a partition of Ω, $\boldsymbol{p}(B) = \sum_j \boldsymbol{p}(B|C_j)\boldsymbol{p}(C_j)$.

External Bayesianity is a mild generalization of commutativity with Bayesian conditionalization. The requirement is that updating the individual probabilities on a common *likelihood function* (as opposed to updating on an event) and then pooling is the same as pooling and then updating the pool on that likelihood function. The likelihood function, $\lambda : \Omega \to [0, \infty)$, is defined on elements of the sample space. In conditionalizing, $\lambda(\cdot)$ serves the same role as the conditional probability $\boldsymbol{p}(B|\cdot)$ in Bayes' theorem above, expressing the degree to which some fixed evidence B is expected on various events. Put roughly, updating on λ results from substituting the likelihood function in for the conditional probabilities on the right hand side of Bayes' theorem. For every $\omega \in \Omega$,

$$\boldsymbol{p}^{\lambda}(\omega) = \frac{\boldsymbol{p}(\omega)\lambda(\omega)}{\sum_{\omega' \in \Omega} \boldsymbol{p}(\omega')\lambda(\omega')}, \text{when } \sum_{\omega' \in \Omega} \boldsymbol{p}(\omega')\lambda(\omega') > 0$$

If $\boldsymbol{p}$ is a probability measure, it must be defined on an algebra including the elements of Ω. Otherwise, take $\boldsymbol{p}$ to be a pmf and obtain a probabiltiy measure on a given algebra by summing over the elements of Ω in each event to obtain the probability of events in the algebra. Comparing the above formula with the version of Bayes' theorem in which the denominator is expanded by the law of total probability makes the relation between $\lambda(\omega)$ and $\boldsymbol{p}(B|\cdot)$ apparent. While not itself a probability distribution, $\lambda(\omega)$ is proportional to $\boldsymbol{p}(B|\omega)$, for fixed data B. And though not a function of general events in $\mathscr{A}$, the likelihood of an event A can be obtained by summing the likelihoods of all $\omega \in A$. Updating on a likelihood function reduces to standard conditionalization on some event, B, when

$$\lambda(\omega) = \begin{cases} 1, \text{ if } \omega \in B \\ 0, \text{ otherwise.} \end{cases}$$

(We verify the claim with routine subsititions in the footnote.[4])

[4]For any $A \in \mathscr{A}, \boldsymbol{p}^{E}(A) = \frac{\boldsymbol{p}(A \cap E)}{\boldsymbol{p}(E)} = \frac{\sum_{\omega \in A \cap E} \boldsymbol{p}(\omega)}{\sum_{\omega \in E} \boldsymbol{p}(\omega)}$. By the definition of a probability measure, $\boldsymbol{p}(A) = \sum_{\omega \in A} \boldsymbol{p}(\omega)$, so $\sum_{\omega \in A} \boldsymbol{p}^{\lambda}(\omega) = \frac{\sum_{\omega \in A} \boldsymbol{p}(\omega)\lambda(\omega)}{\sum_{\omega' \in \Omega} \boldsymbol{p}(\omega')\lambda(\omega')}$ gives us $\boldsymbol{p}^{\lambda}(A)$. We show that these two fractions are equal by showing the equality of both the numerators and denominators. Since, for all $\omega \in A$, $\boldsymbol{p}(\omega)\lambda(\omega) = \boldsymbol{p}(\omega)$ if $\omega \in E$ and 0 otherwise, $\sum_{\omega \in A} \boldsymbol{p}(\omega)\lambda(\omega) = \sum_{\omega \in A \cap E} \boldsymbol{p}(\omega) = \boldsymbol{p}(A \cap E)$. Hence, the numerators are equal. And since, for all $\omega' \in \Omega, \boldsymbol{p}(\omega')\lambda(\omega') =$

Crucially, the likelihood function is assumed to be *common* in the external Bayesianity axiom. So while disagreement concerning the prior is permitted by pooling functions satisfying external Bayesianity, the commutativity of pooling and updating is guaranteed only when there is agreement on the likelihood funciton.

External Bayesianity. For every profile $(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ in the domain of F and every likelihood function λ such that $(\boldsymbol{p}^\lambda, ..., \boldsymbol{p}_n^\lambda)$ remains in the domain of F, $F(\boldsymbol{p}_1^\lambda, ..., \boldsymbol{p}_n^\lambda) = F^\lambda(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$.

A similar axiom requires that a single individual conditionalizing on λ before pooling is the same as conditionalizing the pool on λ (Dietrich and List, 2014b).

Individualwise Bayesianity. For every profile $(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ in the domain of F and every individual k such that $(\boldsymbol{p}_1, ..., \boldsymbol{p}_k^\lambda, ..., \boldsymbol{p}_n)$ remains in the domain, $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_k^\lambda, ..., \boldsymbol{p}_n) = F^\lambda(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$

I also have more to say about individualwise Bayesianity below.

2.2. Types of Pooling Functions. Various concrete pooling functions have been studied in the literature. These functions fare differently on the criteria reviewed just above. Of the commonly discussed pooling operators, linear pooling functions may be the most common and obvious proposal (Stone, 1961; McConway, 1981; Lehrer and Wagner, 1981).

Linear Opinion Pools. $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = \sum_{i=1}^n w_i \boldsymbol{p}_i$, where $w_i \geqslant 0$ and $\sum_{i=1}^n w_i = 1$.

$w_1, ..., w_n$ are fixed non-negative weights summing to 1 that are associated with the n individuals. Linear pooling, then, takes a weighted average of the individual probabilities. Equal weights for the n probability functions specifies one linear pooling function; a *dictatorship* specifies another linear pooling function. In the latter case, all of the weight is accorded to a single individual ($w_i = 1$ for some i) with the result that the pooled probability for any event A is that individual's probability for A: $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) = \boldsymbol{p}_i(A)$. Interestingly, weights $w_i = \frac{1}{n}$ were used in a U.S.

$\boldsymbol{p}(\omega')$ if $\omega' \in E$ and 0 otherwise, we have $\sum_{\omega' \in \Omega} \boldsymbol{p}(\omega')\lambda(\omega') = \sum_{\omega' \in E} \boldsymbol{p}(\omega') = \boldsymbol{p}(E)$. Hence, the denominators are equal, too. So, $\boldsymbol{p}^E = \boldsymbol{p}^\lambda$.

Nuclear Regulatory Commission study of the frequency of nuclear reactor accidents (Ouchi, 2004, p. 5).

Another proposal is to take a weighted *geometric* instead of a weighted arithmetic average of the n probability functions (Madansky, 1964; Bacharach, 1972; Genest et al., 1986).[5]

Geometric Opinion Pools. $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = c\prod_{i=1}^{n} \boldsymbol{p}_i^{w_i}$, where $w_i \geqslant 0$ and $\sum_{i=1}^{n} w_i = 1$, and $c = \frac{1}{\sum_{\omega' \in \Omega}[\boldsymbol{p}_1(\omega')]^{w_1} \cdots [\boldsymbol{p}_n(\omega')]^{w_n}}$ is a normalization factor.

Unlike linear pools, geometric pools specify the collective probabilities of elements of Ω instead of events in general. But as with the likelihood functions above, the probability of any event A is determined by summing the probabilities of $\omega \in A$. Because of the way in which multiplication figures into the geometric pooling recipe, there are profiles for which $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(\omega) = 0$ for all $\omega \in \Omega$, in violation of the probability axioms. If for each $\omega \in \Omega$ there is a $\boldsymbol{p}_i \in (\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ such that $\boldsymbol{p}_i(\omega) = 0$ we have such a violation. To avoid this worry, the domain of geometric pooling operators is restricted to profiles of *regular* pmfs, i.e., those $\boldsymbol{p}$ such that $\boldsymbol{p}(\omega) > 0$ for all $\omega \in \Omega$. We denote the set of regular pmfs $\mathbb{P}'$ making the relevant domain $\mathbb{P}'^n$.[6]

A third, more recent proposal from Dietrich (2010) is given by the following formula.

Multiplicative Opinion Pools. $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(\omega) = c\prod_{i=0}^{n} \boldsymbol{p}_i$, where $\boldsymbol{p}_0$ is a fixed "calibrating" probability function, and $c = \frac{1}{\sum_{\omega' \in \Omega}[\boldsymbol{p}_0(\omega)] \cdot [\boldsymbol{p}_1(\omega')] \cdots [\boldsymbol{p}_n(\omega')]}$ is a normalization factor.

As with geometric pooling functions, the domain of multiplicative pooling functions will be restricted to $\mathbb{P}'^n$. Comments on the interpretation and choice of $\boldsymbol{p}_0$ can be found in (Dietrich and List, 2014b, pp. 17-19)

---

[5] An unweighted geometric pool of n numerical values is given by $\sqrt[n]{x_1 \cdots x_n} = x_1^{\frac{1}{n}} \cdots x_n^{\frac{1}{n}}$.

[6] Rather than assuming regularity or that the algebra contains the elements of Ω, we could make the weaker restriction to the domain of profiles of pmfs such that there is some $\omega \in \Omega$ for which $\boldsymbol{p}_i(\omega) > 0$ for all $i = 1, ..., n$. And pmfs allow us to obtain measures defined on general algebras on Ω.

Various results, both characterization and limitative, for the different pooling operations and axioms have been obtained. For example, SSFP characterizes linear pooling.

THEOREM 1. *(McConway, 1981, Theorem 3.3; Lehrer and Wagner, 1981, Theorem 6.7) Given that the algebra contains at least three disjoint events, a pooling function satisfies SSFP iff it is a linear pooling function.*

THEOREM 2. *(McConway, 1981, Corollary 3.4) Given that the algebra contains at least three disjoint events, F satisfies WSFP and ZPP iff F is a linear pooling function.*

McConway has shown that a pooling function has the WSFP *iff* it has the MP. So linear pooling functions satisfy MP and ZPP.

THEOREM 3. *(Genest, 1984, p. 1104) The geometric pooling functions are externally Bayesian and preserve unanimity.*

Other sorts of pooling functions, such as a certain generalization of geometric pooling, satisfy the conditions of Theorem 3. A characterization of externally Bayesian pooling functions is given in (Genest et al., 1986). Dietrich and List provide a characterization of multiplicative pooling.

THEOREM 4. *(Dietrich and List, 2014b, Theorem 3) The multiplicative pooling functions are the only individualwise Bayesian pooling functions (with domain $\mathbb{P}'^n$).*

There are many Arrovian limitative theorems in the pooling literature. As Robert Nau notes, none of the pooling methods discussed satisfy even unanimity, external Bayesianity, and the marginalization property (2002, p. 266). (As I show below, our proposal in this essay *does* satisfy those properties.) One result that we have occasion to appeal to below follows from results due to Lehrer and Wagner (1983, Theorems 1 and 2) in conjunction with Theorem 1 above:

THEOREM 5. *(Cf. Lehrer and Wagner, 1983) Given that the algebra contains at least three pairwise disjoint events, the only pooling functions that preserve probabilistic independence and satisfy SSFP are dictatorial.*

It follows that non-dictatorial *linear pools* do not preserve probabilistic independence. In general, non-dictatorial pooling methods struggle with independence preservation (Genest and Wagner, 1987). (Here, too, I claim to do better.)

3. Motivations for IP

In general terms, imprecise probabilities (IP) models do not require representing an agent's or group's judgments of subjective probability as numerically precise. Instead, such judgments could be represented by *sets* of probability functions (Kyburg and Pittarelli, 1996), for example, or by *intervals* (Kyburg, 1998).

There are a number of motivations for working with IP models. These include the potential to resolve some of the "paradoxes of decision" (Ellsberg, 1963; Levi, 1986b), allowing for more flexible and less arbitrary models of uncertainty (Gärdenfors and Sahlin, 1982; Walley, 1991), allowing for incomplete preferences (and hence judgments of incomparability) in the subjective expected utility setting (Levi, 1986a; Seidenfeld, 1993; Kaplan, 1996), and increased descriptive realism (Arló-Costa and Helzner, 2010). An overview of these and other motivations for IP can be found in (Bradley, 2014).

Most important for present purposes, IP allows for—what I consider—a very interesting and philosophically well-motivated account of consensus (Levi, 1985; Seidenfeld et al., 1989). Our goal in this section is to present this account of consensus for explicit consideration in the context of pooling. It may help to first consider the case of *full* or *plain* belief. At the outset of inquiry, inquirers may seek consensus as *shared agreement* in their beliefs. This could be achieved by retaining whatever beliefs are common to all parties and suspending judgment on those that are

controversial thereby avoiding question-begging. Importantly, the consensus is generally a *weaker* state of belief. Since inquiry initiating from the consensus view proceeds without begging questions against parties to the consensus, various hypotheses of concern can receive a fair hearing. Such a consensus constitutes a neutral or non-controversial starting point for subsequent inquiry.

The idea that parties to a joint effort in inquiry or decision making should restrict themselves to their shared agreements—as a compromise or as genuine consensus—can be extended to judgments of probability. An analogous sense of suspending judgment concerning what is controversial is available in the IP setting. To suspend judgment among some number of probability distributions is to not rule them out for the purposes of inference and decision making. Put another way, to suspend judgment among some number of distributions is to regard each as *permissible* to use in inference and decision making. If the parties seeking consensus all agree that $\boldsymbol{p}$ is *not* permissible, then the consensus position reflects that agreement and rules it out (this will have to be finessed when we come to the question of convexity below). A *set* of probability functions represents the shared agreements among the group concerning which probability functions *are not* permissible to use in inference and decision making. For example, it is consensus that the probability of some event is not below the minimum of individual assignments.

Many authors refer to the output of a pooling function as a *consensus* (Lehrer and Wagner, 1981; McConway, 1981; Genest and Zidek, 1986). In what way is a precise pool a consensus? Isaac Levi draws a distinction between consensus as the *outcome* of inquiry and consensus at the *outset* of inquiry (1985). At the outset of inquiry, agents may seek common ground upon which to pursue joint inquiry. This is consensus as shared agreement, discussed just above. Disagreement among the parties to the consensus may then be resolved (in the best case) through joint efforts in inquiry—consensus as the outcome of inquiry. Given the restriction that consensus must be representable by a unique probability function, outside of the special case when all individuals are

in total agreement, finding consensus as shared agreement that suspends judgment on unshared probabilistic views is a foreclosed possibility.

The individuals could assume some common, precise probability distribution, but Levi argues this is not consensus as common ground:

> there can be no analogue in contexts of probability judgment of the two senses of consensus I identify. If two or more agents differ in probability judgment, they can all switch either to the distribution adopted by one of them or to some other distribution which is, in a sense, a potential resolution of the conflict between their differing distributions. There is only one kind of consensus to be recognized—namely the resolution of conflict reached through revolution, conversion, voting, bargaining or some other psychological or social process. (1985, pp. 5-6)

Wagner likewise distinguishes between a *compromise* adopted to perform an exercise in group decision making and a *consensus* to which the individuals revise their own beliefs (Section 2). Levi's point in the quotation above is that a precise pool is neither a consensus as shared agreement since, in general, it is not restricted to just the shared probabilistic views; nor is it justified on the basis of inquiry, understood as designing and performing experiments, obtaining evidence, etc. A precise pool might represent the sort of political consensus that a vote does in the case of preferences, or that the output of a judgment aggregation function does in the case of qualitative belief. That is, consensus as bargaining or compromise. Of course, at least one sort of compromise *is* a consensus adopted *for the sake of the argument* rather than genuinely. That is, parties to the compromise could assume the consensus position as Levi identifies it—namely, a convex IP set—for the sake of the argument, or for carrying out some group deliberation or inquiry so long as the consensus position is strong enough for the group's purposes. It must be admitted that there are other compromise positions, including precise pools, that the group might assume.

But Levi's view distinguishes between political and rational consensus. Returning again to the case of full belief, Levi requires that revisions be decomposable into a sequence of contractions and expansions. An inquirer's set of full beliefs constitute her "standard for serious possibility" in the following sense: if A is among her full beliefs, $\neg A$ is not a serious possibility. To change her mind,

the inquirer must first suspend judgment on A by contracting A if she cares to avoid error (where error is judged by her own lights). From the contraction, both A and $\neg A$ are serious possibilities. A *direct* revision to include $\neg A$ involves deliberately importing error from the point of view that rules $\neg A$ out as a serious possibility.

Unlike full beliefs, however, judgments of subjective probability do not bear truth values. So how might one suspend judgment among candidate distributions before changing points of view? As discussed just above, subjective probabilities are used in determining expectations for available acts. Levi's proposal is that to suspend judgment among some number of distributions is to not rule them out for use in the functions that they perform in inquiry and deliberation. Coming to regard a probability distribution as *permissible* is the analogue of opening one's mind to the (serious) *possibility* of $\neg A$ in the case of full belief. Just as assuming a weaker position in full belief avoids begging questions, retreating to a superset of distributions avoids prejudging the issue of determining which distributions are permissible for use in inquiry and deliberation. Moving from a set of probability functions (including a singleton) to a superset is the probabilistic analogue of contraction. Employing sets of probability functions avoids demanding direct revisions to probabilistic judgments the agent regards as *impermissible* from the standpoint of her current probabilistic judgments without first "contracting" to a neutral position that suspends judgment among the relevant probabilistic views (Levi, 1974). As he emphasizes, both reaching common ground at the outset of inquiry and subsequent reasoned changes in probabilistic views are available to groups in the IP setting.

Why has most work on probabilistic aggregation restricted itself to consideration of representations in terms of a single probability function? One reason is that such representation is the standard for individuals and, since we are treating groups as agents in a sense, that representation should extend to groups as well. Genest and Zidek write, "it would be natural to express the consensus judgment in the same form as the originals" (1986, p. 115). But I am not moved by this convention (or in Walley's terminology, by this "Bayesian dogma of precision") for some of the

very reasons as discussed just above. I urge, in what follows, that theorizing concerning IP should be extended to accounts of probabilistic opinion pooling and *vice versa*. Even if the motivations for IP in general, including at the level of individuals, are found less than convincing, one might think that the case for IP at the level of group opinion is more persuasive, say as an account of consensus. I take the motivations above and the propositions that follow as recommending further consideration of IP in the context of pooling and consensus.

4. IP Pooling Formats

I want to make a case that the out-of-the-gate restriction of the codomain of F to $\mathbb{P}$ is unwarranted (just as many have argued that the standard Bayesian assumption that rational individuals are committed to determinate probabilistic judgments is unwarranted). Our strategy is to point to a sensible account that abandons that restriction. Here we assume a representation in terms of a set of probability functions. We make use of *set-valued* functions or *correspondences*. Where F refers to a pooling function that outputs a single probability function, we will use $\mathcal{F}$ to refer to pooling correspondences outputting sets of probability functions:

$$\mathcal{F} : \mathbb{P}^n \to \mathscr{P}(\mathbb{P})$$

4.1. Convex or Not? Much of the work with IP assumes that IP sets of probabilities are *convex* (Smith, 1961; Levi, 1974; Girón and Ríos, 1980; Gilboa and Schmeidler, 1989; Walley, 1991; Moral and Del Sagrado, 1998). A set of probability functions, $\boldsymbol{P}$, is convex if, for any two functions in the set, the set includes every convex combination of those functions.

Convexity. If $\boldsymbol{p}_1, \boldsymbol{p}_2 \in \boldsymbol{P}$, then $\alpha\boldsymbol{p}_1 + (1-\alpha)\boldsymbol{p}_2 \in \boldsymbol{P}$ for $\alpha \in [0,1]$.

Besides some handy computational properties of convex sets of probability functions (Girón and Ríos, 1980), convexity can be motivated philosophically. A set of probability functions represents

the shared agreements among the group members concerning which distributions are ruled out for use in deliberation and inquiry. But is it not common ground that the convex combinations of individual probability functions are ruled out? The idea is that convexity recommends a *weaker* attitude in suspending judgment among some number of probability distributions; fewer distributions are ruled out. Convexity requires keeping an open mind concerning potential compromises or resolutions of conflict (the convex combinations) between various probabilistic views. Levi argues that convex combinations have "all the earmarks of potential resolutions of the conflict; and, given the assumption that one should not preclude potential resolutions when suspending judgment between rival systems [...] all weighted averages of the two functions are thus taken into account" (1980, p. 192).

The normative status of convexity is the subject of outstanding controversy. Seidenfeld, Schervish, and Kadane make a case against convexity in the context of group decision making (1989). They observe that if two Bayesian agents differ in both probability and utility, any compromise position in probability besides *trivial* convex combinations entails a violation of a Pareto constraint on preference. Levi responds in his (1990), arguing against the Pareto condition. Kyburg and Pittarelli lodge some complaints about the property in "Some Problems for Convex Bayesians" (1992). In "Set-Based Bayesianism," they explore relaxing convexity to allow for IP sets in general. Seidenfeld et al.'s theory of coherent choice does not require convexity (2010). They offer a variation of one of Kyburg and Pittarelli's criticisms of convexity, registering a counterexample that exploits the failure of convex combinations to preserve probabilistic independence (but see (Levi, 2009, pp. 373-375) for a response).

Depending on the decision theory, distinctions between pooling formats may or may not be of importance. For example, there are decision rules that cannot distinguish between certain convex and non-convex sets of probabilities (Gilboa and Schmeidler, 1989; Walley, 1991). Such distinctions are meaningful according to other decision rules (Levi, 1980). And there are decision rules that

distinguish between any two sets of probabilities (Seidenfeld et al., 2010). The important point here is that disputes over the format of pooling functions are idle if such distinctions are not decision-theoretically meaningful. On decision theories that cannot distinguish a convex set of probabilities from its extreme points, for example, there is nothing at stake in arguments over whether an IP opinion pool is convex or not.

4.2. Convex Pooling Functions. As a proof of concept, I will investigate aggregation functions that output sets of probability functions. The aggregate is formed by taking the *convex hull* of the n probability distributions:

$$\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = conv\{\boldsymbol{p}_i : i = 1, ..., n\}$$

The convex hull of a set of points is the smallest convex set containing those points. We write $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A)$ as shorthand for the set of probability assignments to A:

$$\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) = \{\boldsymbol{p}(A) : \boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)\}$$

I work with convexity, not because I presume to know of decisive arguments in its favor, but because convexity is a broadly customary assumption, I do not yet feel compelled to dismiss it, and it allows me to make a proof of concept argument for IP pooling. In effect, assuming convexity amounts to making it slightly harder to show some of the propositions below, though the propositions also hold for IP aggregation methods that relax convexity. I return to the issue of convexity below to make good on my earlier promise to clarify how our case for IP in the context of opinion aggregation does not depend entirely on *convex* IP pools (Section 7.3, Proposition 7).

5. Extending Pooling Axioms to the IP Setting

Convex IP pooling functions satisfy the extensions of those axioms to the IP setting. For the SSFP, we replace G with a set-valued function or correspondence: $\mathcal{G} : [0,1]^n \rightarrow \mathscr{P}([0,1])$. $\mathcal{G}$ is a map from n numerical values in $[0,1]$ to a set of probability values, $\mathcal{G}(\boldsymbol{p}_1(\cdot), ..., \boldsymbol{p}_n(\cdot))$. The constraint becomes that there exists a function $\mathcal{G}$ such that, for any event A, $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) = \mathcal{G}(\boldsymbol{p}_1(A), ..., \boldsymbol{p}_n(A))$. WSFP, then, requires a function $\mathcal{G} : \mathscr{A} \times [0,1]^n \rightarrow \mathscr{P}([0,1])$. For unanimity preservation, we do not distinguish between $\boldsymbol{p}$ and $\{\boldsymbol{p}\}$. ZPP is generalized analogously. If $\boldsymbol{p}_i(A) = 0$ for all $i = 1, ..., n$, then $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) = \{0\}$. The MP has a straightforward extension to sets of probability functions: $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) = \mathcal{F}([\boldsymbol{p}_1 \restriction_{\mathscr{A}'}], ..., [\boldsymbol{p}_n \restriction_{\mathscr{A}'}])(A)$, for any $A \in \mathscr{A}'$ There are many ways to generalize constraints. While I offer conservative and natural modifications of the axioms in order to extend them to the imprecise setting, the crucial question is whether I have modified what is *compelling* about the axioms. For example, is representation in terms of a unique probability function crucial to what makes commutativity of conditionalization and pooling compelling, or what is appealing about preserving shared judgments of independence? In each case, I submit, the attractiveness of the axiom does not hinge on whether the output of the aggregation function is a single probability function or a set of them.

First, we note that an analogue of McConway's result holds for IP pooling functions in general.

Proposition 1. *Let $\mathcal{F} : \mathbb{P}^n \rightarrow \mathscr{P}(\mathbb{P})$ be an IP pooling function (not necessarily convex). $\mathcal{F}$ satisfies WSFP iff $\mathcal{F}$ satisfies MP.*

Before stating the next proposition, we record a fact about convex sets of probabilities (simple and familiar to those with a background in geometry) that we will make use of in the proof.[7] Proofs for the lemmas and propositions are recorded in the appendix.

[7] I include a proof of the observation because we appeal to it several times in the other proofs, because it is a simple special case (but all we need) of a more general result concerning convexity (Rockafellar, 1970), and because some readers may not have a conceptual handle on the property.

LEMMA 1. *Let* $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = conv\{\boldsymbol{p}_i : i = 1, ..., n\}$ *for any profile* $(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ *in the domain of* $\mathcal{F}$. *Any* $\boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ *can be expressed as a convex combination of the* n *probability functions, i.e.* $\boldsymbol{p} = \sum_{i=1}^{n} \alpha_i \boldsymbol{p}_i$, *where* $\alpha_i \geqslant 0$ *for* $i = 1, ..., n$ *and* $\sum_{i=1}^{n} \alpha_i = 1$.

PROPOSITION 2. *Convex IP pooling functions satisfy SWFP, WSFP, MP, ZPP, and unanimity preservation.*

As indicated in the proof, SSFP entails both WSFP and ZPP.

While linear pooling functions are not externally Bayesian, convex IP pooling functions satisfy the extension of external Bayesianity to the IP setting. The *convex* or *prior-by-prior* conditionalization of a convex set of probability functions, $\mathcal{F}^B(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$, results from conditionalizing each member of the set. Updating a convex set of probability functions on a common likelihood function is defined analogously:

$$\mathcal{F}^{\lambda}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = \left\{ \boldsymbol{p}^{\lambda} : \boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n), \sum_{\omega' \in \Omega} \boldsymbol{p}(\omega')\lambda(\omega') > 0, \text{ and } \boldsymbol{p}^{\lambda}(\cdot) = \frac{\boldsymbol{p}(\cdot)\lambda(\cdot)}{\sum_{\omega' \in \Omega} \boldsymbol{p}(\omega')\lambda(\omega')} \right\}$$

To show that convex IP pooling functions are externally Bayesian, we first state another observation.

LEMMA 2. *(Cf. Levi, 1978; Girón and Ríos, 1980) Convexity is preserved under updating on a likelihood function, i.e.,* $\mathcal{F}^{\lambda}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ *is convex.*

PROPOSITION 3. *Convex IP pooling functions are externally Bayesian.*

Dietrich and List argue that while geometric pooling is justified on epistemic grounds when individual opinions are based on the same information, multiplicative pooling is justified in cases of *asymmetric* information, when individual opinions are in part based on private information. Their case is built around the individualwise Bayesianity axiom and the fact that multiplicative pooling satisfies it (Theorem 4).

PROPOSITION 4. *Convex IP pooling functions are* ***not*** *individualwise Bayesian.*

I regard Proposition 4, however, as stating a feature and not a bug of convex IP aggregation. At least insofar as the idea is to reach a consensus, it is not desirable for features of one individual's probability distribution to be unilaterally imposed on the group. In the case of full belief, the initial consensus does not adopt just any belief of any member. Better, in our view, for group opinion to change through efforts in intelligently conducted inquiry from initial common ground (in inquiry, a group may designate a subgroup as an information source on a given topic, but this process requires a richer representation). Dietrich and List motivate individualwise Bayesianity by pointing out that if the constraint is not satisfied, then it makes a difference if an individual first learns some information and opinions are then pooled, or if the opinions are pooled and then the information is acquired by the group as a whole. But for consensus, this is as it should be. If the opinions of the group members do not reflect some piece of information, that information is not common ground. The consensus among group members depends on the probabilistic opinions of the members.

None of this is to say, of course, that features of individual probability distributions are irrelevant to group consensus. On the convex IP view, the kernel of truth in individualwise Bayesianity can be formulated by the inequalities below, stated here for standard conditionalization.

$$\begin{aligned} & min\{\boldsymbol{p}(B) : \boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)\} \\ \leqslant\ & min\{\boldsymbol{p}(B) : \boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_i^B, ..., \boldsymbol{p}_n)\} \\ \leqslant\ & min\{\boldsymbol{p}(B) : \boldsymbol{p} \in \mathcal{F}^B(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)\} \end{aligned}$$

And similarly,

$$\begin{aligned} & max\{\boldsymbol{p}(B) : \boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)\} \\ \leqslant\ & max\{\boldsymbol{p}(B) : \boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_i^B, ..., \boldsymbol{p}_n)\} \\ \leqslant\ & max\{\boldsymbol{p}(B) : \boldsymbol{p} \in \mathcal{F}^B(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)\}. \end{aligned}$$

The consensus probabilities for B shift up (at least not down, more precisely) if one individual conditionalizes on B, and shift more if the consensus itself conditionalizes on B. But these inequalities simply reflect facts about what the common ground is and do not reflect group "learning" from one individual's probability function.

Convex IP pooling also inherits some of the challenges facing linear pooling. SSFP conflicts with probabilistic independence preservation. As Theorem 5 states, the only pooling functions that preserve probabilistic independence and satisfy SSFP are dictatorial. The loss of probabilistic independence presents both epistemic challenges as well as decision theoretic ones (Kyburg and Pittarelli, 1992; Seidenfeld et al., 2010).

In the case of convex IP pooling, however, there is more leeway to address the challenges. Several generalizations of the concept of independence for IP have been proposed and studied (de Campos and Moral, 1995; Cozman, 1998). We consider Levi's notion of *confirmational irrelevance*.

Confirmational Irrelevance Preservation. If $\boldsymbol{p}_i(A|B) = \boldsymbol{p}_i(A)$ for $i = 1, ..., n$, then $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) = \mathcal{F}^B(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A)$.[8]

Irrelevance preservation is a generalization of probabilistic independence preservation. It is clear that when $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ is a single probability function, irrelevance preservation reduces to independence preservation. According to some decision theories for IP, it is the whole set $\boldsymbol{P}$ that is relevant for inquiry and decision making (Levi, 1980; Seidenfeld et al., 2010). Irrelevance is a sensible generalization of independence because it allows us to identify when some information will not make a difference to certain inquiries or deliberations, namely, those inquiries and deliberations concerning events to which the information is irrelevant.

[8]This binary case of irrelevance can be generalized to non-binary partitions. Let $A_1, ..., A_k$ be a partition of Ω. In Levi's setup, a question is represented as a partion, each element of which is a potential answer. Information B is pairwise irrelevant to $A_1, ..., A_k$ if B is irrelevant to each cell of the partition.

It also does not take much work to show that confirmational irrelevance preservation is satisfied by any IP pooling function (not necessarily convex) that satisfies *stochastic independence preservation*.

Stochastic Independence Preservation If $\boldsymbol{p}_i(A \cap B) = \boldsymbol{p}_i(A)\boldsymbol{p}_i(B)$, for $i = 1, ..., n$, then, for all $\boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$, $\boldsymbol{p}(A \cap B) = \boldsymbol{p}(A)\boldsymbol{p}(B)$.

It turns out that confirmational irrelevance *is*, but stochastic independence is *not*, preserved by convex IP pooling functions. Suppose that A and B are probabilistically independent according to $\boldsymbol{p}_i$, $i = 1, ..., n$. Since linear pooling does not preserve independence, independence is not preserved at some of the interior, non-extreme points of $\boldsymbol{P}$. However, the whole set of probability *values for* A, $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A)$, is the same before and after conditionalizing on B.

PROPOSITION 5. *Convex IP pooling functions satisfy irrelevance preservation.*

So, while stochastic independence and confirmational irrelevance are equivalent in the precise setting (when $\boldsymbol{p}(B) > 0$), they come apart in the IP context.[9] Because irrelevance preservation reduces to probabilistic independence preservation when the output of the pooling function is a unique probability function, and linear, geometric, and multiplicative pooling functions do not satisfy probabilistic independence preservation in general, we have that linear, geometric, and multiplicative pooling functions do not satisfy irrelevance preservation either. If there are good reasons to require IP pooling functions to satisfy the stronger stochastic independence preservation property, then convex IP pool does not deliver (though there are IP formats that do (Proposition 7)).

[9]Pedersen and Wheeler show how logically distinct independence concepts are teased apart in the context of imprecise probabilities. They write, "there are several distinct concepts of probabilistic independence and [...] they only become extensionally equivalent within a standard, numerically determinate probability model. This means that some sound principles of reasoning about probabilistic independence within determinate probability models are invalid within imprecise probability models" (2014, p. 1307). So IP provides a more subtle setting in which to investigate independence concepts.

Finally, convex IP pooling admits of a simple characterization in terms of the set of *universally admissible means.*[10] We call a function $\mathfrak{m} : [0,1]^n \to [0,1]$ a *mean* on the interval $[0,1]$. We first define a mapping $\mathfrak{M}_n : \mathbb{P}^n \to \mathscr{P}([0,1]^{[0,1]^n})$ by setting for every $(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) \in \mathbb{P}^n$:

$$\mathfrak{M}_n(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = \Big\{\mathfrak{m} \in [0,1]^{[0,1]^n} : \mathfrak{m}\big(\boldsymbol{p}_1(\cdot), ..., \boldsymbol{p}_n(\cdot)\big) \in \mathbb{P}\Big\}.$$

Call a mean *admissible* for $(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ if $\mathfrak{m}\big(\boldsymbol{p}_1(\cdot), ..., \boldsymbol{p}_n(\cdot)\big) \in \mathbb{P}$. Then, $\mathfrak{M}_n(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ is the set of admissible means for $(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$. Using $\mathfrak{M}_n$, we define another mapping $\mathcal{M}_n : \mathbb{P}^n \to \mathscr{P}(\mathbb{P})$ by setting for every $(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) \in \mathbb{P}^n$:

$$\mathcal{M}_n(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = \Big\{\mathfrak{m}\big(\boldsymbol{p}_1(\cdot), ..., \boldsymbol{p}_n(\cdot)\big) : \mathfrak{m} \in \bigcap_{\vec{\boldsymbol{q}} \in \mathbb{P}^n} \mathfrak{M}_n(\vec{\boldsymbol{q}})\Big\}.$$

$\mathcal{M}_n(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ is the set of probability functions that results from composing each *universally* admissible mean with $\big(\boldsymbol{p}_1(\cdot), ..., \boldsymbol{p}_n(\cdot)\big)$.

Proposition 6. *Suppose that $\mathscr{A}$ contains at least three pairwise incompatible events. A mapping $\mathcal{F} : \mathbb{P}^n \to \mathscr{P}(\mathbb{P})$ is a convex IP pooling function—that is, $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = conv(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$—if and only if $\mathcal{M}_n(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ for all $(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) \in \mathbb{P}^n$.*

An interesting line of research to pursue would be to consider pooling from the perspective of an analysis of means (e.g., Kolmogorov, 1930; de Finetti, 1931; Aczél, 1948). Perhaps such an analysis could shed light on issues like the propriety of qualitative conditions on pooling rules, or the function of convexity in reaching a consensus in inquiry and deliberation.

6. Epistemic and Procedural Grounds for

[10] Thanks to Paul Pedersen for suggesting that I include a result along these lines.

IP Accounts

In their review article, Dietrich and List claim that the question of how probabilities should be aggregated admits of no obvious answer, and, ultimately, the appropriateness of the pooling method depends on the purpose and context of aggregation (2014b). They raise the question of whether a pooling method should be justified on *epistemic* or *procedural* grounds. To be justified on epistemic grounds, "the pooling function should generate collective opinions that [...] respect the relevant evidence or track the truth, for example." In order to be justified on procedural grounds, a pooling method should yield a collective opinion that is a "fair representation of the individual opinions" (2014b, p. 2). Dietrich and List claim that, while linear pooling can be justified on procedural grounds, it cannot be justified on epistemic grounds. By satisfying WSFP, linear pooling functions reflect "the democratic idea that the collective opinion on any issue should be determined by individual opinions on that issue" (2014b, p. 6). Geometric pooling, however, can be justified in epistemic terms "by invoking the axiom of external Bayesianity" (2014b, p. 13). The idea seems to be that since updating is a response to the evidence, a pooling method that is well-behaved in the sense of commuting with updating "respects the relevant evidence" by not allowing the order of operations to distort evidential impact.

Convex IP pooling, then, can be justified on Dietrich and List's procedural grounds in virtue of satisfying WSFP. Concerning procedural grounds in general, it is difficult to think of a more fair or democratic representation of individual opinions than a representation that *includes* each opinion and all of the compromises between opinions. But since convex IP pooling functions also satisfy external Bayesianity, it would thus appear that the alleged tension between epistemic and procedural criteria for probabilistic opinion aggregation can be resolved by simply moving to an IP account.

I endorse the basic motivation for Dietirch and List's discussion. Like other deliberate activities, pooling is *goal-directed*. How one should approach pooling depends on, among other things, one's goals. One may have multiple goals, in which case, tensions in jointly satisfying them may require tradeoffs. Nevertheless, I find it rather opaque precisely how WSFP encodes an intuitive procedural constraint on pooling. Similarly, how commutativity of pooling and updating ensures that the collective opinion respects the relevant evidence or tracks the truth stands in need of further clarification. Even if the *philosophical* distinction and interpretation of WSFP and external Bayesianity does not admit of further clarification, however, our point stands that the tension between satisfying the *formal* desiderata can be resolved in the IP setting.

7. Objections to IP Pooling

Perhaps the relative neglect of IP in discussions of pooling can be explained in part by a skepticism concerning the *use* to which IP sets can be put. In their very nice overview of work on pooling, Genest and Zidek write, "the jury remains out on the theory of Walley [...] In particular, it is unclear how [the IP set] could be used 'at the end of the day'" (1986, p. 124). There are essentially two types of uses to which an account of probabilities may be put: those concerning epistemic issues like inference, and those concerning issues in decision making.

7.1. Epistemology. Some degree of the skepticism about the epistemic usefulness of IP may be dispelled by considering recent work. For instance, after Genest and Zidek's article, Walley published his magisterial book addressing applications of IP to issues in statistical reasoning (1991). Fabio Cozman explores the application of IP to issues in Bayesian networks in a number of papers (1998; 2000).

But let us consider some epistemological challenges of a general sort. In reviewing some difficulties for the few available accounts of pooling IP sets of probabilities (accounts allowing imprecision

at the individual level), Robert Nau claims that neither taking the union nor the intersection of convex sets of imprecise probabilities yields a satisfactory account of pooling. He writes,

> As more opinions are pooled, the union can only get larger, and it reflects only the least informative opinions, whereas intuitively there ought to be (at least the possibility of) an increase in precision as the pool gets larger. On the other hand, the intersection of convex sets of measures may be empty if experts are mutually incoherent, and it generally yields too tight a representation of aggregate uncertainty. As more opinions are pooled, the intersection can only shrink, and it reflects only the most extreme among those opinions, whereas intuitively there should be some convergence to an average opinon when the pool gets sufficiently large. Moreover, neither the union nor the intersection provides an opportunity for the differential weighting of opinions, which would be desireable in cases where one individual is considered (either by herslef or by an external evaluator) to be better or worse informed than another individual about a particular event under consideration. (2002, p. 267)

Similar concerns could be expressed about the account of pooling under examination in this essay since uncertainty never decreases by mere pooling on our account. But I think they would misplaced. The appropriateness of the behavior of a pooling function cannot be assessed in abstract, without specifying the *point* of pooling probabilities in the first place. If the point is to find common ground among the opinions being pooled, increasing uncertainty is to be expected. In general, the more opinions among which we try to find common ground, the less common ground there will be.[11] One might not wish to seek consensus among certain opinions, but that is a different matter. On our account uncertainty *can* be reduced, but through inquiry and not through pooling. As the group acquires sufficient information, conditionalization generally leads to a reduction of imprecision. In the IP setting, it is also possible, however, for conditionalization to *increase* imprecision in the short run, a phenomenon known as *dilation* (Seidenfeld and Wasserman, 1993; Wasserman and Seidenfeld, 1994; Herron et al., 1997; Pedersen and Wheeler, 2014, 2015). But our point here is not that conditionalization invariably decreases uncertainty, but that it can and that decreasing uncertainty through conditionalization has familiar Bayesian "learning" foundations whereas pooling (averaging) does not.

[11] I suspect that Nau is not targeting consensus because his models of pooling involve game-theoretic bargaining scenarios pitting the opinions to be aggregated against each other.

Furthermore, in the case of pooling imprecise probabilities, I would not endorse taking intersections for the purpose of finding consensus. In the case of mutual incoherence, intersections yield the empty set. But the lack of any consensus concerning which probability functions can be ruled out means that the group in consensus cannot rule any probability functions out. Taking the convex hull of the union would reflect this, yielding complete uncertainty.[12]

I think it is important to distinguish between finding consensus among some opinions and taking those opinions as evidence. In the latter case, if an agent outside the group considers some members of the group to be less informed than others, *that* opinion should be reflected in conditionalization through the likelihood for the experts' opinions (Cf. the *Supra-Bayesian* approach to pooling (Genest and Zidek, 1986, p. 120)). In the former case, if a group member is considered, by herself or other group members, to be less informed, consensus is often not sought. Finding what common ground the group members share is unproblematic when consensus is sought, regardless of the social, political, or intellectual clout members accord each other. It is also open to, and perhaps rationally obligatory for, the modest group member to allow her opinion concerning her relative informedness to be reflected in her probabilistic opinion before pooling.

Finally, one might object that IP pooling amounts to declining to really aggregate. In a sense, that is true, if pooling is restricted to taking some sort of average of individual probabilities. But, again, what is the theoretical basis for only considering precise averages of subjective probabilities? An IP set clearly *represents* group opinion, and can be employed in inference and decision making.

7.2. Decision Theory. Because decision theory is a very involved topic and I do not treat it in this chapter, I limit myself to pointing out that sophisticated decision theories for IP have been developed and extensively studied. These include Levi's E-admissibility and tie-breaking decision

[12]See Larry Wasserman's review of Walley's book for objections to this representation of complete uncertainty (1993), and Levi's concept of *confirmational commitment* as a potential means of addressing the objections (1974).

rule (1980), Girón and Rios' quasi-Bayesian decision theory (1980), Gilboa and Schmeidler's Γ-Maximin (1989), and Walley's *Maximality* (1991). Seidenfeld, Schervish, and Kadane axiomatize their theory of coherent choice under uncertainty in the framework of set-valued choice functions (2010). Though saying so overcommits me for my project in this essay, I hold the view that theoretical disputes about probability cannot be adjudicated without thorough decision theoretic considerations.

7.3. Convexity Revisited. How much do the results in this essay depend on the convexity of the IP set? Not much! To see why, consider the following very simple IP pooling function, $\mathcal{F} : \mathbb{P}^n \to \mathscr{P}(\mathbb{P})$, such that

$$\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = \{\boldsymbol{p}_i : i = 1, ..., n\}$$

So defined, $\mathcal{F}$ takes as input a profile of probability functions and returns the set of functions in that profile.

Proposition 7. *Let $\mathcal{F} : \mathbb{P}^n \to \mathscr{P}(\mathbb{P})$ be an IP pooling function such that, for each profile in $\mathbb{P}^n$, $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = \{\boldsymbol{p}_i : i = 1, ..., n\}$. Then, $\mathcal{F}$ satisfies SSFP, WSFP, ZPP, MP, unanimity preservation, external Bayesianity, and confirmational irrelevance preservation. Moreover, $\mathcal{F}$ satisfies stochastic independence preservation.*

The proof of Proposition 7 is straightforward and so is omitted here. The upshot is that pooling with imprecise probabilities is promising in a robust sense. So, while the convex IP pooling model is the chief subject of our philosophical approbation and mathematical analysis in this essay, our case for the consideration of IP in the context of pooling does not rest exclusively with that model.

7.4. Dynamics. As I have presented convex IP pooling functions, the input is a profile of individual probability functions and the output is a convex set of probability functions. What if individual probabilities are themselves imprecise? Or what happens if we attempt to pool the

probabilistic opinions of agents that are themselves groups?[13] As it stands, our account is silent. There is, however, a natural extension of the account on offer. Consonant with the philosophical position staked out here, the idea is to convexify the *union* of sets of probability functions.

$$\mathcal{F} : \mathscr{P}(\mathbb{P})^n \to \mathscr{P}(\mathbb{P})$$

Where the profile consists of n *sets* of probability functions, $(\boldsymbol{P}_1, ..., \boldsymbol{P}_n)$, the pool is given by $\mathcal{F}(\boldsymbol{P}_1, ..., \boldsymbol{P}_n) = conv\{\bigcup_i \boldsymbol{P}_i\}$. I leave examination of this more complete account to future work.

8. Conclusion

According to standard Bayesian theory, personal probabilities are *subjective*. One route that has been explored for recovering some objectivity is establishing intersubjective agreement. There are, for example, the famous convergence theorems to the effect that, given non-extreme priors and a suitably large amount of evidence upon which to conditionalize, posteriors converge (Savage, 1954; Gaifman and Snir, 1982). Consensus in the (*very*) long-run, however, is not the only kind of consensus we may seek. Prior to inquiry, consensus as *shared agreement* is still possible, and desirable for joint efforts in inquiry. Convex IP pooling can be philosophically motivated as an account of such consensus.

Our objective has been to undermine the preconception that probabilistic opinion pooling should result in a representative probability function for the group. Our tack has been to explore another option, arguing that, even by the very lights of those working in pooling, this option is promising. We have the following summary (an "X" means the pooling method does not generally satisfy the property):

[13]The problem being raised is similar to one in the literature on AGM belief revision. The *principle of categorical matching* requires that the output of a belief revision operator be of the same format as the input. Otherwise, the account of belief revision, constructed for a certain input format, is silent about iterated belief revision (Gärdenfors and Rott, 1995). In the case of convex IP pooling functions, dynamics *of pooling* are defined so long as we are never pooling sets of probabilities.

TABLE 2. Pooling Method Report Card

	Linear	**Geometric**	**Multiplicative**	**Convex IP**
SSFP	✓	X	X	✓
ZPP	✓	✓	✓	✓
MP	✓	X	X	✓
WSFP	✓	X	X	✓
Unanimity Preservation	✓	✓	X	✓
External Bayesianity	X	✓	X	✓
Individualwise Bayesianity	X	X	✓	X
Irrelevance Preservation	X	X	X	✓

Perhaps the most sensible representation of group opinion, especially when pooling is interpreted as reaching consensus, is not in terms of a single probability function. At the very least, the arguments and results above may be read both as an exploration of extending the mathematical framework of opinion pooling to cover IP pooling, and as a plea for liberalism about pooling formats.

CHAPTER 3

Learning and Pooling, Pooling and Learning

1. Introduction

Bayesian conditionalization is the gold standard of probabilistic learning. Yet several authors advocate modifications of conditionalization for a number of reasons. For example, conditionalization entails assigning probability 1 to the evidence. Dissatisfied with such "dogmatic epistemology," Richard Jeffrey proposed his *probability kinematics* as a way of updating on *uncertain* evidence (Jeffrey, 2004). To take another example, consider probabilistic *imaging*. One widely pursued goal in work on the logic of conditionals is to find a way of identifying the probability of a conditional with the corresponding conditional probability. Attempts to do so have been repeatedly frustrated by a series of triviality results. However, David Lewis introduces imaging and shows that a version of the identity holds when formulated in terms of imaging instead of in terms of conditionalization (Lewis, 1976). And there are other proposals, such as minimizing the Kullback-Leibler divergence (Kullback and Leibler, 1951). Here, the objective is to accommodate the evidence in such a way as to minimize a measure (the K-L divergence) of the difference between posterior and prior.

Probabilistic opinion pooling can be viewed as part of an important strand in Bayesian epistemology and statistics concerned with consensus. The received view is that personal probabilities are *subjective* (Ramsey, 1990; Savage, 1954; de Finetti, 1964), resulting in much fretting about the implications for scientific methodology. The worry is that the objectivity of scientific confirmation,

A version of this chapter is published as a paper, also coauthored with Rush Stewart, under the same title in *Erkenntnis*.

explanation, inference, and the like is compromised to the extent that such probability plays a role. A prominent Bayesian response comes in the form of convergence and merging of opinions theorems, which show that, given agreement about probability 0 events and enough evidence, probabilities converge (almost surely) (Savage, 1954; Gaifman and Snir, 1982).[1] Conditionalization, that is, leads to consensus, "washing out" the problematically subjective priors leaving intersubjective agreement in their place.

Pooling offers a distinct way of reaching a consensus in probabilistic opinion, one available even when the opportunity to collect more evidence is not. Consensus is reached immediately *via* methods for *aggregating* probabilistic judgments instead of in the long run *via* conditionalization (Huttegger, 2015). After all, as Keynes astutely observes, we have reasons not to be particularly concerned with the long run. As with convergence arguments, intersubjective agreement stands in for objectivity in the pooling context. Still, in the literature on probabilistic opinion pooling, one of the constraints of central concern is *external Bayesianity*, which requires that pooling and Bayesian conditionalization on a common likelihood function commute (Madansky, 1964). That is, the result of pooling and then updating is the same as first updating and then pooling. The order of operations does not change the outcome. One natural question to ask in light of the aforementioned alternatives to conditionalization is, what about commutativity with alternative updating policies? This is the question that concerns us in the present essay.

Much of the focus in the pooling literature is on characterization and impossibility results. Such results are not the intended contribution of this chapter (though Proposition 8 generalizes a characterization result due to Wagner to the imprecise probabilities setting). Instead, we continue an exploration of the potential of imprecise probabilities in the context of learning and pooling. In the previous chapter, I argued that collective opinion is more properly represented by imprecise

[1]Not all merging of opinions results require probabilities to converge to certainty (Blackwell and Dubins, 1962). Under certain conditions, Bayesian conditionalizing can bring probabilities close even if they do not converge to 1 or 0.

probabilities (IP) in general. I provide three arguments. First, if pooling is interpreted as reaching a *consensus* in probabilistic opinion, IP pooling is on firmer philosophical ground (Cf. Levi, 1985; Seidenfeld et al., 1989). The point, briefly, is that IP models allow for suspending judgment between some number of probability distributions by not ruling them out for use in deliberation and inquiry, and reflect the common ground among the group concerning which probability distributions *are* ruled out. Such a consensus constitutes a neutral initial position from which to launch further inquiry. Precise pooling functions, on the other hand, do not allow for an analogous suspense of judgment, and may yield collective probabilistic opinions endorsed by none of the group members. Second, there are IP pooling functions that jointly satisfy more of the standard pooling constraints than any precise pooling recipe can. Third, in the IP setting, the tension between a pooling method's being justified on epistemic or procedural grounds (Dietrich and List, 2014b)—reflected in the tension between satisfying certain formal epistemic and procedural constraints—dissipates, an artifact of the assumption of precision.

The results that follow may be taken to contribute to that case for IP. Briefly put, I show that, while the form of updating for a given precise pooling method is quite restricted under the requirement of commutativity, relaxing the assumption that the collective opinion should take the form of a numerically determinate probability function enables us to lift many of those restrictions. Several revision methods are consistent with pooling understood the IP way. After introducing the mathematical pooling framework in the next section, we begin with the gold standard in Section 3. I remain neutral as to whether Jeffrey conditionalization and imaging ultimately admit of sufficient motivation, though I rehearse some of the standard motivations for and reservations about probability kinematics (Section 5) and imaging (Section 6), and state the commutativity results. Motivations for requiring commutativity of learning and pooling are discussed Section 4.

2. Preliminaries

Let Ω denote a sample space, a set of mutually exclusive and exhaustive possible states of the world.[2] In what follows, we assume that Ω is countable. A function $\boldsymbol{p} : \Omega \to [0,1]$ is a *probability mass function* (pmf) *iff* $\sum_{\omega\in\Omega} \boldsymbol{p}(\omega) = 1$. An algebra $\mathscr{A}$ of events over Ω is a set of subsets of Ω closed under complementation and finite unions; closure under countable unions yields a σ-algebra. We assume throughout the essay that $\mathscr{A}$ is a σ-algebra. Given a pmf, we can define a *probability measure*, (abusing notation by using the same symbol as for pmfs) $\boldsymbol{p}$, on events in general: $\boldsymbol{p}(E) = \sum_{\omega\in E} \boldsymbol{p}(\omega)$.

Let $\mathbb{P}$ be the set of all pmfs on Ω. A *precise* pooling function is a function, $F\colon \mathbb{P}^n \to \mathbb{P}$, mapping a profile of n pmfs, $(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$, to a single pmf, $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$. Typically, the n pmfs are taken to represent the opinions of a set N of individuals, and $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ is supposed to represent, in some sense, the aggregate or collective opinion. Various candidate interpretations of an opinion pool exist in the literature: a rational consensus (adopted genuinely by all members or adopted merely "for the sake of the argument"); a compromise adopted for the purpose of group decision making; the opinion a group member adopts after learning the opinion of her "epistemic peers"; the opinion an external agent adopts upon being informed of the n expert opinions, etc. (Genest and Zidek, 1986; Wagner, 2009). There are a number of concrete pooling functions discussed in the literature, but, by far, the two most prominent are linear and geometric pooling functions.

Linear Opinion Pools. $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = \sum_{i=1}^{n} \alpha_i \boldsymbol{p}_i$, where $\alpha_i \geqslant 0$ and $\sum_{i=1}^{n} \alpha_i = 1$.

A linear opinion pool is just a weighted arithmetic average of the n probability functions. A geometric pooling function takes the (weighted) *geometric* average of the n pmfs.

[2] Ω may be thought of as a partition of a space of agent-relative serious possibilities determined by consistency with a state of full belief. As is a state of full belief, Ω is open to being revised, refined, etc., as judged appropriate (Levi, 1980).

Geometric Opinion Pools. $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = c\prod_{i=1}^{n} \boldsymbol{p}_i^{\alpha_i}$, where $\alpha_i \geqslant 0$ and $\sum_{i=1}^{n} \alpha_i = 1$, and $c = \frac{1}{\sum_{\omega' \in \Omega}[\boldsymbol{p}_1(\omega')]^{\alpha_1} \cdots [\boldsymbol{p}_n(\omega')]^{\alpha_n}}$ is a normalization factor.[3]

The focus of this section of the dissertation is on a generalization of pooling functions to the IP setting: $\mathcal{F} : \mathbb{P}^n \to \mathscr{P}(\mathbb{P})$.[4] We use $\mathcal{F}$, opposed to F, to denote set-valued pooling operators. IP pooling functions are maps from profiles of probability measures to sets of probability measures. In particular, we consider pooling functions that map profiles of n pmfs to the convex hull of those functions: $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = conv\{\boldsymbol{p}_i : i = 1, ..., n\}$. The convex hull is the smallest convex set containing $\boldsymbol{p}_i$ for $i = 1, ..., n$. A set, $\boldsymbol{P}$, is convex if it satisfies the following property.

Convexity. If $\boldsymbol{p}_1, \boldsymbol{p}_2 \in \boldsymbol{P}$, then $\alpha\boldsymbol{p}_1 + (1 - \alpha)\boldsymbol{p}_2 \in \boldsymbol{P}$ for $\alpha \in [0, 1]$. [5]

Put another way, $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ is the set of all convex combinations of the individual probability functions. We let $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A)$ denote the set of probability values assigned to A:

$$\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) = \{\boldsymbol{p}(A) : \boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)\}$$

There are alternative IP formats, including set-based formats that do not require convexity, interval-valued probability functions, or, more operationally, sets of desirable gambles, for instance. In my view, all are worthy of extensive study. Convex sets are quite commonly employed in the theory of IP, including in sophisticated decision theories. Since I do not intend to settle complex debates internal to IP theory here, and the results to come do not hinge on convexity in the sense that there are alternative IP formats for which they hold, I will simply restrict my attention to convex sets as an illustration of the potential of IP in theorizing about pooling.

[3] Notice that, due to the way geometric pooling is defined, there are profiles for which $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(\omega) = 0$ for all $\omega \in \Omega$—in violation of the probability axioms. Such a situation arises if for each $\omega \in \Omega$ there is a $\boldsymbol{p}_i \in (\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ such that $\boldsymbol{p}_i(\omega) = 0$. Circumventing this problem, Wagner restricts the domain of pooling operators to the set of profiles for which this does not happen. That is, the domain of a pooling function is the set of profiles such that there is some $\omega \in \Omega$ for which $\boldsymbol{p}_i(\omega) > 0$ for all $i = 1, ..., n$.

[4] See (Schervish and Seidenfeld, 1990; Herron et al., 1997) for studies of convergence relevant to IP.

[5] Within the IP research community, convexity is a matter of some controversy. For attacks on the requirement, see (Seidenfeld et al., 1989; Kyburg and Pittarelli, 1992; Seidenfeld et al., 2010). For defenses, see (Levi, 1990, 2009).

3. External Bayesianity

Essential to Bayesian methodology is the assumption of a *prior* probability distribution on the algebra of events (or propositions) of concern. Learning proceeds by *conditionalizing* the prior on the evidence, yielding a *posterior* distribution. Conditionalization of a probability function, $\boldsymbol{p}$, on evidence, E, results from setting the posterior probability for any event $A \in \mathscr{A}$ equal to the prior conditional probability $\boldsymbol{p}(A|E)$.

$$\boldsymbol{p}^E(A) = \boldsymbol{p}(A|E) = \frac{\boldsymbol{p}(A \cap E)}{\boldsymbol{p}(E)}, \text{ when } \boldsymbol{p}(E) > 0.$$

The posterior, $\boldsymbol{p}^E$, can be thought of as the result of learning E. In the context of sets of probability functions, conditionalization can be generalized by conditionalizing each member of the set:

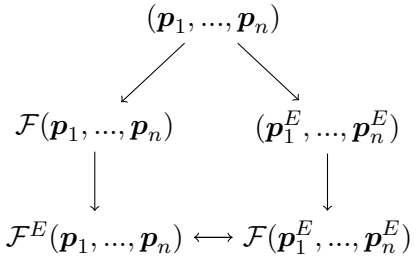
$$\mathcal{F}^E(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = \{\boldsymbol{p}^E : \boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n),\ \boldsymbol{p}(E) > 0, \text{ and } \boldsymbol{p}^E(\cdot) = \boldsymbol{p}(\cdot|E)\}$$

Call $\mathcal{F}^E(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ the *prior-by-prior* conditionalization of $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ (when $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ is convex, $\mathcal{F}^E(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ is called the *convex* conditionalization of $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$).[6] We define prior-by-prior conditionalization generally, allowing $\boldsymbol{p}(E) = 0$ for some $\boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ (Cf. Kyburg, 1987, p. 279). But when we first update the $\boldsymbol{p}_i$, we assume $\boldsymbol{p}_i(E) > 0$ for $i = 1, ..., n$; otherwise, $\mathcal{F}(\boldsymbol{p}_i^E, ..., \boldsymbol{p}_n^E)$ is not defined. Though not stated in the language of probabilistic opinion pooling, proofs of the commutativity of convexifying a set of probability functions and conditionalization exist in the literature.

THEOREM 6. *(Levi, 1978; Girón and Ríos, 1980) Convex IP pooling commutes with conditionalization.*

[6] In the IP setting, conditionalization can actually lead to *greater* uncertainty in the short-run, a very interesting phenomenon known as *dilation* (Seidenfeld and Wasserman, 1993; Pedersen and Wheeler, 2014).

FIGURE 1. Commutativity of Pooling and Conditionalization



Importantly, linear opinion pooling does not commute with conditionalization, though geometric pooling does (Genest, 1984; Russell et al., 2015). As we will see, linear pooling does commute with imaging, though geometric pooling does not.

As standardly pointed out, *external Bayesianity* is a generalization of the requirement that pooling and standard conditionalization commute (Wagner, 2009; Dietrich and List, 2014b; Russell et al., 2015), because conditionalization on a common likelihood function generalizes standard Bayesian conditionalization on an event. A *likelihood function*, $\lambda : \Omega \to [0, \infty)$, is intended to encode, given any $\omega \in \Omega$, how expected some evidence is with the number $\lambda(\omega)$. The conditionalization of a pmf, $\boldsymbol{p}$, on a likelihood function, λ, is give by the following formula.

$$\boldsymbol{p}^{\lambda}(\omega) = \frac{\boldsymbol{p}(\omega)\lambda(\omega)}{\sum_{\omega' \in \Omega} \boldsymbol{p}(\omega')\lambda(\omega')}, \text{ when } \sum_{\omega' \in \Omega} \boldsymbol{p}(\omega')\lambda(\omega') > 0$$

For the special case of Bayesian conditionalization on an event E, define λ as the indicator function for that event:

$$\lambda(\omega) = \begin{cases} 1, \text{ if } \omega \in E \\ \\ 0, \text{ otherwise.} \end{cases}$$

External Bayesianity requires that updating the individual probabilities on a common likelihood function and then pooling is the same as pooling and then updating the pool on that likelihood function.

External Bayesianity. For every profile $(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ in the domain of F and every likelihood function λ such that $(\boldsymbol{p}^\lambda, ..., \boldsymbol{p}_n^\lambda)$ remains in the domain of F, $F(\boldsymbol{p}_1^\lambda, ..., \boldsymbol{p}_n^\lambda) = F^\lambda(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$.

When pooling is presumed to produce a numerically determinate probability function for the group, generalized geometric pooling functions uniquely satisfy external Bayesianity (Genest et al., 1986; Nau, 2002). Extended to the IP setting, the constraint requires $\mathcal{F}(\boldsymbol{p}_1^\lambda, ..., \boldsymbol{p}_n^\lambda) = \mathcal{F}^\lambda(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$. The requirement is still that learning by updating on a common likelihood function and pooling commute, but the format of the pool is altered. What is not altered, we submit, are the compelling aspects of the constraint. Recall the following observation from the previous chapter.

Proposition 3. Convex IP pooling functions are externally Bayesian.

(The *propositions* in this dissertation are my results. When provided, proofs are relegated to the appendix.) And since Bayesian conditionalization is a special case of updating on a likelihood function, it follows that any IP pooling function (not necessarily convex) that is externally Bayesian also satisfies commutativity with conditionalization.

The fact that updating on a common likelihood generalizes updating on an event may serve to show that the assumption of a common likelihood function is not quite as strong as it may appear initially, since the conditionalization of the $\boldsymbol{p}_i$ on some event drops out as a special case. That is, learning the same event is an instance of a shared likelihood function. It is also worth pausing to consider why Bayesians would deal in likelihood functions in the first place if updating with a likelihood function presents ways of learning not reducible to conditionalization.[7] One reason is that, under certain conditions, there is a way of regarding updating with a likelihood function as a case of Bayesian conditionalization by *refining* the algebra so that there is an event such that conditionalizing on it yields the same results as updating with the likelihood function on the coarser

[7] Thanks to Paul Pedersen for emphasizing this point to me.

algebra. We return to this point—which is relevant to Jeffrey Conditionalization as well—at the close of Section 5.

4. Commutativity

But why should it matter if a pooling method is externally Bayesian? More generally, why should we insist on the commutativity of learning and pooling? A few motivations, which I now briefly survey, are offered in the literature.

In introducing the external Bayesianity constraint, Madansky points out that the decisions of a group with common interests employing an externally Bayesian pooling operator will appear to outsiders as the decisions of a single Bayesian agent (1964). How? A Bayesian agent conditionalizes. So, given group opinion, $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$, the updated group opinion should result from the group prior by conditionalization, $F^{\lambda}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$. If the group employs a pooling operator that is not externally Bayesian, and learning happens at the level of individuals, the posterior group opinion may not result from the prior group opinion by conditionalization: $F(\boldsymbol{p}_1^{\lambda}, ..., \boldsymbol{p}_n^{\lambda}) \neq F^{\lambda}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$. If the relevant learning happens at the level of group opinion, then the posterior group opinion may not be the result of applying the (non-externally Bayesian) pooling method that allegedly gives us the way of arriving at group opinion when applied to individual opinions: again, $F(\boldsymbol{p}_1^{\lambda}, ..., \boldsymbol{p}_n^{\lambda}) \neq F^{\lambda}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$.

Of course, there are a number of interpretations of the pool of individual opinions, including as "a rough summary" of the n pmfs (Wagner, 2009). The properties that are appropriate for a pooling function to exhibit depend on the interpretation of the pool and the use to which it is put. When pooling is interpreted as a method of reaching either a compromise or a genuine consensus for use in group decision making, it may be important to ensure adherence of "group opinion" to norms of individual rationality. For views according to which groups can be agents subject to the same rationality conditions, for example, failing to satisfy external Bayesianity raises problems insofar

as Bayesian conditionalization is rationally mandatory. More generally, those problems arise for failures of commutativity of pooling with any rule of learning held to be normatively compelling.

Furthermore, Russell et al. charge pooling methods that fail to commute with conditionalization with vulnerability to a diachronic Dutch book (2015). When the group posterior does not result from the group prior by conditionalizing on what the individuals learn, the conditions for a diachronic Dutch book (at the level of group opinion) are met. Echoing Raiffa (1968, pp. 221-226), Dietrich and List offer other strategic considerations in favor of external Bayesianity. If a pooling method is not externally Bayesian, collective opinion is open to manipulation. By disclosing relevant information at the appropriate time, someone could affect collective opinion by increasing the influence of certain opinions, for example (2014b). There are, in short, possible manipulations besides those of a clever bookie.

Perhaps most uncontroversially, pooling operators for which learning and pooling commute save us the trouble of having to figure out whether updating should come before or after pooling, whether susceptibility to a diachronic book is damning for the pooling method, how and when to safeguard against manipulation, etc. In any event, the main position argued for in this chapter can be understood as a conditional: *if* one finds commutativity of learning (of various types) and pooling compelling, *then* one has reasons to seriously consider IP pooling formats.

5. Jeffrey Conditionalization

As indicated in the introduction, standard Bayesian conditionalization requires that the event conditionalized upon receives probability 1 in the posterior distribution. Jeffrey's point is that not all learning experiences are like that. Sometimes observation leads to a revision in subjective probability even when there is no proposition (event) E that is learned "for certain." Jeffrey's famous candle light example serves to illustrate his point. Suppose you observe your friend's coat, but only under candle light. The coat looks blue, but you are not quite sure. The impact of this

observation is a shift in your subjective probabilities concerning the color of the coat, but none of the options goes to 1. This sort of scenario, some Jeffrey sympathizers claim, is "the normal case" (e.g., Spohn, 2012, p. 38). Improved lighting generally only shifts probabilities a bit more.

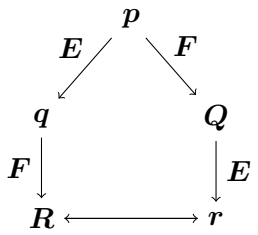
Let $\boldsymbol{E} = \{E_i\}$ be a countable family of pairwise disjoint events partitioning Ω. In the candle light example above, the partition of concern consists of the possible colors of the coat. A posterior, $\boldsymbol{q}$, comes from a prior, $\boldsymbol{p}$, by Jeffrey conditionalization by updating on the new probabilities for the cells of $\boldsymbol{E}, \boldsymbol{q}(E_i)$, in the following way:

$$\boldsymbol{q}(A) = \sum_i \boldsymbol{q}(E_i)\boldsymbol{p}(A|E_i)$$

The $\boldsymbol{q}(E_i)$ express the *direct* effect of an observation on subjective probabilities for the cells of the partition. When $\boldsymbol{q}(E_i) = 1$ for some E_i, Jeffrey conditionalization reduces to standard Bayesian conditionalization.

A fact about Jeffrey conditionalization that many have found problematic is the failure of learning sequences to commute (Skyrms, 1986).

FIGURE 2. Commutativity of Updatings



Suppose that the learning experiences prompting the revision from $\boldsymbol{p}$ to $\boldsymbol{q}$ and the revision of $\boldsymbol{Q}$ to $\boldsymbol{r}$ are the same, as are those leading to the revision of $\boldsymbol{q}$ to $\boldsymbol{R}$ and $\boldsymbol{p}$ to $\boldsymbol{Q}$ (reflected in the updating partitions $\boldsymbol{E}$ and $\boldsymbol{F}$, respectively). Jeffrey conditionalization can yield $\boldsymbol{R} \neq \boldsymbol{r}$. That is, switching the order of two learning experiences can yield different probabilities in the end. Van Fraassen complains:

> Two persons, who have the same relevant experiences on the same day, but in a different order, will not agree in the evening even if they had exactly the same opinions in the morning. Does this not make nonsense of the idea of learning from experience? (1989, p. 338)

It is because of the issue of commutativity of learning experiences (as well as the nice off-the-shelf result of Wagner's (Theorem 8) we appeal to below) that we present here a particular parameterization of Jeffrey conditionalization intended to address the commutativity difficulty.

Hartry Field offers a fix, identifying conditions that are sufficient to ensure that, for finite partitions, $\boldsymbol{R} = \boldsymbol{r}$ in Figure 2 above (Field, 1978). Wagner generalizes the result to countable partitions (Wagner, 2002). We introduce some useful notation. Where A and B are events and $\boldsymbol{q}$ is a revision of $\boldsymbol{p}$, the *Bayes factor* is the ratio of new to old odds:

$$\mathcal{B}(\boldsymbol{q}, \boldsymbol{p}; A : B) = \frac{\boldsymbol{q}(A)/\boldsymbol{q}(B)}{\boldsymbol{p}(A)/\boldsymbol{p}(B)}$$

Instead of being reflected in identical *posteriors*, the proposal on the table is to understand identical learning as reflected in identical Bayes factors. Wagner points out that identifying identical learning with identical Bayes factors has a distinguished pedigree in Bayesian thinking (Good, 1983; Jeffrey, 2004).[8] What Field shows is that Jeffrey conditionalization is commutative when *identical learning* is interpreted as identical Bayes factors.

[8] Wagner contends that identical learning should be thought of as identical Bayes factors rather than identical posteriors. One alleged reason is that posteriors are tainted by the prior, whereas Bayes factors are an uncontaminated measure of the impact of the evidence. How do Bayes factors measure the impact of the evidence in isolation from the prior? Consider the case in which $\boldsymbol{q}$ comes from $\boldsymbol{p}$ by Bayesian conditionalization on E. Then,

$$\boldsymbol{q}(A)/\boldsymbol{q}(B) = \frac{\boldsymbol{p}(A|E)}{\boldsymbol{p}(B|E)}$$

and

$$\mathcal{B}(\boldsymbol{q}, \boldsymbol{p}; A : B) = \frac{\boldsymbol{p}(A|E)/\boldsymbol{p}(B|E)}{\boldsymbol{p}(A)/\boldsymbol{p}(B)}.$$

So, $\mathcal{B}(\boldsymbol{q}, \boldsymbol{p}; A : B)$ is a measure of the change the evidence, E, induces in favor of A over B. $\mathcal{B}(\boldsymbol{q}, \boldsymbol{p}; A : B)$ can also be rearranged using Bayes' theorem.

$$\frac{\boldsymbol{q}(A)}{\boldsymbol{q}(B)} = \frac{\boldsymbol{p}(A|E)}{\boldsymbol{p}(B|E)} = \frac{\frac{\boldsymbol{p}(A)\boldsymbol{p}(E|A)}{\boldsymbol{p}(E)}}{\frac{\boldsymbol{p}(B)\boldsymbol{p}(E|B)}{\boldsymbol{p}(E)}} = \frac{\boldsymbol{p}(A)\boldsymbol{p}(E|A)}{\boldsymbol{p}(B)\boldsymbol{p}(E|B)} = \frac{\boldsymbol{p}(A)}{\boldsymbol{p}(B)} \times \frac{\boldsymbol{p}(E|A)}{\boldsymbol{p}(E|B)}$$

THEOREM 7. *(Wagner, 2002, Theorem 3.1) Consider the revision scheme of Figure 2. If*

$$\mathcal{B}(\boldsymbol{q}, \boldsymbol{p}; E_{i_1} : E_{i_2}) = \mathcal{B}(\boldsymbol{r}, \boldsymbol{Q}; E_{i_1} : E_{i_2}) \text{ for all } i_1, i_2$$

and

$$\mathcal{B}(\boldsymbol{R}, \boldsymbol{q}; F_{j_1} : F_{j_2}) = \mathcal{B}(\boldsymbol{Q}, \boldsymbol{p}; F_{j_1} : F_{j_2}) \text{ for all } j_1, j_2,$$

then $\boldsymbol{R} = \boldsymbol{r}$.

Wagner further shows that Jeffrey conditionalization has an equivalent parameterization in terms of Bayes factors (2009, Theorem 3.2). The function $\boldsymbol{q}$ gives us posteriors for atomic events, $b_k = \mathcal{B}(\boldsymbol{q}, \boldsymbol{p}; E_k : E_1)$, $k = 1, 2, ...$, and $[\omega \in E_k]$ is the characteristic function of the set E_k:

$$[\omega \in E_k] = \begin{cases} 1, \text{ if } \omega \in E_k \\ \\ 0, \text{ otherwise.} \end{cases}$$

Wagner's parameterization, then, is the following.

$$\boldsymbol{q}(\omega) = \boldsymbol{p}^{\boldsymbol{E}}_{J}(\omega) = \frac{\sum_k b_k \boldsymbol{p}(\omega)[\omega \in E_k]}{\sum_k b_k \boldsymbol{p}(E_k)}$$

There are two very nice features of this parameterization that are relevant. First, as I have been explaining, it responds to the complaints about commutativity because the result of a sequence of updates is invariant under permutations of that sequence when Jeffrey conditionalization is understood this way, with identical learning reflected in identical Bayes factors instead of identical posteriors.

Dividing now by $\frac{\boldsymbol{p}(A)}{\boldsymbol{p}(B)}$, the denominator of $\mathcal{B}(\boldsymbol{q}, \boldsymbol{p}; A : B)$, gives us

$$\mathcal{B}(\boldsymbol{q}, \boldsymbol{p}; A : B) = \frac{\boldsymbol{p}(E|A)}{\boldsymbol{p}(E|B)}$$

The quantity $\boldsymbol{p}(E|A)/\boldsymbol{p}(E|B)$ is sometimes referred to as the *likelihood ratio*. So, the Bayes factor is a ratio of the non-prior quantities involved in Bayes' theorem, the quantities that revise the prior.

The second nice feature, as Wagner shows, is that with his parameterization, we can articulate a version of commutativity with Jeffrey conditionalization that provides us with a characterization of externally Bayesian pooling operators in terms familiar to formal epistemologists and philosophers of science. We call Wagner's version of commutativity with Jeffrey CJC_W.

$\boldsymbol{CJC_W}$. For all partitions $\boldsymbol{E} = \{E_k\}$ of Ω, all profiles $(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ in the domain of F, the Jeffrey update of the pool, $F_J^{\boldsymbol{E}}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = \frac{\sum_k b_k F(\boldsymbol{p}_1,...,\boldsymbol{p}_n)[\cdot \in E_k]}{\sum_k b_k F(\boldsymbol{p}_1,...,\boldsymbol{p}_n)(E_k)}$, is identical to $F(\frac{\sum_k b_k \boldsymbol{p}_1[\cdot \in E_k]}{\sum_k b_k \boldsymbol{p}_1(E_k)}, ..., \frac{\sum_k b_k \boldsymbol{p}_n[\cdot \in E_k]}{\sum_k b_k \boldsymbol{p}_n(E_k)}) = F(\boldsymbol{p}_{1J}^{\boldsymbol{E}}, ..., \boldsymbol{p}_{nJ}^{\boldsymbol{E}})$, the pool of the (Jeffrey updated) posteriors.[9]

Crucially, the Bayes factors, b_k for $k = 1, 2, ...$, are held fixed across the $\boldsymbol{p}_i$ (and also used in updating $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$). This is quite different from holding fixed a common posterior distribution, $\boldsymbol{q}$, in Jeffrey conditionalizing the $\boldsymbol{p}_i$. Put another way, he shows that External Bayesianity is equivalent to CJC_W.

THEOREM 8. *(Wagner, 2009, Theorem 3.3) A (precise) pooling operator is externally Bayesian iff it satisfies CJC_W.*

We take $\mathcal{F}_J^{\boldsymbol{E}}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ to be given by Jeffrey conditionalization of each element of $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ on the partition $\boldsymbol{E}$, holding fixed the Bayes factors b_k for $k = 1, 2, ...$ for $\boldsymbol{p}_i$ for $i = 1, ..., n$. A very slight mathematical generalization allows us to extend Wagner's result to IP pooling functions in general.

PROPOSITION 8. *Let $\mathcal{F} : \mathbb{P}^n \to \mathscr{P}(\mathbb{P})$ be an IP pooling function (not necessarily convex). $\mathcal{F}$ is externally Bayesian iff $\mathcal{F}$ satisfies CJC_W.*

[9] Wagner's version of commutativity with Jeffrey conditionalization involves some additional technical assumptions. First, that $\boldsymbol{p}_i(E_k) > 0$ for all i and all k. Second, that $b_1 = 1$ and $\sum_k b_k \boldsymbol{p}_i(E_k) < \infty$ for $i = 1, ..., n$. Third, where $\boldsymbol{q}_i(\omega) = \frac{\sum_k b_k \boldsymbol{p}_i(\omega)[\omega \in E_k]}{\sum_k b_k \boldsymbol{p}_i(E_k)}$, it is the case that $0 < \sum_k b_k F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(E_k) < \infty$. In the IP setting, this last assumption may be adjusted to be a requirement for each $\boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$.

In particular, moving to the convex IP setting does not break the equivalence between external Bayesianity and commutativity with Wagner's parameterization of Jeffrey conditionalization. Putting Propositions 3 and 8 together, we obtain the following.

PROPOSITION 9. *Convex IP pooling satisfies* CJC_W.

We have the following immediate corollary (the proof is trivial given Proposition 9, and is omitted).

PROPOSITION 10. *Convexity is preserved under Jeffrey conditionalization with common Bayes factors.*

To say that convexity is preserved means that, if we start with a convex set, we do not lose convexity in moving to the set of updated probability functions.

However, when $\boldsymbol{q}_i$ comes from $\boldsymbol{p}_i$ by standard Jeffrey conditionalization on some shared posterior distribution, $\boldsymbol{q}$, over a partition, $\boldsymbol{E}$, and the pool is updated likewise by updating each element of $\mathcal{F}(\boldsymbol{p}_i, ..., \boldsymbol{p}_n)$ on that same posterior distribution over $\boldsymbol{E}$, Jeffrey conditionalization and convex IP pooling do *not* commute.

PROPOSITION 11. *Convex IP pooling does* ***not*** *commute with Jeffrey conditionalization on a common posterior.*

Instead of holding fixed common Bayes factors, here we assume a fixed posterior distribution on $\boldsymbol{E}$. Neither is commutativity of Jeffrey conditionalization so formulated and pooling guaranteed in the precise setting (Wagner, 2009, pp. 340-341). In particular, linear and geometric pooling fail to commute with Jeffrey in general. Furthermore, certain "objective" Bayesian updating methods, like minimizing the Kullback-Leibler divergence between posterior and prior, generalize Jeffrey conditionalization (Diaconis and Zabell, 1982a). A corollary of Proposition 11, then, is that minimizing the Kullback-Leibler divergence does not commute with convex IP pooling. Minimizing

the Kullback-Leibler divergence is also a generalization of Jaynes' Maximum Entropy formalism (e.g., Williams, 1980a). While there are many advocates of the Kullback-Leibler approach (e.g., Hartmann, 2014a), even in the precise setting, a number of Bayesian-style objections to MaxEnt methods have been voiced in the literature (see, e.g., Seidenfeld, 1986; Gaifman and Vasudevan, 2012).

How much ground does Jeffrey conditionalization ultimately gain over standard Bayesian conditionalization? In certain cases—characterized by the *superconditioning criterion*—Jeffrey conditionalization can be represented as Bayesian updating on *certain* evidence in an enlarged probability space (Diaconis and Zabell, 1982a, Theorem 2.1). For finite algebras, the effect of Jeffrey conditionalization can always be so represented by Bayesian conditionalization (fn. 1, Wagner, 2002, p. 268). Kyburg attributes an early, informal version of this result to Levi (Levi, 1967; Kyburg, 1987, Lemma A.5). The point, as Kyburg puts it, is not that we make effable the ineffable observational input motivating Jeffrey conditionalization: "This is not to say that we need to *specify* that evidence; it is that there is an algorithm by means of which the impact of the uncertain evidence can be represented as the impact of other 'certain' evidence" (Kyburg, 1987, p. 280). The obvious question is whether or when the formal superconditioning move is a philosophically legitimate one.

6. Imaging

An hypothesis occupying the attention of many scholars working on the logic of conditionals asserts that the probability of a conditional is identical to the relevant conditional probability: $\boldsymbol{p}(A \to B) = \boldsymbol{p}(B|A)$. There are, of course, a number of ways to interpret the components of the claim. There are, for instance, various ways to define a conditional probability, $\boldsymbol{p}(B|A)$, just as there are various ways to specify an interpretation of a conditional connective, $\to$. Suppose we take conditional probability to be defined standardly as in Section 3. Is there an interpretation of $\to$ such that the desired identity holds for all $\boldsymbol{p}$? No, on pain of triviality, as Lewis proved (1976).

That is, a conditional satisfying the identity exists only for trivial probability models. Similar triviality results hold for alternative readings of the identity. For example, "for any $\boldsymbol{p}$ there exists some $\rightarrow$ such that $\boldsymbol{p}(A \rightarrow B) = \boldsymbol{p}(B|A)$" runs into similar problems. An impressive battery of such triviality results has been obtained for different ways of reading the identity. A helpful overview of much of the relevant literature can be found in (Hájek and Hall, 1994). Though it fails in general when formulated in terms of conditionalization, Lewis shows that a version of the identity holds if formulated in terms of *imaging* instead. We turn now to a brief presentation of imaging and Lewis' possibility result.

Robert Stalnaker specified the semantics of the so-called *Stalnaker conditional*, $>$, in terms of possible worlds. Ω is interpreted as a set of possible worlds.[10] Propositions are subsets of Ω. We assume that $\mathscr{A}$ is a σ-algebra of subsets of Ω, the set of relevant propositions. For any $\omega \in \Omega$, let ω_A be the "most similar" possible world at which A holds, the "closest" A-world to ω. Say that $A > B$ is true at ω *iff* B is true at ω_A (when the antecedent is impossible, $A > B$ is taken to be vacuously true at all worlds). Lewis tailors a probabilistic revision scheme to the Stalnaker semantics.

For any non-empty $E \in \mathscr{A}$, imaging shifts the probability from each $\omega' \in \Omega$ to its "image" atom, $\omega \in E$. If $\omega' \in E$, then ω' is its own image atom. Lewis offers an interpretation in terms of possible worlds. On the assumption that for each world there is a unique "most similar" world in E, imaging can be thought of as the process of revising probabilities by shifting the total probability of each world to its most similar world in E. Relaxing the uniqueness assumption results in what is known as *general imaging*. General imaging allows the probability of each $\omega' \in \Omega$ to be shifted to an image *set*, each element of which receives some fraction of the total probability of ω'. Clearly, general imaging reduces to imaging when the image set is a singleton and the total probability of each $\omega' \in \Omega$ is transferred to its image set.

[10] A metaphysically deflationary conception of possible worlds has it that a possible world is just a maximally complete set of sentences in some propositional language, instead of a "possible totality of facts."

Formally, we represent the relevant transfer of probability with a transfer function, $T : \mathscr{A} \times \Omega \times \Omega \to [0, 1]$, such that $\sum_{\omega \in \Omega} T_E(\omega', \omega) = 1$ for all $\omega' \in \Omega$. For any E and all $\omega, \omega' \in \Omega$, $T_E(\omega', \omega)$ (times 100 percent) specifies the percentage of the total probability mass that is transferred from ω' to ω. It is sometimes assumed—e.g., by Lewis but not by Leitgeb (Leitgeb, 2016)—that $\sum_{\omega \in E} T_E(\omega', \omega) = 1$, so that E bears probability 1 after imaging on it. With T in place, we can formulate the recipe for general imaging. Say that $\boldsymbol{q}$ comes from $\boldsymbol{p}$ by *general imaging* if

$$\boldsymbol{q}(\omega) = \boldsymbol{p}(\omega||E) = \sum_{\omega' \in \Omega} \boldsymbol{p}(\omega') T_E(\omega', \omega)$$

The constraint on T of summing to 1 for each ω' ensures that all probability mass is transferred, so no probability mass is created or destroyed, and the result of imaging is again a pmf. As before, the probability of an event $A \in \mathscr{A}$ can be obtained by summing across $\omega \in A$. Lewis claims that conditionalization and imaging are both minimal revisions, but in different senses. While conditionalization "does not distort the profile of probability ratios, equalities, and inequalities among sentences that imply A," imaging "involves no gratuitous movement of probability from worlds to dissimilar worlds" (1976, p. 142). Lewis proves the following possibility result for (sharp) imaging and the probability of conditionals.

THEOREM 9. *(Lewis, 1976, p. 142) The probability of a Stalnaker conditional with a possible antecedent is the probability of the consequent after imaging on the antecedent:* $\boldsymbol{p}(A > B) = \boldsymbol{p}(B||A) = \boldsymbol{q}(B)$.

More important for the purposes of the present essay is that, as Hannes Leitgeb observes, a result about general imaging due to Peter Gärdenfors can be restated in the language of pooling operators (2016). By *update mechanism*, Leitgeb means a function $U : \mathbb{P} \times \mathscr{A} \to \mathbb{P}$ that maps a probability function and a (non-empty) proposition to a probability function. Gärdenfors shows

that general imaging is the unique probabilistic revision method that preserves convex combinations of probability measures. Leitgeb repurposes this result, obtaining the following insight about pooling.

THEOREM 10. *(Cf. Gärdenfors, 1982, Theorem 1) Update by general imaging (with respect to a fixed transfer function T) is the unique update mechanism that commutes with linear pooling with respect to arbitrary coefficients.*

(Here, the transfer function, T, is invariant across priors.) Together with Leitgeb's insight, Gärdenfors' theorem makes showing the next result very easy.

PROPOSITION 12. *Convex IP pooling commutes with general imaging.*

(Commutativity with imaging (CI) could be stated as an axiom for pooling operators.) An analogue of Proposition 10 follows immediately from Proposition 12: convexity is also preserved under general imaging.

Interest in imaging extends beyond natural language semantics and the philosophy of language. Part of the concern with identifying the probability of a conditional with the relevant conditional probability, after all, comes from attempts to give *acceptability* conditions for conditionals. Furthermore, some have argued that, while conditional probability represents *matter-of-fact* supposition in the context of probability, imaging represents *counterfactual* supposition.[11] Lewis himself thought that an interpretation like Stalnaker's is right for *subjunctive* conditionals or counterfactuals, but not for indicative conditionals. Imaging finds crucial employment in James Joyce's account of causal decision theory. Causal relationships are thought to be expressed by subjunctive conditionals, so the probability of such conditionals is of central concern on that view (1999). Baratgin and Politzer contend that empirical evidence indicates that general imaging has some claim as a description of

[11]Others, however, have offered more uniform accounts of supposition (e.g., Levi, 1996).

actual revision of probabilistic judgment in dynamic environments (2010). Many authors, however, complain that a philosophically defensible interpretation of the requisite similarity relation among possible worlds has yet to be provided. Some see both the Stalnaker conditional and imaging as questionable shifts from epistemology to metaphysics (Arló-Costa, 2007). Moreover, beginning at least with Ramsey, an alternative line of research attempts to provide acceptability conditions for counterfactuals in terms of belief revision theory, eschewing construals of counterfactuals as bearing truth values (Levi, 1996). Nevertheless, imaging seems to have captured the fancy of many philosophers and others working on conditionals, counterfactual reasoning, and decision theory.

7. Discussion

Propositions 9 and 10 have important implications outside of the context of opinion pooling. While IP models have been widely and convincingly advanced as superior to precise Bayesian representations of uncertainty, standard conditionalization *via* certain learning has been by and large retained as the relevant updating rule (Levi, 1978; Girón and Ríos, 1980). Proposition 10 shows that there is no *mathematical* necessity in that retention for convex Bayesians. For those compelled by Jeffrey's vision of learning, they can have their convex sets of probabilities and their probability kinematics, too.[12] A similar point holds for imaging. Since convexity is preserved under imaging, imaging constitutes a possible "dynamics" for convex Bayesians.[13]

Furthermore, in the precise setting, only linear opinion pooling commutes with imaging. But linear opinion pooling does not commute with Bayesian conditionalization. It follows that no precise pooling method commutes with both imaging and Bayesian conditionalization. In this way, one's

[12]Though, as Diaconis and Zabell's aforementioned result shows us, in a range of cases there is no mathematical necessity in adopting Jeffrey conditionalization in order to obtain the results of Jeffrey conditionalization.

[13] Though it is not uncontroversial that conditionalization or some other type of updating of represents *learning*. Isaac Levi, for instance, writes, "All conditions of rationality are equilibrium conditions. In a sense they are synchronic conditions [...] Furthermore, in stating conditions of rational equilibrium, no prescription is made regarding the psychological path to be taken in moving from disequilibrium or from one equilibrium position to another. In other words, there are no norms prescribing rational learning processes" (Levi, 1970).

hand is forced on the question of updating methods by commitments to pooling methods, and *vice versa*. Not so in the imprecise setting.

TABLE 3. Summary of Pooling and Updating Commutativity

	Linear	**Geometric**	**Convex IP**
External Bayesianity	X	✓	✓
CJC_W	X	✓	✓
CI	✓	X	✓

Another way to put the point is that, if commutativity of learning and pooling is endorsed *and* more than one updating method is found acceptable (depending on context, say), then there may exist no accommodating precise pooling method.

Further limitations issue from results obtained in the literature. For example, suppose commutativity of pooling with both standard Bayesian conditionalization (or Jeffrey conditionalization) and marginalization is endorsed. In the precise setting, we are out of luck. Again, not so in the IP setting, as the results of this chapter in conjunction with those of the previous chapter attest. Philosophical positions that argue for or assume that pooling should be of a particular format are answerable for the limitations of those methods. For instance, in the epistemological debate about peer disagreement, a prominent position encourages peers to "split the difference" between their probabilistic opinions (Elga, 2007). So-called *conciliatory* views on disagreement generally counsel revising opinions in the direction of the dissenting opinions. The revision goes by equal-weight or near equal-weight linear pooling (Christensen, 2009). Some consequences of the failure of commutativity with conditionalization are highlighted in (Russell et al., 2015). As indicated above, Russell, et al. allege that a variant of a diachronic Dutch book can be made against parties following such a policy of disagreement resolution. Similar points can be made regarding other properties of particular pooling methods. For example, neither linear nor geometric pooling preserves probabilistic independence in general (Genest and Wagner, 1987), though convex IP pooling preserves

Levi's *confirmational irrelevance*, a generalization of probabilistic independence (Proposition 5). Seidenfeld, Schervish, and Kadane offer a decision-theoretic counterexample to the reasonableness of linear pooling on the basis of its failure to preserve independence (2010). Arguing similarly, Elkin and Wheeler present a variant of a Dutch book argument against resolving disagreements according to the equal weight view (Elkin and Wheeler, 2016). I submit that, not only are IP pooling functions more flexible formal tools, but they admit of stronger normative motivations when various prominent pooling criteria (including the commutativity criteria above) are taken as normative yardsticks.

CHAPTER 4

Radical Pooling and Imprecise Probabilities

> Phaedrus: And what is the other principle, Socrates?
> Socrates: That of dividing things again by classes, where the natural joints are, and not trying to break any part, after the manner of a bad carver.
>
> Plato, *Phaedrus*, 265d-265e

1. Introduction

One interpretation for the operation of pooling or aggregating probabilistic judgments amounts to finding a consensus among the parties involved. Much of the literature focused on a constrained form of disagreement among the parties, namely when they disagree *solely* on the degree of belief assigned to events on a *shared* outcome and event space (for a excellent survey see Dietrich and List (2014b); Genest and Zidek (1986); Wagner (2009)). The purpose of this paper is to generalize the question of pooling to cover cases when agents disagree *radically*; not only on their probabilistic judgments but also on the logical space over which those judgments are made. I will further argue that we can import rationality criteria for these cases and they will require the use of imprecise probabilities.

The kind of radical divergence explored here will not involve moral or political disagreement. Following the literature on pooling, agents' attitudes are modeled by probability functions over

a algebra of events. Preferences, or formalisms to represent agents' valuational structure, will be disregarded in the aggregation procedure.[1]

Diverging parties seeking consensus may find that the tension comes not much from the degree of belief they assign to certain events, but rather from *what each takes to be the relevant set of questions and issues*. According to the partition theory initiated by Groenendijk and Stokhof (Groenendijk and Stokhof, 1984), the meaning of a question is identified with a set whose members correspond to each possible answer; i.e. a partition of the logical space where each member of the partition corresponds to a proposition expressing one of the possible answers. Parties in disagreement about which are the relevant questions will therefore parse the space of doxastic possibilities differently.

The paper has four sections. This introductory one will conclude with a motivating example from the history of science. The second section introduces the basic setup, which involves identifying the three possible *loci* of disagreement, and elaborating the notion of pooling as a form of consensus as common ground at the outset of inquiry. Section three contains the substance of the paper. Sections 3.1 and 3.2 argue for a way of understanding the common sample and event spaces, and they make a connection with the issue of conceptual transformation in theory change. Section 3.3 has the main arguments for imprecision. If the arguments about the common sample and event spaces are endorsed, then (i) marginalization, (ii) rigidity, and (iii) divergence accounts of how to strengthen probabilistic judgments to larger algebras will lead to the adoption of imprecise probabilities. Section 3.4 completes the approach by pooling imprecise probabilities. The fourth and final section offers my conclusion and presents a picture of the framework.

1.1. An Example: The Priestley - Lavoisier debate.

[1]The attempt to jointly aggregate probabilities and preferences is arguably impossible (Seidenfeld et al. (1989)).

Finding consensus in the presence of *radical* disagreement resembles questions about (revolutionary) theory change in philosophy of science. Central to the revolutionary view of science advanced by Kuhn (1970, 2000) and Feyerabend (1962) is the claim that language used in a field of science changes so radically during a revolution that the old language and the new one are not inter-translatable. More generally, they espoused the idea that in those situations opposing theories are *incommensurable*. Similarly, agents disagreeing about the logical space of possibilities over which they make their probabilisitic judgments share no common basis on which to *measure* their divergence.

The resemblance has its limits. First, pooling is about finding *rational consensus* between parties. But according to the mainstream reading of Kuhn, in scientific revolutions parties do not have a recourse to a common theoretical language, body of observational evidence, or methodological rules. For these reasons, disputes cannot be resolved on purely epistemic grounds, but rather by revolution, conversion, gestalt switch, bargaining or some other psychological, social, economic or political process. In contrast, I will here argue that finding consensus in radical disagreement can be achieved by invoking epistemic rationality principles akin to those that are endorsed in standard non-radical pooling. Second, pooling is an *information aggregation* procedure. In contrast, Kuhn initially used incommensurability predominately to challenge cumulative characterizations of scientific advance, according to which scientific progress is an improving approximation to the truth.

Despite the differences, the analogy is helpful. Kuhn's notion of incommesurability changed through his career, but the distinction between methodological and taxonomical incommesurability is important here (Kuhn, 2000). Methodological incommensurability is the idea that there are no shared, objective standards of scientific theory appraisal, so that there are no external or neutral standards that univocally determine the comparative evaluation of competing theories. We are *not* concerned with it here. On the other hand, taxonomic or conceptual incommesurability is more

useful for us. During periods of scientific revolution, existing concepts are replaced with new concepts that are incompatible with the old ones because they do not share the same lexical taxonomy, they cross-classify objects according to different sets of kinds (Kuhn, 2000). To paraphrase the *Phaedrus*, they carve nature at *different* joints. The well studied debate between Joseph Priestley and Antoine Lavoisier is revealing.

Priestley advocated the Phlogiston Theory, which attempted to give an account of a number of chemical reactions and in particular combustion. Its basic explanatory hypothesis (H) was that there is a substance which is emitted in combustion, namely *phlogiston*, and which is normally present in the air. For example, wood and other materials that burn easily are rich in phlogiston. Similarly, when a metal is heated, phlogiston is emitted, and we obtain the calx of the metal. Furthermore, the process is reversible in some cases. By heating the red calx of mercury in isolation, Priestley found that he could obtain the metal mercury, and a new kind of *air* which he called dephlogisticated air. Priestley even noticed that dephlogisticated air supported breathing and combustion better than regular air, and explained this by pointing out that when phlogiston is removed from the air, the latter acquired a greater capacity for absorbing the former.

In contrast, Lavoisier's theory of combustion is akin to our modern theory. Below, a helpful table comparing experiments and their descriptions.

Phlogiston Theory		**Modern Theory**	
Input	Output	Input	Output
Metal + air + heat	Calx of metal + phlogisticated air	Metal + air + heat	Metal oxide + air which is poor in oxygen
Red calx (oxide) of mercury + heat	Mercury + dephlogisticated air	Oxide of mercury + heat	Mercury + oxygen

It seems it would be easy to attain some consensus. After all, identifying "dephlogisticated air" with "oxygen" provides a good mapping for these experiments. But the problem is deeper. Phlogiston theory departs from the false presupposition (H) that there is a unique substance which is always emitted in combustion, and this contaminates its terminology and carving of the logical space. This is precisely Kuhn's and Feyerabend's point. Important terms of phlogiston theory are *theory-laden*. Under the guise of H, the term phlogiston refers to that which is emitted in all cases of combustion. For Lavoisier and the modern theory, H is false and therefore it seems that phlogiston fails to refer.

This taxonomical incommesurability leaves us with a historical explanatory gap. Although they recognize that phlogiston theory is incorrect, historians of science might want to explain its success, give credit to their discoveries, and acknowledge some of their truths. But such a revolutionary theory change seems to leave no room for continuity, much less consensus. Kitcher (1978) analysis of this case can save us from this bleak prospect. Through this case study, Kitcher argues for a semantics of theoretical expressions that avoids the problem:

> I suggest that an expression-type used by a scientific community is associated with a set of events such that productions of tokens of that type by members of the community are normally initiated by an event in the associated set. The set which is associated with a particular expression-type will be called the *reference potential* of the expression. Terms which have heterogeneous reference potential, that is, terms whose reference potential contains two or more different initiating events, may reasonably be called *theory-laden*. [pg. 540]

In particular, Kitcher explains how Pristley's use of dephlogisticated air is theory-laden. Priestley's early utterances of the expression, driven by the hypothesis that phlogiston is emitted in combustion, correspond to the substance obtained by removing from air the substance which is emitted in combustion. Those token expressions of dephlogisticated air corresponded to a set of referents that is different fromo posterior token expression. Priestley's later utterances, after his isolation yet misidentification of oxygen through the heating of a red calx (oxide) of mercury, correspond to the gas obtained by heating the red calx of mercury. Hence later tokens of dephlogisticated

air refer to oxygen. Lavoisier found a way to interpret the different tokens Priestley used, and therefore carve the space in a more refined way. Furthermore, in so far as Lavoisier *et al.* used the term oxygen to refer to the late tokens of dephlogisticated air, there was room for communication between theoretical rivals.

One way of representing the problem in the language of Bayesianism is the following. Driven by the hypothesis that phlogiston is always emitted in combustion, Priestley, Cavendish, *et al.* partitioned their event space so that there is an event P that represents the proposition that the output of the experiment includes phlogisticated air. More formally, given an underlying outcome space $\Omega = \{w_1, w_2, ..., w_n\}$, they organized the algebra over Ω so that there is $P = \{w_i : w_i$ s.t. the experiment results in the release of phlogiston into the air$\}$. After heating of a red calx (oxide) of mercury, Priestley discovered a gas that had a great capacity for absorbing phlogiston, and called it dephlogisticated air. More generally, he discovered a method for extracting phlogiston from the air. This corresponds to a proposition stating that that the output of the experiment is *de*phlogisticated air. This proposition corresponds to an event DP in his carving: $DP = \{w_i : w_i$ s.t. the experiment results in the removal of phlogiston from the air$\}$. P and DP are clearly mutually exclusive, but we do not need to assume they are exhaustive to understand the source of radical disagreement with Lavoisier. What matters, nevertheless, is that propositions and events like P and DP were at the core of phlogiston theory.

Lavoisier *et al.* were doing similar experiments so for simplicity I will assume they shared the same sample space $\Omega = \{w_1, w_2, ..., w_n\}$ with Priestley *et al.*. The difference, nevertheless, is that their carving of the algebra was more *refined*. Where Priestley saw a gas without phlogiston, Lavoisier sometimes saw oxygen, *but not always*. In other words, Lavoisier understood that dephlogisticated air was a misnomer, with a coarse heterogeneous reference potential that could be partitioned even more. The proposition DP, then, is also heterogeneous and includes instances of oxygen release but also others. The proposition O that the outcome of the experiment is the release

of oxygen entails but it is not entailed by DP, $O \subsetneq DP$. By rejecting the hypothesis that there is a substance released in every combustion, Lavoisier could group experiments like the heating of the oxide of mercury that result in the release of oxygen in a more refined class. In a further oversimplification, let $\mathscr{A}_P$ be the phlogiston theory algebra, the carving of the logical space of outcomes, and let $\{w_1, w_2\} = DP \subsetneq \mathscr{A}_P$. Furthermore, given H, the absence of phlogiston in the air was at the core of the theory. Tokens of dephlogisticated air with different referents were taken as a unit. If $\{w_1\}$ corresponds of instances with oxygen and $\{w_2\}$ without, then $\{w_1\}, \{w_2\} \notin \mathscr{A}_P$. For Lavoisier, $\{w_1\}, \{w_2\} \in \mathscr{A}_L$, so that at least with respect to this issue he made more relevant distinctions. He carved the event space where the natural joints are.

What, then, would be a good way of aggregating the information? What kind of consensus could we define between parties? For this particular case, taking the meet $\mathscr{A}_P \wedge \mathscr{A}_L$ as a consensus position for the logical carving will ensure that all distinctions are preserve, and it will be the solution defended. But the problem is more general and requires a more systematic presentation.

2. The Framework

2.1. Setup.

The formal representation machinery is the usual. For each agent i, $\Omega_i = \{w_1, w_2, ..., w_n\}$ denotes a sample space, a set of mutually exclusive and exhaustive possible states of the world. For simplicity, Ω_i is assumed to be countable. A function $\boldsymbol{p}_i : \Omega_i \to [0, 1]$ is a *probability mass function* (pmf) *iff* $\sum_{\omega \in \Omega_i} \boldsymbol{p}(\omega) = 1$. The *agenda*, or the set of events under consideration, is assumed to be a algebra $\mathscr{A}_i$ of events over Ω_i - that is, a set of subsets of Ω_i that includes the empty set and is closed under complementation and finite unions (in the general case, closure under countable unions yields a σ-algebra). Given a pmf, we can define a *probability measure*, (abusing notation by using the same symbol as for pmfs) $\boldsymbol{p}_i$, on events in general: $\boldsymbol{p}_i(E) = \sum_{\omega \in E} \boldsymbol{p}_i(\omega)$. Notice, that although a

pmf is sufficient to define a probability measure, it is not necessary. Probability functions can be defined directly via axiomatic means.[2]

There are then three *loci* for disagreement among agents, each corresponding to the concepts just defined: the sample spaces Ω, the algebras $\mathscr{A}$ defined over those sample spaces, and the probability functions $\boldsymbol{p}$ defined over those algebras. *Prima facie*, there seems then to be $2^3 - 1$ possible forms of disagreements (and one complete agreement). Nevertheless, since some definitions depend on others, I will assume here that disagreement about the sample space implies disagreement about the algebra and the probability function (but not the converse), and disagreement about the algebra implies disagreement about the probability function (but not the converse). This leaves us with three fundamental forms of disagreement:[3]

(1) *Full radical* disagreement (when agents differ in all three *loci*),

(2) *Partial radical* disagreement (when agents differ on their algebras and probability functions, but not with respect to their sample space), and

(3) *Non-radical* disagreement (when agents differ only in their subjective probability functions).

---

[2]For completeness, we include the probability axioms. A *probability function* is a mapping $\boldsymbol{p} : \mathscr{A} \to \mathbb{R}$ that satisfies the following conditions:

(i) $\boldsymbol{p}(A) \geqslant 0$ for any $A \in \mathscr{A}$;

(ii) $\boldsymbol{p}(\Omega) = 1$;

(iii) $\boldsymbol{p}(A \cup B) = \boldsymbol{p}(A) + \boldsymbol{p}(B)$ for any $A, B \in \mathscr{A}$ such that $A \cap B = \varnothing$.

If, in addition, $\mathscr{A}$ is a σ-algebra and $\boldsymbol{p}$ satisfies the following condition, $\boldsymbol{p}$ is called *countably additive*:

(iv) If $\{A_n\}_{n=1}^{\infty} \subseteq \mathscr{A}$ is a collection of pairwise disjoint events, then $\boldsymbol{p}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \boldsymbol{p}(A_n)$.

[3] Much of this depends on the accurate representation of the agents' sample and event spaces by the modeler. Consider the following case. Agent A has $\Omega_A = \{w_1, w_2\}$, $\mathscr{A}_A = \mathcal{P}(\Omega_1)$, and $\boldsymbol{p}_1(w_1) = \frac{1}{2}$. Agent B has $\Omega_B = \{w_1, w_2, w_3\}$, $\mathscr{A}_B = \{\{w_1\}, \{w_2, w_3\}, \varnothing, \{w_1, w_2, w_3\}\}$, and $\boldsymbol{p}_2(w_2) = \boldsymbol{p}_2(w_3) = \frac{1}{4}$. There is a sense in which both agents disagree about everything, just because they depart from different outcome or sample spaces. Nevertheless, there is no clear epistemic or doxastic sense in which agent B distinguishes between w_2 and w_3, since those elements are bundled in the algebra. The choice of representing the seconds' agent set of doxastic possibilities this way needs to be justified on grounds of disagreement resolution, but it is not. Furthermore, there is a bijection f between $\mathscr{A}_A$ and $\mathscr{A}_B$ that preserves both algebraic structure and probabilistic values - $f(\{w_2\})=\{w_2, w_3\}$. So it is may not clear to asses what the disagreement is about if the representation is inaccurate.

As formalized here, the Priestley-Lavoisier debate falls under the second category. But with a bit of imagination, we could picture cases of *full* radical disagreement. Suppose, for the sake of the argument, that Lavoisier had performed more replications of the kind of experiments Priestley was doing. Assume, furthermore, that such experiments have sufficiently uncontrolled variables so that Lavoisier iterations led to an experimental outcome that Priestley *never observed*. Let us call it w^o. Lavoisier's sample space would include w^o and Priestley's would not, *not because of the lack of refinement*, but because such observation is not acknowledged as a possibility.

Sample and event spaces offer different ways for modelers to represent disagreement on doxastic or epistemic possibilities.[4] Each agent i has an agenda of questions they want to investigate. This induces a partition of the space of possibilities, with each element of the partition corresponding to one of the different complete answer (to all of the questions) that i is deeming relevant, and formally represented in the framework by the atoms of the agent's algebra $\mathscr{A}$. Agents may disagree on the set of questions, and/or on their possible answers. They can have different degrees of refinement. An agent i may be simply more refined than j, and that would be captured by the fact that their $\mathscr{A}_j$ is a subalgebra of $\mathscr{A}_i$, like in our analysis of the Priestley-Lavoisier debate.[5] It is also possible for i to be more refined in some respects, but more coarse in others.[6] Another option, like our imaginary case before, is for i to include an answer (outcome) that j did not consider, and is also incompatible with what j considers possible. This is represented in the model by $\Omega_j \subsetneq \Omega_i$. The three forms of disagreement allow to model a wide variety of cases.

[4] I make no distinction between the two concepts. Doxastic possibilities are the objects towards which an agent has doxastic attitudes, epistemic possibilities are the objects towards which an agent has epistemic attitudes. Here they are assumed to be the same.

[5] We say that $\mathscr{A}'$ is a subalgebra of $\mathscr{A}$ if $\mathscr{A}' \subseteq \mathscr{A}$ and $\mathscr{A}'$, with the distinguished operations of $\mathscr{A}$, is a (σ-)algebra. That is, the operations must be the restrictions of the operations of the whole algebra; being a subset that is a boolean algebra is not sufficient for being a subalgebra of $\mathscr{A}$ (Halmos, 1963).

[6] Say $\Omega_1 = \Omega_2 = \{w_1, w_2, w_3, w_4\}$, while $At(\mathscr{A}_1) = \{\varnothing, \{w_1\}, \{w_2\}, \{w_3, w_4\}\}$ and $At(\mathscr{A}_2) = \{\varnothing, \{w_3\}, \{w_4\}, \{w_1, w_2\}\}$.

2.2. Pooling and consensus as common ground.

The problem of opinion aggregation is the problem of determining a sensible formula for representing the opinions of a group (Genest and Zidek, 1986). Probabilistic opinion pooling is one proposal for finding such consensus. One general way of aggregating the probabilistic opinions of a group to form a collective opinion is to employ some *pooling* function. Say P is the set of probability functions defined over a common algebra $\mathscr{A}$. Formally, a *precise* pooling method for a group of n individuals is a function:

$$F\colon P^n \to P$$

mapping profiles of probability functions for the n agents (or simply the n distributions under consideration), $(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$, to *single* probability functions intended to represent group opinion, $F(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$. The probabilities are assigned to events, represented by a *common* algebra $\mathscr{A}$ defined over a *shared* sample space Ω.

Various interpretations of pooling are proposed in the literature (Lehrer and Wagner, 1981; McConway, 1981; Genest and Zidek, 1986). Wagner (2009), for example, offers the following: (a) A rough summary of the current probabilities of the n individuals; (b) a compromise adopted by the individuals for the purpose of group decision making; (c) a rational consensus to which the individuals revise their probabilities after discussion; (d) the opinion a decision maker external to the group adopts upon being informed of the n expert opinions in the group; and (e) the opinion an individual in the group adopts upon being informed of the $n-1$ opinions of his "epistemic peers in the group. These five interpretations do not exhaust the possibilities. The target interpretation here is rational consensus (c) adopted either *for the sake of the argument* in order to perform some task in group inference or decision making. A secondary and related interpretation, (e) the opinion of an individual group member after learning the opinions of the others, will also be relevant in some

of the procedures defined here. This latter interpretation leads to the issues of (peer) disagreement and expert opinion.

Important for our purposes here, imprecise probabilities allow for a very interesting and philosophically well-motivated account of consensus (Levi, 1985; Seidenfeld et al., 1989). Consider first the case of *full* or *plain* belief. For each of the propositions in the doxastic space of possibilities, agents are said to have three attitudes: acceptance, rejection, or suspension of judgment. Agents also have three standard forms of belief change, following Alchourrn et al. (1985) [AGM]: expansion, contraction, and revision. *At the outset of inquiry*, inquirers may seek consensus as *shared agreement* in their beliefs. This could be achieved by retaining whatever beliefs are accepted by all parties and suspending judgment on those that are not shared, i.e. contracting to the strongest belief set shared by all. Inquiry initiating from the consensus view can proceed without begging questions against parties in the consensus, thereby allowing various hypotheses of concern to receive a fair hearing. Such a consensus constitutes a neutral or non-controversial starting point for subsequent inquiry.

Levi (1985) draws a distinction between consensus as the *outcome* of inquiry and consensus at the *outset* of inquiry. At the outset of inquiry, agents may seek a common ground upon which to pursue joint inquiry. This is consensus as shared agreement, discussed just above. Disagreement among the parties to the consensus may then be resolved (in the best case) through joint efforts in inquiry - consensus as the outcome of inquiry. A precise pool, namely a pooling function with a unique probability as its outcome, is neither a consensus as shared agreement, nor justified on the basis of joint inquiry. In the case of full belief, an analogous process to pooling would be for each party to either endorse one party's beliefs or to switch to some compromise set of beliefs that none endorse originally. Without first suspending judgment, it is difficult to understand how an individual could be justified in switching to a point of view their beliefs rule out. As Levi emphasizes,

both reaching common ground at the outset of inquiry and reasoned changes in probabilistic views are available to groups in the imprecise probabilities setting.

The idea that parties joining effort in inquiry or decision making should restrict themselves to their shared agreements can be extended to judgments of probability. An analogous sense of suspending judgment concerning what is controversial is available in the imprecise probabilities setting. To suspend judgment among some number of probability distributions is to not rule them out for the purposes of inference and decision making. If the parties seeking consensus all agree that $\boldsymbol{p}$ is *not* permissible, then the consensus position reflects that agreement and rules it out. A *set* of probability functions represents the shared agreements among the group concerning which probability functions *are not* permissible to use in inference and decision making. This avenue was explored by Stewart and Quintana (2016, 2018).

Consensus as shared agreement also provides a general heuristic to cope with full and partial *radical* disagreement.

Returning to the case of plain belief, the case of radical disagreement extends the attitudes agents can have; not only can they accept, reject or suspend judgment on propositions, but they can also be *unaware* of their possibility. In our account, Priestley was *unaware* that dephlogisticated air had a heterogeneous reference class. We can therefore recognize three forms of structural change for the agents: radical strengthening, radical weakening, and radical revision. An agent strengthens radically when they extend their set of doxastic possibilities, be it at the level of the outcome space or the event space. Similarly, an agent weakens radically when they forget or now deem irrelevant some of the original possibilities. Radical revision involves both a strengthening and a weakening, so that the new doxastic space of possibilities is neither a strict superset nor a strict subset of the original. These types of structural beliefs changes have been explored by Bradley (2017) in the

probabilistic case, by Cresto (2008) in an AGM-style, and within game theory in the literature on unawareness by Fagin and Halpern (1987); Halpern (2001); Dekel et al. (1998); Modica and Rustichini (1999).

Notice that a structural *strengthening* does not necessarily entail a structural *expansion* in the usual sense. Bringing formerly inaccessible events into consideration does not in principle involve any doxastic attitude of acceptance, rejection, or suspension of judgment towards them, nor does it require a credal probabilistic judgment. On the contrary, there is a sense in which a structural strengthening is more akin to a *contraction*. In the plain belief case, suspending judgment about a previously accepted or rejected proposition (i.e. contracting) means leaving open the possibility that it would be rejected or accepted in the future. In structural strengthening, new doxastic possibilities are incorporated that will require acceptance, rejection, suspension of judgment, or a probabilistic judgment, in the future.

Going back to consensus, finding a common ground at the outset requires weakening to the strongest position compatible with everyone's views. In other words, the consensus position is the strongest position weaker than all of than the individual ones. Taken as an *rule of thumb*, this is the heuristic that will be employed in the next sections. Furthermore, if, as I will argue later, this form of consensus is taken to be rationally grounded, then it may explain Lavoisier's victory in the debate as a reasonable outcome.

3. Radical Pooling

This section will proceed constructively. I will begin by focusing on *full* radical disagreement and then move to *partial* radical disagreement, and then proper pooling. Each subsection will solve part of the puzzle, and the last one will summarize the picture.

3.1. A common sample space.

Suppose there are $i, j \in I$ such that $\Omega_i \neq \Omega_j$. As in our sample imaginary case, Lavoisier made observations that Priestley did not consider possible outcomes of his experiments. What is a suitable candidate for a *common ground sample space*? I will argue that the union of the individual sample spaces, i.e. $\Omega^* = \bigcup_{i \in I} \Omega_i$, is the best candidate.

Finding a common ground requires not begging the question about which are the relevant possibilities, so that all hypothesis in this respect are given a fair hearing. Excluding outcomes deemed possible by some of the agents would entail a controversial starting point of inquiry, one in which the consensus position is identified with a dictatorial one, or with some compromise that no agent originally endorsed but still excludes some. At the outset, epistemic grounds cannot justify that exclusion. Fairness concerns demand that no view is in principle excluded; if some agent is not regarded as an equal there would be no rational grounds to call the position a *consensus*.

Like in the case of suspension of judgment, expanding any Ω_i to Ω^* is in fact a particular kind of *contraction*. It amounts considering *more* propositions possible and open for acceptance, rejection, suspension of judgment or probabilistic evaluation. Suspending judgment over a previously endorsed or rejected proposition is the act of opening the possibility that it can be (believed to be) either false or true, i.e. no doxastic judgement is made about its truth value. By structurally strengthening any Ω_i to Ω^* new outcomes are certainly judged *possible*, but no judgment is made about the truth or probabilistic value of those new possibilities. By taking Ω^* as the common ground sample space agents are required to weaken their position to the strongest position compatible with everyone's possibility space.

In our imaginary case, if Priestley is committed to finding a *consensus* with Lavoisier, then the starting position should incorporate Lavoisier's (claimed) observations as *possible*, even if Priestley never had access to them. Failing to include w^o *at the outset* would imply excluding Priestley's contribution without any evidence or reasons against it, and it also entail a loss of information.

Nevertheless, the consensus position developed so far requires no judgment about the likelihood of an outcome, and in so far as it was never observed by him, Priestley can justifiable require it to have null probability.

The present view of consensus is as common ground *at the outset* of inquiry. Future investigation or deliberation among agents might lead them to conclude that some of the outcomes originally endorsed as possible by some of the agents are in fact irrelevant. This might very well lead to a structural weakening in which Ω^* is reduced to a subset. If after *jointly* repeating the experiment multiple times, Priestley's alleged outcome w^o is never observed, nothing precludes them to weaken their sample space *structurally* and remove w^o. But that amounts to a form of consensus *at the outcome* of some rational procedure. At the outset, without bringing into consideration new evidence or arguments to support the exclusion of some outcomes as irrelevant or impossible, nothing less than Ω^* can be accepted on rational grounds. No reasons support expanding the consensus view to a superset of Ω^*, that would entail endorsing possibilities *ex nihilo*.

Finding a common sample space is still not enough at this point, since agents have their event space defined over their original outcome space. Each agent i starts with a sample space Ω_i and an event space $\mathscr{A}_i$ defined over Ω_i. Once the common ground sample space Ω^* is reached, agents need to extend $\mathscr{A}_i$ to a sample space $\mathscr{A}_i^*$ defined over Ω^*. The suggestion here is for $\mathscr{A}_i^*$ to be the coarsest algebra over Ω^* such that $\mathscr{A}_i \subseteq \mathscr{A}_i^*$.

Observation 1. *There is a unique coarsest algebra $\mathscr{A}_i^*$ over Ω^* such that $\mathscr{A}_i \subseteq \mathscr{A}_i^*$. $\mathscr{A}_i^*$ is the closure under countable unions of $\mathscr{A}_i \cup \{w^* : w^* \in \Omega^*/\Omega_i\}$.*[7]

[7]The algebras over Ω^* are partially ordered by the subalgebra relation defined before. $\mathscr{A}$ is coarser than $\mathscr{A}'$ if the former is a subalgebra of the latter. An algebra satisfying some condition (here $\mathscr{A}_i \subseteq \mathscr{A}_i^*$) is the coarsest if no subalgebra of it satisfies the condition.

Let $\overline{X} = \Omega^*/\Omega_i$. Consider $\mathscr{A}_i^*$ to be the closure under countable unions of members of $\mathscr{A}_i \cup \{\overline{X}\}$. Notice first that $\mathscr{A}_i^*$ is a σ-algebra over Ω^*. It is closed under countable unions by definition. Since any $x_k \in \mathscr{A}_i^*$ can be decomposed as the countable union of members of $\mathscr{A}_i \cup \{\overline{X}\}$, either $\overline{X} \subseteq x_k$ or $\overline{X} \cap x_k = \varnothing$. Furthermore, let $y_k = x_k/\overline{X} \in \mathscr{A}_i$. Consider the countable intersection $\bigcap_k x_k$ of members of $\mathscr{A}_i^*$. Let $\bigcap_k x_k = (\bigcap_m x_m) \cap (\bigcap_n x_n)$

The *rationale* for such a choice is as expected. Agents are required to move to the weakest position compatible with others. By taking the *coarsest* algebra that includes theirs, they are not giving up on any of their own consideration while including others doxastic possibilities without committing to any probabilistic judgment about them. In our example, Priestley would have to simply accept the event $\{w^o\}$ as possible, and minimally include it into his carving of the logical space.

3.2. A shared event space.

We have now reduced the problem of *full* radical disagreement to the problem of *partial* radical disagreement. In the previous section agents started diverging in their sample space and ended up finding a common ground. But although they agree on the expanded new sample space Ω^*, they each had a new event space $\mathscr{A}_i^*$ which will most likely differ for different agents. So now agents agree on the sample space but disagree about the event space. How should they then find a *common ground event space*?

As anticipated before, the suggestion here is that they ought to take the *meet* of the $\mathscr{A}_i^*$. The *meet* of a set of algebras is the coarsest algebra that is finer than all of the ones in the set. In detail, $\mathscr{A}^* = \bigwedge_i \mathscr{A}_i^*$ is such that (a) for all i, $\mathscr{A}_i^* \subseteq \mathscr{A}^*$, (b) $\mathscr{A}^*$ is still an algebra, and (c) no subalgebra of $\mathscr{A}^*$ satisfies (a). By taking the *meet* as the shared event space agents are precisely following the heuristic defended here: taking as common ground the coarsest space that is compatible with everyone's original position. Taking any reduced algebra would imply excluding events deemed

---

where $\overline{X} \subseteq x_m$ for all m, and $\overline{X} \cap x_n = \varnothing$ for all n. Now, $\bigcap_k x_k = (\bigcap_m x_m) \cap (\bigcap_n x_n) = (\bigcap_m (y_m \cup \overline{X})) \cap (\bigcap_n x_n) = ((\bigcap_m y_m) \cup \overline{X}) \cap (\bigcap_n x_n)$. If n is empty, then $\bigcap_k x_k = (\bigcap_m y_m) \cup \overline{X} \in \mathscr{A}_i^*$ since $\bigcap_m y_m \in \mathscr{A}_i$. If n is non-empty, then $\bigcap_k x_k = ((\bigcap_m y_m) \cup \overline{X}) \cap (\bigcap_n x_n) = (\bigcap_m y_m) \cap (\bigcap_n x_n) \in \mathscr{A}_i \subseteq \mathscr{A}_i^*$. Now for complementation. Take any $x \in \mathscr{A}_i^*$. If $\overline{X} \cap x = \varnothing$, then $x \in \mathscr{A}_i$ and $\Omega_i/x \in \mathscr{A}_i$. But $\overline{x} = \Omega^*/x = (\overline{X} \cup \Omega_i)/x = (\overline{X}/x) \cup (\Omega_i/x) = \overline{X} \cup (\Omega_i/x) \in \mathscr{A}_i^*$. The second case is when $\overline{X} \subseteq x$. Now $\overline{x} = \overline{y \cup \overline{X}} = \overline{y} \cap \overline{\overline{X}} = \overline{y} \cap \Omega_i \in \mathscr{A}_i \subseteq \mathscr{A}_i^*$. So we have that there *exists* a σ-algebra over Ω^* that includes $\mathscr{A}_i$. It remains to show that it is the unique coarsest.

Suppose $\mathscr{A}_i^{@}$ is a σ-algebra over Ω^* such that $\mathscr{A}_i \subseteq \mathscr{A}_i^{@}$. Hence $\Omega_i \in \mathscr{A}_i^{@}$, and therefore $\overline{X} \in \mathscr{A}_i^{@}$ by closure under complementation. It is clear then that $\mathscr{A}_i^* \subseteq \mathscr{A}_i^{@}$, and in consequence $\mathscr{A}_i^*$ is a subalgebra of $\mathscr{A}_i^{@}$.

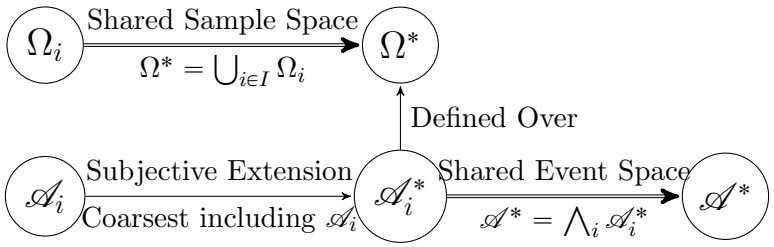
possible by some agent. Taking any finer algebra would be rationally unjustified, in so far as the refinement would not be grounded on the available evidence.

If there is some accuracy in my presentation and formalization of the Priestley-Lavoisier debate, then the proposed solution aligns with the scientific consensus that resulted from the dispute. By rejecting the hypothesis that there is a substance that is released in combustion, namely phlogiston, Lavoisier studied the properties of the gas resulting from heating a red calx of (oxide of) mercury as properties of a substance of its own kind, oxygen. Priestley, on the other hand, studied them as properties of a gas depleted from phlogiston and therefore started with a coarser algebra. Once again, Lavoisier carved the event space where the natural joints are. But if this is the case, then the meet $\mathscr{A}_P \wedge \mathscr{A}_L$ is just $\mathscr{A}_L$ and the proposed common ground is precisely Lavoisier's carving.

Not all cases of partial radical disagreement fit the example I've been exploring, namely when one individual is simply more refined than another. Nothing precludes the alternative that *all* of the parties involved considered possibilities that none of the other did. For example, for two agents A and B we could reasonably have $\Omega^* = \Omega_A = \Omega_B = \{w_1, w_2, w_3, w_4\}$, $\mathscr{A}_A$ is generated by $\{\{w_1\}, \{w_2\}, \{w_3, w_4\}\}$, while $\mathscr{A}_B$ is generated by $\{\{w_1, w_2\}, \{w_3\}, \{w_4\}\}$. Here A distinguishes between $\{w_1\}$ and $\{w_2\}$ while B does not, and conversely for $\{w_3\}$ and $\{w_4\}$. Furthermore, notice that by taking the meet here, there will be new events not present in any of the individual's event space brought into consideration. The consensus position acknowledge as doxastically possible events that *no one* considered possible before. $\mathscr{A}_A \wedge \mathscr{A}_B = \mathscr{A}^* = \mathscr{P}(\Omega^*)$, and in particular $\{w_2, w_3\} \in \mathscr{A}^*$ but $\{w_2, w_3\} \notin \mathscr{A}_A$ and $\{w_2, w_3\} \notin \mathscr{A}_B$. I take this to be a positive feature of the proposal. It shows that seeking consensus in one of its weakest forms, as common ground and at the outset of inquiry, already can lead to conceptual refinement and innovation.

3.3. Extending the probability functions: A case for imprecision.

So far so good, but not much was said about probabilistic judgments. The moves so far were the following:



Consensus first required to find a common sample space Ω^*, which I argued should be the union of the sample spaces of all of the members of the group. After that, each agent extended their event space $\mathscr{A}_i$ to the new common sample space by taking $\mathscr{A}_i^*$, the coarsest algebra that includes their original event space. The third step was to find a common event space, and I defended the idea of taking the meet $\mathscr{A}^*$ of all the $\mathscr{A}_i^*$. Probability functions were left far behind, only defined for the original Ω_i and $\mathscr{A}_i$. How should the agent extend their subjective probability, originally defined for narrower sample and event spaces, to the new common ground *of doxastic possibilities*?

Let us start with an easy case, when all the pmf $\boldsymbol{p}_i$ are defined over the power set algebra of a Ω_i. There is, I claim, one very natural way of extending each $\boldsymbol{p}_i$ over Ω_i to a pmf $\boldsymbol{p}_i^*$ over Ω^*:

$$\boldsymbol{p}_i^*(w) = \begin{cases} \boldsymbol{p}_i(w) & \text{if } w \in \Omega_i \\ 0 & \text{if otherwise} \end{cases}$$

There are reasons to assign null probabilistic weight to outcomes that were not considered before. Recall our imagined scenario before. Lavoisier claims to have observed some new outcome

w^o from his experiments, an outcome that Priestley never observed. I argued before that seeking consensus required Priestley to accept that outcome, and its corresponding event, into the common ground sample and event spaces. How should Priestley adjust his *personal subjective* probabilities $\boldsymbol{p}_P$ to the new space of possibilities? The suggestion is $\boldsymbol{p}^*_P(w^o) = 0$. After all, he *never* observed the alleged experimental outcome.

Notice first that this is *not* yet the output of a *pooling* function. The pooling function is the result of aggregating all such $\boldsymbol{p}^*_i$; that would be the probabilistic consensus of the group. By itself, $\boldsymbol{p}^*_i$ is the extension of $\boldsymbol{p}_i$ to the common ground sample and event spaces. On the one hand, i did not originally acknowledged all the $w \in \Omega^*$ but $w \notin \Omega_i$ as relevant doxastic possibilities, so in principle there seems to be no reason to give them any probabilistic weight. On the other hand, giving them some probabilistic weight implies modifying - most likely reducing - the probabilistic weight of doxastic possibilities they originally considered relevant to the problem; which seems *prima facie* unjustified. Giving null weight to the doxastic possibilities originally regarded as irrelevant seems to be the most cost effective way of taking into account the common sample and event spaces - though not yet others' probabilistic judgments.

Furthermore, if this solution is endorsed, now the problem of radical pooling is reduced to the problem of non-radically pooling the $\boldsymbol{p}^*_i$, since they are all defined over the same spaces. Somewhere else (Stewart and Quintana, 2016, 2018) I defended the use of imprecise probabilities for non-radical pooling; but non-radical aggregation is not the subject of this essay. Imprecise probabilities will be vindicated *again* after the following considerations.

The previous solution is only partial, since it is unjustified to assume that all the subjective probabilities are defined over the power set algebras. We want a procedure that is defined *in general*. Our guiding example was precisely one in which Priestley's original event space was coarser than

Lavoisier's. Modelling a situation in which one agent is more doxastically refined than another requires making use of sub-algebras of the power set.

Marginalization leads to imprecision

Each probability measure $\boldsymbol{p}_i$ is defined over $\mathscr{A}_i$, which may or may not be the power set of Ω_i, and the task is to extend that measure to $\mathscr{A}^*$. The suggestion here is to use imprecise probabilities. In particular, there are several admissible extensions of $\boldsymbol{p}_i$ to $\mathscr{A}^*$ and there is no reason for i to prefer one over the other.

$$\mathbb{P}_i^* = \{\boldsymbol{p}_i^* : \boldsymbol{p}_i^* \restriction_{\mathscr{A}_i} = \boldsymbol{p}_i\}$$

So $\mathbb{P}_i^*$ is the set of all probability measures that extend $\boldsymbol{p}_i$ - i.e. all the $\boldsymbol{p}_i^*$ defined over $\mathscr{A}^*$ such that restricted to $\mathscr{A}_i$ gives back $\boldsymbol{p}_i$. More precisely: $\boldsymbol{p}_i^* \restriction_{\mathscr{A}_i} = \boldsymbol{p}_i$ if and only if for each $H \in \mathscr{A}_i$, $\boldsymbol{p}_i^*(H) = \boldsymbol{p}_i(H)$.[8]

Observation 2. *$\mathbb{P}_i^*$ is convex for all i.*[9]

Notice, furthermore, that this is a generalization of the previous case where probability measures were defined in terms of pmfs. New events will have non-null probability weight only if they have old events as subsets. Completely new events, which were not regarded doxastically possible *at all* before, will have null probabilistic weight. In the imagined scenario, Priestley would assign zero probability to Lavoisier's alleged observation.

Any precise $\boldsymbol{p}_i^*$ satisfying the condition will be arbitrary and unjustified.

[8] In general the marginalization operation $\boldsymbol{p}_i^* \restriction_{\mathscr{A}_i}$ is defined only when $\mathscr{A}_i$ is a subalgebra of $\mathscr{A}^*$, but this needs not to be the case here and that is why the definition is necessary.

[9] Suppose $\boldsymbol{p}^1, \boldsymbol{p}^2 \in \mathbb{P}_i^*$. I want to show that for $\alpha \in [0,1]$, $\boldsymbol{p}^* = \alpha.\boldsymbol{p}^1 + (1-\alpha).\boldsymbol{p}^2 \in \mathbb{P}_i^*$.
It is enough to show that $\boldsymbol{p}^* \restriction_{\mathscr{A}_i} = \boldsymbol{p}_i$. Given any $E \in \mathscr{A}_i$, $\boldsymbol{p}^*(E) = \alpha.\boldsymbol{p}^1(E) + (1-\alpha).\boldsymbol{p}^2(E) = \alpha.\boldsymbol{p}_i(E) + (1-\alpha).\boldsymbol{p}_i(E) = \boldsymbol{p}_i(E)$.

Using imprecise probabilities here follows the general heuristic and conception of consensus as common ground that I have been defending. Arguably, imprecision amounts to a weakening and an increase in uncertainty. That is fine, since we are requiring each of the agents to extend their probabilistic judgment to events that they never even considered doxastically possible.

Rigidity leads to imprecision

A standard, yet contested, principle of Bayesian epistemology is *rigidity*: conditional probabilities are kept fixed throughout updating procedures. Let $\boldsymbol{p}(\cdot)$ be defined over some algebra $\mathscr{A}$ of events be a representation of some agent's (coherent) credal state. Suppose now the agent learns an event E and updates their credal state to $\boldsymbol{q}(\cdot) = \boldsymbol{p}(\cdot|E)$ according to Bayesian standards. For all events $H \in \mathscr{A}$ we have:

- **Rigidity:** $\boldsymbol{q}(H|E) = \boldsymbol{p}(H|E \cap E) = \boldsymbol{p}(H|E)$

Richard Jeffrey (Jeffrey, 1990) famously argued against this form of rigidity, and cases like the problem of old evidence show it can be rational to change one's initial conditional probabilities. I do not intend to discuss this form of rigidity, but one relevant for the structural strengthening case, and in line with some ideas presented by Bradley (Bradley, 2017).

The idea behind rigidity is that we should extend our relational attitudes to the new set in such a way as to conserve all prior relational credences. Becoming aware of the new events should not affect the relative credibility of previously considered events. Becoming aware of such new events does not provide any justification for such a change, and therefore it would not be a conservative structural strengthening. For all $E, H \in \mathscr{A}_i$ with $\boldsymbol{p}_i(E) \neq 0 \neq \boldsymbol{p}_i^*(E)$:

- **Strong Structural Rigidity [SSR]:** $\boldsymbol{p}_i^*(H|E) = \boldsymbol{p}_i(H|E)$

SSR can arguably be regarded as too strong of a requirement. If so, consider the principle that demands that the strengthened conditional probabilities of *old* events given the *old* set of possibilities should be kept the same. Namely, if after becoming aware of new possibilities agents learn that the old ones are the relevant ones, they ought to return to the original probabilistic judgments. For all $H \in \mathscr{A}_i$:

- **Weak Structural Rigidity [WSR]:** $\boldsymbol{p}_i^*(H|\Omega_i) = \boldsymbol{p}_i(H|\Omega_i) = \boldsymbol{p}_i(H)$

The following observation holds trivially.

Observation 3. *Strong Structural Rigidity and Weak Structural Rigidity are equivalent.*[10]

Going back to imprecision, it is easy to observe that all the $\boldsymbol{p}_i^* \in \mathbb{P}_i^* = \{\boldsymbol{p}_i^* : \boldsymbol{p}_i^* \restriction_{\mathscr{A}_i} = \boldsymbol{p}_i\}$ satisfy SSR,[11] yet the set of all $\boldsymbol{p}_i^*$ satisfying SSR is strictly larger than $\mathbb{P}_i^*$.[12] The following condition suffices to make the two sets coextentional.

- **Structural Conservatism:** $\boldsymbol{p}_i^*(\Omega_i) = \boldsymbol{p}_i(\Omega_i) = 1$.

Structural Consevatism requires that outcomes originally inaccessible to the agent will get null probability, but new events compatible with the previous set of doxastic possibilities need not have null probabilistic value. Notice that once again that a structural strengthening is *not* yet the output of a *pooling* function. It is not yet taking into account any (probabilistically representable) evidence in support of the new possibilities, nor seeking probabilistic consensus. Structural strengthening should be distinguished from some form of *expansion*, the case in which not only the space of possibilities is extended but also new possibilities are accepted.

[10] The proof relies on conditional probabilities being defined rather than primitive. SSR entails WSR tivially, since $\Omega_i \in \mathscr{A}_i$ and $\boldsymbol{p}_i(\Omega_i) = 1$. WSR and the definition of conditional probabilities secure that for all $E, H \in \mathscr{A}_i$, $\boldsymbol{p}_i(H|E) = \frac{\boldsymbol{p}_i(H \cap E)}{\boldsymbol{p}_i(E)} = \frac{\boldsymbol{p}_i(H \cap E|\Omega_i)}{\boldsymbol{p}_i(E|\Omega_i)} = \frac{\boldsymbol{p}_i^*(H \cap E)}{\boldsymbol{p}_i^*(E)} = \boldsymbol{p}_i^*(H|E)$.

[11] If $\boldsymbol{p}_i^* \in \mathbb{P}_i^*$, then for all $H \in \mathscr{A}_i$, $\boldsymbol{p}_i^*(H) = \boldsymbol{p}_i(H)$. Hence, $\boldsymbol{p}_i^*(H|E) = \frac{\boldsymbol{p}_i^*(H \cap E)}{\boldsymbol{p}_i^*(E)} = \frac{\boldsymbol{p}_i(H \cap E)}{\boldsymbol{p}_i(E)} = \boldsymbol{p}_i(H|E)$.

[12] Let $\Omega_i = \{w_1, w_2\}$, $\mathscr{A}_i = \mathscr{P}(\Omega_i)$, and $\boldsymbol{p}_i(w_1) = \frac{1}{2}$. On the other hand, let $\Omega^* = \{w_1, w_2, w_3\}$, $\mathscr{A}^* = \mathscr{P}(\Omega^*)$, and $\boldsymbol{p}_i^*(w_1) = \boldsymbol{p}_i^*(w_2) = \boldsymbol{p}_i^*(w_3) = \frac{1}{3}$. Then $\boldsymbol{p}_i^*$ satisfies SSR but $\boldsymbol{p}_i^* \notin \mathbb{P}_i^*$.

OBSERVATION 4. *The set of all* $\boldsymbol{p}_i^*$ *defined over* $\mathscr{A}^*$ *satisfying SSR and Structural Conservatism is just* $\mathbb{P}_i^* = \{\boldsymbol{p}_i^* : \boldsymbol{p}_i^* \restriction_{\mathscr{A}_i} = \boldsymbol{p}_i\}$.[13]

Divergence measures lead to imprecision

When extending $\boldsymbol{p}_i$ from $\mathscr{A}_i$ to $\mathscr{A}^*$ it is worth noticing that although $\mathscr{A}_i \subseteq \mathscr{A}^*$, the former need not be a subalgebra of the latter [i.e. when $\Omega_i \subsetneq \Omega^*$]. $\boldsymbol{p}_i$ can therefore be interpreted as an *incoherent* credence function over $\mathscr{A}^*$, one in which there are no values assigned for some elements of the algebra. Hence the question of how to structurally strengthen $\boldsymbol{p}_i$ to $\mathscr{A}^*$ is akin to the issue of how to fix incoherent credences, a topic extensively developed by Pettigrew (Pettigrew, 2017) and Predd *et al* (Predd et al., 2009). A well argued solution involves using divergence measures.

Let $\mathbb{P}$ be the set of *credence* functions defined on $\mathscr{A}$.[14] A divergence is a function $\mathscr{D}_{\mathscr{A}} : \mathbb{P} \times \mathbb{P} \to [0, \infty]$ such that for all credences $\boldsymbol{p}, \boldsymbol{p}' \in \mathbb{P}$, (i) $\mathscr{D}_{\mathscr{A}}(\boldsymbol{p}, \boldsymbol{p}) = 0$, and (ii) $\mathscr{D}_{\mathscr{A}}(\boldsymbol{p}, \boldsymbol{p}') > 0$ if $\boldsymbol{p} \neq \boldsymbol{p}'$. Symmetry is not required, nor the satisfaction of triangle inequality. The following are two classical examples:

Squared Euclidean Distance [SED]

$$SED_{\mathscr{A}}(\boldsymbol{p}, \boldsymbol{p}') = \sum_{E \in \mathscr{A}} [\boldsymbol{p}(E) - \boldsymbol{p}'(E)]^2$$

Generalized Kullback-Leibler [GKL]

[13] A previous note showed that the $\boldsymbol{p}_i^* \in \mathbb{P}_i^*$ satisfy SSR, and they trivially satisfy Structural Conservatism. Suppose $\boldsymbol{p}_i^*$ satisfies SSR and Structural Conservatism. Let $H \in \mathscr{A}_i$ (so $H \subseteq \Omega_i$) and consider $\boldsymbol{p}_i^*(H)$. By Structural Conservatism $\boldsymbol{p}_i^*(H) = \boldsymbol{p}_i^*(H|\Omega_i)$. By SSR and Observation 4, $\boldsymbol{p}_i^*$ satisfies WSR and therefore $\boldsymbol{p}_i^*(H|\Omega_i) = \boldsymbol{p}_i(H)$.

[14] Notice that nothing in the definition requires the elements of $\mathbb{P}$ to be *probability* functions, just that they assign values in [0,1] to the elements of $\mathscr{A}$

$$GKL_{\mathscr{A}}(\boldsymbol{p}, \boldsymbol{p}') = \sum_{E \in \mathscr{A}} \left[\boldsymbol{p}(E) log \left(\frac{\boldsymbol{p}(E)}{\boldsymbol{p}'(E)} \right) - \boldsymbol{p}(E) + \boldsymbol{p}'(E) \right]$$

Divergence measures are generally used under the methodological assumption of minimal mutilation, which states that any shift from a prior to a posterior credal states should accommodate the posterior condition by minimally changing the prior. For example, Bayesian conditionalization satisfies the minimal mutilation principle according to GKL: $\boldsymbol{p}(\cdot|E) = \boldsymbol{p}'$ minimizes $GKL_{\mathscr{A}}(\boldsymbol{p}, \boldsymbol{p}')$ on the condition that the $\boldsymbol{p}'$ is a probability function (i.e. coherent credal functions) that assigns value 1 to E. Similarly, Kullback-Leibler divergence generalizes Jeffrey conditionalization (Diaconis and Zabell, 1982b). GKL is also a generalization of Jaynes Maximum Entropy formalism (Williams, 1980b). The principle of minimal mutilation, and GKL in particular, has also gained support in solving issues like the problem of old evidence (e.g., Hartmann, (Hartmann, 2014b)).

To fix $\boldsymbol{p}_i$ amounts then to finding the (set of) coherent $\boldsymbol{p}_i^*$ defined over $\mathscr{A}^*$ that minimizes some appropriate divergence measure. We can avoid the question on whether to use SED or GKL by looking at a generalization of both:

Additive Bregman Divergence [ABD]

Suppose $\phi : [0, 1] \to \mathbb{R}$ is a strictly convex function that is twice differentiable on (0,1) with a continuous second derivative. Let $\mathbb{P}$ be the set of *credence* functions defined on $\mathscr{A}$. Suppose $\mathscr{D} : \mathbb{P} \times \mathbb{P} \to [0, \infty]$. Then $\mathscr{D}$ is the additive Bregman divergence generated by ϕ if, for any $\boldsymbol{p}, \boldsymbol{p}' \in \mathbb{P}$,

$$\mathscr{D}_{\mathscr{A}}(\boldsymbol{p}, \boldsymbol{p}') = \sum_{E \in \mathscr{A}} [\phi(\boldsymbol{p}(E)) - \phi(\boldsymbol{p}'(E)) - \phi'(\boldsymbol{p}'(E))(\boldsymbol{p}(E) - \boldsymbol{p}'(E))]$$

SED is the ABD generated by $\phi(x) = x^2$, and GKL is the ABD generated by $\phi(x) = xlogx - x$. The additive Bregman divergence can also be justified on accuracy-first grounds. Predd *et al* (Predd et al., 2009) show that if $\boldsymbol{p}$ is an incoherent credence function, then the coherent credence function $\boldsymbol{p}'$ that minimizes ABD with respect to it is more accurate than $\boldsymbol{p}$ at all possible worlds. Therefore fixing incoherent credence functions using ABD increases accuracy.

The following observation is important for our purposes:

OBSERVATION 5. *For all $\boldsymbol{p}_i^*$ coherent credence functions defined over $\mathscr{A}^*$:*

$$\boldsymbol{p}_i^* \in \mathbb{P}_i^* = \{\boldsymbol{p}_i^* : \boldsymbol{p}_i^* \restriction_{\mathscr{A}_i} = \boldsymbol{p}_i\} \textit{ iff } \mathscr{D}_{\mathscr{A}_i}(\boldsymbol{p}_i, \boldsymbol{p}_i^*) = 0 = \mathscr{D}_{\mathscr{A}_i}(\boldsymbol{p}_i^*, \boldsymbol{p}_i)^{15}$$

This section explored the structural strengthening of $\boldsymbol{p}_i$ to $\mathscr{A}^*$ and argued for imprecision, namely that the strengthening ought to be $\mathbb{P}_i^* = \{\boldsymbol{p}_i^* : \boldsymbol{p}_i^* \restriction_{\mathscr{A}_i} = \boldsymbol{p}_i\}$. This was shown to be *compatible* with marginalization, the principle of minimal mutilation expressed by divergence measures (as well as some accuracy-first intuitions), and some conservative constrains involving rigidity. Any of the $\boldsymbol{p}_i^* \in \mathbb{P}_i^*$ will satisfy those principles. Furthermore, excluding any of the $\boldsymbol{p}_i^*$ satisfying the conditions would be in principle arbitrary and unjustified. In other words, if we take any or all of those conditions to be exhaustive principles for structural strengthening, then we would be required to adopt $\mathbb{P}_i^* = \{\boldsymbol{p}_i^* : \boldsymbol{p}_i^* \restriction_{\mathscr{A}_i} = \boldsymbol{p}_i\}$.

3.4. Pooling (imprecise) probabilistic judgments.

[15]The observation follows from the fact that ABD is a divergence measure as defined before. If $\boldsymbol{p}_i^* \in \mathbb{P}_i^*$ then for all $E \in \mathscr{A}_i$, $\boldsymbol{p}_i^*(E) = \boldsymbol{p}_i(E)$, and hence each element of the summation will be null. On the other hand, if $\boldsymbol{p}_i^* \notin \mathbb{P}_i^*$, then there is $E \in \mathscr{A}_i$ such that $\boldsymbol{p}_i^*(E) \neq \boldsymbol{p}_i(E)$. I will rely on the well known fact that ϕ is strictly convex if and only if $\forall x \neq y \in [0,1], \phi(x) > \phi(y) + \phi'(y)(x - y)$. This implies that for any $E \in \mathscr{A}_i$, if $\boldsymbol{p}_i^*(E) = \boldsymbol{p}_i(E)$, then $\phi(\boldsymbol{p}(E)) - \phi(\boldsymbol{p}'(E)) - \phi'(\boldsymbol{p}'(E))(\boldsymbol{p}(E) - \boldsymbol{p}'(E)) = 0$; but also, if $\boldsymbol{p}_i^*(E) \neq \boldsymbol{p}_i(E)$, then $\phi(\boldsymbol{p}(E)) - \phi(\boldsymbol{p}'(E)) - \phi'(\boldsymbol{p}'(E))(\boldsymbol{p}(E) - \boldsymbol{p}'(E)) > 0$, which is sufficient to complete the proof.

Each agent now has a convex set of probability functions $\mathbb{P}_i^*$ defined over a common algebra $\mathscr{A}^*$. The view defended is that pooling means *finding a common ground* - excluding that and only that which all agree should be excluded. Say $\mathbb{P} = \mathscr{P}(P)$, the power set of the set of all probability functions defined on an algebra $\mathscr{A}^*$. A generalized imprecise pooling function is defined as:

$$\mathcal{F} : \mathbb{P}^n \rightarrow \mathbb{P}$$

The general question about how to aggregate profiles of imprecise probabilities is not settled by the literature, and I do not intend to provide definitive arguments about it here. Rather, I am more interested in providing arguments for the feasibility of some options and the unfeasibility of others in the present context. There are several natural candidates for imprecise pooling. Let $\mathbb{P}_i^*$ be the set of probability functions corresponding to agent i.

Proposal 1:: $\mathcal{F}(\mathbb{P}_1^*, ..., \mathbb{P}_n^*) = \bigcap_{i=1}^n \mathbb{P}_i$.

Proposal 2: $\mathcal{F}(\mathbb{P}_1^*, ..., \mathbb{P}_n^*) = \bigcup_{i=1}^n \mathbb{P}_i$.

Proposal 3: $\mathcal{F}(\mathbb{P}_1^*, ..., \mathbb{P}_n^*) = conv\left\{ \bigcup_{i=1}^n \mathbb{P}_i \right\}$.

In the present view, agents strengthening their original probabilistic judgments will assign zero probability to events or outcomes incompatible with the original sample space. This entails that in most cases adopting Proposals 1 would trivialize the account. It suffices for an agent to assign non null probability to an outcome not considered by others to make the intersection empty. Furthermore, about the first two proposals, Robert Nau writes:

> As more opinions are pooled, the union can only get larger, and it reflects only the least informative opinions, whereas intuitively there ought to be (at least the possibility of) an increase in precision as the pool gets larger. On the other hand, the intersection of convex sets of measures may be empty if experts are mutually incoherent, and it generally yields too tight a representation of aggregate uncertainty. As more opinions are pooled, the intersection can only shrink, and it reflects only the most extreme among those opinions, whereas intuitively there should be some convergence to an average opinion when the pool gets sufficiently large. Moreover, neither the union nor the intersection provides an opportunity for the differential weighting of opinions, which would be desirable in cases where

> one individual is considered (either by herself or by an external evaluator) to be better or worse informed than another individual about a particular event under consideration. (Nau, 2002)

Nau's reservations about taking the intersection and the triviality argument mentioned before suffice to reject Proposal 1.

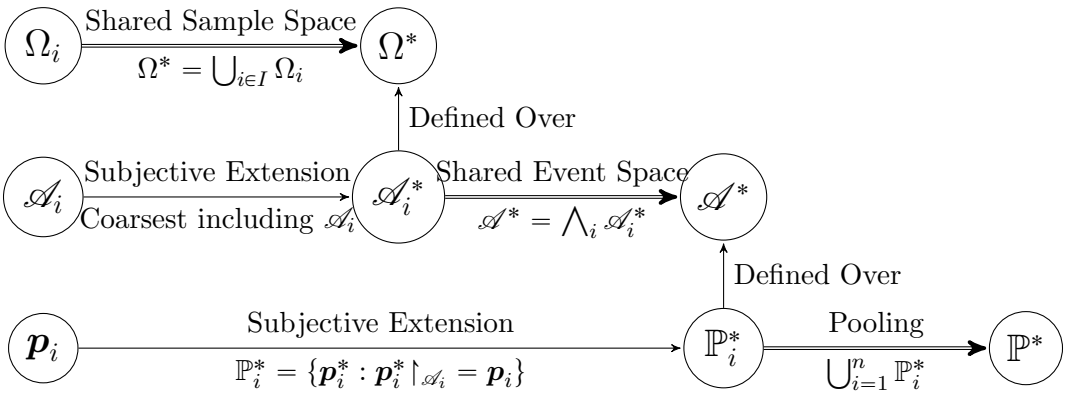
Nau's reservations about taking the union are less persuasive here. If, as suggested, pooling is interpreted as consensus as common ground, then more precision is not expected as the pool gets larger. Similarly, in finding common ground individuals are all given a fair standing and *at the outset* they are not assumed to be unequally informed. Such epistemic priority can be revealed *during* inquiry. Furthermore, a Bayesian account of inquiry secures that joint investigation and updating leads to convergence and uncertainty reduction. In principle, the arguments reviewed are not compelling against Proposal 2.

On a different note, agents may want to preserve all those probability functions that represent a potential resolution of the conflict between their differing distributions. If we assume that these potential resolutions have the form of convex combinations as suggested by (Stewart and Quintana, 2016, 2018) and the literature on linear pooling, then Proposal 3 seems a feasible alternative.

To conclude, taking unions ensures that all of the agents probabilistic judgments are giving a fair hearing, and that only those probabilities rejected by everyone are excluded. Taking the convex set of unions might lead to greater imprecision, but this might be warranted if all potential resolutions are to be preserved at the outset. Proposals 2 and 3 are feasible alternatives for pooling.

4. Conclusion

The general procedure can be visualized in the following diagram:



The nature of this essay is programmatic. It is focused in assessing a form of disagreement that has been neglected in the literature, disagreement about which are the relevant sample or event spaces. The position defended here is that pooling can be interpreted as a technical and philosophical characterization of a form of consensus, namely consensus as common ground at the outset of inquiry. This view was explained and defended in Section 2, and developed formally in Section 3. Furthermore, one of upshots of such an account is that (i) marginalization, (ii) rigidity, and (iii) divergence accounts of how to strengthen probabilistic judgments to larger algebras will lead to the adoption of imprecise probabilities.

Nowhere in this essay it is claimed that taking consensus as common ground is the *only* solution to the problem, but rather that it is a sound one. On the one hand, there might be other formal approaches. For example, considering product algebras and probabilistic extensions to them. On the other hand, there might be other philosophical approaches, by taking different notions of consensus. One might think, for example, that an agent that has a *prior* defined over a more refined event or sample space has some kind of epistemic priority over a *coarser* (less sophisticated) agent. Hence, in some instances, resolving radical disagreement may resemble a case of expert testimony.

The Priestley and Lavoisier debate served as a motivating example of the kind of disagreements that can be found in science, or inquiry in general. I argued that conceptual incommesurability can be modeled as a disagreement about what are the relevant sample and event spaces, what are the best natural joints to carve. But even in the face of such discrepancy, some form of consensus can be defined. At the outset, disagreements about what variables are relevant for explanation should begin by giving a fair treatment to all the possibilities. During inquiry, some of those possibilities may be discarded. For example, statistical methods of model selection like variance analysis or the AIC algorithm may deem some variables insignificant. But a pooling function cannot be expected to give an armchair account of the outcome of inquiry. In order to resolve radical disagreement rationally agents need start the conversation by taking others' world views as possible.

CHAPTER 5

Harsanyi meets Smith's Altruism and Spite

1. Introduction

Harsanyi (1977) originally thought of his Aggregation and Impartial Observer Theorems in the context of Adam Smith's moral sentimentalism. Nevertheless, subsequent discussion focused on whether they provide an accurate representation of utilitarian social welfare functions. The purpose here is to go back to the original spirit of the results and show that they help to formally characterize Smith's tripartite division of the passions between social, unsocial, and selfish [here represented by altruism, spite, and self-interest]. Furthermore, I suggest that Smith's recognition of the value of unsocial passions [and their proper taming] in his Theory of Moral Sentiments (Smith, 2002) makes the *Das Adam Smith Problem* even worse.

This introduction will present Harsanyi's results and summarize the basics of Smith's moral sentimentalism. Section 2 will be devoted to a discussion of *altruism*, where a *definition* is provided following Kitcher (2010), as well as a *characterization* using Harsanyi's result. Section 3 will proceed analogously with the notion of *spite*, now using an adapted version of the classic theorem. The concluding section will discuss *Das Adam Smith Problem* and his tripartite division of society.

1.1. Harsanyi, between moral sentiments and utilitarianism.

Harsanyi (1955, 1977) famously presented two distinct results popularized as the Aggregation Theorem and the Impartial Observer Theorem.

THEOREM 11. *Aggregation Theorem*

Suppose $u, u_1, ..., u_n$ are von Neumann-Morgenstern (VNM) utility functions on the set L of all lotteries generated from a non empty set of pure prospects and, for all $p, q \in L$

(1) *$u(p) = u(q)$ whenever $u_i(p) = u_i(q)$ for all $i = 1, .., n$*[1]

Then there are[2] *real numbers $\alpha_1, ..., \alpha_n$, and β*[3]

$$u(p) = \sum_{i=1}^{n} \alpha_i u_i(p) + b \tag{1}$$

If, in addition, for all $p, q \in L$

(2) *$u(p) > u(q)$ whenever $u_i(p) \geqslant u_i(q)$ for all i, and $u_i(p) > u_i(q)$ for some i.*

Then $\alpha_i > 0$ for all i.

Moving towards the Impartial Observer Theorem, Harsanyi (1977) presented his results in the context of Smith's moral sentimentalism:

> Since Adam Smith, moral philosophers have often pointed out that the moral point of view is essentially the point of view of a *sympathetic* but *impartial* observer. It is the point of view of a person taking a positive sympathetic interest in the welfare of *each* participant but having no partial bias in favor of *any* participant. (1977, p. 48-49)

He provides a particular interpretation of Smith's *sympathy*:

> This must obviously involve his imagining himself to be placed in individual j's *objective position*, i.e., to be placed in the objective positions (e.g., income, wealth, consumption level, state of health, social position) that i would face in

[1] In all truth, this is a presentation of a weaker version of the result, since Harsanyi arrives to the same conclusions but departing from (vNM) preference orderings rather than (vNM) utilities.

[2] Furthermore, if the $u_1, ..., u_n$ are affinely independent, then the α_i and β are unique.

[3] A further assumption here involves what Weymark Weymark (1991) calles *Independent Prospects*: For each $i = 1, ..., n$, there exists $p^i, q^i \in P$ such that such that for all $p \in P$, $p^i \nsucc_j q^i$ for all $i \neq j$ and $p^i \sim q^i$. This will be Axiom 5 below.

> social situation A. But it must also involve assessing these objective conditions in terms of j's own *subjective attitudes* and *personal preferences.* (1977, p. 52)[4]

In order to account for such sympathetic judgments of the imagination, Harsanyi introduces lotteries over *extended alternatives* of the form $[A_j, P_j]$, where A_j is j's personal prize (or personal position in social situation A), and P_j denotes j's subjective attitudes towards that prize or position.

According to Harsanyi's reading of Smith, an individual's choice among social situations is *moral* if it satisfies the requirement of impartiality and impersonality; this is, if they *did not know in advance* what their own social position would be in each social situation [Harsanyi (1977), pg.49]. He first models this uncertainty by requiring that they treat a social situation A as if it were an equiprobable mixture of the n extended alternatives $[A_1, P_1], ..., [A_n, P_n]$, where $i = 1, .., n$ are all the members of society. Later, he expresses the impartiality and impersonality making use of an assumption of symmetry.

THEOREM 12. *Impartial Observer Theorem*

Suppose $W, u_1, ..., u_n$ are von Neumann-Morgenstern (VNM) utility functions on the set L of all lotteries generated from extended alternatives and, for all $p, q \in L$ (1) and (2) are satisfied. Furthermore, assume the following holds

(3) *If $u_1, ..., u_n$ are expressed in the same utility unit*[5]*, then W must be a symmetric function of these individual utility functions.*

, then

$$W(p) = \frac{1}{n} \sum_{i=1}^{n} u_i(p) \tag{2}$$

[4]This latter account on the basis of *consumer's sovereignty*: The interests of each individual must be defined in terms of their *own* personal preference and not in terms of what somebody else thinks is good for them. We will return to this idea when assessing *non-paternalistic altruism.*

[5]For the time being, I will bypass the issue of interpersonal utility comparisons.

Although Harsanyi (1977) framed the interpretation of the results in terms of Smith's moral sentimentalism and sympathy, later discussion revolved around their significance for social welfare and utilitarianism. In particular, he took equation (2) to express utilitarianism, namely the claim that social welfare ought to be understood as the equally weighted sum of individual utilities. His result was challenged most notably by Sen (1974, 1977, 2005) and Diamond (1967). They objected the appropriateness of the expected utility axioms (mainly Independence) for social decision-making, and the fact that the linearity result depends on whether or not VNM utility representations are used at the individual level.[6] In a nutshell, they questioned whether Harsanyi's theorems provide an accurate axiomatization of (weighted) utilitarianism or just another representation theorem. The literature discussing these issues extends way beyond the scope of this essay. For an excellent, though outdated, survey paper see Weymark (1991); for quality examples of the state of the contemporary debate, see Fleurbaey and Mongin (2016), Mongin (2001), or Fleurbaey (2010).

The suggestion now is to return to Harsanyi's original interpretative line. As it will be developed in the next section, Adam Smith's development of his *Impartial Spectator* is built over a moral psychology that takes into account *individual's* judgments of sympathy as well as their social, unsocial, and selfish passions. Harsanyi's Aggregation Theorem, independently of its relevance with respect to utilitarianism or social welfare, will become useful as a technique to represent and characterize those other-directed subjective attitudes.

1.2. Smith's moral passions.

In *The Theory of Moral Sentiments* (TMS, Smith (2002)), Smith argues that social psychology is a better guide to moral action than reason. One way of reading Smiths TMS is as a reconstruction of virtue ethics, based on the study of human moral sentiments. Virtuous actions are

[6] In principle, it is theoretically permissible to replace an individual's VNM utility function by any non-affine increasing transformation, see Fleurbaey and Mongin (2016) for a resolution of this point.

those actions that have merit and are praiseworthy (propriety), vicious are those that are improper and punishable. In order to determine which actions are proper or improper we rely on the moral sentiment of sympathy, the tripartite division of the moral passions, plus other cognitive abilities like imagination. His famous *Impartial Spectator* emerges as a moral standard over the base of individuals exercising their subjective moral/sympathetic judgments in society.

Smith's most fundamental concept in TMS is that of sympathy, in particular *situational* sympathy, whereby we imagine how we would feel in the circumstances of others (a rich discussion of Smith on sympathy can be found in Griswold (1998), ch.2).[7] Situational Sympathy therefore involves not only the capacity for empathy (recognizing and sharing the emotions of others), but also a judgment. If after imagining ourselves in other's shoes, we picture ourselves feeling the same emotions they express, then we judge their sentiments appropriate; otherwise we judge them inappropriate. Furthermore, if there is concordance between the emotions experienced by the the agent and the emotions that arise form the situational sympathy, then we achieve *mutual sympathy*. And this state is not just the sharing of positive emotions, but "any passion whatever":

> We enter into their gratitude towards those faithful friends who did not desert them in their difficulties; and we heartily go along with their resentment against those perfidious traitors who injured, abandoned, or deceived them (2002, p. 13)

For Smith, we not only *supply*, but we also *demand* sympathy. As agents, we make constant efforts to adjust our feelings so that sympathetic spectators would judge our actions as appropriate. It is this process of mutual emotional adjustment that gives rise to virtue. We will get back to this in a moment.

[7] "Sympathy, therefore, does not arise so much from the view of the passion, as from that of the situation which excites it. We sometimes feel for another, a passion of which he himself seems to be altogether incapable; because, when we put ourselves in his case, that passion arises in our breast from the imagination, though it does not in his from the reality." (2002, p. 15)

Mutual Sympathy, then, is "our fellow-feeling with any passion whatever" (pg.13), which begs for classification of emotions. In TMS, Smith recognizes three types of (moral) passions brought by the habit of the imagination: Unsocial passions, social passions, and selfish passions.[8] The unsocial passions are "hatred and resentment, with all their different modifications" (pg.41); the social passions are "generosity, humanity, kindness, compassion, mutual friendship and esteem, all the social and benevolent affections"(pg.47); and the selfish passions are "never either so graceful as is sometimes the one set, nor is ever so odious as is sometimes the other" (pg.49), which Smith exemplifies them with grief and joy.

One way of reading this tripartite distinction among the moral passions is as inducing for each individual a partition of society in three distinct groups. A group towards which they are spiteful, a group towards which they are altruistic, and a group towards which they are selfish.[9] This is the picture pursued in this essay. The next sections will provide definitions and formal characterizations of these other-directed attitudes, as well as some discussion of their significance.

One biased approach to the *impartial spectator* is through the *Wealth of Nations* (WN, Smith Adam 1723-1790 (2000)). Out of the individuals' sympathetic judgments of propriety and impropriety, and their demand for sympathy, a marketplace comes into place where agents negotiate and adjust their moral passions. The impartial spectator comes about as the standard for judgment:[10]

> Whatever judgment we can form concerning them [our own sentiments and conduct], accordingly, must always bear some secret reference, either to what are, or to what, upon a certain condition, would be, or to what, we imagine, ought to be

[8]This account is developed in Part I, Section II, Chapters III-V.

[9]We should also recognize the possibility of context-dependence. For each individual j there might be situations in which I'm altruistic towards j, situations in which I'm spiteful towards j, and situations in which I'm simply self-interested in dealings with j. For the purposes of this essay we will avoid this issue.

[10]This account of the impartial spectator as a form of equilibrium point in the marketplace of judgments of sympathy is in no way canonical.

> the judgment of others. We endeavour to examine our own conduct as we imagine any other fair and impartial spectator would examine it. (2002, p. 128-129)
>
> There exists in the mind of every man, an idea of this kind, gradually formed from his observations upon the character and conduct both of himself and of other people. It is the slow, gradual, and progressive work of the great demigod within the breast, the great judge and arbiter of conduct. (2002, p. 219)

Smith offers several interpretations of this demigod, sometimes as an ideal of exact propriety and perfection under perfect information, sometimes by an approximation to this idea as it is exemplified by the best of society,[11] and others. I do not pretend an exhaustive presentation here. Furthermore, it is doubtful that Harsanyi's account of the impartial observer as an individual who is uncertain with respect to their social position was intended to be an accurate characterization of Smith's. The purpose for invoking the impartial observer here is to emphasize that for Smith it is a *bottom up* emerging concept, "gradually formed from his observations upon the character and conduct of both himself and of other people". Hence the study of Smith's three types of moral passions at the *individual* level (rather than the social welfare level) is a necessary precondition, and this will be the focus of the present work.

2. Altruism

Smith's opening lines of TMS speak to the importance that the social passions have to his moral project:

> How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it.

[11] "In estimating our own merit, in judging of our own character and conduct, there are two different standards to which we naturally compare them. The one is the idea of exact propriety and perfection, so far as we are each of us capable of comprehending that idea. The other is that degree of approximation to this idea which is commonly attained in the world, and which the greater part of our friends and companions, of our rivals and competitors, may have actually arrived at." (Smith (2002) pg. 219)

For Smith, social passions are such that *even in excess* they do not induce any form of aversion or resentment, but rather some compassion;[12] very much unlike unsocial passions. Furthermore, when they are appropriate, they are the object of gratitude and require reward.[13] The proper objects of either gratitude or resentment constitute the basis for merit and demerit, which are the building blocks of virtue and vice.

The focus of the present section is on benevolence, the concern for the well being of others. It will be treated her under with the concept of *altruism*; since this is the technical term adopted by contemporary literature.

2.1. A counterfactual definition.

Allow me to begin with an intuitive definition of behavioral altruism:

> An individual agent behaves altruistically whenever it performs an action that it is costly for itself but beneficial for another party.

This definition is behavioral in so far as there is no appeal to intentions like beliefs or desires, nor to emotions like the warm glow felt by those that enjoy helping others. It is defined just in terms of actions in an environment or decision context. The first task is to specify what is meant by "costly" and "beneficial". This will be done by refining (and generalizing) the definition.

Kitcher (2010) introduced a counterfactual account of altruism, an approach that we could trace back to the *Groundwork* (Kant and Gregor, 1998).[14] He compares the attitudes that agents have

[12] "Those amiable passions, even when they are acknowledged to be excessive, are never regarded with aversion. There is something agreeable even in the weakness of friendship and humanity. The too tender mother, the too indulgent father, the too generous and affectionate friend, may sometimes, perhaps, on account of the softness of their natures, be looked upon with a species of pity, in which, however, there is a mixture of love, but can never be regarded with hatred and aversion, nor even with contempt, unless by the most brutal and worthless of mankind". (pg. 49)

[13] "Actions of a beneficent tendency, which proceed from proper motives, seem alone to require reward; because such alone are the approved objects of gratitude, or excite the sympathetic gratitude of the spectator". (pg. 91)

[14] In the first section of his *Groundwork*, Kant is interested in knowing whether an altruist ("friend of humanity") is acting out of *duty* (i.e. respect for the moral law) or out of *inclination* (i.e. the "warm glow"). In order to solve this puzzle, he suggests a counterfactual account by exploring what would the agent do if it were not affected by the pleasure of helping:

> "Suppose, then, that the mind of that friend of humanity were beclouded by his own grief, which extinguishes all compassion for the fate of others; that he still had the means to benefit others in

in two different contexts, a solitary choice context C^* in which the payoffs are defined *only* for the decision maker; and a corresponding social choice context C with choices involving the same payoff for him or her, but payoffs are also defined for other agents.[15] Furthermore, in context C agents have information about the preferences of other agents. The basic idea is that to assess whether an agent is behaving altruistically we need to compare what that agent would have done if others were not involved. If, when moved from the solitary situation to the social one, the agent is prepared to change their choice in a way that benefits the other agent, then they are behaving altruistically. The underlying rationale is that the agent is paying an opportunity cost by changing the choice to a less preferred option, and the most salient reason they are doing it is to benefit others. An agent i *is an altruist* with respect to another agent j if[16] *there are* solitary and social counterpart choice contexts C^* and C such that:

(a) i's choice in C is different from their choice in C^*.

(b) i's choice in C offers a better payoff to j than the payoff corresponding to the choice that i makes in C^*.

The two contexts C and C^* need not necessarily be presented as *counterfactual* counterparts. One possible reading is purely epistemic. The agent is presented with two decision problems. Although they are informed of the outcomes for themselves in both, in the solitary context they have no idea about the payoffs for others - they simply don't enter the picture.

---

need, but the need of others did not touch him because he is sufficiently occupied with his own; and that now, as inclination no longer stimulates him to it, he were yet to tear himself out of this deadly insensibility, and to do the action without any inclination, solely from duty; not until then does it have its genuine moral worth." (4:398)

We are not interested in *duty* here, but this counterfactual method fits nicely when trying to asses altruistic attitudes.

[15]The reader might fairly object that there is an unjustified assumption here. Namely, the idea that we can in most contexts separate the outcomes affecting different agents. In particular, that cases involving altruism are such that we can always posit these type of solitary and social counterparts. I take the complaint to be a fair one, but I do not wish to address it here. Rather, I urge the reader to accept it now for the sake of the argument, and later assess its merits by its conceptual advantages.

[16]For the time being, allow me to take the following condition as *sufficient*.

Consider the following example. Suppose first that you are not originally willing to invest in a risky gamble (like buying the national lottery ticket) [context C^*]. Nevertheless, assume now you are willing to invest but only after learning that the earnings will go to a charity [context C]. Then we can safely claim that you are being altruistic. If you would have chosen to play in both contexts, then we would not be able to decide whether you are are acting altruistically you might just enjoy gambling. Furthermore, if provided with similar solitary and social context counterparts including all available lotteries, you always choose the same in both contexts; then it is safe to say that you do not care about the recipients of the charity.

Although Kitcher requires also assessing the intentional structure (beliefs and desires) of the agent, I will here disregard this dimension and treat it using the utility theory formal apparatus. This is because there are some methodological difficulties with *psychological* altruism.

On the one hand, psychological altruism is hard to judge, simply because it is in general difficult to determine the agent's intentional and motivational structure. This is an empirical enterprise beyond the scope of this essay, if it is even feasible.[17]

On the other, there are particular difficulties in determining whether the agent has the right intentional structure in *strategic* situations - namely, Game Theory. Issues like expected reciprocity, strategic helping, reputation building, or signaling weight as some of the many reasons agents could have in iterated or one-shot games. For an excellent critical essay on this matter, see Elster (2006).

[17] To bring back some Kantian themes, and with the proviso that Kant is concerned about duty, he stresses that experience will *never* be able to determine the cause of the will:

> In fact, it is absolutely impossible by means of experience to make out with complete certainty a single case in which the maxim of an action that otherwise conforms with duty rest solely on moral grounds and on the representation of ones duty. (...) from this it cannot be inferred with certainty that the real determining cause of the will was not actually a covert impulse of self-love under the mere pretense of that idea; for which we then gladly flatter ourselves with the false presumption of a nobler motive, whereas in fact we can never, even by the most strenuous examination, get entirely behind our covert incentives, because when moral worth is at issue what counts is not the actions, which one sees, but their inner principles, which one does not see. [4:407]

The present essay will make use of the formal apparatus of utility theory and its representation theorems. They have been usually framed in behavioristic terms. Agents are not required to provide reasons for their preference structure, and the framework has no room for capturing their intentional structure. It only minimally requires agents to provide a preference ordering, constrained by (questionable) rationality principles. The usual upshot is that even with this minimal information about the agents, they can still be regarded *as* maximizing expected utility. Nevertheless, the apparatus of expected utility theory *can* be used to ascribe to agents intentional states, especially preferences and beliefs, and therefore this essay lays *between* psychological and behavioral altruism.

2.2. Formalizing the definition.

I will proceed in the usual Von Neumann and Morgenstern (1947) framework, in the same way as Harsanyi does. A context is nothing more than a choice among a set of lotteries. Given an agent i, a context C^* is *solitary* in the sense that it specifies outcomes *only for agent* i; while C is *social* because it specifies outcomes for i but also for other agents. We say that C is the social counterpart of C^* if for each lottery in the latter there is exactly one lottery in the former with the same expected outcomes for i, and vice-versa. Clearly, a solitary context for a single agent i can have several counterpart social contexts corresponding to it. A situation is a pair (C^*, C) of solitary and social counterpart contexts.

More formally, for each of the agents $i \in I$ we have a finite set of *personal* prizes or outcomes $X_i = \{A_i^1, .., A_i^n\}$[18], and a set of lotteries or prospects $Y_i = \Delta X_i$ defined over those outcomes. On the other hand, there are *social* prizes or outcomes $X = \prod_i X_i$ represented as a Cartesian product of the personal outcomes[19]. In this way, a social outcome can be thought as a profile of individual outcomes. Also, there are social lotteries or prospects $Y = \Delta X$. Lotteries $Y = \Delta X$ and $Y_i = \Delta X_i$ reflect the two largest social and personal contexts of the previous definition.

[18] Nothing precludes each agent to consider *different* sets of outcomes.

[19] For simplicity, I here avoid using *extended lotteries* as defined before. All the formalisms hold analogously.

Following Harsanyi (1977), the first set of assumptions of the model are rationality assumptions. Agents are sufficiently (vNM) rational with respect to their preferences over personal and social lotteries.

Axiom 1: Individual rationality over personal preferences

Each of the agents $i \in I$ has a preference ordering $\succcurlyeq_i$ over Y_i that satisfies the von Neumann-Morgenstern axioms[20]. Here $L_1^i \sim_i L_2^i$ is $L_1 \succcurlyeq_i L_2$ and $L_2^i \succcurlyeq_i L_1^i$, while $L_1^i >_i L_2^i$ is $L_1^i \succcurlyeq_i L_2^i$ but $L_1^i \succ_i L_2^i$.

Axiom 2: Individual rationality over social preferences

Each of the agents $i \in I$ has a preference ordering $\gtrapprox_i$ over Y that satisfies the von Neumann-Morgenstern axioms. Here $L_1 \approx_i L_2$ is $L_1 \gtrapprox_i L_2$ and $L_2 \gtrapprox_i L_1$, while $L_1 \gg_i L_2$ is just $L_1 \gtrapprox_i L_2$ but $L_1 \not\approx_i L_2$.

Each of the agents $i \in I$ preference ordering $\succcurlyeq_i$ over personal lotteries Y_i can be extended to a preference ordering $\succcurlyeq_i$[21] over the social lotteries Y in the following way: For all $L_i \in Y$, $L_1 \succcurlyeq_i L_2$ if and only if $L_1^i \succcurlyeq_i L_2^i$. Here a lottery of the form L_j^i is the marginal personal lottery for agent i in social lottery L_j.

Observation 6. *Each of the agents $i \in I$ extended preference ordering $\succcurlyeq_i$ over social lotteries Y also satisfies the von Neumann-Morgenstern axioms.*

[20] I here group a set of axioms as one single proposition for simplicity. In the interest of completeness, I will list them:

- Axiom 1 (Ordering): $\succcurlyeq_i$ is a reflexive, complete, and transitive binary ordering.
- Axiom 3 (Continuity): Given lotteries L, M and K: If $L \succcurlyeq M \succcurlyeq K$, then there is $p \in [0,1]$ such that $< L, p; K, (1-p) > \sim M$.
- Axiom 4 (Independence): Given lotteries L, M and K: If $L > M$, then for any K and $p \in (0,1]$, $\langle L, p; K, (1-p)\rangle > \langle M, p; K, (1-p)\rangle$.

[21] There is some abuse of notation here.

The observation follows trivially. The idea here is that the new preference ordering $\succcurlyeq_i$ represents how the agent would organize their **social** preferences if they were not informed about how the outcomes affect others; namely by only taking into account their **personal** preferences.[22]

Now let us move to the altruistic conditions. The main idea here is to transition from the intuitive definitions presented before to an interpretation in terms of Pareto-like requirements.

The gist of Kitcher's definition was that an agent is altruistic towards another if, for some personal and social context pair, there is shift in choices when moving from personal to social contexts such that it favors the beneficiary. Altruists are willing to pay some (opportunity) cost so long as the outcome will benefit someone else. There are may ways in which this definition is unsatisfactory, in particular when we take into account multiple agents and sets of contexts.

Suppose you are indifferent to Bob but you care about Alice. Furthermore, assume that in a particular context choice pair you are willing to change your choice once you learn it will benefit Alice. But by a mere accident your social choice also benefits Bob. According to the current definition, that would suffice to make you an altruist towards Bob; which we rejected by assumption. Nevertheless, if Kitcher's condition is taken as a *necessary* requirement for altruism (rather that a sufficient one), then it seems fitter. For if Alice is among the people you care about, then there are surely many situations in which you are willing to pay an opportunity cost to help her.

Furthermore, it is also unsatisfactory to look just to particular context choice pairs. After all, you might be willing to pay to benefit Alice in some situation, but also pay to harm her in another. To be altruistic towards Alice means that you have the *general* or *systematic* disposition to help her, and that you are *never* disposed to harm her.

[22]The reader might fairly object that there is an unjustified assumption here. Namely, the idea that we can in most contexts separate the outcomes affecting different agents; as if they were independent. Although one might argue that actual decision-making sometimes recognizes effects on other people and sometimes not; I take the complaint to be fair, but I do not wish to address it here. Rather, I urge the reader to accept it now for the sake of the argument, and later assess its merits by its conceptual advantages.

But the previous principle is also problematic if stated carelessly. There is no inconsistency in caring *more* about Bob than about Alice; while also being an altruist towards Alice. There might be contexts in which you might prefer to harm Alice (although you care about her), so long as Bob receives a substantial benefit.

The way of cashing out the previous intuition without falling into the problems just presented is to say that an agent i^* is an altruist with respect to an agent j if *all other things being equal (for the relevant set of agents)*, i^* would rather (a) not harm, or (b) benefit, j.[23] This is precisely what is captured by the Pareto-like principles below.

Axiom 3: Weak Pareto (No Harming Altruism) [WA]

We say that i^* is weakly altruistic with respect to the agents in $J \subseteq I$ when:

If for all agents $j \in J$ we have that $L_1^j \succcurlyeq_j L_2^j$, then $\prod_j L_1^j \gtrapprox_{i^*} \prod_j L_2^j$.[24]

Axiom 4: Strong Pareto (Beneficent Altruism) [SA]

We say that i^* is strongly altruistic with respect to the agents in $J \subseteq I$ when two conditions hold:

(1') If for all agents $j \in J$ we have that $L_1^j \sim_j L_2^j$, then $\prod_j L_1^j \approx_{i^*} \prod_j L_2^j$.

[23] A word needs to be said with respect to *paternalistic* and *non-paternalistic* altruism. This in turn is related with what Harsanyi Harsanyi (1977) calls the *principle of consumer's sovereignty*:

> The interest of each individual must be defined fundamentally in terms of his *own* personal preferences and not in terms of what somebody else thinks is "good for him". (pg. 52)

The distinction relies on an ambiguity of what it means to believe that an action benefits others. On the one hand, this could mean that Bob chooses what he thinks is best for Alice *according to his preference system, and disregarding what he knows about Alice's.* This is paternalistic altruism, since he is disregarding Alice's preferences and acting according to what he believes is beneficial for her. On the other hand there is non-paternalistic altruism, when *the altruist takes the recipient as a sovereign subject of his or her own good.*

[24] There is some abuse of notation here. If $J \subset I$, then $\prod_j L_1^j \gtrapprox_{i^*} \prod_j L_2^j$ means that for any two lotteries $\prod_i L_1^i$ and $\prod_i L_2^i$, if the former coincides with $\prod_j L_1^j$ in all the $j \in J$ values, the second one with $\prod_j L_2^j$, then $\prod_i L_1^i \gtrapprox_{i^*} \prod_i L_2^i$.

(2') If for all agents $j \in J$ we have that $L_1^j \geqslant_j L_2^j$, but also there is at least one agent $k \in J$ such that $L_1^k >_k L_2^k$, then $\prod_j L_1^j \gg_{i*} \prod_j L_2^j$.

Condition (1') is usually called *Pareto Indifference.* The antecedents of the first definition and of clause (2') of the second ensure that if everyone agrees that the *personal* prizes they get in one lottery are not worse than the personal prize they get in the other, then *social* preferences of the altruist should find the first lottery as preferred as the second. Furthermore, notice that these altruistic Pareto axioms are expressed conditionally but not biconditionally. An altruist may prefer a social lottery over another even though the Pareto condition is not met for the personal prizes in those lotteries. In particular, suppose i^* cares about both j_1 and j_2 and there are two lotteries L_1 and L_2. Also assume that j_1 strictly prefers the personal prize of L_1 over that of L_2, and the converse holds for j_2. Nothing precludes i^* to *prefer* j_1 over j_2 by ranking $L_1 \gg_{i*} L_2$. This is not inconsistent with i^* being an altruist towards j_2, just as in the case of Bob and Alice mentioned before.

The previous observation is relevant since I will assume throughout the essay that agents care about themselves. Formally, if $J \subseteq I$ is the collection of agents that i^* cares about (as expressed in the altruistic axioms), then $i^* \in J$. This is not required by the formalism, but by the concept of altruism being characterized. As it will become clear later, if i^* is (strongly) altruistic towards j, then there will be two lotteries L_1 and L_2 such that i^* strictly prefers the personal prize in the former but the social outcomes of the second, just because j is better off there. In a nutshell, i^* is willing to pay a (opportunity) cost to benefit j. This shows that Kitcher's *necessary* requirement is met.

2.3. Characterizing Altruism.

The main result offered here is a corollary of Harsanyi's Aggregation Theorem (Harsanyi, 1955, 1977), but modified to fit the current framework. Although Harsanyi's result holds, his published presentations are in some way flawed. In particular, Resnik (1983) introduced a difficulty and Fishburn (1984) provided an amended proof. I will nevertheless follow and adjust Harsanyi (1977) terminology for its simplicity and naturalness. The following innocuous axiom is necessary:

Axiom 5: Richness of the space[25]

For each of the agents $i \in I$, there are $y_i, z_i \in Y_i$ such that: $y_i \nsucc_i z_i$.

This just requires that all agents are not indifferent between some two personal outcomes.

This result can be obtained from the axioms above:

PROPOSITION 13. *Characterizing Altruistic Utilities as weighted Solitary Utilities:*[26]

If Axioms 1 and 2 are satisfied, then each $i \in I$ has a solitary utility function su_i that represents their solitary preferences, and an altruistic utility function au_i such that that represents their social preferences:

- *For any two lotteries $L_1^i, L_2^i \in Y_i$, $L_1^i \gtrapprox_{i*} L_2^i$ if and only if $su_i(L_1^i) \geqslant su_i(L_2^i)$. Also, su_i satisfies the expected utility property: $su_i(\langle A_i^k, p_k \rangle) = \sum_k su_i(A_i^k).p_k$. Here the A_i^k are personal prizes for i.*
- *For any two lotteries $L_1, L_2 \in Y$, $L_1 \geqslant_i L_2$ if and only if $au_i(L_1) \geqslant au_i(L_2)$. Also, au_i satisfies the expected utility property: $au_i(\langle B^k, p_k \rangle) = \sum_k au_i(B^k).p_k$. Here the B^k are social prizes.*

[25] In the literature around Harsanyi's result, a similar principle is usually called *Independent Prospects*; but the name does not suit the current framework.

[26] I call it a *characterization*, rather than a *representation*, because the counterpositive of the proposition holds trivially.

More interestingly, if i^* also satisfies Weak Altruism with respect to the $j \in J$, as well as Axiom 5, then there are $\alpha_1, ..., \alpha_n \in [0,1]$ such that for any social lottery $\prod_j L^j$:

- $au_{i^*}(\prod_j L^j) = \sum_{j \in J} \alpha_j . su_j(L^j)$
- *Furthermore, if i^* satisfies Strong Altruism then for all $j \in J$, $\alpha_j \in (0,1]$ and $\sum_j \alpha_j = 1$.*

A short proof is given in the appendix. The upshot of this result is that altruistic utilities can be characterized as a weighted averages of solitary utilities.

We concluded that to be an altruist means that all other things being equal (for the relevant set of agents), our altruist would help their beneficiary. This was captured by the Pareto axioms stated before. With them we used Harsanyi's result to show the proposition. Notice that the necessary condition imposed by Kitcher's criteria is met. Consider an agent $j \in J$ such that $\alpha_j > 0$. It is a simple exercise to construct two social lotteries, L_1 and L_2, such that all agents except j and i^* receive the same prize; but i^* *personally* prefers his or her personal prize in L_1 over the one in L_2, and j *personally* prefers his or her personal prize in L over the one in L_1. Nevertheless, since i^* is an altruist towards j, he or she would *socially* prefer L_2 over L_1; this is, they would be willing to pay a personal cost in order to benefit j.

To conclude, notice that this way of characterizing altruism is consistent with the standard general requirements of monotonicity imposed over altruistic utilities. For example, Kolm (2006) (pg. 7-8) insists that if we are to follow Smith, we should represent the level of happiness of an individual i^* by a utility function $u_{i^*} = u_{i^*}(u_{-i^*}, x_{i^*})$, where $u_{-i^*} = \{u_j\}_{j \in J}$ is the set of levels u_j for all individuals $j \in J$, and x_{i^*} denotes other factors of i^*'s happiness. The general reuirement is that each u_{i^*} must be an increasing function of each u_j, for all $j \in J$. Kitcher (2010, 1993) also presents constrains requiring at least monotonicity. The Pareto axioms guarantee that present

formalization in terms of *positive* weighted averages satisfies them. Furthermore, if we think those axioms accurately capture the requirement, then Harsanyi's result secures that the linearity is necessary.

3. Spite

Maybe surprisingly, Smith ascribed a fundamental and positive role to the unsocial passions of hatred and resentment; *so long as they are properly tamed*. This can be viewed as a general interpretative key for Smith: Recognizing that apparently negative human attributes have a positive overall effect in society.[27] The *Wealth of Nations* shows that self-interest can lead to economic progress. Almost twenty years before, Smith argued in his *Theory of Moral Sentiments* that resentment is necessary for individual survival, and it is at the base of justice and punishment.

Smith is emphatic in that these passions are in general disagreeable.[28] They are particularly disagreeable if compared with the social passions. Yet, Smith argues that, if properly "brought down to a pitch much lower than that to which undisciplined nature would raise them" (pg. 41) they could serve society and the individual. This is, if hatred and resentment are *appropriate*, in the sense that we can in fact *sympathize* with them, then they bare some use:

> But though the utility of those passions to the individual, by rendering it dangerous to insult or injure him, be acknowledged; and though their utility to the public, as the guardians of justice, and of the equality of its administration, be not less considerable, as shall be shewn hereafter; yet there is still something disagreeable in the passions themselves. (2002, p. 43)

A functional member of society must be capable to bare resentment and hatred, even if they never exercise it. Without that capability, Smith seems to argue, they would be unable to defend

[27] If we take into account Mendeville's *Fable*, this could be seen as a topic of the times.

[28] "Too violent a propensity to those detestable passions, renders a person the object of universal dread and abhorrence, who, like a wild beast, ought, we think, to be hunted out of all civil society". (pg. 49)

themselves and deter their attackers.[29] Furthermore, those who are capable of exercising *proper* resentment are those who exercise the virtue of justice.[30] Finally, Smith sometimes even seem to suggest that the impartial spectator, as a standard of moral behavior, must also be capable of expressing (or sympathize with) unsocial passions.[31]

The topic of *spite*, or more generally negative social emotions, seem to be under-treated in the literature. In economics, Kolm (1995, 2002) is the prime example of the study of *envy*. In biology, shortly after his influential papers on the evolutionary advantages of altruism (Hamilton, 1963, 1964; Axelrod and Hamilton, 1981), Hamilton (1970) set himself to provide an evolutionary explanation of *spite* among social animals. It would be disingenuous to deny that spite is part of human psychology. Following Smith's, the proper challenge is to understand them, and train them in a way that become beneficial for society at large.

3.1. Definition, Formalization and Characterization.

I will here take spite as altruism's evil twin. If, intuitively, altruism is the willingness to pay some cost in order to benefit someone; spite is the willingness to pay a cost in order to *harm* someone. Since they are so structurally similar, most of the subtleties of its formalization are similar to that of the formalization of altruism, which were discussed in Sections 2 and 3. I will then proceed to formalize the notion analogously:

Axiom 6: Weak Spite (No Help Spite) [WS]

We say that i^* is weakly spiteful with respect to the agents in $J \subseteq I$ when:

[29] "Resentment seems to have been given us by nature for defence, and for defence only. It is the safeguard of justice and the security of innocence".(pg. 92)

[30] "This virtue is justice: the violation of justice is injury: it does real and positive hurt to some particular persons, from motives which are naturally disapproved of. It is, therefore, the proper object of resentment, and of punishment, which is the natural consequence of resentment".(pg. 93)

[31] "The heart of every impartial spectator rejects all fellow- feeling with the selfishness of his motives, and he is the proper object of the highest disapprobation". (pg. 81)

If for all agents $j \in J$ we have that $L_1^j \succcurlyeq_j L_2^j$, then $\prod_j L_2^j \gtrapprox_{i*} \prod_j L_1^j$.

Axiom 7: Strong Spite (Harmful Spite) [SS]

We say that i^* is strongly spiteful with respect to the agents in $J \subseteq I$ when two conditions hold:

(a) If for all agents $j \in J$ we have that $L_1^j \sim_j L_2^j$, then $\prod_j L_2^j \approx_{i*} \prod_j L_1^j$.

(b) If for all agents $j \in J$ we have that $L_1^j \succcurlyeq_j L_2^j$, but also there is at least one agent $k \in J$ such that $L_1^k \succ_k L_2^k$, then $\prod_j L_1^j \gg_{i*} \prod_j L_2^j$.

PROPOSITION 14. *Characterizing Spiteful Utilities as weighted Solitary Utilities:*

If Axioms 1 and 2, then each $i \in I$ has a solitary utility function su_i that represents his/her solitary preferences, and a spiteful utility function tu_i such that that represents his/her social preferences:

- *For any two lotteries $L_1^i, L_2^i \in Y_i$, $L_1^i \gtrapprox_{i*} L_2^i$ if and only if $su_i(L_1^i) \geqslant tu_i(L_2^i)$. Also, tu_i satisfies the expected utility property: $su_i(\langle A_i^k, p_k \rangle) = \sum_k su_i(A_i^k).p_k$. Here the A_i^k are personal prizes for i.*
- *For any two lotteries $L_1, L_2 \in Y$, $L_1 \succcurlyeq_i L_2$ if and only if $tu_i(L_1) \geqslant tu_i(L_2)$. Also, tu_i satisfies the expected utility property: $tu_i(\langle B^k, p_k \rangle) = \sum_k tu_i(B^k).p_k$. Here the B^k are social prizes.*

More interestingly, if i^* also satisfies Weak Spite with respect to the $j \in J$, as well as Axiom 5, then there are $\beta_1, ..., \beta_m \in [-1, 0]$ such that for any social lottery $\prod_j L^j$:

- $tu_{i*}(\prod_j L^j) = \sum_{j \in J} \beta_j . su_j(L^j)$
- *Furthermore, if* i^* *Satisfies Strong altruism then for all* $j \in J$, $\beta_j \in [-1, 0)$ *and* $\sum_j \beta_j = -1$.

The fact that Harsanyi's Aggregation Theorem can imply *negative* weights seems to have been neglected. In particular, if only Pareto Indifference is required, nothing precludes *all* parameters to be negative. When presenting an Impartial Observer, Harsanyi and others usually imagine a *beneficent* agent that serves as a standard that takes everyone as equal. This is very much in line with utilitarianism, but it might be in tension with Smith's moral sentimentalism. For him, the Impartial Spectator is capable of praising virtue and blaming vice, sympathize with altruism and antagonize with improper spite, reward heroes and punish criminals.

4. Das Adam Smith Problem

Das Adam Smith Problem is an argument that arose among German scholars during the second half of the nineteenth century concerning the compatibility of Smith's conceptions of human nature advanced in *Theory of Moral Sentiments* and *Wealth of Nations*. Section 4 of Kitcher (2010) articulates this problem in detail. Roughly, the question is how to reconcile the emphasis on sympathy and benevolence in Smith's first book, contained in the opening quote of this paper, with the emphasis on self-interest in Smith's later work, as testified by the famous:

> It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest.

The previous observations about Smith's treatment of the *unsocial* passions show that the tension is not just between benevolent attitudes and a self-interested driven market; but on how to harmonize the latter both with *altruism* and with *spite*. As explained before, unsocial passions like resentment, anger and hatred play a fundamental role in Smith's moral sentimentalism. Without the recognition of negative emotions towards others, Smith's account of human nature in TMS is

simply incomplete. The purpose here is not to solve *Das Adam Smith Problem* (DASP), but to make it worse.

Furthermore, if we take Harsanyi's framework to be an accurate interpretation of Smith, then his Aggregation Theorem is *precisely a worsening of DASP*. Smith's distinction between the unsocial, social and self-interested moral passions lead the way to a three-fold partition of society between those we care about, those towards which we hold resentment, and those for which we do not have any systematic concern in various contexts. The first part of Harsanyi's theorem implies linearity in the presence of Pareto Indifference; but nothing secures that the real weights α_i will all be positive or negative. Without loss of generality, we can divide those weights in those two groups. Positive values will correspond to individuals the agent cares about, negative to those they hold resentment to, and those individuals with zero weight represent those the agent has no particular regard for.

We can, on the other hand, depart from such a societal partition, impose the pareto conditions to the groups, and obtain the desired parameters using Harsanyi]s technique. Suppose now that for each agent $i^* \in I$, society is divided between a group $J_1 \subseteq I$ of individuals towards which they are altruistic, a group $J_2 \subseteq I$ towards which they are spiteful, and a group $J_3 \subseteq I$ towards which they have no systematic attitude. Furthermore, for simplicity, assume $J_1 \cup J_2 \cup J_3 = I$ and that groups are disjoint with each other.[32]

Axiom 8: Strong Societal Division [SSD]

Let $J = J_1 \cup J_2$.

[32]The assumption that groups are disjoint is an inescapable idealization in the context of Harsanyi's theorem, which assigns a unique weight to each individual. Yet it is questionable whether people have systematic either altruistic or spiteful attitudes towards individuals, or even to themselves. Those attitudes would require a completely different formal approach.

(a) If for all agents $j \in J$ we have that $L_1^j \sim_j L_2^j$, then $\prod_j L_2^j \approx_{i*} \prod_j L_1^j$.

(b) If (i) for all agents $j_1 \in J_1$ we have that $L_1^{j_1} \succcurlyeq_{j_1} L_2^{j_1}$, (ii) for all agents $j_2 \in J_2$ we have that $L_2^{j_2} \succcurlyeq_{j_2} L_1^{j_2}$, and (iii) there is at least one $j_1 \in J_1$ such that $L_1^{j_1} \succ_{j_1} L_2^{j_1}$ **or** there is at least one $j_2 \in J_2$ such that $L_2^{j_2} \succ_{j_2} L_1^{j_2}$; then $\prod_j L_1^j \gg_{i*} \prod_j L_2^j$.

Axiom 8 implicitly contains the altruistic and spiteful attitudes characterized by the previous axioms. All other things being equal, i^* would be willing to benefit the $j_1 \in J_1$ and harm the $j_2 \in J_2$. The following proposition follows using the same techniques:

PROPOSITION 15. *Characterizing Other-Related Utilities as weighted Solitary Utilities:*

If Axioms 1 and 2, then each $i \in I$ has a solitary utility function su_i that represents his/her solitary preferences, and an social utility function tu_i such that that represents his/her social preferences:

- *For any two lotteries $L_1^i, L_2^i \in Y_i$, $L_1^i \gtrapprox_{i*} L_2^i$ if and only if $su_i(L_1^i) \geqslant su_i(L_2^i)$. Also, su_i satisfies the expected utility property: $su_i(\langle A_i^k, p_k \rangle) = \sum_k su_i(A_i^k).p_k$. Here the A_i^k are personal prizes for i.*
- *For any two lotteries $L_1, L_2 \in Y$, $L_1 \succcurlyeq_i L_2$ if and only if $u_i(L_1) \geqslant u_i(L_2)$. Also, u_i satisfies the expected utility property: $u_i(\langle B^k, p_k \rangle) = \sum_k u_i(B^k).p_k$. Here the B^k are social prizes.*

If i^* also satisfies Axiom 8 with respect to the $j_1 \in J_1$ and $j_2 \in J_2$ respectively, then there are $\alpha_1, ..., \alpha_n \in [0, 1]$ and $\beta_1, ..., \beta_m \in [-1, 0]$ such that for any social lottery of the form $L = \prod_j L^j$:

- $u_{i*}(\prod_j L^j) = \sum_{j \in J_1} \alpha_j.su_j(L^j) + \sum_{j \in J_2} \beta_j.su_j(L^j)$

- *Furthermore, for all* $j \in J_1$, $\alpha_j \in (0, 1]$*; and for all* $j \in J_2$, $\alpha_j \in [-1, 0)$.[33]

Kitcher (2010) approaches DASP by imagining spheres in our social interactions. We engage with others within some of those spheres: we have a degree of altruism towards family, some different towards friends, and another towards acquaintances; we might prefer to buy the groceries at the local store out of sympathy, but we will be more selfish if their prices are too expensive, and we might completely disregard the well being of big company stockholders if we are doing business. There is no account for *spite* in that work.

The results here presented favor that picture of spheres, since there is substantial freedom in the α_j and β_j parameters; sufficient to model those familiar situations. Nothing precludes an agent to be particularly beneficent towards their family and friends, while minimally altruistic towards acquaintances; this can be easily modeled by assigning larger αs to former and smaller to the latter. Similarly, if our agent holds a particular spite against a group of individuals, or their general misanthropy makes them slightly resentful towards all of humanity, this can be easily modeled by adjusting the βs appropriately. Altruism and spite also come in a variety of *degrees* that can be parametrized.

Finally, what about the $j_3 \in J_3$, the agents towards which i^* seems to have no particular sympathetic attitude? In fact, the result ensures that with respect to them, i^* is *self-interested*. On the assumption that $i^* \in J_1$, the idea that i^* cares about themself, it is easy to show that *all other things being equal for the (other)* $j \in J_1 \cup J_2$, i^* would put their personal interest over those of the $j_3 \in J_3$.

[33]Notice, in particular, that if $J_i \neq \varnothing \neq J_2$ the previous entailment that $\sum_j \alpha_j = 1$ and $\sum_j \beta_j = -1$ does not hold any more.

The Smithian project of moral sentimentalism, or for that matter that of social psychology, requires the recognition of other-directed attitudes beyond that of beneficence and self-interest; and the different degrees in which these attitudes are expressed. Throughout the paper I presented definitions and characterization results for both altruism and spite that distinguish for each individual their private utilities from their social other-directed ones. Self-interest came at the end. In the final picture, agents divide society between friends, foes, and neutral; and each individual's social utilities are differently responsive to the interests of the members of each group.

Concepts like spite and negative emotions like hatred or resentment have been under-theorized in the literature. Maybe because, as Smith acutely notes, these are naturally disagreeable passions. But they need recognition, for they play a fundamental role in morality and politics. The hope is that by properly understanding them and, as Smith suggests, taming them, we can shape them into something positive for society at large.

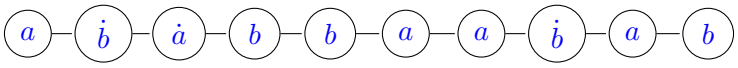
CHAPTER 6

Schelling Segregation in Online Social Networks

1. Introduction

Schelling (2006, 1969, 1971a) models of segregation are recognized as a seminal (Hatna and Benenson, 2015) in agent-based modelling and one of the first contributions in computational social sciences and complex systems dynamics (Epstein and Axtell, 1996). Schelling observed that segregation might emerge from the mutual adaptation of people who seek to live close to members of their own group within spatial constraints but without assuming negative attitudes towards other groups. This provided a new explanatory framework in which undesirable macro-behavior could be explained as the result of otherwise well-intended micro-motives (Schelling, 2006). The most popular models are often cited as a prototype of checkerboard models (Zhang, 2009), and tipping behavior in residential dynamics (Card et al., 2008).

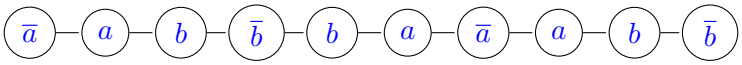
Let us start with the original and simplest model in 1969 Schelling (1969). There, Schelling imagines a line along which two types of agents[1] have been distributed in equal numbers.



Notice that all nodes have an immediate neighbor of a different type. If each of the agents is happy to live together in a ratio of about fifty-fifty, then they are all satisfied if their neighborhood is defined to be the whole line. If instead we define the neighborhood to be their immediate nodes on

[1]Blacks and whites, men and women, natives and immigrants, or any socially reductive dualism of your choice.

their side (and themselves), then some agents will be unhappy. In the illustration above, dissatisfied nodes appear with a dot above. Since they are unhappy, Schelling imagines agents will move to the nearest point at which at least half of their neighbors are like them. This is done stepwise, starting from the left. Once nodes start moving, other nodes can become dissatisfied. If so, mark the new unhappy nodes with a dot, and restart the process. This dynamic will stabilize here:



Now the four nodes with a line above are segregated: they have no immediate neighbors of their type. Schelling's brilliance lies in recognizing that these kind of emergent phenomena, where stable equilibria are highly segregated, can result from the most tolerant micromotives.

Schelling's model as been generalized and modified in multiple ways. For example, there may be more than two agent-types (Wilensky, 1997). Hatna and Benenson (2015) allow for agents to have a preference for diversity, something we do here too. Gandica et al. (2016) generalize by exploring different topologies, allowing for superposition. Paolillo and Lorenz (2018) consider segregation across two dimensions, which lead to cross-contagion effects. Rogers and J. McKane (2011) present a unified framework for segregation with an even more general topology. Examples abound, yet most of them focus on *segregation* and therefore assume a fixed topological space representing a geography which agents inhabit and explore. Even if the topology is generalized by using networks, since it is meant to represent a geography, the Schelling dynamics implemented in those studies does not modify the network in any way. Here, that is precisely what we do. The object of this essay is *Online Segregation*, and the dynamics runs over the social relations that agents have with each other in the context of social networks.

Much like a geography, recent technologies created a new space for human interaction and socialization. In fact, the past decade saw a *replacement* of some geographical spaces of socialization by digital ones. Where a few decades ago gamers gathered together to play chess, cards or board games, now the online gaming community reaches several millions of individuals. Romantic and sexual encounters that before developed spontaneously as a result of a shared space have been heavily influenced by dating apps like Tinder, OkCupid, etc. Local clubs of interest have been virtually replaced by specialized online forums where people can discuss the most varied topics. Friendships and family relations can be held together internationally by WhatsApp group chats. Personal experiences that before were shared in a common event can be now communicated through Facebook or Instagram. Political engagement, which before was done *in situ* or through conventional media, has been massively disrupted by Online Social Networks like Twitter. All of this brings us to central questions: Is the Internet truly bringing us *together*? In particular, are Online Social Networks really making us more *connected*? It does not seem so.

Online polarization, the tendency of agents to adopt more extreme views and attitudes, can be understood as a byproduct of tribalism. In popular culture, tribalism refers to a way of behaving and thinking in which individuals draw much of their personal identity from the tribe they belong and are therefore loyal to their social group while they have animosity towards other groups. In the sciences, tribalism has been widely studied cross culturally and in its evolutionary history (Isaacs and Pye, 1989). We are here interested in *Internet Tribalism* of the kind that emerges from interactions through social media like Facebook, Twitter, Instagram, etc. For the purposes of this essay, we will treat online tribalism and online segregation as very much the same phenomena. After all, segregated networks operate very much like tribalism: agents interact mostly with members of their own group, for which they have a preference. Much like racial segregation, internet tribalism is widely recognized. It has been long reported by the media (Cragg, 2011), diagnosed by NGO's dealing with political polarization (More in Common (2018)), and heavily studied by data scientists

and programmers for the purposes of identifying tribe membership (Bryden et al., 2013; Gloor et al., 2018).

Online tribalism can be very problematic. The connection with political polarization is obvious, but there are other issues. Another example is its relation with *fake news*. William Clifford, who coined the expression "tribal self", famously argued in Clifford (1879) that "it is wrong always, everywhere, and for anyone, to believe anything upon insufficient evidence." Agents within a tribe are expected to be like-minded and biased towards the world view of their tribe. Clifford observed that this makes individuals *credulous*, and willing to spread information without sufficient evidence. What is worse, he observed, is that it such a cultural norm reduces the epistemic standards of others, fostering a credulous *character*. But that, he pointed, is not the end of its evils. By lowering the standards and promoting epistemic vices, the whole social network becomes vulnerable to manipulation: "The credulous man is father to the liar and the cheat." This is to say that problems like online segregation and *fake news*, as well as other issues that result from digital interaction, are most likely interconnected.

Here we are not interested in identifying tribes, much less manipulating them. Rather, we want to provide an explanation of the sort that Schelling provided for segregation, and with it some suggestions on how to increase integration. Abstracting from racism and other forms of in-group bias, Schelling showed that segregation is *cheap*, it can emerge from very weak conditions. In a similar vein, we intend to show that online tribalism can emerge from mild micro preferences, even when we abstract the well established human tribal tendencies.

The model hopes to resemble the kind of dynamic a user of social media would find familiar. Agents are part of a social network in which they are connected with other agents. They also have some agency with respect to whom they connect: they can befriend or follow other agents, and they can also sever their links - very much like in Twitter, Facebook, Instagram, etc. More

precisely, directed networks in which agents decide whether follow others are akin to Twitter or Instagram. Undirected networks in which connections are reciprocal resemble Facebook's friending and unfriending options. At each stage of the dynamic, each individual can assess their neighbors' (friends or following) tribe. Much like in Schelling, agents will be satisfied if a sufficient amount of their neighbors are of their type. If they are not satisfied, they will (randomly) sever some of their links and (randomly) seek for new friends to follow. This dynamic on the network stabilizes when all agents are satisfied, or when cycles emerge. How *segregated* or *tribal* a network is depends on how much agents of one tribe are connected with agents of the other. In highly polarized networks, members of different tribes do not communicate enough, which might lead to misunderstandings, information bubbles, empathy gaps, etc. The purpose of this work is to assess how *cheap* this kind of polarization is, under what conditions it emerges, and up to which degree.

2. Model and Simulation Experiments

We implemented our model by (quite substantially) modifying Netlogo's version of Schelling's model (Wilensky, 1997). The original model is defined over a grid and it was here adapted to run over networks. The study was done both with directed and undirected networks. In the latter, agents have reciprocal relations they can sever - representing Facebook's "unfriend" option -, but they can also create friendships with other agents without their consent - unlike Facebook. In directed networks agents can only create or sever outgoing connections, representing those who they follow and therefore appear in their feed - very much like in Twitter or Instagram. The initial networks are Erdös-Rényi random networks $G(N, p)$ with N nodes and where each edge has an independent probability p of appearing.[2] For all the simulations we took $N = 100$ and the parameter p sweeping across [0.05, 0.10, 0.15]. Given a node i, $N(i)$ is its set of neighbor. In the case of directed networks, $N(i)$ is the set of nodes i points to. Some initial network properties

[2]Just to clarify, in directed networks and edge from node a to a node b is distinct from an edge from b to a and both appear with independent probability p.

can be expected by the link probability p: density[3], global clustering coefficient, and average path length.

Agents belong to one of two tribes, modeled by a color tag (blue and red), so that for agent i, $T(i) \in \{B, R\}$. The model allows for up two five different tribes, but the experiments here were done with just two for simplicity. The total number of agents was fixed in one hundred, but the population of the blue tribe was allowed to vary in each simulation with a sweeping parameter of [25, 50, 75]. For example, suppose the population of blue is fixed at 25, then 25 nodes are selected uniformly at random and colored blue, while the rest is colored red.

In Network Science, segregation and tribalism are connected with the notion of homophily. Homophily is the tendency to associate with those that are considered similar McPherson et al. (2001); Currarini et al. (2016). Heterophily, in contrast, is the tendency to value diversity and associate with that which is different. In the model, homophily [Σ] and heterophily [Δ] preference thresholds are global parameters shared by all agents.[4] The homophily threshold sweep was [20%, 25%, 30%, 35%, 40%, 45%], and the heterophily was [0%, 15%, 30%, 45%]. Each agent i has similarity measure σ_i and a diversity measure δ_i given by the proportion of neighbors of the same and different type:

$$\sigma_i = \frac{\#\{j \in N(i) : T(j) = T(i)\}}{\#N(i)}$$

$$\delta_i = \frac{\#\{j \in N(i) : T(j) \neq T(i)\}}{\#N(i)}$$

[3]The density of an undirected network is the amount of actual connections $[\sum_i \#\{j \in N(i)\}]$ over the amount of possible connections $[\binom{N}{2} = \frac{N*(N-1)}{2}]$. For a directed network, it is that number divided by two, since in the original definition each undirected link counted twice - once for each of the nodes it connects.

[4]One might expect that such preferences may vary across groups, and such an implementation can be easily coded in further experiments. The original Netlogo implementation assigns to each agent i a personal homophily Σ_i which is a random number in [0,Σ), and similarly for heterophily. Since I do not think this constitutes any real conceptual improvement, while it adds computational demands, I required all agents to have the same thresholds.

Each network $G(N,p)$ also has a similarity (σ_G) and a diversity (δ_G) measure. How tribal or segregated a network is was measured using σ_G, which is defined by the proportion of similar neighbors across the network over the total number of neighbors[5], while δ_G is just $1-\sigma_G$:

$$\sigma_G = \frac{\sum_i \#\{j \in N(i) : T(j) = T(i)\}}{\sum_i \#N(i)}$$

At each stage of the dynamic, an agent i is happy in case $\Sigma < \sigma_i$ *and* $\Delta < \delta_i$.[6] A network's happiness H_G at a certain stage is defined by the proportion of happy agents over the total amount of agents - while the unhappiness parameter U_G is just $1-H_G$. Happy agents do not change their set of neighbors, but unhappy agents do. We study two dynamics. On the one hand, unhappy agents randomly sever one of their (outgoing) links and randomly seek for one other friend (to follow). On the other hand, unhappy agents sever *all* of their links and they randomly seek friends until they have the same amount of connections as before. Some observations are in place here. First, much like in most of Schelling's original models (Schelling, 1971a, 1969, 2006), agents are not actively seeking for locations (or friends) that satisfy their homophily/heterophily threshold preferences. In our case, they not even sever connections selectively. Rather, they randomly sever and then randomly explore the space of possibilities to hopefully stabilize whenever those preferences are satisfied. The implicit hypothesis is that if segregation or tribalism emerges from random procedures, we can be sure that it will emerge from more selective micro behaviors.[7] Second, the purpose of the sever one vs. sever all dynamics is to assess variation in tribalism degree and speed of convergence. It is clear that real agents do not sever all their connections at once. Third, both dynamics ensure that the total amount of links in the networks is constant (hence density is constant). Furthermore, in the case of directed networks, agents always have the same amount of outgoing links. The idea here is to study how the dynamic affects the network structure, in

[5] In the implementation, this is further multiplied by 100 so to obtain percentiles.

[6] In the case that $0 < \Sigma, \Delta$, I assumed for simplicity that nodes with a single neighbor are happy.

[7] Another natural dynamic would be for agents to be able to connect *only* with friends of current friends.

particular whether clustering and average path length change.[8] For this reason, global clustering coefficient and average path length were measured at the start and end of each of the experiments.

For each point in the parameter sweep space, we ran 100 simulation repetitions, for a total of 86400 observations. Each simulation was run either until all agents were happy, or until 500 time steps passed. The tables below summarize the sweep parameters, and dynamic and output measures just explained:

Directed Network	**Probability p**	**Blue Population**	**Homophily Threshold**	**Heterophily Threshold**	**Sever All?**
true, false	0.05, 0.1, 0.15	25%, 50%, 75%	20%, 25%, 30%, 35%, 40%, 45%	0%, 15%, 30%, 45%	true, false

TABLE 4. Parameter sweeping space

Dynamic Variables for Agents	**Computation**	**Range**
Neighbor similarity (σ_i)	$\frac{\#\{j \in N(i) : T(j) = T(i)\}}{\#N(i)}$	[0,1]
Neighbor diversity (δ_i)	$\frac{\#\{j \in N(i) : T(j) \neq T(i)\}}{\#N(i)}$	[0,1]
Happiness	$\Sigma < \sigma_i$ and $\Delta < \delta_i$	true, false

TABLE 5. Variables for Agents

For each simulation, the values of each of the sweeping parameters were recorded. Furthermore, for the table below, initial and final measures were recorded for all variables except those which are constant (link amount, network density, and average degree).

Global Output Measures	**Computation**
Number of Links	Trivial
Network Density	$\frac{\sum_i L(i)}{990}$ (*2 for undirected networks)
Average Degree	$\frac{\sum_i L(i)}{100}$ (*2 for undirected networks)
Global Clustering	The number of closed triplets in a network divided by the total number of triplets
Mean Clustering	Mean of local clustering among all agents
Path Length	Average shortest-path length between all distinct pairs of agents in the network
Network Segregation (σ_G)	$\frac{\sum_i \#\{j \in N(i) : T(j) = T(i)\}}{\sum_i \#N(i)}$
Network Unhappiness (U_G)	Proportion of unhappy agents over the total

TABLE 6. Global Measures Recorded

[8] It would be more interesting to study whether there are changes in degree distribution, and whether social hierarchies emerge from this dynamic in the same way as they emerge from the Barabási-Albert preferential attachment algorithm.

3. Results

3.1. The Schelling Dynamic Induces Segregation.

Schelling's original models had no heterophily threshold, so we will start by taking a look at the data corresponding to those simulations.

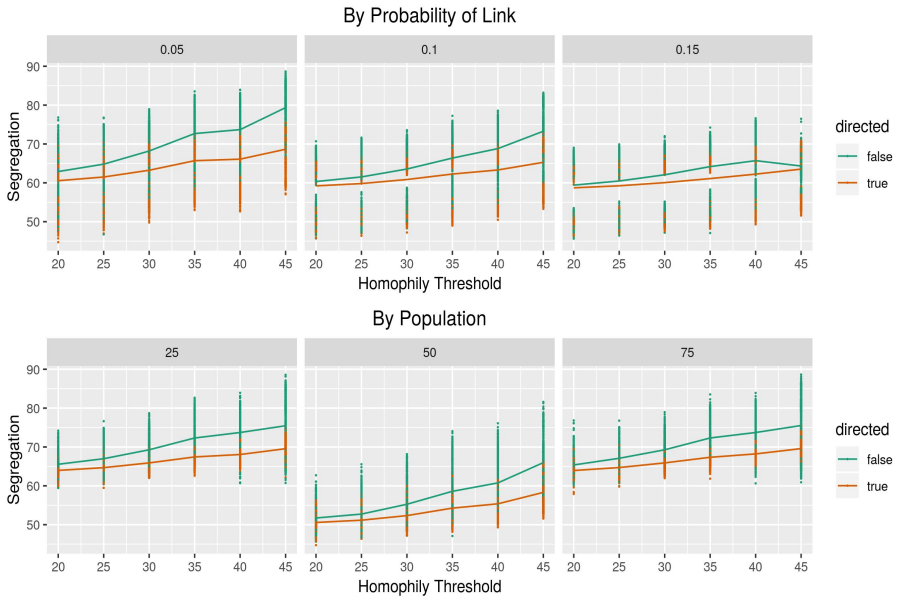


FIGURE 3. Segregation vs Homophily without Heterophily

Figure 1 displays the increase of segregation as we increase the homophily threshold when there was no heterophily required. Data is partitioned according to link probability, population, and whether the network is directed or not. The lines correspond to observation means, and the dots to observation instances.

To begin, imposing different homophily thresholds did affect how much segregation was obtained. For example, when link probability $p = 0.1$, a homophily treshold of 30% led to an average segregation of 60% across all populations for directed networks, and about 63% for undirected. And when blue and red populations were equal, a threshold of 40% led to a segregation of 55% and 60%.

A noticeable feature is that undirected networks were in general more affected by homophily than directed ones. This pattern will emerge systematically, and developed later. Different link probability (density) also had an effect. Notice that as it increases, directed and undirected network mean lines get closer. Finally, the shape of the mean curves across populations is similar, but their starting point is different. This is because different proportions of blue-red will induce different base rate segregation at the initial state. In a society where the major type is a fraction x of the total population, the baseline segregation in random initial conditions is: $x * x + (1 - x) * (1 - x)$. If blue is in the majority and $x = \frac{3}{4} = 0.75$, then each blue individual can expect 75% of their neighbors to be blue, and each red for 25% of their neighbors to be red. Hence for our cases, the expected baseline segregation for 25% and 75% blue populations is of 62.5%, and when population are equal it is expected to be of 50%. Thus, it is misleading to consider *absolute* segregation values. Rather, Figure 2 presents segregation *difference*, namely how much segregation changed between the initial state and the final after the implementation of the dynamic. From now on, we will stick to segregation difference as a measure for tribalism.

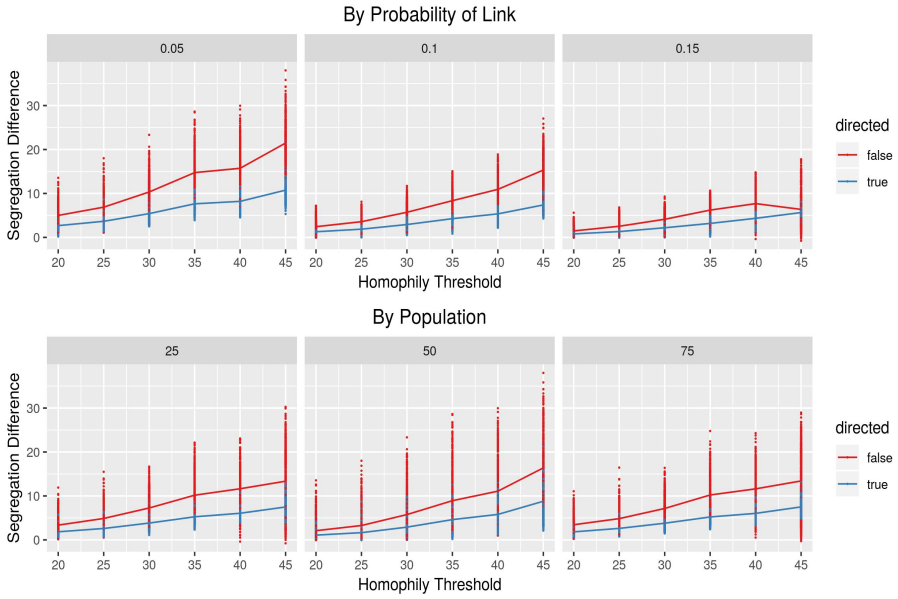


FIGURE 4. Segregation Difference vs Homophily without Heterophily

The observations made before still hold. Undirected networks are more suceptible to the dynamic, an increase in density (link probability) reduces the effect of the dynamic, but differences between population proportions do not.

Now is the time to make a central observation. When the population is equal, the expected base segregation at initiation is of 50%. For all the homophily thresholds considered here, agents are expected to be happy with that initial segregation. In the long run, factoring out random fluctuations in initial conditions by taking large data sets, networks with equal population should have all of their agents happy *at initiation* when the homophily threshold is less than the fixed baseline segregation of 50%. Yet, the dynamics of repeated relinking triggers the evolution of an equilibrium configuration with a positive increase in segregation. For example, for a threshold of 30%, there was a increase in mean segregation of about 5% for undirected networks, and of 3% for directed. Let us take a closer look to this phenomenon in Figure 3:

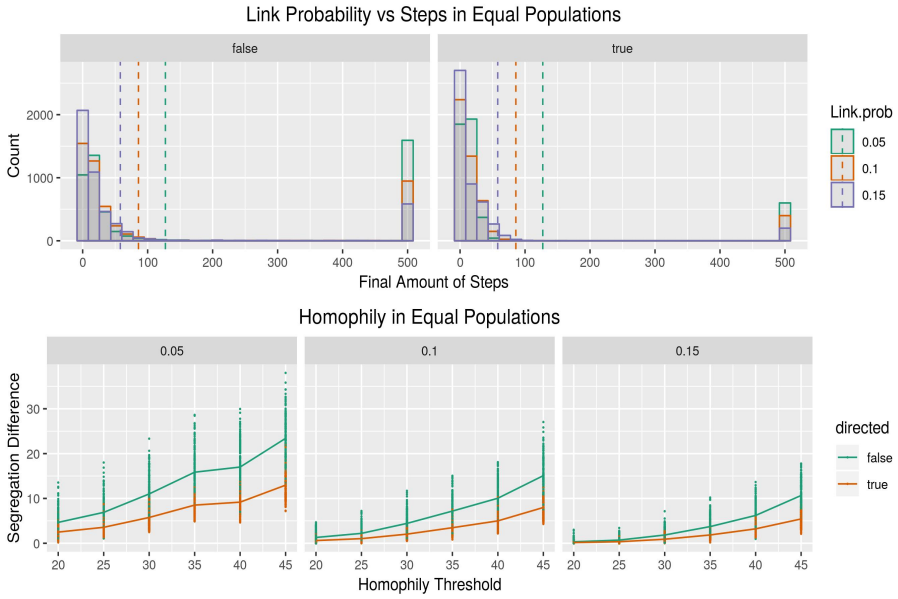


FIGURE 5. Segregation Difference and Speed vs Homophily in Equal Populations

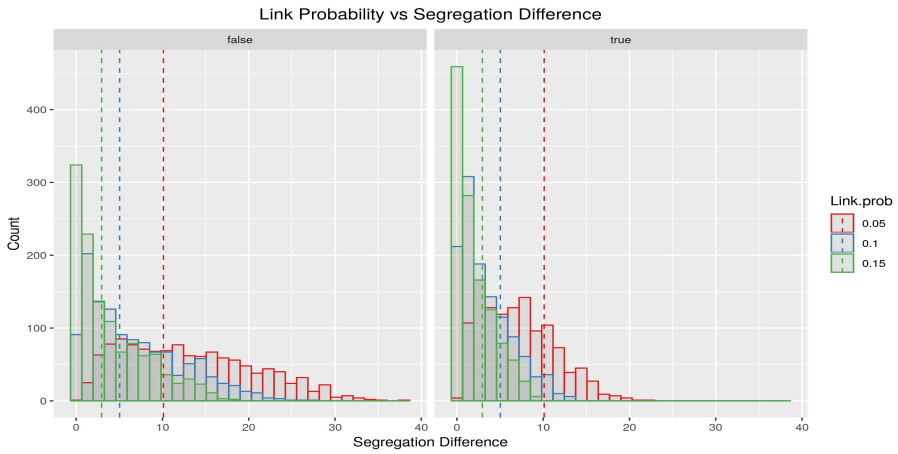


FIGURE 6. Segregation Difference Densities (Equal Populations and no Heterophily)

An increase in density led to an increase in speed of conversion. The dashed lines in the plots above show the mean of steps for both the undirected and directed networks (false true), and for different link probabilities (i.e. expected densities). The higher the proportion of connections, the more likely is that an individual agent's neighbors will express the initial baseline segregation. Furthermore, slight changes would be sufficient to reach equilibrium. Conversely, low density networks will exhibit more agents with not enough connections to secure their desired homophily, so that the dynamic does not converge and it is forced to stop after 500 steps.

Density also affected how much segregation was obtained by the dynamics, as observed in the plots below of Figure 3. With a base segregation of 50%, a 30% homophily threshold should be easily satisfied. But with low density networks this led to a mean increase in segregation of 10% for undirected networks and 5% for directed. To explore this, let us take a closer look at the segregation difference distribution with respect to different densities and network types in Figure 4 on the next page.

In Figure 4, the mean segregation difference for both directed and undirected networks (false true) with a density of 5% was of ten points, while for more dense networks it was below five, and decreasing as the network has more connections. Furthermore, density had an effect on variance.

Across network with 5% density (directed and undirected), the variance of the segregation difference was about 46, while for a 10% density it was about 22, and a density of 15% it was about 12.

To conclude, the results clearly showed that micro-motives created a macro-behavior: homophily thresholds induce increasing segregation. Our Schelling dynamics may explain some of the segregation we observe in Online Social Networks.

It is worth pointing that the effect of the dynamic may *seem* minor. After all, an increase in mean segregation of about 5% or 3% in equal populations across all networks could be disregarded. But this is a misleading conclusion. As Figures 2-4 show, the effect of micro behavior heavily depends on the density of the network, with lower densities leading to more segregation effect (although slower). For a 5% density the effect was of a 10 and 5 point increase for undirected and directed networks respectively. But a 5% density means that on average each agent is connected with 5% of the population. In a community of 100 individuals, each agent would have 5 friends; in a community of 1000, 50. In an Online Social Network of one billion individuals, it means that each agent has on average 50 million friends; an absurd assumption. According to *Pew Reserch*, in 2013 the average Facebook user had 338 friends, and the network had about 1.2 billion active users total. This corresponds to a density of about 0.00003%. The segregation effect of micro motives in networks with such a low density would be extremely large.[9]

3.2. Heterophily beats Homophily in Unequal Populations.

If the reader is concerned about the social implications of results of the previous section, do not despair. Segregation induced by agents' preferences for their own group can be effectively countered by praising diversity. In this section we will study the interaction between homophily and heterophily, love for what is different.

[9]Unfortunately, computational limitations restrict the possibility of running experiments with individuals in the thousands, much less millions, and very low densities.

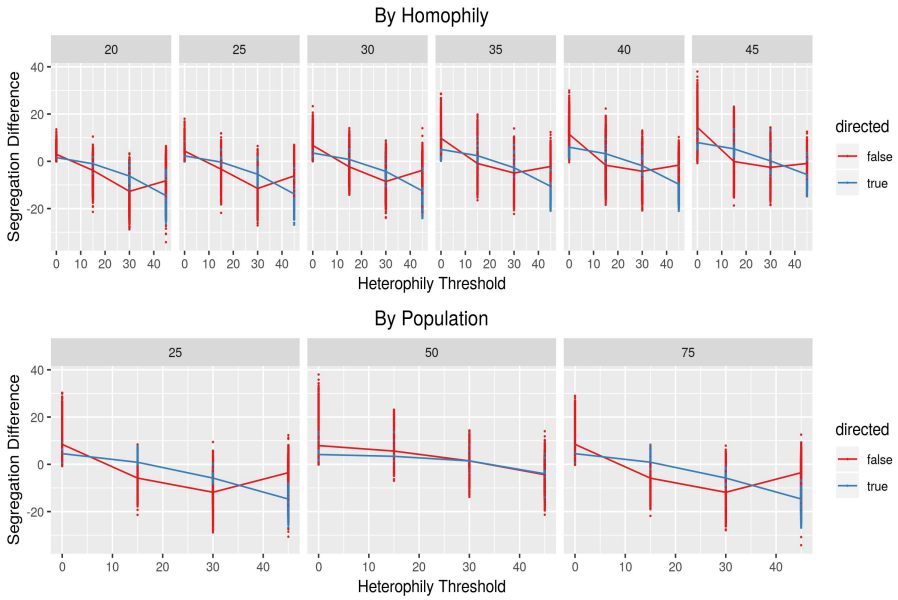


FIGURE 7. Heterophily vs Homophily on Segregation

The top plots in Figure 5 next page present the segregation difference induced by the heterophily threshold for different values of the homophily threshold. As it is expected, heterophily reduces segregation in all cases for both directed and undirected networks, with a slightly higher effect for directed networks. The second thing to observe is that as homophily increases, the base segregation difference increases too. When the homophily threshold is in 20% (top left), heterophily reduces segregation from about 2 to around -10 points; but when homophily is 45% (top right), this goes from about 10 to about 5. Yet the heterophily graphs are *roughly* similar in *shape*. This indicates that the interaction between homophily and heterophily in the dynamics is *roughly* linear. Although this will prove to be false in not long, it is a helpful heuristic. Let us call the "Philiac Difference" of an experiment the difference between the homophilic and heterophilic thresholds. This will become handy soon.

The bottom plots of Figure 5 reveal something even more interesting. It seems that for unequal populations the effect of heterophily was larger than for equal populations. Let us take a closer look, using the Philic Difference as a parameter in Figure 6.

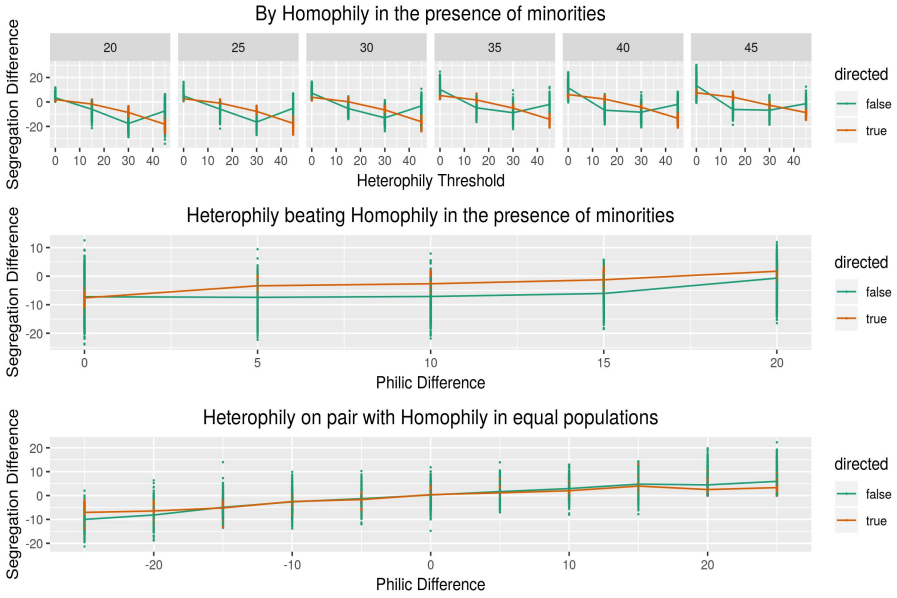


FIGURE 8. Heterophily Beats Homophily

When populations are unequal, and therefore there is a minority, heterophily and homophily have different power. It will be easier for the majority to satisfy their homophily than their heterophily, and the converse for the minority. But since segregation is measured considering everyone equally and independently of their group, that which affects the majority more will have a larger effect overall.

The top plots in Figure 6 show that for all homophily thresholds, a small amount of heterophily is sufficient to induce a segregation reduction - namely integration. The middle plot exemplifies that even when homophily is larger than heterophily by a margin of 0 to 20, integration is still obtained. Interestingly, once again directed networks are more affected by the interaction than undirected ones. The bottom plot shows that in equal population proportions, both threshold demands have the same force. In particular, when the Philic Difference is of zero, no segregation nor integration was observed.

The observation is hopeful: In the presence of minorities, convincing individuals of the value of diversity will have a greater effect on integration than emphasizing their in-group preference has on segregation.

3.3. Heterophily makes us Happier in Unequal populations.

So far we explored the effects of homophily and heterophily with respect to segregation, but not with respect to happiness. Following the trend in the previous section, a praise for diversity is correlated with higher increase in happiness for unequal populations. Here "Happiness Difference" is the difference between the final proportion of satisfied individuals and the initial proportion.

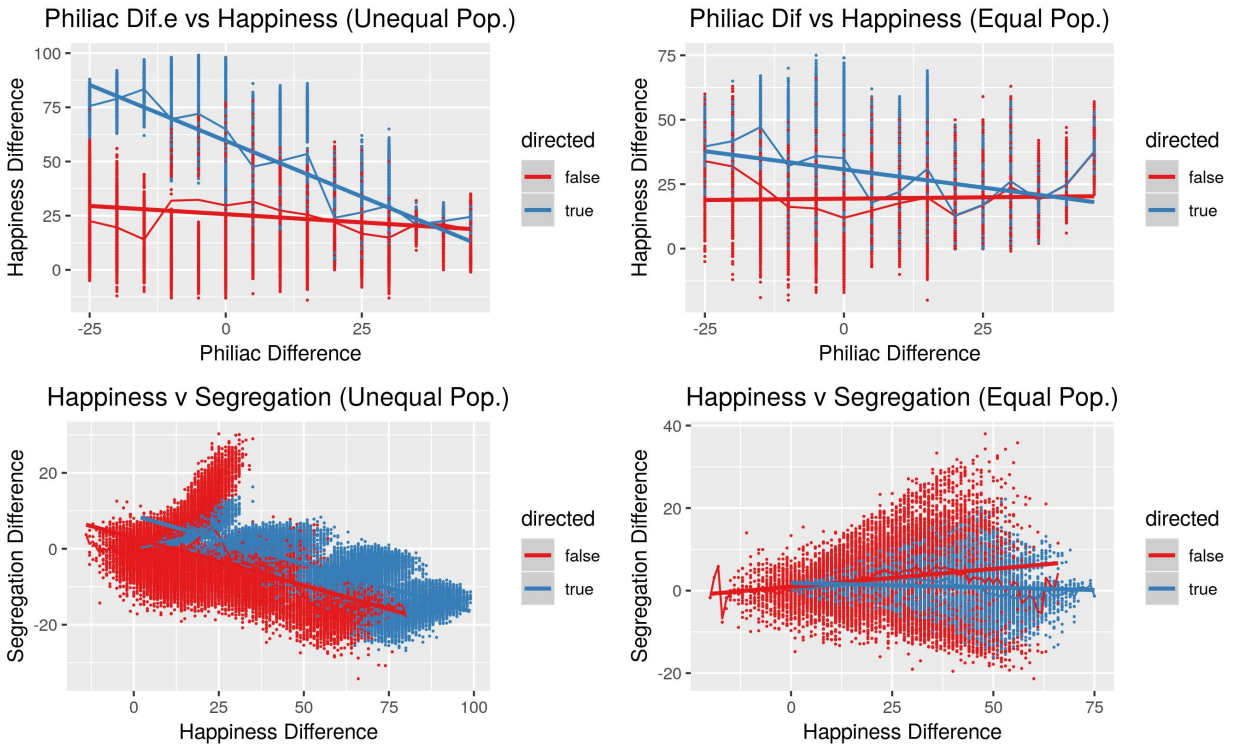


FIGURE 9. Happiness

The plots on top of Figure 7 show that in almost all cases there was a gain in happiness after the dynamics was implemtented (i.e. the happiness difference is positive). More interestingly, as the philiac difference increases in unequal populations, namely as they become more homophilic, the gain in happiness decreases. This is particularly pronounced in directed networks. The straight

lines are regression lines, while the others are the means. The effect is also observed in directed networks with equally distributed populations. For an explanation of this, consider a network with unequal populations. At the initial state, its harder for individuals in the majority to satisfy their heterophily, and for individuals in the minority to satisfy their homophily. Heterophily has a greater effect on how (un)happy the overall population is, relative to the effect that homophily has. As the dynamic is implemented, there is more happiness gained by satisfying heterophily than by satisfying homophily.

The plots at the bottom of Figure 7 are also informative. First, in unequal populations happiness is correlated with *integration*. This once again depends on the fact that heterophily has more effect when there are minorities. On the other hand, in equal populations happiness induces Heteroscedasticity: as happiness increases the segregation difference variance increases. In the graph this is is shown by the fact that although the mean of the points is always around zero, as happiness increases they are more spread. This is because happiness can increase by satisfying homophily or/and by satisfying heterophily, and these different ways will translate in different degrees of segregation increase and reduction.

Again we encounter an encouraging conclusion. Celebrating diversity, i.e. heterophily, not only is more powerful than celebrating homogeneity, it also makes people happier. There is hope for integration.

3.4. Interaction between Homophily and Heterophily.

In developing the last two points, we assumed that philiac difference was a good way of measuring the interaction between homophily and heterophily. This was a helpful heuristic to represent some interesting phenomena, but it is not entirely correct. To be clear, the points made in the last subsection hold: heterophily beats homophily in unequal populations both in respect to segregation

and happiness. But the graphical representations were *slightly* misleading. To observe this, let us look at two data sets. On the one hand, we go back to the set with no heterophilia, but where the homophily thresholds were just 20%, 25%, and 30%. On the other, let us look at the data sets in which the philic difference has those values.[10] If the relation between homophilia and heterophilia with respect to segregation difference were linear, these two data sets should look relatively alike, but they do not.

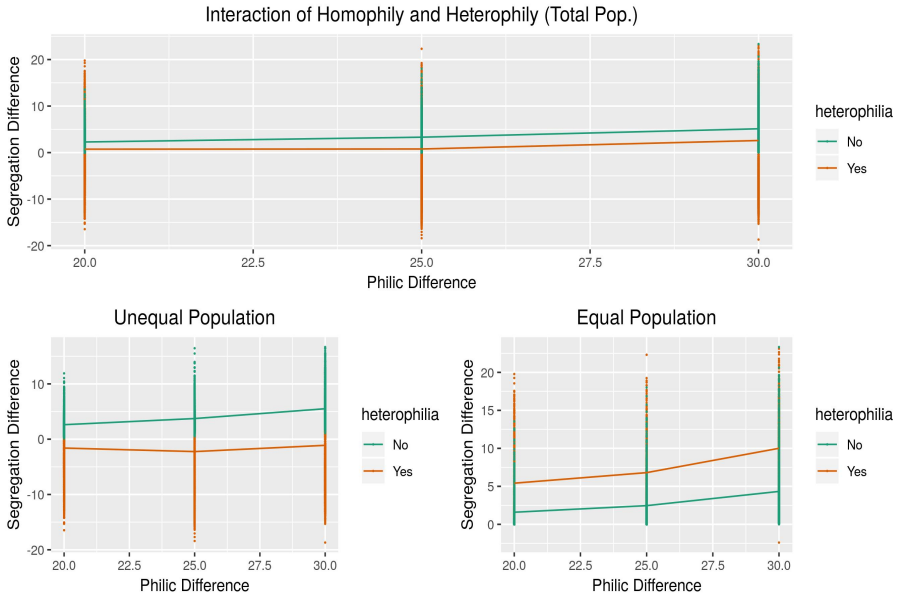


FIGURE 10. Non Linear Interaction

These figures show, for example, that its not the same to have (a) a 25% threshold for homophily with no heterophily, and (b) a 45% threshold for homophily and a 20% threshold for heterophily; even in equal populations. For example, in the bottom right plot, case (a) induced a mean segregation increase of about 2.5% (across all networks), while case (b) induced about 7%. The fact that such a difference manifests even in the case of equal populations is surprising, and suggests that the relationship between the two constraints is more complicated than expected.

3.5. Direction and Relink Dynamic.

[10]Furthermore, we disregarded the data with no homophilia.

Now we take a closer look at the differences between directed and undirected networks, and also briefly reflect on the two different dynamics presented (with a single random relinking, or a complete relinking).

We already observed that in general the undirected networks were more affected when there was no heterophily, but directed networks were more affected when there was [Figures 1-7], and that the relation between happiness and philiac difference was larger on directed networks. Let us now look at the segregation difference and steps distributions (by link probability).

On the left of Figure 9 next page, means appear to be close to each other. This might be due to the fact that undirected networks seem to be more affected by homophily, while directed networks by heterophily [i.e. Figures 1-7]; and when taking means across all values their mean segregation evens out. A glance at the segregation distribution suggests that directed networks are more concentrated around the mean. This can be verified by looking at the variance: The segregation difference variance for the full data set is 66.66, while for directed networks is 54.55, and for undirected is 78.41. On the other hand, directed networks were clearly faster in attaining an equilibrium, and were substantially more successful also. Many undirected networks fail to achieve a stable equilibria, as is shown by the fact that almost half of them lie within 500 steps - meaning that the dynamic was coerced to end even when there were unhappy individuals.

Once again, Figure 10 shows that the segregation difference means are almost the same for the relink all vs relink one dynamics. This may be misleading in that properly partitioning the data could show that the different dynamics have a significant effect on segregation and integration. But to avoid boring the reader with more graphs, this study will be left for another time. More relevantly, different dynamics do have an effect on the variance. The segregation difference variance for the relink all dynamic is 72.83, while for relink one is 60.12. On the other hand, unexpectedly relink all had a *slower* mean rate of convergence to equilibrium than relink one; and no significant difference in equilibrium failure is observed.

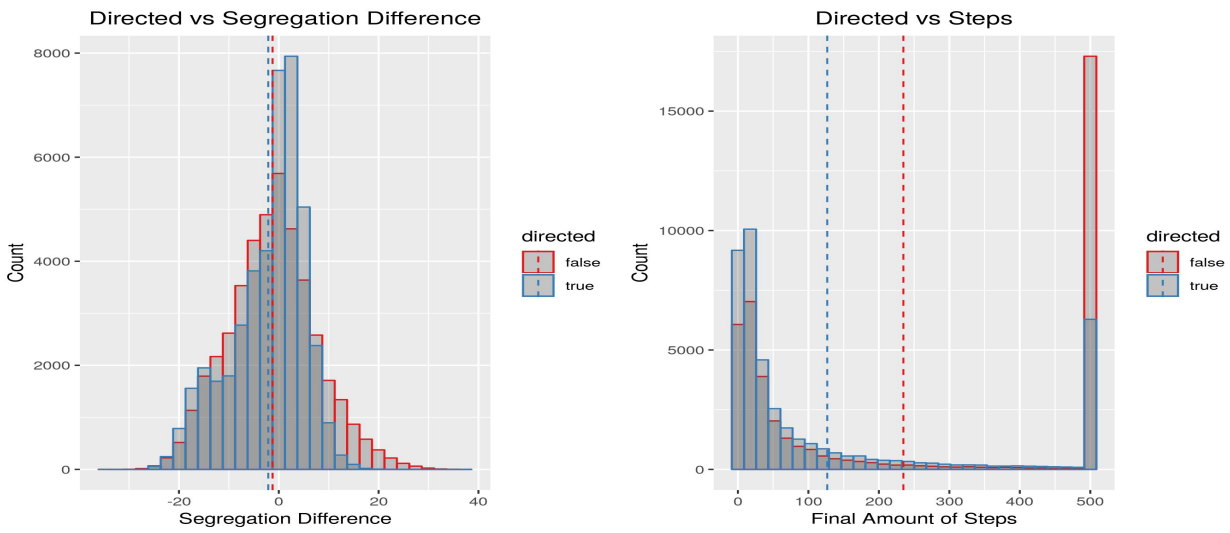


FIGURE 11. Directed vs Undirected

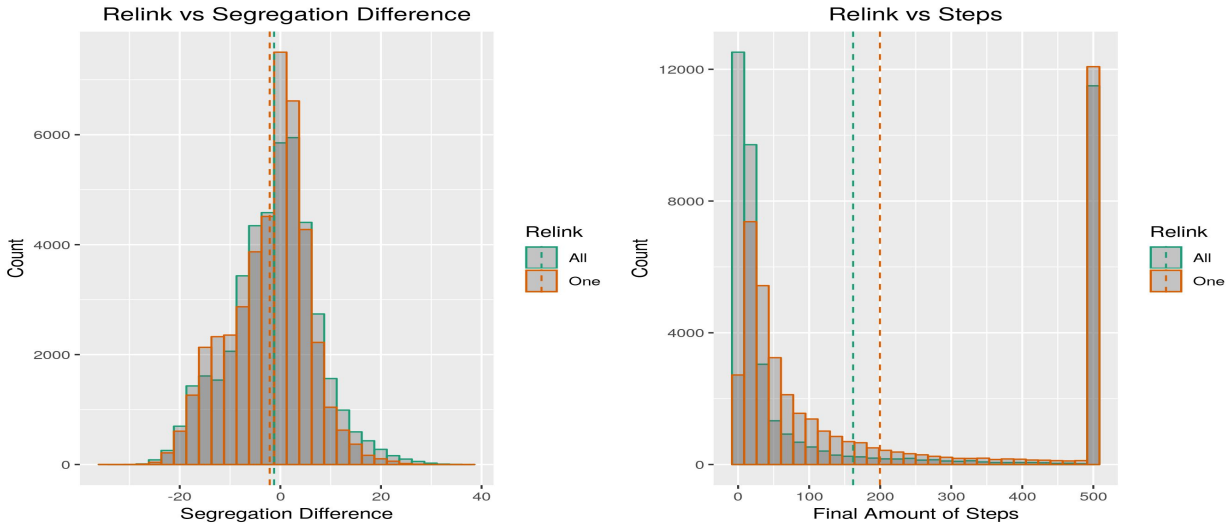


FIGURE 12. Relink One vs Relink All

To sum up, undirected networks in which connections are reciprocal have a higher variance and segregation effect but lower convergence speed, relative to directed networks. Whether the relinking was done one at a time or not also had an effect on variance, and speed. Relinking all was faster but with a larger spread.

4. Conclusions and Discussion

Schelling (2006, 1969, 1971a) offered a new type of explanation for macro social arrangements in terms of individual preferences and micro behavior. The upshot of the study was that geographical ethnic segregation can emerge even when agents have minor preferences for their own group. The first purpose of this essay was to show that the same kind of explanation can be given for segregation in the context of Online Social Networks.

Communities were represented via Erdös-Rényi random networks $G(N, p)$ in which individual agents satisfy their homophily and heterophily thresholds by severing and creating links with other agents. Like in Twitter or Instagram, we studied directed networks where individuals can follow or unfollow other agents. Like in Facebook, undirected networks have reciprocal connections or friends. The results show that imposing a Schelling-like dynamic on these networks generates segregation and integration, according to different parameters.

In general, undirected networks were more affected by homophily and directed networks by heterophily, and the latter had higher speed of convergence and less variance. More importantly for our explanatory purposes, differences in network density affected how the dynamic changed segregation, with lower densities implying higher effect. The density of a network is the amount of actual connections over the amount of total connections, and this was studied by looking at the link probability p, which corresponds to the expected density. The lowest valued studied was 0.05, corresponding to an expected density of 5%. In this case, results show that the dynamic induced a non negligible segregation increase of up to 23 points (in undirected networks with equal populations and no heterophily). As pointed before, real life Online Social Networks have a *substantially* lower densities, which suggests that the effect of a Schelling-like dynamic will be much

larger. Unfortunately, computational limitations restrict the study of more representative sizes.[11] Finally, although lower densities increased the effect, they also slowed it down.

We also studied dynamics with heterophily thresholds in which agents enjoy diverse company. It was encouraging to observe that heterophily had a greater effect than homophily in populations with minorities, leading to positive integration. This is because when there are minorities the majority is affected more heavily by heterophily than homophily. Relatedly, the results show that populations that prefer diversity end up more happy, in particular when there are minorities and networks are directed. Furthermore, happiness *increase* is correlated with integration in unequal populations. Both these results are a natural consequence of the first observation: heterophily beats homophily in unequal populations.

The model presented here can be modified and adapted in different ways. Higher computational power would be useful in studying more representative sizes. Agents' types may be increased to more than two - in fact the Netlogo model attached to this essay allows for uniform distribution of up two five tribes. Homophily and heterophily thresholds may not be universal, allowing them to vary across different groups - it is not hard to imagine that some tribes are more prone to diversity than others. Severing connections and finding new ones may not be random, but maybe involving in and out-group biases. Agents may not easily identify the tribe of their neighbors. Initial networks need not be random Erdös-Rényi. Feasible alternative are Barabási-Albert and Watts-Strogatz networks. Hierarchical networks like those generated using Barabási-Albert preferential attachment algorithm are more representative of actual Online Social Networks because their degree distribution follows a power law (Pareto distribution) in its tail (long tails, with a few nodes concentrating most of the connections). Alternatively, Watts-Strogatz networks have short average path length[12] but,

[11] Models $G(N, p)$ with N=100 and significantly lower densities will almost always fail to converge. Models with larger N, even in the few hundreds, are hard to compute.

[12] Watts-Strogatz networks have the small world property according to which the typical distance between two randomly chosen nodes grows proportionally to the logarithm of the number of nodes N in the network. This captures

unlike Erdös-Rényi, they have high clustering. Conversely, and maybe more interestingly, natural Schelling-dynamics like the ones introduced here may be defined on networks so that they explain the distribution and segregation of *actual* internet communities. A minor step in this direction is carried away in the Appendix, where we present the effect that the dynamic had on network structural properties like clustering and average path length.

Tribalism and segregation in Online Social Networks is a well documented phenomena which, very much like geographical racial segregation, can have negative social effects. The results obtained in this study are both disappointing and encouraging. Given the low density of online networks, even very low homophily thresholds may lead to high degrees of segregation. Nevertheless, the results also show that inducing heterophily, in particular with regards to minorities, can very well counter the segregation effects of homophily and generate integration. Furthermore, it might increase overall population happiness.

These results can be taken as informative towards policies, or towards ways of improving our social relations online. For example, divisive speeches and narratives may increase homophily since individuals are forced to take a stand on the matter, identify themselves with a tribe and signal belonging. This paper shows how easy is to generate segregation from low levels of homophily, and therefore a cultural non-partisan effort to reduce generalizing and divisive outbursts would contribute to integration - provided that is what parties are seeking. More moderate discourse may even make it harder for agents to identify others as belonging to certain tribes, a capacity that we assumed here. Furthermore, the model shows that praising diversity and heterophily in the presence of minorities has a strategic advantage over homophily and leads to integration and

---

the *six degrees of separation* intuition in ER. For those networks it can be shown that for large n, average path length and diameter (largest shortest path) are approximately proportional to log(n)/log(d). If n = 6.7 million, and each person knows around 50 people so that d = 50, log(n)/log(d) is precisely about 6.

satisfaction. The new digital technologies are creating new forms of socialization, and with them new problems emerge. This paper is trying to understand and address one of them.

CHAPTER 7

Social Centrality in the Contagion of Rumors

> The danger to society is not merely that it should believe wrong things, though that is great enough; but that it should become credulous, and lose the habit of testing things and inquiring into them; for then it must sink back into savagery.
>
> Clifford, *The Ethics of Belief*

n

1. Introduction

The spread of rumors is a very interesting practice from an epistemological perspective. One of the earliest and most famous treatment of them is due to Clifford (1879), "The Ethics of Belief." Clifford was concerned with the evils brought by *credulity*, the habit of endorsing believes without sufficient justification. His dictum is memorable: "it is wrong always, everywhere, and for anyone, to believe anything upon insufficient evidence." His central argument is fundamentally *social*.

For Clifford, personal beliefs are not a private matter, but they concern society and even humanity at large.[1] Second, he is concerned with epistemic practices and virtues rather particular cases of belief or disbelief. For him, what is more worrisome is the development of a credulous *character*, "when a habit of believing for unworthy reasons is fostered and made permanent." Third, Clifford argues that such epistemic character flaws can be adopted by others or society at large, which brings us back to the epigraph. Here is where *rumors* come into play. A credulous character not only endorses propositions without sufficient evidence, but they also communicate them to others,

[1] He states:

> Belief, that sacred faculty which promts the decisions of our will, and knits into harmonious working all the compacted energies of our being, is ours not for ourselves but for humanity.

How this view would play with the question of censorship is unclear. But Clifford seems to be espousing personal freedom to belive but moral and social responsibility to believe with evidence.

sometimes even making use of their epistemic authority or some enforcing mechanism. This makes others credulous, and the disease continues to spread. Finally, the worse of all evils. As society becomes credulous, it becomes manipulable and susceptible to deception and *fake* information: "It may matter little to me, in my cloud-castle of sweet illusions and darling lies; but it matters much to Man that I have made my neighbours ready to deceive. The credulous man is father to the liar and the cheat."

Such cliffordian worries about the evils of rumors may be reasonably exacerbated by the rise of the Internet and digital social media. Mößner and Kitcher (2017) discuss some of the epistemic problems that emerge from what they identify as the democratization of knowledge and information through the internet. The question whether the internet *democratizes* information, in the sense that the production and consumption of knowledge or data is more evenly distributed in the overall population, will be discussed later. More related to our concern with rumors, Mößner and Kitcher (2017) discuss the concept of epistemic *opacity*. They attribute this property to *sources* of information: "A source is opaque for a seeker when the seeker cannot apply the markers available as to vouch for the reliability of the source." Sources can be opaque for different reasons. Anonymous sources are usually opaque, but also unknown news outlets or even academic journals. More generally, the strategies that agents developed to reliably assign epistemic authority are not useful or applicable for the source at hand. This does *not* mean that the sources are *untrustworthy* according to those standards, but that they cannot be evaluated by them.

As I see them, rumors are *opaque* in a different but related way. While the opaqueness of a source has to do with how reliable it is, the opaqueness of a rumor has to do with its corroboration. For different reasons, it is hard for agents to deploy strategies to assess the veracity of a rumor. One recent example of rumor that circulated online was the claim that '600 Murders Took Place in Chicago during the second weekend of August 2018', originated in a television show and sparking

anxiety about such a large number of violence in the city. This was later *factchecked* as false [the reported amount for that city in that weekend was just one]. But at the time the rumor began, official statistics were not released and there were no easy sources for falsification.

In the face of such rumor opacity, Clifford suggested suspension of judgment, since there is not enough evidence for its endorsement. But Clifford's core value was *truth*, and there might be other concerns at play when it comes to rumors, namely *justice*. Consider the recent the #MeToo movement. Following the story, it could be argued that initial public outbreak of the Weinstein atrocities was precisely the result of the spread of rumors about him. Furthermore, soon after many lists of alleged sexual predators were distributed online, and in the face of them our strategies of corroboration are either useless or too slow. This is not to say that the allegations are false. The point is that Clifford's suggestion of skepticism and suspension of judgment in the face of insufficient evidence may be trumped by other concerns. Following James (1979) "The Will To Believe," some beliefs need to be pursued *first*, even with insufficient evidence, in order for them to turn out to be true. Rumors' opacity may not be completely bad after all. Rumors may be a way of speaking truth *behind* power.

Last paragraph argued that in some occasions rumor opacity can be defended on the grounds of justice, which inevitable leads us to the issue *epistemic injustice* famously developed by Fricker (2007). The issue is complex enough to be treated fairly in this introduction. Suffices to say that Fricker identifies two important forms of epistemic injustices suffered by marginalized groups. *Testimonial injustice* occurs when a speaker is given less credibility than deserved on the basis of their identity. *Hermeneutical injustice* involves a structural prejudice in the economy of collective hermeneutical resources, so that marginalized groups and individuals may lack the conceptual frameworks that allows them to understand and interpret some of their experiences. Fricker exemplifies the latter with the concept of sexual harassment, and our recount of the #MeToo movement

may illustrate the first, since women spread rumors about Weinstein because their public statements were not given enough credibility.

Nevertheless, it may be reasonably argued that epistemically marginalized groups are particularly vulnerable to stereotyping or defamatory rumors. In the examples discussed, opacity had a strategic advantage in the pursue of justice, but this may not be generalizable. Returning to Clifford's intuitions, it is by no means clear that the *habit* or *practice* of spreading rumors would be conducive to justice in the long run. This is a serious issue that deserves examination, but will be left for another time.

Rumors present a further interesting case for epistemic injustice, since the social position that an agent enjoys in the communication structure may affect how successful they are in spreading rumors. This is the subject of the present essay. It is a study of how different social arrangements (egalitarian, hierarchical, and random) and the centrality that individual agents have within those arrangements affects the spread of rumors. This is done by providing a formal representation of social arrangements in terms of networks, and of rumors in terms of epidemiological models of contagion. The upshot will be that how successful agents are can be partially explained in terms of both properties of the networks and how central they are to those arrangements. But the results are surprising and interrelated, central agents are not always the most successful, and this varies across different networks. Whether network centrality is a form of epistemic privilege is still unclear, but this work is a step forward in answering the question.

The next section presents the technical details of the model. First, the epidemiological model of contagion that will be used as well as some other variants. Second, a detailed presentation of different social networks and centrality measures. Section 3 presents the results of the study. The last section is conclusive.

2. Formal Models of Rumor Contagion

The past few years saw a significant increase in the computational study of rumors, most likely due to the emergence of *fake news* in the context of Online Social Networks. Shelke and Attar (2019) is an excellent recent review paper that details many techniques, in particular for rumor *source detection*. Three formal features will be presented here. First, as explained in Shelke and Attar (2019), the spread of rumors is usually modelled using an epidemiological framework. Much like diseases, there is a *contagion* of rumors. Second, rumors circulate in different networks which may have a variety of properties. Here we will organize the study by considering broadly egalitarian, hierarchical and random networks. Finally, we will call 'influencers' those individuals that originate the rumor. We are interested in assessing how the network centrality of the influencers affect how successful they are in spreading it. For this, several notions of network centrality will be used and explained below.

2.1. The Epidemiology of Rumor.

Epidemiology is the study of the distribution and determinants of health-related states or events, and the application of this study to the control of diseases and other health problems. Two of the most famous mathematical models of disease spread developed in the field are the SIS and SIR models. Their basic idea is the following. We start with a population of N individuals, each of which can be in two (SIS) or three (SIR) states: Susceptible, Infected, and Resistant. $S(t)$ represents the total amount of susceptible individuals at time t, and analogously for $I(t)$ and $R(t)$. In the SIS model there are no resistant individuals and the population dynamics is modeled by the following set of equations:

- $\frac{dS}{dt} = \gamma I - \frac{\beta . IS}{N}$
- $\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I$

Here β is the infectious rate, it models the probability of getting the disease in a contact between a susceptible and an infectious subject. γ captures the recovery rate, the probability that an infected individual recovers and becomes susceptible again. At each transition stage t, (a) the rate of susceptible individuals decreased by $\frac{\beta.I(t-1)S(t-1)}{N}$ and increases by $\gamma I(t-1)$, and (b) the rate of infected is increased by $\frac{\beta.I(t-1)S(t-1)}{N}$ and is reduced by $\gamma I(t-1)$. In the SIR model, infected agents can also become resistant with probability δ:

- $\frac{dS}{dt} = \gamma I - \frac{\beta.IS}{N}$
- $\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I - \delta I.$
- $\frac{dR}{dt} = \delta I$

In this work we will only focus on the SIS model in the interest of space, but the experiments and analysis were done for the SIR model too. This formal representation does not take into account vital dynamics (individuals being born and dying), immunity loss, the category of disease carrier (but not infected), and other variants. In the next section we will enumerate different ways in which the formalism can be extended.

More importantly for us here, models do not take into account *network structure*. They assume that transmission is uniform across all individuals, so that the contagion is proportional to the quantity of infected agents. But this simplifying assumption is not realistic. In general, agents interact do not interact uniformly with everyone but they have a small network of friends and acquaintances with whom they share most of their social time. Furthermore, some agents may be more social and have more connections, while others have less. Some may be more central and influential than other in the network arrangement. The spread of rumors will be affected by these subtleties.

2.2. Contagion on Networks.

Formally, a network or graph $G(N, E)$ is defined as a set of nodes (or vertices) N and a set of edges E between those nodes. We will only consider non-weighted undirected edges. For our purposes, nodes represent individuals and edges offer a representation of a connection between them. In the simplest interpretation, two individuals are connected if they each belong to the social circle of the other, namely if they interact on a moderately regular basis.

Given a network $G(N, E)$, the following is a discrete implementation of the Susceptible-Infected-Susceptible model of contagion on them. For each node i and stage t, the state of i in t, $s(i, t)$ is defined by the state of i and all its neighbors at stage $t - 1$. A susceptible node will become infected with probability equal as β [infectious rate] times the amount of infected neighbors, and an infected node will become susceptible with probability equal to γ [recovery rate].

The previous model is deliberately simple because it is devised to study the relation between very fundamental network properties and rumor contagion, as it will be explained in the next section. Extensions and modifications on the model can be done either on the contagion dynamic or the network definition. Regarding the latter, for example:

(1) Edges could be *directed*, so that contagion is not symmetric between agents.

(2) Edges could be *weighted*, so that contagion may encounter some kind of resistance, or because the communication medium is defective.

(3) Nodes could be *typed* at each step, to reflect in-group and out-group biases. In this way, contagion may operate differently between agents of the same type than between agents of different types. In the literature, this is the study of homophily. Alternatively, this could help model the kind of testimonial injustice explained before.

(4) Node population could vary, reflecting generational dynamics of birth, death, and inheritance.

(5) Edges, *friendships* between nodes, may change in time responding to a dynamic. For example, following Schelling's Schelling (1971b) famous dynamic of segregation, nodes may rewire according to their type.

On the contagion dynamic, the model could be extended to the Susceptible-Infected-Resistant (SIR) dynamic, but also the following:

(1) Rather than considering doxastic *states*, agents could have credal *levels*. After all, the endorsement of a proposition may come in degrees.

(2) The model could have several rumors operating simultaneously. Furthermore, rumors could interact with each other positively or negatively. Several testimonial rumors pointing in the same direction may be a good reason to spread them all, while conflicting rumors may reduce each other's diffusion. Alternatively, rumors could mutate in different ways, by switching one to another or changing their levels.

(3) Many other parameters could be included. For example, following Bosse et al. (2014), this could mean adding how expressive or communicative an agent is, how sensitive a recipient is, whether recipient amplifies or reduces the rumor (adoption level), etc.

(4) Contagion could be represented along the lines of social norms, where the adoption of a rumor depends on whether a certain proportion of neighbors adopt it. This is sometimes called *complex* contagion in the literature Centola and Macy (2007).

We focus on the simplest model because it allows us to study the behavior of rumor contagion in well studied and classic network structures and properties. Establishing conclusions on highly attuned models may depend strongly on their particularities, while establishing them on the simplest cases sets up the basis for future study of more involved representations.

2.3. Networks and Network Properties.

In this essay I will model contagion over three types of networks, and some of their combinations: (a) Random Erdös-Rényi networks, (b) hierarchical networks using Barabási-Albert preferential attachment algorithm, and (c) egalitarian networks using Watts-Strogatz networks.

The cannonical random network is the Erdös-Rényi (ER) model. There are two ways of presenting them. In the $G(N, p)$ model, a network is constructed by connecting nodes randomly where every pair of nodes is connected with (independent) probability p. Any given network with N nodes and M edges has a probability $p^M.(1-p)^{\binom{N}{2}-M}$ of being generated. The second presentation of the model is due to Gilbert Gilbert (1959). Given N nodes and M edges, $G(N, M)$ is a network chosen uniformly at random from the collection of all networks which have N nodes and M edges. Both models are related since the expected number of edges in $G(N, p)$ is $\binom{N}{2}.p$. By the law of large numbers any graph in $G(N, p)$ will almost surely have approximately this many edges. Hence, $G(N, p)$ should behave similarly to $G(N, M)$ with $M = \binom{N}{2}.p$ as N increases.

The second class of networks we are going to study are generated using Barabási-Albert preferential attachment (PA) algorithm Barabási and Albert (1999). The algorithm is the following: (1) At stage 0 we begin with m fully connected nodes and (2) at each new stage, a single node is created which forms m edges to existing nodes. New nodes create links preferentially, giving priority to nodes with more links. In particular, at stage t the probability of attaching to node i is given by $d_i(t)/2tm$, where $d_i(t)$ is i's degree (the amount of edges it has), and $2tm = \sum_i d_i(t)$ is the total degree of the network. Networks generated this way are said to be hierarchical because their degree distribution is such that older nodes receive substantially more connections than newer nodes. The process generates a long-tailed distribution that follows a Pareto distribution (or power law) in its tail, which is heavily studied in the social sciences.

There are ways of combining hierarchical and random networks of the kinds that were presented. One straightforward way would be to take a PA network and replace a number or percentage of its

edges with the same number or percentage of random edges as explained before. A more involved hybrid model modifies the Barabási-Albert procedure so that at each stage, new nodes still assign m connections, but only a certain percentage of them are assigned by preference and the rest are assigned at random with the available nodes. The difference between the two hybrids is that the former results in a random network when the percentage is 100, while the latter preserves some of the features of preferential attachment.

The third type of networks are egalitarian in that, for the base case, all nodes share the same degrees and centrality measures (see below). In a ring each node connects to exactly m other adjacent nodes [$\frac{m}{2}$ on each side], forming the expected circle-shaped arrangement. Building on rings, I will here consider Watts-Strogatz Watts and Strogatz (1998) networks. Intuitively, these networks are hybrids that begin with a ring and for each node they replace a certain percentage of its edges [on one side] with random edges. As before, when the percentage is 100 then we obtain (something very close to) a random network. The beauty of Watts-Strogatz is that they manage to secure short average path lengths and high clustering, roughly by inheriting properties both from rings (the clustering) and form Erdös-Rényi (the short path) networks. Not only the average path is short, but these networks (as well as ER)[2] have the small world property according to which the typical distance between two randomly chosen nodes grows proportionally to the logarithm of the number of nodes N in the network.[3]

To summarize, Erdös-Rényi random networks provide a good comparison class and standard to study contagion on networks. The Barabási-Albert preferential attachment algorithm accurately captures hierarchical arrangements because its degree distribution follows the power law (in its

[2]The question whether there are preferential attachment networks satisfying this property is not *trivial*.

[3]This captures the *six degrees of separation* intuition in ER. For those networks it can be shown that for large n, average path length and diameter (largest shortest path) are approximately proportional to log(n)/log(d). If n = 6.7 million, and each person knows around 50 people so that d = 50, log(n)/log(d) is precisely about 6.

tail), and have been shown to emerge in many real like cases.[4] Finally, Watts-Strogatz networks represent more egalitarian networks with nice structural properties.

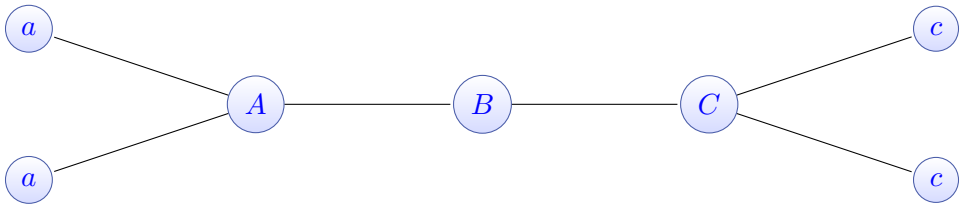
We will also study two well known network properties: Global Clustering Coefficient and Average Path Length. A path between two nodes is the *shortest* sequence of edges that begins in one and end in the other. The length of the path is the amount of edges is contains. If two nodes are not connected, this is somewhat represented with an infinite path, or an N/A value. The average path length is then the average of the length of the paths within a network. If the network is not connected, then it has no average path length. If the network is connected, then the average path length gives us an idea of how fast the contagion effect can manifest. For example, a simple ring is connected but it has a long average path length $[\frac{N}{2} : \frac{m}{2} = \frac{N}{m}]$, while (connected) ER networks exhibit short average path length and follow the small world property exhibited before. In a ring, it takes the "disease" or rumor several steps to reach other nodes from the initial infected; in a ER, it takes less. Watts-Strogatz networks help us explore this features, since higher rewiring rate reduces average path length.

The clustering coefficient of a node i is defined as the number of edges between the i's neighbors divided by the total number of possible edges between its neighbors. Intuitively, the proportion of i's friends that are friends with each other. Clustering then helps measure the density of ties between nodes. The Global Clustering Coefficient (GCC) of a network uses the notion of a triplet. A triplet consists of three nodes that are connected by either two (open triplet) or three (closed triplet) undirected edges. The GCC is the number of closed triplets over the total number of triplets. Once again Watts-Strogatz becomes useful. Rings are highly clustered, while ER networks have very low GCC. As the rewiring-rate in Watts-Strogatz increases, GCC goes down.

[4]Examples abound, but Price De Solla Price (1976) first studied such patterns in scientific bibliographic citations. Proving his point, I reference that paper for being the first one with a concrete application but I am not citing contemporary work.

Finally, the variables of our interest: nodes' Power, or Centrality measures.

The most natural measure of power is the **normalized degree** of a node, namely the amount of edges it has relative to the average amount of edges in the whole network. Intuitively, a node with more "friends" will be able to influence more people. But this notion of centrality is somewhat naive. Consider the following network:



Node B has less degree than both A and C, but B's position in the network is very strategic. Notice, for example, that B is at most two edges from every node, while A and B are at most three. Also, any contagion from the A crowd must go through B in order to reach the C crowd, and vice-versa. Relatedly, in this connected network consider the following centrality measure. For each node, **Closeness Centrality** is the inverse of the sum of it's distances to all other turtles.[5] So for both A and C this is $\frac{1}{11}$, while for B it is $\frac{1}{10}$. So B is more central than both A and C. There are ways of defining Closeness Centrality when the network is not connected. The implementation here measures it by considering all the nodes that are accessible from the given node (i.e. its largest component).

Another well studied measure of power is **Betweeness Centrality**. The betweenness centrality for a node is the number of the shortest paths (between two other nodes) that go through the node. Formally:

[5]Here we will take the inverse of the average rather than the sum, because its easier to implement and preserves the relevant features.

$$b(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Here σ_{st} is the total number of shortest paths from node s to node t, and $\sigma_{st}(v)$ is the number of those paths that pass through v.[6] So in our case, A and B both have a betweeness centrality of $\frac{7}{15}$, while B has a betweeness centrality of $\frac{9}{15}$. Again B is more central in this regard.

The fourth centrality measure is **page rank**, famously implemented by Larry Page, one of the founders of Google, in the beginnings of the seach engine. The version of page rank used here can be thought of as the proportion of time that an agent walking forever at random on the network would spend at the node being measured. The agent has an equal chance of taking any of a nodes edges, and will jump around the network completely randomly 15% of the time. In practice, nodes that are connected to a lot of other turtles that are themselves well-connected (and so on) get a higher page rank.

The final centrality measure is **Eigenvector Centrality**. The intuitive idea is that centrality is a somewhat self-referential notion. The centrality of a node is proportional to the summed centralities of its neighbors. This can be defined using some linear algebra. A network $G(N, E)$ can be represented by its adjacency matrix $A = (a_{v,t})$ where $a_{v,t} = 1$ if node v is linked to node t, and zero otherwise. The centrality x_v of a node v is defined as:

$$x_v = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t$$

Here we see that the centrality measure x_v is proportional to the summed centralities of v's neighbors. If we take $\mathbf{x} = (x_1, ..., x_N)$ to be a vector of the measures for the N nodes, the system of equations is:

[6] In the model, I will be using its normalized version defined:

$$normal(g(v)) = \frac{g(v) - min(g}{max(g) - min(g)}$$

Otherwise, the measure will depend heavily on the number of nodes and edges in the network.

$$\mathbf{x} = \frac{1}{\lambda}\mathbf{A}\mathbf{x}$$

And rearranging:

$$\lambda\mathbf{x} = \mathbf{A}\mathbf{x}$$

Hence $\mathbf{x}$ is an eigenvector with eigevalue λ. The system has several solutions, but using the PerronFrobenius theorem we can be sure that there is a largest real eigenvalue and that the corresponding eigenvector can be chosen to have strictly positive values in its components - so we normalize in a way that its entries in $\mathbf{x}$ have a maximum value of one.

Wrapping up, in these experiments I considered as measures (a) normalized degree, (b) closeness centrality, (c) normalized betweeness centrality, (d) page rank, and (e) eigenvector centrality. It should be noted that closeness and betweeness centralities do depend on whether the network is connected.

3. Some Surprising Results

The purpose of this section is to summarize the results obtained. The Netlogo model is available for assessment, as well as several R-markdown files with the detailed regression analysis.

Section 2 described the Suceptibe-Infected-Suceptible model, its implementation on networks, the types of networks considered, and different sets of properties. To refresh, given a network $G(N, E)$ the procedure begins by randomly infecting a single node, which we called "influencer".

We consider two dynamics, one in which the influencer is capable of recovering, and one in which they can not. In the first case, the influencer is not *intentionally* trying to spread the rumor but rather *became* randomly infect and has the urge to spread it. In the second case, the influencer

is *intentionally* spreading the rumor. In the literature, Shelke and Attar (2019), this is sometimes conceptualized as the difference between *misinformation* and *disinformation* - intentionality does matter and the behavioral pattern might help identify the source of the rumor. As explained before, the state of a node i in stage t, $s(i,t)$ is defined by the state of i and all its neighbors at stage $t-1$. A susceptible node will become infected with probability equal to β [infectious rate] times the amount of infected neighbors, and an infected node will become susceptible with probability equal to γ [recovery rate]. Experiments were done with Erdös-Rényi networks, (hybrid) Barabási-Albert generated networks, and the Watts-Strogatz model.

In doing the study, the initial hypothesis was that the centrality of the influencer would have some effect on the success of the contagion. The results offer positive evidence in this direction. Furthermore, as it is already been shown Centola (2010), network variables like average path length and global clustering coefficient explain part of the variance in diffusion. What is surprising is that in many scenarios centrality measures like PageRank and Eigenvector are in many cases **negatively** correlated with contagion success.

3.1. Preeliminaries.

For each type of network we ran 216000 simulations, as explained below. Since the closeness and betweeness centrality measures used here depend on whether the network is connected, we only looked at the experiments that generated connected networks. This left us with more than 150000 experiments for each network type, enough to make informed observations.

As Figure 1 shows, in many experiments contagion was not successful, with a percentage of infected nodes below 1% at the end of the process. For this reason, the data was subsetted to study the instances of successful contagion - when the percentage of infected nodes was larger or smaller than 1% respectively. The plots in Figure 1 have the distribution of percentage of infection, i.e.

the number of experiments for which such percentage was recorded at the end of the simulation. The top plot corresponds to the total data and the bottom to the successful data.

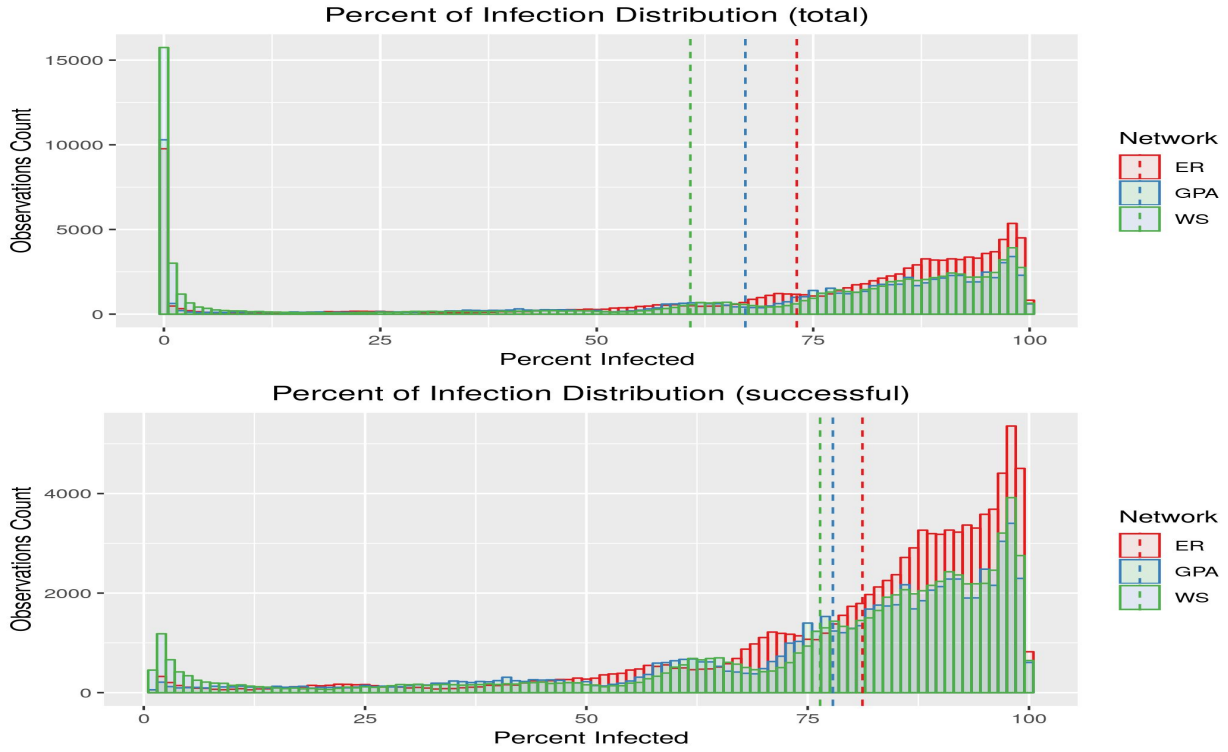


FIGURE 13. Distribution of Contagion

Although the experiments with different network types are not strictly comparable because they depend on different sweeping parameters, it is noticeable that in general random networks [ER] were the most successful in spreading the rumor contagion and generalized preferential attachment [GPA] was second. Variance was also different for the three networks, with WS having one of 635, GPA of 449, and ER of 369 in the successful contagion data set. This will be presented in more detail later.

The purpose of the study is to evaluate how much network structure and the centrality of the initial influencer affects contagion. In order to do this, each experiment stored the values of the following three kinds of variables:

Contagion Variables	Network Variables	Centrality Variables
Percentage Infected, Percentage Recovered, Spread %, Recovery %, Resistance %, Influencer Recovery.	Number of Nodes, Number of Links Link % [ER], Rewire % [GPA, WS], Average Degree, Global Clustering Coefficient, and Average Path Length.	Normalized Degree, Eigenvector Centrality, Closeness Centrality, Page Rank, and (norm) Betweeness Centrality.

TABLE 7. Variable Sets

The next step was to ran several multivariate regression analysis over these data sets. The guiding question was how much of the infection variance can be explained by the different sets of variables. Infection percentage was taken as the predicted, dependent, variable and different models were explored for the complete and the successful data sets. The next sections summarize the results in detail.

3.2. Erdös-Rényi Random Networks.

For random networks, experiments were performed by considering 500 repetitions the following sweeping parameters:

Spread % (β)	Recovery % (γ)	Population	Link %	Recovery?
1, 3, 5	1, 4, 7, 10	100, 200, 300	5, 10, 15	true, false

TABLE 8. Erdös-Rényi Parameters

For example, we start by fixing the population N in 100 and the link % in 5 for the code to generate an Erdös-Rényi $G(100, 0.05)$ network. Second, contagion parameters like spread, recovery, and resistance chance are fixed; as well as whether the influencer can recover. Third, an initial random node (the influencer) is infected and the experiment is performed until the network stabilizes or 1000 steps passed. Finally, the data of the experiment is stored.

Table 9 summarizes the percentage of the variance that can be explained by each of the ten regression models. For example, by looking at the data set that corresponds to the successful SIS experiments (bottom right), we can explain with statistical significant 14% of the infection by the centrality variables. In other words, socially powerful individuals have an effect on contagion.

Data Set	***Best Model***	***Full Model***	***Contagion***	***Network***	***Centrality***
Complete SIS	59%	59%	44%	15%	15%
SIS with Success	72%	72%	53%	15%	14%

TABLE 9. Erdös-Rényi Variance Explanation

The Full Model corresponds here to the model that takes into account *all* recorded variables. The Best Model begins with the Full Model and chooses a model by Akaike Information Criterion (AIC) in a stepwise algorithm. Fundamentally, this procedure disregards statistically uninformative variables. In most, but not all cases, the best and the full models are the same. Also, in most but not all cases, all variables in all models were statistically significant with a p-value less than 0.05. In many of the experiments in the complete data set contagion failed, so much of the dependant variable (percentage of infected) is centered at zero and therefore variance is harder to explain.

The results provide evidence to the common sense claim that the centrality or power of an influencer has an effect on contagion. What is surprising is that they also suggest that some centrality measures have a *negative* effect, so that the more central an influencer is, the less the contagion is expected to spread. For example, Figure 2 plots the percentage of infected over page rank and eigenvector centrality in the successful data set, by the expected amount of links the networks had.

FIGURE 14. Some Centralities and Contagion

The lines here represent linear regressions. The more dense the network was, i.e. the higher the link probability, the more the infection. It is clear that page rank had a negative effect on now successful the contagion effect was. Eigenvector, on the other hand, seems to have had a positive effect. A more subtle analysis reveals that the first negative correlation holds under scrutiny, but not the seconds. To understand this, let us look at the coefficients[7] of the best regression model for the successful SIS data set.

This model explains 72% of the contagion variance and has a p-value $< 10^{-16}$. The first thing to notice is that not all variables are included, some of them were deemed uninformative by the Akaike Information Criterion (AIC). Yet variables from the three categories appear, suggesting that they have independent explanatory power.

The dependent or predicted variable here is the infection percentage. The estimate coefficient for each independent variable is telling us the size and direction of the effect that it is having on the

[7]Some of the coefficients were rounded for simplicity.

Variable	Estimate	Standard Error	T-Value	P-Value
Intercept	0.01	1.18	86	$< 10^{-16}$
Number Nodes	0.05	≈ 0	26	$< 10^{-16}$
Number Links	-0.003	≈ 0	-18	$< 10^{-16}$
Recovery? (true)	2	0.07	30	$< 10^{-16}$
Recovery % (γ)	-3.5	0.01	-343	$< 10^{-16}$
Spread % (β)	5.75	0.02	275	$< 10^{-16}$
Average Degree	0.6	0.03	18	$< 10^{-16}$
Glob. Clust. Coeff.	-8.32	4	-2	0.04
Avg. Path Lenght	-16.7	0.31	-53	$< 10^{-16}$
Eigenvector C.	-1.1	0.5	-2.3	0.02
Page Rank	-1	0.37	-2.7	0.006
N. Betweeness C.	1.2	0.35	3.4	0.0006

TABLE 10. Best Model for the SIS data set with success

dependent variable. For example, in this model each 1% increase in spread chance (β) will increase the infection by 5.75%, and each 1% increase in the recovery chance (γ) will decrease 3.5 points in the infection percentage. The standard error is an estimate of the standard deviation of the coefficient, the amount it varies across cases. It measures the precision of the coefficient estimate, so that smaller values represent more precise estimates. The T-Value is the coefficient divided by its standard error, and is used here mostly to compute the P-Value. The P-Value is the probability of seeing a result as extreme as the obtained (the T-Value) in a collection of random data in which the variable had no effect. In other words, assuming a (random) distribution in which the variable had no effect (i.e. the null hypothesis is true), the value is telling us how likely it is that the data obtained came from that distribution. With a P-Value of 0.05 there is only a 5% chance that the observed results come up in a random distribution, so you can say with a 95% probability of being correct that the variable is having some effect (i.e. rejection of the null hypothesis). In this model, all variables appear with a P-Value < 0.05, which is usually regarded as the threshold for statistical significance.

The linear model also offers some intuitive support to the idea that there were no gross mistakes in the experimental setup, or the underlying code. It is natural to expect that recovery chance

will decrease contagion, and spread will increase it. Similarly, the greater the average degree of the network, the greater the contagion effect. Furthermore, there is a well established literature initiated by Watts and Strogatz (Watts and Strogatz, 1998; Centola, 2010; Granovetter, 1973) revealing that average path length and global clustering coefficients are negatively correlated with contagion success (and speed). Intuitively, the longer the average path, the more steps it would take for contagion to reach nodes distant from the original infection, and therefore higher chance of recovery in the mean time. High clustering means, in many cases, that nodes are segregated so that only those clustered around the original infection will end up infected. On the other hand, results are also unintuitive. Whether the influencer recovers affects the overall contagion, but somewhat conversely: If it recovers, that increases the percent of contagion positively by two points. Each extra node increases contagion by 0.05%, and number of links have a negative effect, albeit negligible.

The most unexpected observation is that Page Rank and Eigenvector Centrality have a negative effect on contagion. The best model suggests that each 1% increase in page rank[8] ***decreases*** 1% the contagion effect. Similarly, the more central the initial node is with respect to eigenvector centrality, the less successful the contagion is expected to be.[9] In contrast, the normalized betweeness centrality shows a positive effect on contagion. Also surprisingly, influencer normalized degree and closeness centrality were deemed uninformative by the stepwise AIC algorithm.

The next sections will show that the effect that centrality variables and network variables have on contagion are not negligible for other types of networks. Furthermore, the surprising result that

[8]The Page Rank measure is normalized so that the sum of all page rank values is approximately one hundred, hence each point increase in Page Rank can be interpreted as a 1% increase of the influencer centrality

[9]Eigenvector and (normalized) Betweeness centrality are normalized so that the most central nodes gets a value of 1 and the lease central a value of 0, but nothing secures that sum of all is 1. This is why the percentile interpretation given for page rank cannot be given to these measures.

some of the centrality measures have a negative effect on contagion, and some others are irrelevant will be a persistent emergent phenomenon.

3.3. Preferential Attachment.

Here we run 100 repetitions of the following combinations of parameters:

Spread %	Recovery %	Population	Rewire Prob	Recovery?	Neighbors
1,3,5	1,4,7,10	100, 200, 300	0, 0.25, 0.5, 0.75, 1	true, false	1,6,11

TABLE 11. Barabási-Albert Preferential Attachment Parameters

Some variables were kept the same as before: Spread %, Recovery %, Resistance %, Population, and Recovery. The Neighbors variable refers to the amoung m of edges created at each stage of the Barabási-Albert PA algorithm. Finally, the Rewire Prob is a hybridization parameter: the percentage of the edges of the network were replaced by random edges. Variables were stored and categorized in the same way as before. Once again, in most of the experiments the contagion was not successful and hence the data was again divided into two groups. This is how much of the variance is explained by the regression analysis:

Data Set	***Best Model***	***Full Model***	***Contagion***	***Network***	***Centrality***
Complete SIS	66%	66%	43%	29%	21%
SIS with Success	78%	78%	52%	18%	16%

TABLE 12. Barabási-Albert Preferential Attachment Variance Explanation

As with the case with Random Networks, the models explain a substantial amount of the variance, and both network and centrality variables show up as significant. The surprising result of the previous section still holds: some measures of centrality appear to have a negative effect on contagion. Let us look at the Best Model for the SIS data set with success.

Once again all variables of the best model are statistically significant. The new rewiring probability variable positively affects contagion. This shows that the more random (and less hierarchical) the networks are, the more successful the contagion is. Also, like in random networks, page rank

	Estimate	**Standard Error**	**T-Value**	**P-Value**
Intercept	77.6	1.33	58	$< 10^{-16}$
Number Nodes	0.036	0.0017	20	$< 10^{-16}$
Number Links	-0.0018	≈ 0	-8	$< 10^{-16}$
Recovery? (true)	2	0.08	25	$< 10^{-16}$
Recovery %	-3.93	0.01	-340	$< 10^{-16}$
Spread %	6.52	0.02	272	$< 10^{-16}$
Rewire Prob	3.5	0.28	12	$< 10^{-16}$
Average Degree	0.64	0.03	18	$< 10^{-16}$
Glob. Clust. Coeff.	9.6	2	4.8	$< 10^{-6}$
Avg. Path Lenght	-11.1	0.18	-56	$< 10^{-16}$
Eigenvector C.	-3.6	0.7	-5.2	0.000001
Page Rank	-2	0.28	-7.5	$< 10^{-13}$
Norm. Degree	0.62	0.16	4	0.00006
N. Betweeness C.	1.33	0.5	2.7	0.007

TABLE 13. Best Model for the SIS data set with success

and eigenvector centrality also have a negative effect on contagion. Furthermore, the effect seems to be more pronounced. For each 1% increase in the page rank of a node, the percentage of infected is expected to decrease by 2%. In the same set for random networks eigenvector had an coefficient of -1.1 while now it is -3.6, more than three times larger. Normalized betweeness degree has again a positive effect, and we now find the influencer's normalized degree as significant.

Like in the random network case, page rank and eigenvector centrality have negative coefficients in the best model for the full data set and the SIS model (both the complete and the successful subsets). They also have negative coefficient for those data sets in the full model and the model using only centrality variables. Before drawing any general conclusions, let us explore the results in small world egalitarian networks.

3.4. Small World Watts-Strogatz.

Again, we run 100 repetitions of the following combinations of parameters:

Spread %	Recovery %	Population	Rewire Prob	Recovery?	Neighbors
1,3,5	1,4,7,10	100,200,300	0, 0.25, 0.5, 0.75, 1	true, false	1,6,11

TABLE 14. Watts-Strogatz Parameters

Data Set	***Best Model***	***Full Model***	***Contagion***	***Network***	***Centrality***
Complete SIS	73%	73%	27%	43%	37%
SIS with Success	78%	78%	28%	34%	25%

TABLE 15. Watts-Strogatz Variance Explanation

Variables are organized in the same way as in the previous section. Given that we are now familiar with the methodology, let us move to the variance explanation table.

Table 16 shows that in Watts-Strogatz, network and centrality variables explain more of the contagion effect than they did in the previous cases. Let us take a look at the best model regression:

	Estimate	**Standard Error**	**T-Value**	**P-Value**
Intercept	30.1	0.6	49	$< 10^{-16}$
Number Nodes	0.0037	0.0025	1.5	0.14
Number Links	0.003	0.0025	14	$< 10^{-16}$
Recovery? (true)	3	0.1	32.4	$< 10^{-16}$
Recovery %	-3.82	0.013	-283	$< 10^{-16}$
Spread %	6.65	0.028	238	$< 10^{-16}$
Rewire Prob	10.75	0.25	42	$< 10^{-16}$
Average Degree	-0.35	0.032	-11	$< 10^{-16}$
Glob. Clust. Coeff.	49.2	0.7	70	$< 10^{-6}$
Avg. Path Lenght	-0.67	0.011	-58	$< 10^{-16}$
Eigenvector C.	10.84	0.36	30	$< 10^{-16}$
Closeness C.	78	1.06	73	$< 10^{-16}$
Page Rank	-1113	51.28	-21	$< 10^{-16}$
Norm. Degree	-1.22	0.51	-2	0.0166
N. Betweeness C.	-4.57	0.38	-12	$< 10^{-16}$

TABLE 16. Best Model for the SIS data set with success

All variables, except the number of nodes, are statistically significant. As before, the rewiring probability variable affects contagion positively, so that in the more random (and less egalitarian) networks the percentage of infected was higher. But now the story about centrality measures is not the same. Page rank still has a negative effect, and *very* substantial. Eigenvector is now positive, while it was negative in the other networks. Betweeness and normalized degree are now negatinve

and they were positive before. Closeness centrality is now a relevant variable according to the AIC, and it has a positive effect.

4. Conclusion

4.1. Technical Upshot of the Results.

So what have we learned from these experiments? For starters, that network contagion is not as simple as expected. Yet, some claims are supported by the evidence.

(1) The centrality of the initial infection (influencer) has a significant effect on contagion. More interestingly, the effect of different centralities varies across different networks.

(2) Network properties have a significant effect on contagion.

(3) How hierarchical or egalitarian a network is has an effect on the contagion.

The fist claim (1) is grounded on the fact that in all of the models presented, some of the centrality variables were statistically significant (with a p-value of 0.05). The surprise, nonetheless, was that in some cases centrality had a negative effect. To summarize some examples:

- Page Rank: Had a negative estimate in all of the models and all of the data sets.
- Eigenvector Centrality: Had a negative estimate in random and preferential attachment networks, while a positive in Watts-Strogatz.
- Betweeness Centrality: Had a positive effect in random and preferential attachment networks, while a negative in Watts-Strogatz.
- Normalized Degree: Was positive in preferential attachment, negative in Watts-Strogatz, and irrelevant in random networks.

In all cases, (2) general network properties had an significant effect on contagion. In particular, the average path length of the networks was negatively correlated, since the longer the paths the harder it would be for the rumor to be transmitted further away in the network. Global clustering coefficient was more interesting, since it has a positive effect in hierarchical and egalitarian networks

but a negative one in Erdös-Rényi. Average degree was statistically significant but had only a minor effect.

More interestingly, (3) how egalitarian, hierarchical or random the network was did affect contagion. The boxplot[10] in Figure 3 presents the mean percent of infection by how randomized WS and GPA networks were. Surprisingly, when there was no randomization so that networks were purely egalitarian (a ring) or hierarchical (pure Barabási-Albert procedure) the mean percent infected was the same for both kinds of networks. Nevertheless, as soon as randomization is introduced contagion increases overall and it is more effective in hierarchical networks - although the difference is small.

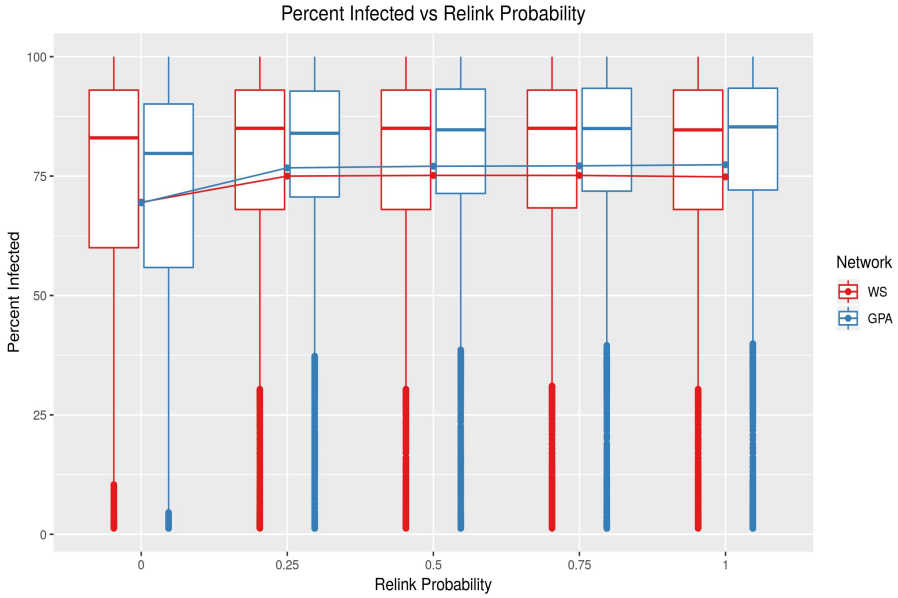


FIGURE 15. Randomization and Contagion

4.2. Philosophical Upshot of the Results.

Unfortunately, the results do not offer a moral that can be truthfully summarized in an eloquent or provocative slogan. This is not surprising, since in many ways this work is propaedeutic for

[10]Means are represented by the dots, median by the horizontal lines at each parameter, boxes represent the overall spread of the sample, and dots at the bottom are outlier data points.

future research. The contagion of rumors in a network is more complex than expected, and further research may be necessary before definite conclusions about epistemic justice or about Clifford's worries on the spread of misinformation. Yet, since Ioannidis (2005), it is well known that negative or inconclusive investigations are hard to publish, so some preeliminary observations are on point.

To begin, the experiments suggest that Clifford's worries about the dangers of the spread of misinformation are warranted. The regression models before showed that infection rate (β) was positively correlated with infection, and the recovery rate (γ) was negatively correlated. But this was not truly informative about absolute contagion rates. In Figure 4, the Net Contagion Effect variable in the horizontal axis is defined as $\beta - \gamma$, namely the probability of infection minus the probability of recovery. The graph takes into account all data, not just the successful contagion cases. Notice that in its lower rate, when $\beta - \gamma = -9$, the mean population infection was between 7% and 27%. When the net contagion effect is zero, the mean population infection was above 70% for all networks. This graph shows that it takes very little to get most of the population infected with a rumor. Clifford's fears seem vindicated.

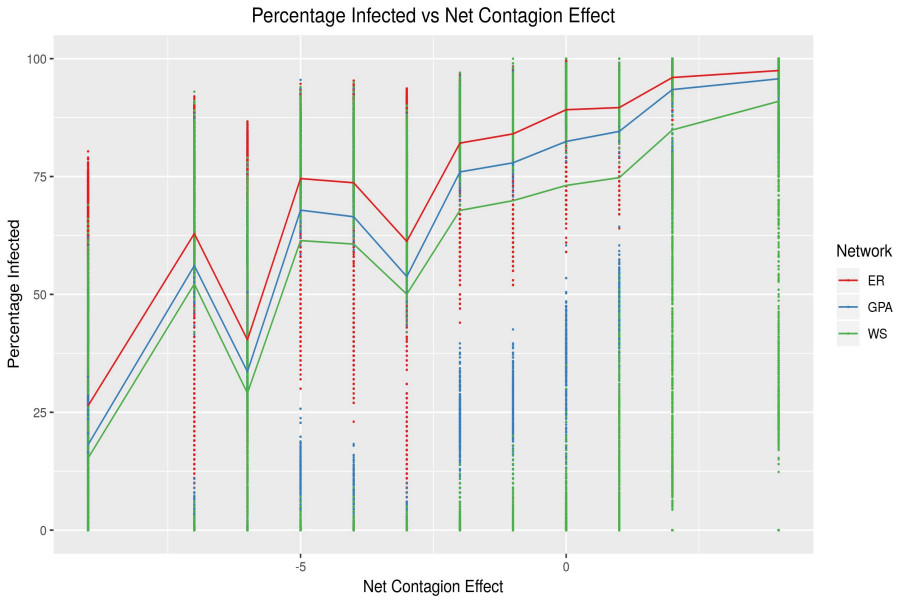


FIGURE 16. Net Contagion Effect

Let us move to epistemic injustice. The results showed that social hierarchies, represented by centrality measures, do have a statistically significant effect in the power to spread rumors. So it is fair to conclude that, *in so far those hierarchies are morally or epistemologically unjustified*, we might be in the presence of a new kind of epistemic injustice. Yet the picture is not straightforward. Very surprisingly, *PageRank* centrality was *negatively* correlated with contagion. It is unclear what to do with this. More interestingly, the effect that centrality had depended on the overall structure of the network; whether they are more egalitarian (Watts-Strogatz), hierarchical (Barabási-Albert), or random (Erdös-Rényi). Yet whether networks were hierarchical or egalitarian made no *gross* difference in mean contagion success or variance, while random networks were more infected on average and had higher variance. This suggests that attempting to reduce contagion effects by making social networks less hierarchical and more egalitarian may not have a significant effect; although more research can be revealing in this respect.

To conclude, it may seem, as Mößner and Kitcher (2017) suggested, that technologies like the Internet or Online Social Networks are an epistemically democratizing force in which individuals are now capable of producing and consuming more information freely. Following Clifford's epistemic conservatism, I do not think there is enough evidence to support that claim. Furthermore, this paper suggests that new, different, and sophisticated epistemic hierarchies may be emerging. Hierarchies in which those with a privileged position in the network can heavily influence others. Yet, in agreement with Mößner and Kitcher (2017), the issue about epistemic hierarchies is not that they are negative in themselves, but that they may not be conducive to the kind of values and goals that we are pursuing.

Bibliography

Aczél, J. (1948). On mean values. *Bulletin of the American Mathematical Society 54*(4), 392–400.

Aczél, J. and H. Oser (2006). *Lectures on Functional Equations and Their Applications*. Courier Corporation.

Alchourrn, C. E., P. Grdenfors, and D. Makinson (1985). On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic 50*(2), 510–530.

Arló-Costa, H. (2007). The logic of conditionals. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2014 ed.).

Arló-Costa, H. and J. Helzner (2010). Ambiguity aversion: The explanatory power of indeterminate probabilities. *Synthese 172*(1), 37–55.

Axelrod, R. and W. D. Hamilton (1981). The evolution of cooperation. *Science 211*(4489), 13901396.

Bacharach, M. (1972). Scientific disagreement. *Unpublished Manuscript*.

Barabási, A.-L. and R. Albert (1999). Emergence of scaling in random networks. *Science 286*(5439), 509–512.

Baratgin, J. and G. Politzer (2010). Updating: A psychologically basic situation of probability revision. *Thinking & Reasoning 16*(4), 253–287.

Blackwell, D. and L. E. Dubins (1962). Merging of opinions with increasing information. *The Annals of Mathematical Statistics*, 882–886.

Bordley, R. F. (1982). A multiplicative formula for aggregating probability assessments. *Management Science 28*(10), 1137–1148.

Bosse, T., R. Duell, Z. A. Memon, J. Treur, and C. N. van der Wal (2014). Agent-based modeling of emotion contagion in groups. *Cognitive Computation 7*, 111–136.

Bradley, R. (2017). *Decision Theory with a Human Face*. Cambridge University Press.

Bradley, S. (2014). Imprecise probabilities. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2014 ed.).

Bryden, J., S. Funk, and V. A. Jansen (2013, Feb). Word usage mirrors community structure in the online social network twitter. *EPJ Data Science 2*(1), 3.

Card, D., A. Mas, and J. Rothstein (2008, 02). Tipping and the Dynamics of Segregation*. *The Quarterly Journal of Economics 123*(1), 177–218.

Centola, D. (2010). The spread of behavior in an online social network experiment. *Science 329*(5996), 1194–1197.

Centola, D. and M. Macy (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology 113*(3), 702–734.

Christensen, D. (2009). Disagreement as evidence: The epistemology of controversy. *Philosophy Compass 4*(5), 756–767.

Clemen, R. T. and R. L. Winkler (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis 19*(2), 187–203.

Clifford, W. K. (1879). *Lectures and Essays*, Volume II. London: Macmillan.

Cozman, F. G. (1998). Irrelevance and independence relations in quasi-bayesian networks. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pp. 89–96. Morgan Kaufmann Publishers Inc.

Cozman, F. G. (2000). Credal networks. *Artificial Intelligence 120*(2), 199–233.

Cragg, M. (2011). The rise of the twitter tribes. *The Guardian*.

Cresto, E. (2008). A model for structural changes of belief. *Studia Logica 88*(3), 431–451.

Currarini, S., J. Matheson, and F. Vega-Redondo (2016). A simple model of homophily in social networks. *European Economic Review 90*, 18 – 39. Social identity and discrimination.

de Campos, L. M. and S. Moral (1995). Independence concepts for convex sets of probabilities. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pp. 108–115. Morgan Kaufmann Publishers Inc.

de Finetti, B. (1931). Sul concetto di media. *Giornale dell'Istituto Italiano degli Attuari 2*, 369–396.

de Finetti, B. (1964). Foresight: Its logical laws, its subjective sources. In H. E. Kyburg and H. E. Smoklery (Eds.), *Studies in Subjective Probability*. Wiley.

De Solla Price, D. (1976, 09). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science 27*, 292 – 306.

Dekel, E., B. Lipman, and A. Rustichini (1998). Standard state-space models preclude unawareness. *Econometrica 66*(1), 159–174.

Diaconis, P. and S. L. Zabell (1982a). Updating subjective probability. *Journal of the American Statistical Association 77*(380), 822–830.

Diaconis, P. and S. L. Zabell (1982b). Updating subjective probability. *Journal of the American Statistical Association 77*(380), 822–830.

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

Dietrich, F. (2010). Bayesian group belief. *Social Choice and Welfare 35*(4), 595–626.

Dietrich, F. and C. List (2011). A model of non-informational preference change. *Journal of Theoretical Politics 23*(2), 145–164.

Dietrich, F. and C. List (2014a, October). Probabilistic opinion pooling.

Dietrich, F. and C. List (2014b). Probabilistic opinion pooling. In A. Hájek and C. Hitchcock (Eds.), *Oxford Handbook of Probability and Philosophy*. Oxford University Press.

Dietrich, F. and C. List (2017, Apr). Probabilistic opinion pooling generalized. part two: the premise-based approach. *Social Choice and Welfare 48*(4), 787–814.

Elga, A. (2007). Reflection and disagreement. *Noûs 41*(3), 478–502.

Elkin, L. and G. Wheeler (2016). Resolving peer disagreements through imprecise probabilities. *Noûs*, DOI: 10.1111/nous.12143.

Ellsberg, D. (1963). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics 77*(2), 327–336.

Elster, J. (2006). Chapter 3 Altruistic Behavior and Altruistic Motivations. In S.-C. Kolm and J. M. Ythier (Eds.), *Foundations*, Volume 1 of *Handbook of the Economics of Giving, Altruism and Reciprocity*, pp. 183206. Elsevier.

Epstein, J. and R. Axtell (1996, 01). *Growing Artificial Societies: Social Science from the Bottom Up*, Volume 76.

Fagin, R. and J. Y. Halpern (1987). Belief, awareness, and limited reasoning. *Artificial Intelligence 34*(1), 39 – 76.

Feyerabend, P. K. (1962). Explanation, reduction and empiricism. In H. Feigl and G. Maxwell (Eds.), *Crítica: Revista Hispanoamericana de Filosofía*, pp. 103–106.

Field, H. (1978). A note on jeffrey conditionalization. *Philosophy of Science*, 361–367.

Fishburn, P. C. (1984, Jul). On Harsanyi's utilitarian cardinal welfare theorem. *Theory and Decision 17*(1), 2128.

Fleurbaey, M. (2010). Assessing risky social situations. *Journal of Political Economy 118*(4), 649–680.

Fleurbaey, M. and P. Mongin (2016, August). The utilitarian relevance of the aggregation theorem. *American Economic Journal: Microeconomics 8*(3), 289–306.

French, S. (1985, September). Group consensus probability distributions: A critical survey. In D. L. J.M. Bernardo, M.H. DeGroot and A. Smith (Eds.), *Bayesian Statistics: Proceedings of*

the Second Valencia International Meeting, Volume 2, pp. 183–201. North-Holland.

Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press.

Gaifman, H. and M. Snir (1982). Probabilities over rich languages, testing and randomness. *The Journal of Symbolic Logic 47*(03), 495–548.

Gaifman, H. and A. Vasudevan (2012). Deceptive updating and minimal information methods. *Synthese 187*(1), 147–178.

Gandica, Y., F. Gargiulo, and T. Carletti (2016). Can topology reshape segregation patterns? *Chaos, Solitons & Fractals 90*, 46 – 54. Challenges in Data Science.

Gärdenfors, P. (1982). Imaging and conditionalization. *The Journal of Philosophy*, 747–760.

Gärdenfors, P. and H. Rott (1995). *Handbook of Logic in Artificial Intelligence and Logic Programming: Epistemic and temporal reasoning*, Volume 4, Chapter Belief Revision. Oxford University Press, Oxford.

Gärdenfors, P. and N.-E. Sahlin (1982). Unreliable probabilities, risk taking, and decision making. *Synthese 53*(3), 361–386.

Genest, C. (1984). A characterization theorem for externally bayesian groups. *The Annals of Statistics*, 1100–1105.

Genest, C., K. J. McConway, and M. J. Schervish (1986). Characterization of externally bayesian pooling operators. *The Annals of Statistics*, 487–501.

Genest, C. and C. G. Wagner (1987). Further evidence against independence preservation in expert judgement synthesis. *Aequationes Mathematicae 32*(1), 74–86.

Genest, C. and J. V. Zidek (1986). Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, 114–135.

Gilbert, E. N. (1959, 12). Random graphs. *Ann. Math. Statist. 30*(4), 1141–1144.

Gilboa, I. and D. Schmeidler (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics 18*(2), 141–153.

Girón, F. J. and S. Ríos (1980). Quasi-bayesian behaviour: A more realistic approach to decision making? *Trabajos de Estadística y de Investigación Operativa 31*(1), 17–38.

Gloor, P., A. Fronzetti Colladon, J. Marcos de Oliveira, and P. Rovelli (2018, 09). Identifying tribes on twitter through shared context.

Good, I. J. (1983). *Good Thinking: The Foundations of Probability and Its Applications*. U of Minnesota Press.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology 78*(6), 1360–1380.

Griswold, Jr, C. L. (1998). *Adam Smith and the Virtues of Enlightenment*. Modern European Philosophy. Cambridge University Press.

Groenendijk, J. and M. Stokhof (1984). Studies on the Semantics of Questions and the Pragmatics of Answers.

Hájek, A. and N. Hall (1994). The hypothesis of the conditional construal of conditional probability. In E. Eells and B. Skyrms (Eds.), *Probability and Conditionals: Belief Revision and Rational Decision*, pp. 75–112. Cambridge University Press.

Halmos, P. R. (1963). *Lectures on Boolean Algebras*. Van Nostrand, Princeton.

Halpern, J. Y. (2001). Alternative semantics for unawareness. *Games and Economic Behavior 37*(2), 321 – 339.

Hamilton, W. D. (1963). The Evolution of Altruistic Behavior. *The American Naturalist 97*(896), 354356.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. II. *Journal of Theoretical Biology 7*(1), 1752.

Hamilton, W. D. (1970, dec). Selfish and Spiteful Behaviour in an Evolutionary Model. *Nature 228*, 1218.

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

Harsanyi, J. C. (1977). *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge books online. Cambridge University Press.

Hartmann, S. (2014a). A new solution to the problem of old evidence. In *Philosophy of Science Association 24th Biennial Meeting*, Chicago, IL.

Hartmann, S. (2014b, July). A new solution to the problem of old evidence.

Hatna, E. and I. Benenson (2015, 06). Combining segregation and integration: Schelling model dynamics for heterogeneous population. *Journal of Artificial Societies and Social Simulation*.

Herron, T., T. Seidenfeld, and L. Wasserman (1997). Divisive conditioning: Further results on dilation. *Philosophy of Science*, 411–444.

Herzberg, F. (2014). Aggregating infinitely many probability measures. *Theory and Decision*, 1–19.

Huttegger, S. M. (2015). Merging of opinions and probability kinematics. *The Review of Symbolic Logic 8*(04), 611–648.

in Common, M. (2018). The hidden tribes of america. *https://hiddentribes.us/*.

Ioannidis, J. (2005). Why most published research findings are false. *PLoS medicine*.

Isaacs, H. and L. Pye (1989). *Idols of the Tribe: Group Identity and Political Change*. Harvard University Press.

James, W. (1979). *The Will to Believe and Other Essays in Popular Philosophy*, Volume II. Harvard University Press.

Jeffrey, R. (1990). *The Logic of Decision*. McGraw-Hill series in probability and statistics. University of Chicago Press.

Jeffrey, R. (2004). *Subjective Probability: The Real Thing*. Cambridge University Press.

Joyce, J. M. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press.

Kant, I. and M. Gregor (1998). *Kant: Groundwork of the Metaphysics of Morals*. Cambridge Texts in the History of Philosophy. Cambridge University Press.

Kaplan, M. (1996). *Decision Theory as Philosophy*. Cambridge University Press.

Kitcher, P. (1978). Theories, theorists and theoretical change. *The Philosophical Review 87*(4), 519–547.

Kitcher, P. (1993). *The Advancement of Science: Science without Legend, Objectivity without Illusions*. Oxford University Press New York.

Kitcher, P. (2010). Varieties of altruism. *Economics and Philosophy 26*(2), 121148.

Kolm, S. (2002). *Modern Theories of Justice*. MIT Press.

Kolm, S.-C. (1995, Jan). The economics of social sentiments: the case of envy. *Japanese economic review 46(1)*, 63–87.

Kolm, S.-C. (2006). Introduction to the Economics of Giving, Altruism and Reciprocity. Volume 1, Chapter 01, pp. 1–122. Elsevier.

Kolmogorov, A. N. (1930). Sur la notion de la moyenne. *Atti della R. Academia Nazionale dei Lincei 12*(9), 388–391.

Kuhn, T. S. (1970). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Kuhn, T. S. (2000). *The Road since Structure*. University of Chicago Press.

Kullback, S. and R. A. Leibler (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 79–86.

Kyburg, H. E. (1987). Bayesian and non-bayesian evidential updating. *Artificial Intelligence 31*(3), 271–293.

Kyburg, H. E. (1998). Interval-valued probabilities. *Imprecise Probabilities Project*.

Kyburg, H. E. and M. Pittarelli (1992). Some problems for convex bayesians. In *Proceedings of the Eighth international conference on Uncertainty in artificial intelligence*, pp. 149–154. Morgan

Kaufmann Publishers Inc.

Kyburg, H. E. and M. Pittarelli (1996). Set-based bayesianism. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 26*(3), 324–339.

Lehrer, K. and C. Wagner (1981). *Rational Consensus in Science and Society: A Philosophical and Mathematical Study*, Volume 21. Springer.

Lehrer, K. and C. Wagner (1983). Probability amalgamation and the independence issue: A reply to laddaga. *Synthese 55*(3), 339–346.

Leitgeb, H. (2016). Imaging all the people. *Episteme* (DOI: 10.1017/epi.2016.14).

Levi, I. (1967). Probability kinematics. *British Journal for the Philosophy of Science*, 197–209.

Levi, I. (1970). Probability and evidence. In M. Swain (Ed.), *Induction, Acceptance, and Rational Belief*, pp. 134–156. New York: Humanities Press.

Levi, I. (1974). On indeterminate probabilities. *The Journal of Philosophy 71*(13), 391–418.

Levi, I. (1978). Irrelevance. In C. Hooker, J. Leach, and E. McClennen (Eds.), *Foundations and Applications of Decision Theory*, Volume 1, pp. 263–273. Boston: Springer.

Levi, I. (1980). *The Enterprise of Knowledge*. MIT Press, Cambridge, MA.

Levi, I. (1985). Consensus as shared agreement and outcome of inquiry. *Synthese 62*(1), pp. 3–11.

Levi, I. (1986a). *Hard choices: Decision making under unresolved conflict*. Cambridge University Press.

Levi, I. (1986b). The paradoxes of allais and ellsberg. *Economics and Philosophy 2*(1), 23–53.

Levi, I. (1990). Pareto unanimity and consensus. *The Journal of Philosophy 87*(9), 481–492.

Levi, I. (1996). *For the Sake of the Argument: Ramsey Test Conditionals, Inductive Inference and Nonmonotonic Reasoning*. Cambridge University Press.

Levi, I. (2009). Why indeterminate probability is rational. *Journal of Applied Logic 7*(4), 364–376.

Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *The Philosophical Review 85*, 297–315.

List, C. and P. Pettit (2002). Aggregating Sets of Judgments: An Impossibility Result. *Economics and Philosophy 18*(1), 89–110.

List, C. and P. Pettit (2011). Oxford University Press.

List, C. and Pettit, P. (2011). *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford University Press.

Madansky, A. (1964). Externally bayesian groups. Santa Monica, CA: RAND Corporation.

McConway, K. J. (1981). Marginalization and linear opinion pools. *Journal of the American Statistical Association 76*(374), 410–414.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology 27*(1), 415–444.

Modica, S. and A. Rustichini (1999). Unawareness and partitional information structures. *Games and Economic Behavior 27*(2), 265 – 298.

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

Mongin, P. (2001). The impartial observer theorem of social ethics. *Economics and Philosophy 17*(2), 147179.

Moral, S. and J. Del Sagrado (1998). Aggregation of imprecise probabilities. In B. Bouchon-Meunier (Ed.), *Aggregation and Fusion of Imperfect Information*, pp. 162–188. Springer.

Mößner, N. and P. Kitcher (2017, Mar). Knowledge, democracy, and the internet. *Minerva 55*(1), 1–24.

Nau, R. F. (2002). The aggregation of imprecise probabilities. *Journal of Statistical Planning and Inference 105*(1), 265–282.

Ouchi, F. (2004). A literature review on the use of expert opinion in probabilistic risk analysis.

Paolillo, R. and J. Lorenz (2018, 09). How different homophily preferences mitigate and spur ethnic and value segregation: Schelling's model extended. *Advances in Complex Systems 21*, 1850026.

Pedersen, A. P. and G. Wheeler (2014). Demystifying dilation. *Erkenntnis 79*(6), 1305–1342.

Pedersen, A. P. and G. Wheeler (2015). Dilation, disintegrations, and delayed decisions. In *Proceedings of the 9th.*

Pettigrew, R. (2017, Nov). Aggregating incoherent agents who disagree. *Synthese*.

Predd, J. B., R. Seiringer, E. H. Lieb, D. N. Osherson, H. V. Poor, and S. R. Kulkarni (2009, October). Probabilistic coherence and proper scoring rules. *IEEE Trans. Inf. Theor. 55*(10), 4786–4792.

Raiffa, H. (1968). *Decision Analysis: Introductory Lectures on Choices under Uncertainty.* Random House.

Ramsey, F. P. (1990). Truth and probability. In D. H. Mellor (Ed.), *Philosophical Papers*, pp. 52–109. Cambridge University Press.

Redacted. Redacted.

Resnik, M. D. (1983, Dec). A restriction on a theorem of Harsanyi. *Theory and Decision 15*(4), 309320.

Rockafellar, R. T. (1970). *Convex Analysis*. Number 28. Princeton University Press.

Rogers, T. and A. J. McKane (2011, 04). A unified framework for schelling's model of segregation. *Journal of Statistical Mechanics: Theory and Experiment 2011.*

Russell, J. S., J. Hawthorne, and L. Buchak (2015). Groupthink. *Philosophical Studies 172*(5), 1287–1309.

Savage, L. (1972, originally published in 1954). *The Foundations of Statistics*. New York: John Wiley and Sons.

Schelling, T. (2006). *Micromotives and Macrobehavior*. W. W. Norton.

Schelling, T. C. (1969). Models of segregation. *American Economic Review* (59), 488–493.

Schelling, T. C. (1971a). Dynamic models of segregation. *The Journal of Mathematical Sociology 1*(2), 143–186.

Schelling, T. C. (1971b). Dynamic models of segregation. *The Journal of Mathematical Sociology 1*(2), 143–186.

Schervish, M. and T. Seidenfeld (1990). An approach to consensus and certainty with increasing evidence. *Journal of Statistical Planning and Inference 25*(3), 401–414.

Seidenfeld, T. (1986). Entropy and uncertainty. *Philosophy of Science*, 467–491.

Seidenfeld, T. (1993). Outline of a theory of partially ordered preferences. *Philosophical Topics 21*(1), 173–189.

Seidenfeld, T., J. B. Kadane, and M. J. Schervish (1989). On the shared preferences of two bayesian decision makers. *The Journal of Philosophy 86*(5), 225–244.

Seidenfeld, T., M. J. Schervish, and J. B. Kadane (2010). Coherent choice functions under uncertainty. *Synthese 172*(1), 157–176.

Seidenfeld, T. and L. Wasserman (1993). Dilation for sets of probabilities. *The Annals of Statistics 21*(3), 1139–1154.

Sen, A. (1974, Feb). Rawls versus Bentham: An axiomatic examination of the pure distribution problem. *Theory and Decision 4*(3), 301309.

Sen, A. (1977). *Non-Linear Social Welfare Functions: A Reply to Professor Harsanyi*, pp. 297302. Dordrecht: Springer Netherlands.

Sen, A. (2005). Social choice theory. In K. J. Arrow and M. Intriligator (Eds.), *Handbook of Mathematical Economics* (2 ed.), Volume 3, Chapter 22, pp. 1073–1181. Elsevier.

Shelke, S. and V. Attar (2019). Source detection of rumor in social network a review. *Online Social Networks and Media 9*, 30 – 42.

Skyrms, B. (1986). *Choice and Chance: An Introduction to Inductive Logic* (3rd Edition ed.). Belmont: Wadsworth Publishing Company.

Smith, A. (2002). *Adam Smith: The Theory of Moral Sentiments*. Cambridge Texts in the History of Philosophy. Cambridge University Press.

Smith, C. A. B. (1961). Consistency in statistical inference and decision. *Journal of the Royal Statistical Society. Series B (Methodological) 23*(1), pp. 1–37.

Smith Adam 1723-1790 (2000). *The wealth of nations / Adam Smith ; introduction by Robert Reich ; edited, with notes, marginal summary, and enlarged index by Edwin Cannan.* New York : Modern Library, 2000. Originally published: An inquiry into the nature and causes of the wealth of nations / Adam Smith ; edited ... by Edwin Cannan. 1994. With new introd. by Robert Reich.;Includes bibliographical references and index.

Spohn, W. (2012). *The Laws of Belief: Ranking Theory and Its Philosophical Applications.* Oxford University Press, USA.

Stewart, R. T. and I. O. Quintana (2016). Probabilistic Opinion Pooling with Imprecise Probabilities. *Journal of Philosophical Logic*, Forthcoming.

Stewart, R. T. and I. O. Quintana (2018, Jun). Learning and pooling, pooling and learning. *Erkenntnis 83*(3), 369–389.

Stone, M. (1961). The opinion pool. *The Annals of Mathematical Statistics 32*(4), 1339–1342.

Uzquiano, G. (2018). Groups: Toward a theory of plural embodiment. *Journal of Philosophy 115*(8), 423–452.

van Fraassen, B. C. (1989). *Laws and Symmetry.* Clarendon Press Oxford.

Von Neumann, J. and O. Morgenstern (1947). *Theory of Games and Economic Behavior.* Princeton University Press.

Wagner, C. (2002). Probability kinematics and commutativity. *Philosophy of Science 69*(2), 266–278.

Wagner, C. (2009). Jeffrey conditioning and external bayesianity. *Logic Journal of IGPL 18*(2), 336–345.

Walley, P. (1982). The elicitation and aggregation of beliefs. Technical Report 23, Department of Statistics, University of Warwick, Coventry CV4 7AL, England.

Walley, P. (1991). *Statistical reasoning with imprecise probabilities*. Chapman and Hall London.

Wasserman, L. (1993). Review: Statistical reasoning with imprecise probabilities by peter walley. *Journal of the American Statistical Association 88*(422), pp. 700–702.

Wasserman, L. and T. Seidenfeld (1994). The dilation phenomenon in robust bayesian inference. *Journal of Statistical Planning and Inference 40*(2), 345–356.

Watts, D. J. and S. H. Strogatz (1998, June). Collective dynamics of 'small-world' networks. *Nature 393*(6684), 440–442.

Weymark, J. A. (1991). A reconsideration of the harsanyi–sen debate on utilitarianism. In J. Elster and J. E. Roemer (Eds.), *Interpersonal Comparisons of Well-Being*, pp. 255. Cambridge University Press.

Wilensky, U. (1997). Netlogo segregation model. *Center for Connected Learning and Computer-Based Modeling*, 1850026.

Williams, P. M. (1980a). Bayesian conditionalisation and the principle of minimum information. *British Journal for the Philosophy of Science*, 131–144.

Williams, P. M. (1980b). Bayesian conditionalisation and the principle of minimum information. *The British Journal for the Philosophy of Science 31*(2), 131–144.

Zhang, J. (2009, 01). Tipping and residential segregation: A unified schelling model. *Institute for the Study of Labor (IZA), IZA Discussion Papers 51*.

Netlogo Instructions and Repositories

Setting up the Software

In order to download the software you can click *here*, or copy-paste the following link in your browser:

- https://ccl.northwestern.edu/netlogo/download.shtml

You should select version 6.0.4 or higher. For the latest version you can go *here*, or copy-paste the following link:

- https://ccl.northwestern.edu/netlogo/6.1.0-RC2/

That page also describes the installation process according to your operating system.

Once you were able to open the program, you should be ready to go to the next section.

The model makes use of the networks extension [nw] which is included in versions 5.0.1 or higher. Furthermore, the extension's grammar and applications changed slightly in the transition to newer versions. The code included in this dissertation works for versions 6.0.4 or 6.1.0-RC2 (latest). Nothing secures future versions will keep the grammar the same.

Simulations' Repository

You can find the repository for the simulations *here*, or once again copy-pasting the following link in your browser:

- https://drive.google.com/drive/folders/1LJfoCq3sijNFbvmbxT3S260vqoEacWPr?usp=sharing

In these repositories you will find different types of files:

- Extension ".nlogo" corresponds to the NetLogo models, which will be briefly described in the next sections.
- Extension ".csv" corresponds to the data sets of the simulation results. They should be reproducible, modulo some random parameters, using the NetLogo models.
- Extensions ".R" and ".Rmd" correspond to the R and R-Markdown codes that I used to make the statistical analysis and plotting. In their present form, they are not at all elegant. That is why I excluded them from the material. They will be cleaned up in not long, and included only if the committee thinks it is relevant.
- Extension ".html" allows you to visualize the R-Markdowns.

The next sections will give a very brief introduction on how to use and explore the models.

Schelling on Social Networks Model

Begin by downloading the file "Schelling Segregation Final.nlogo" from the repository, and opening it using NetLogo.

Here a brief example on how to explore the model.

(1) Begin by setting the "save-data" switch at the top right to "Off"; otherwise experimental results would be stored.
(2) Set up the desired number of nodes using the "num-nodes" slider on the top left.
(3) Set up the "link-prob" slider near the center, with the ER random parameter. Intuitively, each node is expected to have link-prob * 100 links.
(4) Choose whether the network to be generated is going to be directed or not.

(5) Click “General Setup”. Alternative, the buttons for setting up a network or a directed network.

(6) At this point you should have a network with colorful agents around. Now its time to divide them into tribes.

(7) Below the Schelling Online Segregation Model banner, there are two sliders corresponding to tribes. In the two-tribe model, you get to decide what the proportion of blue (vs red) agents is. In the multi tribe, how many tribes you want. So decide accordingly.

(8) Click the “Setup” button below the Two Tribe or Multi Tribe notes.

(9) So we now have a network with tribes. It is time to set up the parameters of the dynamic. “Tolerance” and “Intolerance” correspond to heterophily and homophily. The Relink Dynamic is explained in the paper.

(10) Click “General Go” to observe the evolution of the dynamic.

(11) The plot shows the evolution of the proportion of (un)happiness and segregation. Unfortunately, it is not easy to design a code for the visualization in the main display that shows the segregation layout. This is only cosmetic, but hopefully I can develop it later.

This is not a full description of the model, but it will be helpful for an initial exploration.

Rumor Contagion on Social Networks Model

Begin by downloading the file “Rumor Contagion - Complete.nlogo” from the repository, and opening it using NetLogo.

Once again, a brief example on how to explore the model. This model is a bit more involved than the previous one because it includes a wide variety of networks, not just random networks. We

will exemplify with WattsStrogatz networks because its nice to visualize. You are free to inquire about other types.

(1) Begin by setting the "save-data" switch at the top right to "Off"; otherwise experimental results would be stored.

(2) Set up the desired number of nodes using the "num-nodes" slider on the top left.

(3) Set up the "n-links2" slider under the "Hybrid Models" title. Also set up the "rewire-prob" slider. The first corresponds to how many links each node will have, and the second with how much randomization will be inposed on the WattsStrogatz network.

(4) Click "WattsStrogatz model". You are free to experiment changing the parameters fixed before.

(5) You should have a network with colorful agents around. Now its time to divide them into the contagion categories.

(6) Below the Contagion banner, there are several sliders and one switch that you can use to determine recovery rate, spread chance, whether agents can gain resistance (this is the SIR model), the amount of initial infected agents, and whether those initial infected agents can recover (and gain resistance).

(7) Click the "Setup" button below those sliders. You will see a change of color that corresponds to whether agents are Suceptible (blue) or Infected (red) at initiation.

(8) Click "Go" to observe the contagion effect.

(9) The plot shows the evolution of the proportion of the population that corresponds to Susceptible, Infected or Recovered.

Appendices

APPENDIX A

Appendix to Chapter 2: Proofs

Proof of Proposition 1

PROOF. We carry out McConway's proof with minimal adjustments made for our framework (pp. 411-412 1981, Theorem 3.1).

WSFP $\Rightarrow$ MP. Assume that $\mathcal{F}$ has the WSFP, i.e., there is a function $\mathcal{G} : \mathscr{A} \times [0,1]^n \rightarrow \mathscr{P}([0,1])$ such that $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) = \mathcal{G}(A, \boldsymbol{p}_1(A), ..., \boldsymbol{p}_n(A))$. By WSFP, we have $\mathcal{F}([\boldsymbol{p}_1 \restriction_{\mathscr{A}'}], ..., [\boldsymbol{p}_n \restriction_{\mathscr{A}'}])(A)$ $= \mathcal{G}(A, [\boldsymbol{p}_1 \restriction_{\mathscr{A}'}](A), ..., [\boldsymbol{p}_n \restriction_{\mathscr{A}'}](A))$. Since $\mathcal{G}$ is a function and $\boldsymbol{p}_i(A) = [\boldsymbol{p}_i \restriction_{\mathscr{A}'}](A)$ for any $A \in \mathscr{A}'$ (all such $A \in \mathscr{A}'$ are also in $\mathscr{A}$), it follows that $\mathcal{G}(A, [\boldsymbol{p}_1 \restriction_{\mathscr{A}'}](A), ..., [\boldsymbol{p}_n \restriction_{\mathscr{A}'}](A)) = \mathcal{G}(A, \boldsymbol{p}_1(A), ..., \boldsymbol{p}_n(A))$ $= \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A)$. Hence, $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) = \mathcal{F}([\boldsymbol{p}_1 \restriction_{\mathscr{A}'}], ..., [\boldsymbol{p}_n \restriction_{\mathscr{A}'}])(A)$.

MP $\Rightarrow$ WSFP. Assume that $\mathcal{F}$ has the MP. Let $A \in \mathscr{A}$. We want to show that $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A)$ depends only on A and $\boldsymbol{p}_i(A), i = 1, ..., n$.

First, if $A = \varnothing$ or $A = \Omega$, then, since the range of $\mathcal{F}$ is $\mathscr{P}(\mathbb{P})$, $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A)$ depends only on A and $\boldsymbol{p}_i(A), i = 1, .., n$, for any profile because, setting $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) = \mathcal{G}(A, \boldsymbol{p}_1(A), ..., \boldsymbol{p}_n(A))$ and $\mathcal{F}(\boldsymbol{p}'_1, ..., \boldsymbol{p}'_n)(A) = \mathcal{G}(A, \boldsymbol{p}'_1(A), ..., \boldsymbol{p}'_n(A))$, it follows that $\mathcal{G}(A, \boldsymbol{p}_1(A), ..., \boldsymbol{p}_n(A)) = \mathcal{G}(A, \boldsymbol{p}'_1(A), ..., \boldsymbol{p}'_n(A))$.

Next, suppose that $\varnothing \neq A \neq \Omega$. Consider the σ-algebra $\mathscr{A}' = \{\varnothing, A, A^c, \Omega\}$. $\mathscr{A}$ contains A and has $\mathscr{A}'$ as a sub-algebra. By MP, then

$$\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) = \mathcal{F}([\boldsymbol{p}_1 \restriction_{\mathscr{A}'}], ..., [\boldsymbol{p}_n \restriction_{\mathscr{A}'}])(A).$$

$\mathscr{A}'$ is uniquely defined by A and any probability over $\mathscr{A}'$ is uniquely determined by the probability of A under that distribution. So the righthand side of the equation above is determined by A and $\boldsymbol{p}_i\restriction_{\mathscr{A}'}(A) = [\boldsymbol{p}_i\restriction_{\mathscr{A}'}](A) = \boldsymbol{p}_i(A)$. □

1. Proof of Lemma 1

PROOF. Let $Y = \{\boldsymbol{p} : \boldsymbol{p} = \sum_{i=1}^{n} \alpha_i \boldsymbol{p}_i$ such that $\alpha_i \geqslant 0$ for $i = 1, ..., n$ and $\sum_{i=1}^{n} \alpha_i = 1\}$. We want to show the following:

$$\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = conv\{\boldsymbol{p}_i : i = 1, ..., n\} = Y$$

The first equality we have by definition. In order to show the second equality, we have to show that Y is the smallest convex set containing $\{\boldsymbol{p}_i : i = 1, ..., n\}$. To show convexity, we show that for any two functions in Y, any convex combination of those functions is in Y. Suppose that $\boldsymbol{p}, \boldsymbol{p}' \in Y$. By assumption, $\boldsymbol{p} = \sum_{i=1}^{n} \alpha_i \boldsymbol{p}_i$ and $\boldsymbol{p}' = \sum_{i=1}^{n} \beta_i \boldsymbol{p}_i$. Consider $\boldsymbol{p}^\star = \gamma\boldsymbol{p} + (1-\gamma)\boldsymbol{p}' = \gamma(\sum_{i=1}^{n} \alpha_i \boldsymbol{p}_i) + (1-\gamma)\sum_{i=1}^{n} \beta_i \boldsymbol{p}_i$.

$$\begin{aligned}
\boldsymbol{p}^\star = &\ \gamma \textstyle\sum_{i=1}^{n} \alpha_i \boldsymbol{p}_i + (1-\gamma)\sum_{i=1}^{n} \beta_i \boldsymbol{p}_i \\
= &\ \textstyle\sum_{i=1}^{n} \gamma\alpha_i \boldsymbol{p}_i + \sum_{i=1}^{n}(1-\gamma)\beta_i \boldsymbol{p}_i \\
= &\ \textstyle\sum_{i=1}^{n}[\gamma\alpha_i \boldsymbol{p}_i + (1-\gamma)\beta_i \boldsymbol{p}_i] \\
= &\ \textstyle\sum_{i=1}^{n}[\gamma\alpha_i + (1-\gamma)\beta_i]\boldsymbol{p}_i \\
= &\ \textstyle\sum_{j=1}^{n} \delta_j \boldsymbol{p}_j
\end{aligned}$$

where $\delta_j = \gamma\alpha_j + (1-\gamma)\beta_j$. $\delta_j \geqslant 0$ for $j = 1, ..., n$ because every term is nonnegative. $\sum_{j=1}^{n} \delta_j = \sum_{i=1}^{n}[\gamma\alpha_i + (1-\gamma)\beta_i] = \sum_{i=1}^{n} \gamma\alpha_i + \sum_{i=1}^{n}(1-\gamma)\beta_i = \gamma\sum_{i=1}^{n} \alpha_i + (1-\gamma)\sum_{i=1}^{n} \beta_i = \gamma(1) + (1-\gamma)1 = 1$.

Hence, $\boldsymbol{p}^\star \in Y$, so Y is convex. If Y were not the smallest such set, then there would be some convex $Z \subsetneq Y$ such that $\{\boldsymbol{p}_i : i = 1, ..., n\} \subseteq Z$. But for any $\boldsymbol{p} \in Y, \boldsymbol{p}$ is a convex combination of the elements in $\{\boldsymbol{p}_i : i = 1, ..., n\}$. Since Z is convex and contains the $\boldsymbol{p}_i$, it follows that $\boldsymbol{p} \in Z$, which is a contradiction. □

2. Proof of Proposition 2

PROOF. Since $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) = \{\boldsymbol{p}(A) : \boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)\}$, we let $\mathcal{G}$ of the SSFP be the convex hull operation applied to $\{\boldsymbol{p}_i(A) : i = 1, ..., n\}$. It is clear that $\mathcal{G}$ depends just on the individual probabilities for A. We need to show that

$$\{\boldsymbol{p}(A) : \boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)\} = conv\{\boldsymbol{p}_i(A) : i = 1, ..., n\}.$$

Trivially, the lefthand side includes $\{\boldsymbol{p}_i(A) : i = 1, ..., n\}$. Suppose $\boldsymbol{p}(A), \boldsymbol{p}'(A) \in \{\boldsymbol{p}(A) : \boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)\}$. Since $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ is convex, it follows immediately that any convex combination of $\boldsymbol{p}(A), \boldsymbol{p}'(A)$ is in $\{\boldsymbol{p}(A) : \boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)\}$. Finally, suppose that there is some convex $Z \subsetneq \{\boldsymbol{p}(A) : \boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)\}$ which contains $\{\boldsymbol{p}_i(A) : i = 1, ..., n\}$. But for any $\boldsymbol{p}(A) \in \{\boldsymbol{p}(A) : \boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)\}$, $\boldsymbol{p}(A)$ is a convex combination of the $\boldsymbol{p}_i(A)$ since every $\boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ is such a convex combination of the $\boldsymbol{p}_i$ (Lemma 1). Hence, $\boldsymbol{p}(A) \in Z$, contrary to our supposition. So, the equality holds and the SWFP is satisfied.

But since SWFP clearly implies WSFP, WSFP is satisfied, too. By Proposition 1, it follows immediately that $\mathcal{F}$ has the MP.

Because $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ is a set of probability functions, $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(\varnothing) = \{0\}$. Let $\boldsymbol{p}_i(A) = 0$, $i = 1, ..., n$. Since there is a function, $\mathcal{G}$, such that $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) = \mathcal{G}(\boldsymbol{p}_1(A), ..., \boldsymbol{p}_n(A))$, we have it that

$$\begin{aligned}\{0\} &= \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(\varnothing)\\ &= \mathcal{G}(\boldsymbol{p}_1(\varnothing), ..., \boldsymbol{p}_n(\varnothing))\\ &= \mathcal{G}(\boldsymbol{p}_1(A), ..., \boldsymbol{p}_n(A))\\ &= \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A)\end{aligned}$$

So, ZPP follows from SWFP.

For any profile $(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) \in \mathbb{P}^n$, if all $\boldsymbol{p}_i$ are identical, then the convex hull is just $\{\boldsymbol{p}_i\}$. So $\mathcal{F}$ satisfies *unanimity preservation.*

□

3. Proof of Lemma 2

We generalize a proof of a result due originally to Girón and Rios and Levi (Levi, 1978; Girón and Ríos, 1980) for updating on an *event* to updating on a common likelihood function.

Proof. We want to show that $\mathcal{F}^\lambda(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ is convex. That is, given any two members, $\boldsymbol{p}^\lambda, \boldsymbol{p}'^\lambda \in \mathcal{F}^\lambda(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ and $\alpha \in [0, 1]$, $\boldsymbol{p}^\star = \alpha\boldsymbol{p}^\lambda + (1-\alpha)\boldsymbol{p}'^\lambda$ is in $\mathcal{F}^\lambda(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$. If there is a convex combination of $\boldsymbol{p}$ and $\boldsymbol{p}'$, $\boldsymbol{p}_\star$, such that $\boldsymbol{p}_\star^\lambda = \boldsymbol{p}^\star$, then the convexity of $\mathcal{F}^\lambda(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ is established as a consequence of the convexity of $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$. Where $\boldsymbol{p}_\star^\lambda(\cdot) = \frac{\boldsymbol{p}_\star(\cdot)\lambda(\cdot)}{\sum_{\omega'\in\Omega}\boldsymbol{p}_\star(\omega')\lambda(\omega')} = \frac{\beta\boldsymbol{p}(\cdot)\lambda(\cdot)+(1-\beta)\boldsymbol{p}'(\cdot)\lambda(\cdot)}{\beta\sum_{\omega'\in\Omega}\boldsymbol{p}(\omega')\lambda(\omega')+(1-\beta)\sum_{\omega'\in\Omega}\boldsymbol{p}'(\omega')\lambda(\omega')}$, for any α we want to find some β such that

$$\boldsymbol{p}^\star(\cdot) = \alpha\boldsymbol{p}^\lambda(\cdot) + (1-\alpha)\boldsymbol{p}'^\lambda(\cdot) = \frac{\beta\boldsymbol{p}(\cdot)\lambda(\cdot) + (1-\beta)\boldsymbol{p}'(\cdot)\lambda(\cdot)}{\beta\sum_{\omega'\in\Omega}\boldsymbol{p}(\omega')\lambda(\omega') + (1-\beta)\sum_{\omega'\in\Omega}\boldsymbol{p}'(\omega')\lambda(\omega')} = \boldsymbol{p}_\star^\lambda(\cdot).$$

For $\beta = \frac{\alpha\sum_{\omega*\in\Omega}\boldsymbol{p}'(\omega*)\lambda(\omega*)}{\alpha\sum_{\omega*\in\Omega}\boldsymbol{p}'(\omega*)\lambda(\omega*)+(1-\alpha)\sum_{\omega*\in\Omega}\boldsymbol{p}(\omega*)\lambda(\omega*)}$, the equality is verifiable with some tedious algebra. □

4. Proof of Proposition 3

PROOF. We must show that convex IP pooling functions are externally Bayesian, i.e., $\mathcal{F}(\boldsymbol{p}_1^\lambda, ..., \boldsymbol{p}_n^\lambda) = \mathcal{F}^\lambda(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ (provided the relevant profiles are in the domain of $\mathcal{F}$).

$\mathcal{F}(\boldsymbol{p}_1^\lambda, ..., \boldsymbol{p}_n^\lambda) \subseteq \mathcal{F}^\lambda(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$. Trivially, for each $i = 1, ..., n$, $\boldsymbol{p}_i^\lambda \in \mathcal{F}^\lambda(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$. By Lemma 2, $\mathcal{F}^\lambda(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ is convex. It follows that $conv\{\boldsymbol{p}_i^\lambda : i = 1, ..., n\} = \mathcal{F}(\boldsymbol{p}_1^\lambda, ..., \boldsymbol{p}_n^\lambda) \subseteq \mathcal{F}^\lambda(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$.

$\mathcal{F}^\lambda(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) \subseteq \mathcal{F}(\boldsymbol{p}_1^\lambda, ..., \boldsymbol{p}_n^\lambda)$. By Lemma 1, any $\boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ can be expressed as the convex combination of the n extreme points generating $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$, i.e., $\boldsymbol{p} = \sum_{i=1}^n \alpha_i \boldsymbol{p}_i$ where $\alpha_i \geqslant 0$ for $i = 1, ..., n$ and $\sum_{i=1}^n \alpha_i = 1$. By definition,

$$\mathcal{F}^\lambda(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = \{\boldsymbol{p}^\lambda : \boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) \text{ and } \boldsymbol{p}^\lambda(\cdot) = \frac{\boldsymbol{p}(\cdot)\lambda(\cdot)}{\sum_{\omega' \in \Omega} \boldsymbol{p}(\omega')\lambda(\omega')}\}$$

We show that any member of $\mathcal{F}^\lambda(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ is identical to some member of $\mathcal{F}(\boldsymbol{p}_1^\lambda, ..., \boldsymbol{p}_n^\lambda)$.

$$\begin{aligned}
\boldsymbol{p}^\lambda(\omega) &= \frac{\boldsymbol{p}(\omega)\lambda(\omega)}{\sum_{\omega' \in \Omega} \boldsymbol{p}(\omega')\lambda(\omega')} && \text{[Definition]} \\
&= \frac{\sum_{i=1}^n \alpha_i \boldsymbol{p}_i(\omega)\lambda(\omega)}{\sum_{\omega' \in \Omega} \sum_{i=1}^n \alpha_i \boldsymbol{p}_i(\omega')\lambda(\omega')} && \text{[Lemma 1]} \\
&= \frac{\sum_{i=1}^n \alpha_i \boldsymbol{p}_i^\lambda(\omega) \cdot \sum_{\omega' \in \Omega} \boldsymbol{p}_i(\omega')\lambda(\omega')}{\sum_{\omega' \in \Omega} \sum_{i=1}^n \alpha_i \boldsymbol{p}_i(\omega')\lambda(\omega')} && [\boldsymbol{p}_i(\omega)\lambda(\omega) = \boldsymbol{p}_i(\omega)^\lambda \cdot \textstyle\sum_{\omega' \in \Omega} \boldsymbol{p}_i(\omega')\lambda(\omega')] \\
&= \sum_{j=1}^n \beta_j \boldsymbol{p}_j^\lambda(\omega) \in \mathcal{F}(\boldsymbol{p}_1^\lambda, ..., \boldsymbol{p}_n^\lambda) && \text{[Algebra]}
\end{aligned}$$

where $\beta_j = \frac{\alpha_j \cdot \sum_{\omega' \in \Omega} \boldsymbol{p}_j(\omega')\lambda(\omega')}{\sum_{\omega' \in \Omega} \sum_{i=1}^n \alpha_i \boldsymbol{p}_i(\omega')\lambda(\omega')}$ with $\beta_j \geqslant 0$ for all $j = 1, ..., n$ and $\sum_{j=1}^n \beta_j = 1$.

□

5. Proof of Proposition 4

PROOF. We provide a very simple type of counterexample to individualwise Bayesianity, though counterexamples are plentiful. Consider the profile $(\boldsymbol{p}_1, \boldsymbol{p}_2)$ for $n = 2$ agents such that $\boldsymbol{p}_1 = \boldsymbol{p}_2$. Individualwise Bayesianity requires that $\mathcal{F}(\boldsymbol{p}_1, \boldsymbol{p}_2^{\lambda}) = \mathcal{F}^{\lambda}(\boldsymbol{p}_1, \boldsymbol{p}_2)$ (provided both $(\boldsymbol{p}_1, \boldsymbol{p}_2)$ and $(\boldsymbol{p}_1, \boldsymbol{p}_2^{\lambda})$ are in the domain of $\mathcal{F}$). By Proposition 3 (external Bayesianity), it follows that $\mathcal{F}^{\lambda}(\boldsymbol{p}_1, \boldsymbol{p}_2) = \mathcal{F}(\boldsymbol{p}_1^{\lambda}, \boldsymbol{p}_2^{\lambda})$. But since $\boldsymbol{p}_1 = \boldsymbol{p}_2$, it follows that $\boldsymbol{p}_1^{\lambda} = \boldsymbol{p}_2^{\lambda}$. By unanimity (Proposition 2), then, we have $\mathcal{F}(\boldsymbol{p}_1^{\lambda}, \boldsymbol{p}_2^{\lambda}) = \{\boldsymbol{p_i}^{\lambda}\}$, where $\boldsymbol{p}_i^{\lambda} = \boldsymbol{p}_1^{\lambda} = \boldsymbol{p}_2^{\lambda}$. However, in general $\boldsymbol{p}_i \neq \boldsymbol{p}_i^{\lambda}$ and so $\mathcal{F}(\boldsymbol{p}_1, \boldsymbol{p}_2^{\lambda})$ is *not* a singleton. It follows that, in general, $\mathcal{F}(\boldsymbol{p}_1, \boldsymbol{p}_2^{\lambda}) \neq \mathcal{F}^{\lambda}(\boldsymbol{p}_1, \boldsymbol{p}_2)$.

□

6. Proof of Proposition 5

PROOF. Suppose that $\boldsymbol{p}_i(A|B) = \boldsymbol{p}_i(A)$ for $i = 1, ..., n$. We want to show that $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) = \mathcal{F}^B(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A)$. Consider $\boldsymbol{p}^{\star}(A) \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A)$ and $\boldsymbol{p}_{\star}(A) \in \mathcal{F}^B(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A)$. By Lemma 1, $\boldsymbol{p}^{\star}(A) = \sum_{i=1}^{n} \alpha_i \boldsymbol{p}_i(A)$, for appropriate α_i. By Proposition 3 (external Bayesianity), $\mathcal{F}^B(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) = \mathcal{F}(\boldsymbol{p}_1^B, ..., \boldsymbol{p}_n^B)(A)$ (Proposition 3 holds for standard conditionalization since standard conditionalization is a special case of updating on a likelihood function, as noted in the body of the essay). So, we have $\boldsymbol{p}_{\star}(A) = \sum_{i=1}^{n} \beta_i \boldsymbol{p}_i^B(A)$, for appropriate β_i, again by Lemma 1. By hypothesis $\boldsymbol{p}_i^B(A) = \boldsymbol{p}_i(A)$ for $i = 1, ..., n$. Hence, $\boldsymbol{p}_{\star}(A) = \sum_{i=1}^{n} \beta_i \boldsymbol{p}_i^B(A) = \sum_{i=1}^{n} \beta_i \boldsymbol{p}_i(A)$. Letting $\alpha_i = \beta_i$, it follows that $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A) = \mathcal{F}^B(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)(A)$. □

7. Proof of Proposition 6

PROOF. We show first that $\mathcal{M}_n(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) \subseteq conv(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ for all $(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) \in \mathbb{P}^n$. Since there are at least three disjoint events, $A_1, A_2, A_3 \in \mathscr{A}$, following (Lehrer and Wagner, 1981, Theorems 6.4, 6.7) and (McConway, 1981, Theorem 3.3), we can exploit techniques and results

for functional equations. For any numbers $a_i, b_i \in [0,1]$ with $a_i + b_i \in [0,1]$, define a sequence of probability measures, $\boldsymbol{p}_i,\ i = 1, ..., n$ by setting

$$\begin{aligned}\boldsymbol{p}_i(A_1) &= \quad a_i \\ \boldsymbol{p}_i(A_2) &= \quad b_i \\ \boldsymbol{p}_i(A_3) &= \quad 1 - a_i - b_i\end{aligned}$$

Since it is the case that $\mathfrak{m}\big(\boldsymbol{p}_1(\cdot), ..., \boldsymbol{p}_n(\cdot)\big) \in \mathbb{P}$ for all $(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) \in \mathbb{P}^n$ and every $\mathfrak{m}\big(\boldsymbol{p}_1(\cdot), ..., \boldsymbol{p}_n(\cdot)\big) \in \mathcal{M}_n(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$, we have it that $\mathfrak{m}\big(\boldsymbol{p}_1(A), ..., \boldsymbol{p}_n(A)\big) = \boldsymbol{p}(A)$, for some $\boldsymbol{p} \in \mathbb{P}$ and all $A \in \mathscr{A}$. Now, by the additivity of probability measures, $\boldsymbol{p}(A_1 \cup A_2) = \boldsymbol{p}(A_1) + \boldsymbol{p}(A_2)$. Hence, $\mathfrak{m}(a_1 + b_1, ..., a_n + b_n) = \mathfrak{m}\big(a_1, ..., a_n\big) + \mathfrak{m}\big(b_1, ..., b_n\big)$. So, $\mathfrak{m}$ satisfies Cauchy's multivariable functional equation. For each $i = 1, ..., n$, define $\mathfrak{m}_i(a) = \mathfrak{m}(0, ..., a, ..., 0)$, where a occupies the i-th position of the vector $(0, ..., a, ..., 0)$. It is clear that $\mathfrak{m}_i(a + b) = \mathfrak{m}_i(a) + \mathfrak{m}_i(b)$ for all $a, b \in [0,1]$ with $a + b \in [0,1]$. Because $\mathfrak{m}$ is nonnegative, so is $\mathfrak{m}_i,\ i = 1, ..., n$. By Theorem 3 of (Aczél and Oser, 2006, p. 48), it follows that there exists a nonnegative constant α_i such that $\mathfrak{m}_i(a) = \alpha_i a$ for all $a \in [0,1]$. By the Cauchy equation we have

$$\begin{aligned}\mathfrak{m}(a_1, ..., a_n) &= \mathfrak{m}(a_1, 0, ..., 0) + \mathfrak{m}(0, a_2, ..., a_n) \\ &= \mathfrak{m}(a_1, 0, ..., 0) + \mathfrak{m}(0, a_2, 0, ..., 0) + ... + \mathfrak{m}(0, ..., 0, a_n)\end{aligned}$$

So we have $\mathfrak{m}(a_1, ..., a_n) = \mathfrak{m}_1(a_1) + ... + \mathfrak{m}_n(a_n) = \alpha_1 a_1 + ... + \alpha_n a_n$. And since $\mathfrak{m}(1, ..., 1) = 1$ (by consideration of the probability of Ω), it follows that $\sum_{i=1}^n \alpha_i = 1$. Thus, $\mathfrak{m}$ is a convex combination.

Now, we want to show that $conv(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) \subseteq \mathcal{M}_n(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$. Let $\boldsymbol{p}$ be an element of $conv(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$. It is clear that there exists an $\mathfrak{m} \in \mathfrak{M}_n(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ such that $\mathfrak{m}\big(\boldsymbol{p}_1(\cdot), ..., \boldsymbol{p}_n(\cdot)\big) = \boldsymbol{p}$. And since $\boldsymbol{p}$ is just a convex combination, there exist weights $\alpha_1, ..., \alpha_n \in [0,1]$ such that $\sum_{i=1}^n \alpha_i = 1$ and $\boldsymbol{p} = \sum_{i=1}^n \alpha_i \boldsymbol{p}_i$. But for any other profile $(\boldsymbol{q}_1, ..., \boldsymbol{q}_n) \in \mathbb{P}^n$, taking any convex combination yields a probability measure. In particular, $\sum_{i=1}^n \alpha_i \boldsymbol{q}_i \in \mathbb{P}$. It follows that $\mathfrak{m} \in \bigcap_{\vec{\boldsymbol{q}} \in \mathbb{P}^n} \mathfrak{M}_n(\vec{\boldsymbol{q}})$. So, $\boldsymbol{p} = \mathfrak{m}\big(\boldsymbol{p}_1(\cdot), ..., \boldsymbol{p}_n(\cdot)\big) \in \mathcal{M}_n(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$, as desired.

The two inclusions above show that $\mathcal{M}_n(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = conv(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$. Hence, $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = conv(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ is equivalent to $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = \mathcal{M}_n(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$. □

APPENDIX B

Appendix to Chapter 3: Proofs

Proof of Proposition 8

PROOF. We follow through Wagner's proof for the precise case (2009, Theorem 3.3), adapting it for IP where necessary.

($\Rightarrow$) Assume that $\mathcal{F}$ is externally Bayesian, i.e., for all profiles and any likelihood function, $\mathcal{F}^{\lambda}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = \mathcal{F}(\boldsymbol{p}_1^{\lambda}, ..., \boldsymbol{p}_n^{\lambda})$. We want to show that, for all partitions $\boldsymbol{E} = \{E_k\}$ of Ω and all profiles in $\mathbb{P}^n$,

$$\begin{aligned}
\mathcal{F}_J^{\boldsymbol{E}}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) &= \left\{ \frac{\sum_k b_k \boldsymbol{p}[\cdot \in E_k]}{\sum_k b_k \boldsymbol{p}(E_k)} : \boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) \right\} \\
&= \mathcal{F}\left(\frac{\sum_k b_k \boldsymbol{p}_1[\cdot \in E_k]}{\sum_k b_k \boldsymbol{p}_1(E_k)}, ..., \frac{\sum_k b_k \boldsymbol{p}_n[\cdot \in E_k]}{\sum_k b_k \boldsymbol{p}_n(E_k)} \right) \\
&= \mathcal{F}(\boldsymbol{p}_{1J}^{\boldsymbol{E}}, ..., \boldsymbol{p}_{nJ}^{\boldsymbol{E}})
\end{aligned}$$

where the first and last equalities are definitional. Recall the definition of b_k: $b_k = \mathcal{B}(\boldsymbol{q}, \boldsymbol{p}; E_k : E_1) = \dfrac{\boldsymbol{q}(E_k)/\boldsymbol{q}(E_1)}{\boldsymbol{p}(E_k)/\boldsymbol{p}(E_1)}$, $k = 1, 2, ...$ Set $\lambda(\omega) = \sum_k b_k[\omega \in E_k]$. Wagner observes the following chain of equalities then obtains for $\boldsymbol{p}_i, i = 1, ..., n$ (2009, (3.10), p. 342):

$$(\star) \sum_{\omega \in \Omega} \lambda(\omega) \boldsymbol{p}_i(\omega) = \sum_{\omega \in \Omega} \boldsymbol{p}_i(\omega) \sum_k b_k[\omega \in E_k] = \sum_k b_k \sum_{\omega \in \Omega} \boldsymbol{p}_i(\omega)[\omega \in E_k] = \sum_k b_k \boldsymbol{p}_i(E_k)$$

Since each of the terms $b_k \boldsymbol{p}_i(E_k)$ is positive and $\sum_k b_k \boldsymbol{p}_i(E_k) < \infty$, λ is a likelihood function for $\boldsymbol{p}_i$, with $\boldsymbol{p}_i^\lambda$ a defined, updated pmf for $i = 1, ..., n$. Using $(\star)$, we can obtain

$$\mathcal{F}(\boldsymbol{p}_{1J}^{\boldsymbol{E}}, ..., \boldsymbol{p}_{nJ}^{\boldsymbol{E}}) = \mathcal{F}\left(\frac{\boldsymbol{p}_1 \lambda(\cdot)}{\sum_{\omega' \in \Omega} \boldsymbol{p}_1(\omega')\lambda(\omega')}, ..., \frac{\boldsymbol{p}_n \lambda(\cdot)}{\sum_{\omega' \in \Omega} \boldsymbol{p}_n(\omega')\lambda(\omega')} \right)$$

by substituting, for each $i = 1, ..., n$, $\lambda(\cdot)$ for $\sum_k b_k[\omega \in E_k]$ in the numerator and $\sum_{\omega' \in \Omega} \boldsymbol{p}_i(\omega')\lambda(\omega')$ for $\sum_k b_k \boldsymbol{p}_i(E_k)$ in the denominator. But by definition,

$$\mathcal{F}\left(\frac{\boldsymbol{p}_1 \lambda(\cdot)}{\sum_{\omega' \in \Omega} \boldsymbol{p}_1(\omega')\lambda(\omega')}, ..., \frac{\boldsymbol{p}_n \lambda(\cdot)}{\sum_{\omega' \in \Omega} \boldsymbol{p}_n(\omega')\lambda(\omega')} \right) = \mathcal{F}(\boldsymbol{p}_1^\lambda, ..., \boldsymbol{p}_n^\lambda)$$

and by assumption $\mathcal{F}(\boldsymbol{p}_1^\lambda, ..., \boldsymbol{p}_n^\lambda) = \mathcal{F}^\lambda(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$. By definition, $\mathcal{F}^\lambda(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = \{\boldsymbol{p}^\lambda : \boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)\}$. But, for all $\boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$, $\boldsymbol{p}^\lambda = \frac{\sum_k b_k \boldsymbol{p}[\cdot \in E_k]}{\sum_k b_k \boldsymbol{p}(E_k)}$. Hence, $\mathcal{F}^\lambda(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = \mathcal{F}_J^{\boldsymbol{E}}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$. So, $\mathcal{F}_J^{\boldsymbol{E}}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = \mathcal{F}(\boldsymbol{p}_{1J}^{\boldsymbol{E}}, ..., \boldsymbol{p}_{nJ}^{\boldsymbol{E}})$ follows from the assumption.

($\Leftarrow$) Suppose that $\mathcal{F}$ satisfies CJC_W and that λ is a likelihood function for $\boldsymbol{p}_i, i = 1, ..., n$. Let $(\omega_1, \omega_2, ...)$ be a list of all of those $\omega \in \Omega$ such that $\lambda(\omega) > 0$, and let $\boldsymbol{E} = \{E_1, E_2, ...\}$, where $E_i := \{\omega_i\}$. Setting $b_k = \frac{\lambda(\omega_k)}{\lambda(\omega_1)}$ for $k = 1, 2, ...$, it follows that $b_k > 0$ and that $b_1 = 1$. Since λ is a likelihood for $\boldsymbol{p}_i, i = 1, ..., n$, we have $\sum_k b_k \boldsymbol{p}_i(E_k) < \infty, i = 1, ..., n$, and that $(\boldsymbol{q}_1, ..., \boldsymbol{q}_n) \in \mathbb{P}^n$, where $\boldsymbol{q}_i(\omega) := \frac{\sum_k b_k \boldsymbol{p}_i(\omega)[\omega \in E_k]}{\sum_k b_k \boldsymbol{p}_i(E_k)}$. From CJC_W, it follows that 1) $0 < \sum_k b_k \boldsymbol{p}(E_k) < \infty$ for all $\boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$, and that 2) $\mathcal{F}_J^{\boldsymbol{E}}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = \mathcal{F}(\boldsymbol{p}_{1J}^{\boldsymbol{E}}, ..., \boldsymbol{p}_{nJ}^{\boldsymbol{E}})$. 1) implies that $0 < \sum_{\omega \in \Omega} \lambda(\omega)\boldsymbol{p}(\omega) < \infty$ for all $\boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$, and 2) implies that $\mathcal{F}^\lambda(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) = \mathcal{F}(\boldsymbol{p}_1^\lambda, ..., \boldsymbol{p}_n^\lambda)$ (since substituting the definition of b_k in terms of λ in $\frac{\sum_k b_k \boldsymbol{p}_i(\omega)[\omega \in E_k]}{\sum_k b_k \boldsymbol{p}_i(E_k)}$, the formula for obtaining the $\boldsymbol{q}_i$, reduces that formula to the formula for updating on that λ).

□

Proof of Proposition 11

PROOF. We provide a case in which convex IP pooling and Jeffrey conditionalization *as standardly construed* do not commute. Let $\boldsymbol{q}_i$ come from $\boldsymbol{p}_i$ by Jeffrey conditionalization, and let $\boldsymbol{q}$ be a common posterior distribution over partition $\boldsymbol{E}$ for $\boldsymbol{p}_i$, $i = 1, ..., n$. Let $\mathcal{F}_J^{\boldsymbol{E}}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ come from $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ by Jeffrey conditionalizing each $\boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ using $\boldsymbol{q}$, the common posterior distribution over $\boldsymbol{E}$. We offer a counterexample to commutativity in which $\mathcal{F}_J^{\boldsymbol{E}}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n) \neq \mathcal{F}(\boldsymbol{q}_1, ..., \boldsymbol{q}_n)$.

Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, and consider the following two pmfs.

TABLE 16. Priors

	ω_1	ω_2	ω_3	ω_4
$\boldsymbol{p}_1$	1/4	1/4	1/4	1/4
$\boldsymbol{p}_2$	1/8	1/2	1/4	1/8

Let $\boldsymbol{E} = \{E_1, E_2\}$ with $E_1 = \{\omega_1, \omega_2\}$ and $E_2 = \{\omega_3, \omega_4\}$ be a partition of Ω. Jeffrey updating both pmfs using $\boldsymbol{q}$, where $\boldsymbol{q}(E_1) = 2/3$ and $\boldsymbol{q}(E_2) = 1/3$, we obtain the following posteriors.

TABLE 17. Posteriors

	ω_1	ω_2	ω_3	ω_4
$\boldsymbol{q}_1$	1/3	1/3	1/6	1/6
$\boldsymbol{q}_2$	2/15	8/15	2/9	1/9

Consider the $.50 - .50$ mixture of $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$, $\boldsymbol{p}^\star = 0.5\boldsymbol{p}_1 + 0.5\boldsymbol{p}_2$. It is clear that $\boldsymbol{p}^\star \in \mathcal{F}(\boldsymbol{p}_1, \boldsymbol{p}_2)$. Jeffrey conditionalizing $\boldsymbol{p}^\star$ with $\boldsymbol{q}$ gives us $\boldsymbol{q}^\star$. In particular, $\boldsymbol{q}^\star(\omega_1) = 2/9$ and $\boldsymbol{q}^\star(\omega_3) = 4/21$. It is clear that $\boldsymbol{q}^\star \in \mathcal{F}_{\boldsymbol{E}}^J(\boldsymbol{p}_1, \boldsymbol{p}_2)$. Any $\boldsymbol{q}_\star \in \mathcal{F}(\boldsymbol{q}_1, \boldsymbol{q}_2)$ is of the form $\boldsymbol{q}_\star = \alpha\boldsymbol{q}_1 + (1 - \alpha)\boldsymbol{q}_2$ for $\alpha \in [0, 1]$.

Suppose that $\mathcal{F}_J^{\boldsymbol{E}}(\boldsymbol{p}_1, \boldsymbol{p}_2) = \mathcal{F}(\boldsymbol{q}_1, \boldsymbol{q}_2)$. Then, there is a $\boldsymbol{q}_\star \in \mathcal{F}(\boldsymbol{q}_1, \boldsymbol{q}_2)$ such that $\boldsymbol{q}^\star = \boldsymbol{q}_\star$. In particular, $\boldsymbol{q}_\star(\omega_1) = 2/9$ and $\boldsymbol{q}_\star(\omega_3) = 4/21$. Letting $\boldsymbol{q}_\star(\omega_1) = 2/9$, we can compute α.

$$2/9 = \boldsymbol{q}_\star(\omega_1) = \alpha\boldsymbol{q}_1(\omega_1) + (1 - \alpha)\boldsymbol{q}_2(\omega_1) = \alpha 1/3 + (1 - \alpha)2/15$$

Solving, we get $\alpha = 4/9$. However, we are supposed to have $\boldsymbol{q}_\star(\omega_3) = 4/21$. For $\alpha = 4/9$, that is not the case.

$$\boldsymbol{q}_\star(\omega_3) = \alpha \boldsymbol{q}_1(\omega_3) + (1-\alpha)\boldsymbol{q}_2(\omega_3) = 4/9(1/6) + 5/9(2/9) = 16/81 > 4/21 = \boldsymbol{q}^\star(\omega_3)$$

It follows that $\mathcal{F}_J^{\boldsymbol{E}}(\boldsymbol{p}_1, \boldsymbol{p}_2) \neq \mathcal{F}(\boldsymbol{q}_1, \boldsymbol{q}_2)$.

□

Proof of Proposition 12

Proof. We want to show that $\mathcal{F}(\boldsymbol{q}_1, ..., \boldsymbol{q}_n) = \mathcal{F}_I^E(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$, where $\boldsymbol{q}_i$ comes from $\boldsymbol{p}_i$ by general imaging on E, and $\mathcal{F}_I^E(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ comes from $\mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ by general imaging each $\boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ on E. Again, we show both inclusions. In the proofs, we appeal to the fact any element of a convex set is some convex combination of the generating, extreme points: For any $\boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n), \boldsymbol{p} = \sum_{i=1}^n \alpha_i \boldsymbol{p}_i$, where $\alpha_i \geqslant 0$ for $i = 1, ..., n$, and $\sum_{i=1}^n \alpha_i = 1$ (see, e.g., Redacted, Redacted, Lemma 1).

Let $\boldsymbol{q} \in \mathcal{F}(\boldsymbol{q}_1, ..., \boldsymbol{q}_n)$. So, $\boldsymbol{q} = \sum_{i=1}^n \alpha_i \boldsymbol{q}_i$. Since $\boldsymbol{q}$ is a linear pool of $\boldsymbol{q}_i$ for $i = 1, ..., n$, by Gärdenfors' result, Theorem 10, $\boldsymbol{q}$ is also the result of imaging $\boldsymbol{p} = \sum_{i=1}^n \alpha_i \boldsymbol{p}_i$ on E, because linear pooling and general imaging commute. Since $\boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$, it follows that $\boldsymbol{q} \in \mathcal{F}_I^E(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$.

For the other direction, assume that $\boldsymbol{q} \in \mathcal{F}_I^E(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$. So, $\boldsymbol{q}$ is the result of general imaging some $\boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n)$ on E. For any $\boldsymbol{p} \in \mathcal{F}(\boldsymbol{p}_1, ..., \boldsymbol{p}_n), \boldsymbol{p} = \sum_{i=1}^n \alpha_i \boldsymbol{p}_i$. By Gärdenfors' result, $\boldsymbol{q} = \sum_{i=1}^n \alpha_i \boldsymbol{q}_i$, where the $\boldsymbol{q}_i$ come from the $\boldsymbol{p}_i$ by general imaging on E, because general imaging and linear pooling commute. But then it follows that $\boldsymbol{q} \in \mathcal{F}(\boldsymbol{q}_1, ..., \boldsymbol{q}_n)$.

□

APPENDIX C

Appendix to Chapter 5: Proofs

Proof of Proposition 1: Characterizing Altruistic Utilities as weighted Solitary Utilities

PROOF. The existence of the solitary and altruistic utility functions su_i and au_i that represent the solitary and altruistic preferences respectively follow directly from the standard von Neumann - Morgenstern result. Furthermore, this result secures that we can use any positive affine transformation as a utility representation of those preferences; this will become relevant later.

As Observation 1 points, we can extend each agent i's **solitary** preferences $\succcurlyeq_i$ to an ordering $\succcurlyeq_i$ over the **social** lotteries that also satisfies the von Neumann - Morgenstern axioms. So we get a utility function su_i^* that represents those preferences. Furthermore, since these are also defined up to affine transformations, we can assume without lost of generality that for all $L \in Y, L^i \in Y_i$, $su_i(L) = su^*(L^i)$ [where L^i is the i-marginal of L].

The rest of the proof is a simple applicatin of Harsanyi's theorem as stated in this essay.

Axiom 3 of Weak Altruism [Pareto] secures the satisfaction of (1), Pareto Indifference, and therefore guarantee the linear decomposition.

Axiom 4 of Strong Altruism [Pareto] secures the satisfaction of (2), and therefore that the α_j are positive. □

Proof of Proposition 2: Characterizing Spiteful Utilities as weighted Solitary Utilities

PROOF. Notice that in the context of Axioms 1,2 and 5, both Weak and Strong Spite entail Pareto Indifference. So by Harsanyi's Aggregation Theorem we are guaranteed the linear decomposition.

$$tu_{i*}(\prod_j L^j) = \sum_{j \in J} \beta_j . su_j(L^j)$$

It only remains to be shown that the parameters β_j are non-positive (and negative) and add up to -1.

For each $i \in I$, let O^i be such that $su_i(O^i) = 1$. We can safely define this because of Axiom 5. Also, we stipulate that $su_i(Z^i) = 0$ and $tu_{i*}(\prod_i Z^i) = 0$. In fact this last stipulation is used in the proof of Harsanyi's original result. Each su_i is now defined *uniquely* (not just up to affine transformations).

Now, let $O^j_* = \prod_{i \neq j} Z^i x O^j$. This is the lottery in which j gets O^j and the rest of the agents get their respective neutral lottery. Since $\forall i \ su_i(Z^i) = 0$, $su_j(O^j) = 1 > 0$ and the su_i are utility representations, we can make use of Weak (and Strong) *Spite*. In particular, WS secures that $0 \geqslant tu_{i*}(O^j_*) = \beta_j$ and SA that $0 > \beta_j$. □

Proof of Proposition 3: Characterizing Other-Related Utilities as weighted Solitary Utilities

PROOF. Once again, in the context of Axioms 1,2 and 5, Axiom 8 entails Pareto Indifference. Again linear decomposition is ensured.

$$tu_{i*}(\prod_j L^j) = \sum_{j \in J} \beta_j . su_j(L^j)$$

Using the same stipulations, su_i is now defined *uniquely*. We also have that for each $i \in I$, let O^i be such that $su_i(O^i) = 1$; that $su_i(Z^i) = 0$, and that $tu_{i*}(\prod_i Z^i) = 0$.

As before, for all $j \in J_1 \cup J_2$, let $O^j_* = \prod_{i \neq j} Z^i x O^j$. This is the lottery in which j gets O^j and the rest of the agents get their respective neutral lottery. Again, $\forall i \; su_i(Z^i) = 0$, $su_j(O^j) = 1 > 0$. So for $j \in J_1$, using Axiom 8 we get $\alpha_j =\geqslant tu_{i*}(O^j_*) > 0$; and for $j \in J_2$, it secures $0 \geqslant tu_{i*}(O^j_*) = \beta_j$. $\square$

APPENDIX D

Appendix to Chapter 6: Exploration

Network Structural Properties: Clustering and Average Path Lenght

The purpose of this appendix is to briefly study the effect that the Schelling dynamic had on the networks global and mean clustering, as well as its average path length.

Let us start with clustering. We measured each network's global clustering coefficients before and after introducing the dynamic, and we stored the difference. The global clustering coefficient of a network is defined based on the types of triplets in the network. A triplet consists of a central node and two of its neighbors. If its neighbors are also connected, its a closed triplet. If its neighbors are not connected, its an open triplet. The global clustering coefficient is simply the number of closed triplets in a network divided by the total number of triplets. We also defined a variable, "Degree of Intervention", that is simply the addition of the homophily and heterophily thresholds.

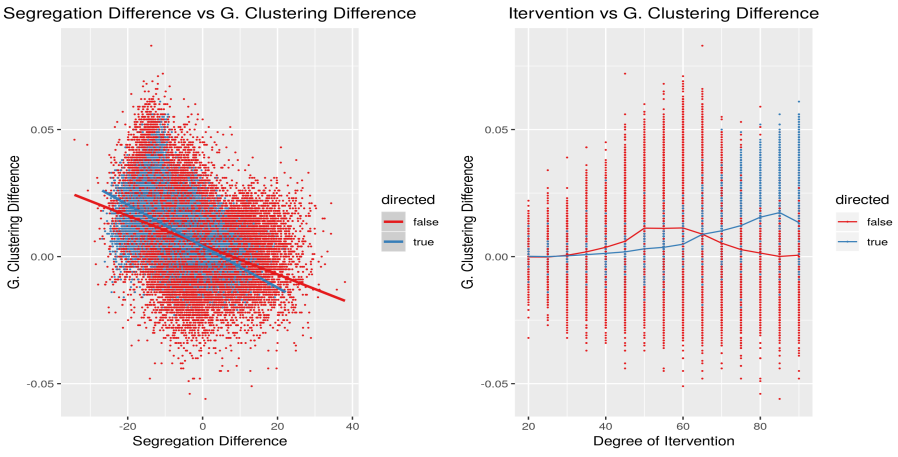


FIGURE 17. Clustering Change

Although the intervention degree does not seem to have a significant effect on the change in global clustering degree, the segregation difference is negatively correlated with clustering difference. In the plot to the left of Figure 11, networks that were *de*segregated by the dynamic *gained* clustering, while networks that were segregated lost it. This holds both for directed and undirected networks. This observation is *prima facie* unintuitive, in more segregated networks agents of each type are expected to have more friends of their own type. But this would be a misunderstanding, for *in group* clustering may consistently increase while reducing *overall* clustering, as suggested by the graph. In order to asses the value of this explanation, in group clustering measures would be required which were not done in this study. This task is left for future investigation.

Let us move to average path length. This is the average number of steps along the shortest paths for all possible pairs of network nodes. If the network is not connected, i.e. if there are at least two nodes with no path between them, then the average path length is undefined.

First we study the effect the dynamics had on whether the networks gain or lost connectivity. Out of the 86400 simulations, in 3494 cases the dynamic transformed a disconnected network into a connected one, and in 6212 cases it transformed a connected network into a disconnected one. The table below summarizes the information for these categorical changes:

Total	**Directed (true,false)**	**Density (5,10,15)**	**Relink (all,one)**	**Mean Seg Diff**
3494	1024, 2470	3424, 70, 0	1844, 1650	1.5784
6212	2442, 3770	5906, 304, 2	3258, 2954	-4.523

TABLE 20. Connectivity Change

It should be clear that random networks with low density (5%) are less likely to be connected than those with high density. Since by definition the dynamics do not change density, low density unconnected networks will appear both before and after its implementation. Yet, there were almost

twice as many networks that lost connectivity than those which gain it. This suggests that implementing the dynamics might be detrimental to connectivity. Surprisingly, networks that gained connectivity on average increased seggregation difference, while the opposite is the case for those that lost it.

In the remaining 63231 cases in which connectivity was neither gained nor lost, average path length of directed networks decreased with segregation difference, as illustrated by Figure 12. It is not clear why this is the case. On the other hand, as with clustering, degree of intervention does not seem to have a significant effect on average path length change.

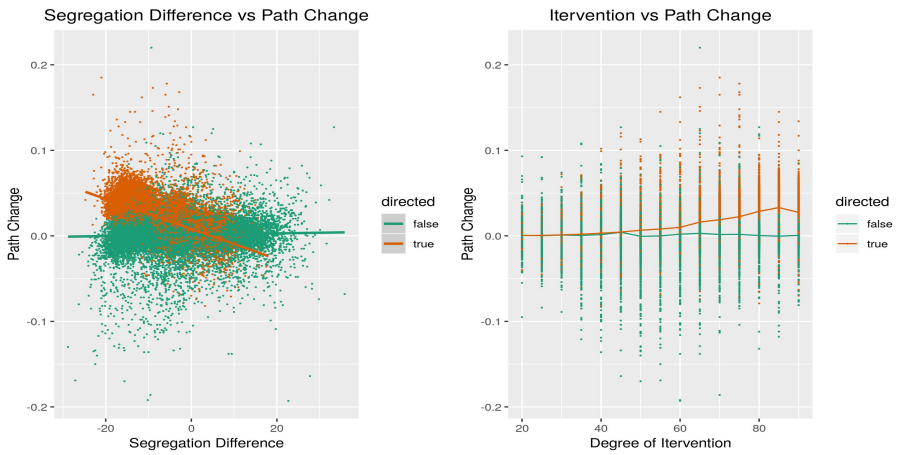


FIGURE 18. Relink One vs Relink All

This appendix is mostly propaedeutic. The purpose was to study the effects of the Schelling dynamics on some broad network properties. Part of the purpose of formal models is to provide an explanation of actual phenomena. Schelling presented a type of dynamics that, we argued, can explain segregation in Online Social Networks. Barabási-Albert preferential attachment algorithm, to give another, provides a dynamics that explains observed degree distributions following the power law. In general, defining a dynamics that explains *both* social arrangements like segregation *and* observed network properties like degree distribution (or average path length and global clustering) would be optimal. The exercise in this appendix was a step in that direction.