Consensus and disagreement in small committees

Martini, C.

Publication date:
2011

Document Version
Publisher's PDF, also known as Version of record

Consensus and Disagreement in Small Committees

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University, op gezag van de rector magnificus, prof.dr.Ph. Eijlander, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op vrijdag 16 december 2011 om 10.15 uur door

Carlo Martini

geboren op 22 september 1983 te Padova, Italië.

Promotiecommissie

Promotores:	Prof. dr. S. Hartmann
	Prof. dr. P.H.M. Ruys
Overige leden:	Prof. dr. J.J. Graafland
	Prof. dr. A.P. Thomas
	Prof. dr. M.V.B.P.M. van Hees
	Prof. dr. J.J. Vromen
	Dr. J.J. Reiss

FSC
www.fsc.org
MIX
Paper from
responsible sources
FSC® C014810

ISBN: 978-94-6191-092-9

This work is dedicated to Alberto Giopp

Questo lavoro è dedicato al Cav. Uff. Alberto Giopp

Acknowledgements

From the start of my Ph.D. at Tilburg University in 2008, I have been extremely fortunate to be working with two excellent supervisors. It is for merit and not for custom that my first and foremost thanks go to Stephan Hartmann and Pieter Ruys, for their constant supervision during the writing of this thesis. This is a work on agreement and disagreement; it is thanks to the countless agreements and disagreements (with no pun intended), during the meetings and casual conversations we had in the past three years, that this work has reached its present state.

Especially in its final stages, this thesis has greatly benefited from the the comments and criticisms of all the members of the Ph.D. Committee. My thanks go, in alphabetical order, to Johan Graafland, Julian Reiss, Alan Thomas, Martin van Hees, and Jack Vromen for reading this manuscript and providing invaluable feedback and suggestions.

A large part of this thesis deals with the Lehrer-Wagner model for consensus; a special thank goes to Carl Wagner for helping me understand and explore the model during his visit at Tilburg University in 2008.

At different stages of my Ph.D. I have benefited from two visiting fellowships, the first one at the P&E Program at the University of Pennsylvania, for which I owe my thanks to Cristina Bicchieri; her supervision while at UPenn was extremely helpful in developing my ideas for the second part of this thesis. My second visiting fellowship was spent in the Sydney Centre for the Foundations of Science, at the University of Sydney, for which my thanks go to Mark Colyvan. Numerous discussions with Mark on the Lehrer-Wagner model and epistemic disagreement resulted in a joint paper, coauthored also with Jan Sprenger. To Jan go my thanks for our numerous chats on all topics in philosophy, as well as for being a helpful and neat housemate for most of my time as a Ph.D. student.

Throughout my undergraduate and graduate career I have been fortunate to enjoy academic conversations with so many people that it would be impossible to honor everyone here. I wish, however, to mention my colleagues at Tilburg University, especially the members of the Tilburg Center for Logic and Philosophy of Science (TiLPS), and the participants of the TiLPS research seminars. I have equally benefited from conversations with the many visiting fellows to the TiLPS, among which special thanks go to Jonah Schupbach, for several instructing chats as well as our common taste for *La Trappe.*

For their precious feedback, I wish to thank the organizers and the participants in the Current Projects seminar at the University of Sydney, the PhilSoc seminar at the Australian National University, and the PPE research seminar at the University of Pennsylvania. I have presented my papers, mostly due to Stephan Hartmann's zealous and constant encouragement, at about thirty conferences and workshops. I could not possibly remember all those who gave me precious feedback and criticisms and, for fear of failing to mention some, I wish to thank all of them collectively.

For almost thirty years now, I have received the unconditional support of my family. My special thanks go to Alessandra, Mario, Alberto, Paolo, as well as to all the other members of my family. A memory goes to the late Angelica and Romilda (Nela).

Nei quasi trent'anni ormai passati, ho sempre ricevuto il supporto incondizionato dei miei famigliari. Un ringraziamento speciale va ad Alessandra, Mario, Alberto, Paolo, e a tutti gli altri membri della mia famiglia che mi sono stati vicini in vari modi in tutti questi anni. Un ricordo va alle scomparse Angelica e Romilda (Nela).

I am blessed with the gift of long lasting friendship from a number of people around the world. While sometimes we don't meet or talk for several months or even years, their friendship is one of the most valuable fortunes in my life, which neither time nor distance manage to weaken.

My friends at Tilburg made the time I wasn't dedicating to my work a good and pleasant one. While I owe those good times to all of them, none excluded, my special thanks go to Chiara, for turning me into an IKEA master, Gaia, for the fish recipes and the instructive walks at the farmers' market, Salvatore, for our conversations on Life, the Universe and Everything, and Ying, for the sushi, the pastries, and especially for the innumerable laughs.

My final and wholehearted thanks go to Eugene and Daisy, for making the Valley my second home, and to Lauren, for filling that home with warmth and joy.

Contents

Chapter 1

Introduction

Situations of disagreement are a very common occurrence, possibly even the norm, in most types of human interaction. At the same time humans spend a great deal of energy seeking to eliminate disagreement and reach a consensus. From the most private kinds of interactions, e.g. a group of friends planning to go to the movies, to the most difficult scenarios, like global diplomacy, consensus is looked for among all classes and occupations: politicians, physicians, businessmen, and also scientists.

In a large part of the contemporary world, large crowds resolve their disagreement by democratic means, such as voting, political representation, and other mechanisms. That is not the case, however, for small groups where the room for debate is larger, and disagreement can be resolved by other means than the complex procedures in place in democracies and other large electorates. In most situations, in small and informal groupings of people, disagreement is resolved "naturally" by the dynamics of interaction among the members of the group.

Imagine a group of friends planning to visit the Van Gogh museum in Amsterdam; under normal conditions, if there is initial disagreement it is easily resolved, although the larger the group becomes, the harder it may be to accommodate everyone's preferences. Nonetheless, there are groups where for the magnitude of what is at stake, or the complexity of the subject of disagreement, it is harder, in some cases even impossible, for the natural mechanisms of human interaction to provide a resolution of disagreement which is good for all or for most.

These are circumstances where the decision that has to be taken has important consequences for the decision makers or those whom they represent, where the interests and personal or group preferences are very high, or else where the issue is particularly complex to be decided on. In all those cases resolving disagreement can be a very difficult task, one that requires sophistication and analysis both of the subject of disagreement and of the possible mechanisms for resolving it.

How do small groups, in particular purposive groups — groups with a specific intent or goal — in situations where the state of disagreement to be resolved is complex and rests on strong interests, resolve their conflict and reach a consensus? This will be one of the central question of this thesis.

It is clear that there are a number of ways to approach that question. Epistemology mostly takes the problem of disagreement and consensus independently on the topic or matter that is the subject of disagreement. This is a quite abstract approach, but the analyses contained in the literature on epistemic disagreement and consensus have provided very valuable analytic tools to the investigation of the question mentioned above. Among many of these tools are the so called consensus models, as well as a taxonomy of the possible stances, and their correlated justifications, that an epistemic agent, who is faced with a situation of disagreement, may adopt. The first part of this thesis will deal with the problem of disagreement from a mostly abstract viewpoint, the chosen viewpoint of most contemporary epistemology.

Disagreement and consensus, however, are particularly interesting in practical contexts. The question "What, exactly, is the subject of disagreement?" is an important one for most practical considerations on how to resolve a situation of disagreement, and possibly achieve a consensual resolution of a conflict. In the light of that, the possible choices were many, as to which specific context to address; therefore some arbitrariness was necessary. I chose, for this work, to focus my analysis on disagreement in economics. While, as I said, the choice is clearly in part arbitrary and dictated by personal inclinations, it is nonetheless the choice of a subject that is highly debated in contemporary civil society.

Everyone would like to have more agreement in economics, or at least in that part of the science that is concerned with policy making. Among the many reasons why, is that fact that everyone would like to be able to *see more clearly* what the connection is between the science itself and the policy and decision making it supports. Disagreement, and especially widespread and methodological disagreement, however, do not help that cause. In the second part of this work I analyze the formation of consensus in the economic sciences and among the so-called "economic experts". The problem, I will argue, is equivalent to the search for a core of institutional knowledge in economics; to wit, a core of principles and facts around which institutions can build their economic policies and strategies.

More consensus, at least according to some, implies also more responsibility. The very last part of the thesis stands as a corollary analysis on the question of how small groups can resolve disagreement. The thesis analyzed in the final section is that groups who possess a specific mechanism for reaching consensus, or at least convergence, of views, should also be held responsible for the consequences of their views. Such groups, in broad strokes, possess many of the features that we normally attribute to individual agents and, like individuals, should be deemed responsible for their actions. This is the third part of this thesis, and while it constitutes a relatively minor contribution to the rest of the work, its goal is to open a window on some issues that are related to the capacity of small groups to form a consensus. In the following, I will introduce each part of the thesis in more detail.

I start here from the first approach to consensus and disagreement, the epistemological one. One can not fail to notice that there are many reasons of philosophical interest in the topics of consensus and disagreement, as the literature testifies. Recently, in the 1970s, at least two important philosophical results pointed to a puzzling conclusion: "Reasonable" people cannot disagree. More precisely, once a group of rational epistemic agents discover that they are in disagreement, and engage in exchanging evidence and opinions, they *should not* disagree any longer than the time it takes for a consensus to be formed. In other words, resilient disagreement is irrational.

The results just mentioned, which point to the irrationality of resilient disagreement, are Aumann's *agreement theorem* (Aumann 1976), and Lehrer and Wagner's model for consensus (Lehrer and Wagner 1981). While Aumann's theorem and the Lehrer-Wagner model are based on very different frameworks (the former on Bayesian and the latter on linear updating) they reach the same conclusion, namely that "actual disagreement among experts must result either from an incomplete exchange of information, individual dogmatism, or a failure to grasp the mathematical implications of their initial stage [of disagreement]." (Lehrer 1976, 331). Aumann, on the same note, concludes that "people with the same priors cannot agree to disagree." (Aumann 1976, 1236)

According to Aumann, Lehrer and Wagner, it is the mathematics and the principles of rationality which bind us to a resolution of our situation of disagreement. After their work, however, the problem of disagreement did not see much development, apart from the technical development related to Aumann's results (see the review article, Bonanno and Nehring 1997), until the past decade when the literature on disagreement started growing significantly in epistemology. As of January 2011, the entry 'epistemology of disagreement', on the popular philosophy website www.philpapers.org, counts 52 entries, most of which have been published in the past ten years.

Epistemologists of disagreement would like to know what role their disagreement with another epistemic agent has, with respect to what they should believe on the subject they are disagreeing about. The central question they try to answer is "what should you do when you discover that someone firmly disagrees with you on some claim P?" (Frances 2010, 1).

There are three main answers to that question. The *steadfast position* (see Kelly 2005) claims that, when disagreeing with an epistemic peer[1], you should simply stick to your own beliefs, because disagreement does not provide any type of evidence to the fact that you might be wrong in holding whatever beliefs you have. Instead, according to the *precautionary position* (see Feldman 2007), upon disagreement with an epistemic peer on a certain matter you should suspend your judgment on that matter. While prescribing an *epistemic* attitude, the precautionary position does not provide a *practical* course of action; it may be legitimate, for the purposes of decision theory, to take different stances with respect to how to act, upon suspension of belief on a certain problem.

Finally, the *conciliatory position* (see Christensen 2010) — the one on which this thesis will focus, in chapters 2, 3, and 4 — defends the claim that a rational agent should take disagreement as evidence and update her opinion accordingly, moving closer to the disagreeing partner's own view. In order to defend such view one needs to assume that beliefs come in degrees. Obviously some cases of disagreement are not compatible with conciliatory positions, when one's beliefs allow only binary values — 0/1 or 'yes'/'no' —. Recall for a moment the story of king Solomon and the two mothers in 1 Kings 3:16-28. Both mothers were claiming a newborn as their own, one of the two mother's newborn having died shortly after birth. In conciliatory spirit, King Solomon suggested to split the baby in two halves, one for each of the mothers. This is an evident case where the conciliatory spirit fails to give the right answer, as King Solomon understood in all his wisdom.

Despite the aforementioned case and similar others — see for instance Sen's story of the three children and a flute (Sen 2009, 13-15) — there are many examples of situations where disagreement is on a continuous value, thus open to the mathematical treatment presented in this thesis. What rate of inflation a government should aim at, what a feasible and significant CO_2 emission-cut goal is; those and several other scenarios allow for the type of treatment the conciliatory position defends.

The first question that this thesis will attempt an answer to, then, is on

[1]More on the notion of "epistemic peerhood" will be said in chapter 3 (see, in particular, section 3.4). In general, an epistemic peer is someone who has your same cognitive abilities, access to evidence, and so on, on the matter that is the subject of disagreement.

the issue of disagreement and "rational consensus"[2].

Query 1. Is it possible to form *rational consensus* by applying the principles of mathematical rationality, when faced with a situation of disagreement?

Query 1 can be divided into several subquestions. In the first place, what is the difference between consensus and compromise? Not all resolutions of disagreement are the product of a consensus; they can be imposed with force, negotiated in a compromise, and so on. Moreover, is disagreement evidence for changing one's beliefs, as Christensen and Kelly claim? (Kelly 2005; Christensen 2009) And if so, what type of evidence is disagreement?

All of the above are mostly epistemological problems, and their formulation is by necessity somehow idealized in order to make them formally tractable. On the other hand, it was said earlier in this introduction that disagreement and consensus are also very practical issues in many disparate fields, among which is the field of science. In science, disagreement is normally about theories and/or methodology. Consensus instead, for instance in the phrase "scientific consensus", is often used to mean a number of accepted propositions about a specific scientific problem. Such propositions are accepted by the relevant scientific community, without implying that each single scientist of that community has personally endorsed the proposition word by word, or has personally investigated the issues and come to endorse their truthfulness.

> "The Scientific Consensus represents the position generally agreed upon at a given time by most scientists specialized in a given field."[3]

According to Kuhn, a consensus is the set of propositions accepted at a certain stage of a science and agreed upon among the majority of the scientists that are part of a so-called *scientific paradigm*. "Men whose research is based on shared paradigms are committed to the same rules and standards for scientific practice. That commitment and the apparent consensus it produces are prerequisites for normal science [...]" (Kuhn 1970, II - The Route to Normal Science). But how does a paradigm form?

[2]The phrase "rational consensus" is used in the work of Lehrer and Wagner, which will be central to the first part of this thesis. It will be clarified in the following chapters what the meaning (or, as we will see, meanings) or rational is.

[3]From the website of the non-profit international organization *GreenFacts*. URL: http://www.greenfacts.org/glossary/abc/consensus.htm.

Are there so-called *rational factors* alone — viz. evidence and testing — that play a part in the formation of such consensus, as the ideal of the scientific method implies? Is the scientific method, by which disagreement disappears (at least provided sufficient evidence) and science progresses, suitable for all sciences as a normative theory of consensus formation?

In economics, the methodological prescriptions of the rational view and the ideal of the scientific method, have gone a long way in influencing the way economics is done. Some paradigmatic methodological prescriptions are contained in the views expressed in Friedman (1953). Even though his concept of 'positive economics' has been extensively investigated and refined[4], the essential mechanisms of hypothesis formulation and empirical testing have been interpreted mostly in the sense typical of the natural sciences, and of physics in particular. That is to say that hypothesizing and testing are based, respectively, on mathematical and computational means, and on statistics and experiments.

Modeling and testing are by large considered the *rational* criteria by which a certain scientific consensus should be evaluated. Although critiques and corrections have come from many sides of philosophical and scientific inquiry, the appeal of a *rational* method — one based on a common language and shared and verifiable evidence — is hard to deny, also perhaps because it seems to work so well in so many sciences. Despite the appeal, it is clear that economic knowledge comes from a variety of sources, experiments (in some cases and according to some even thought experiments), modeling, statistic analysis, economic history and the list could continue. Is there a method then for discriminating and ranking among these sources?

The second question at the core of this thesis is about economic methodology and its relation to the formation of economic consensus.

Query 2. Can we formulate some criteria on the basis of which a specific economic consensus can be evaluated as acceptable or not acceptable, by the community of scientists, and possibly policy makers, who are involved with that received consensus?

The ultimate depository of scientific consensus, at any rate, are the scientists themselves. But what is the relation between the science itself, the scientist, and the application of economic knowledge, regardless of whether it derives from theories, experiments, history, experience or other sources?

[4]A number of scholars of economic methodology have contributed in 2009 to a volume dedicated to an analysis and review of Friedman's famous essay *The methodology of positive economics* (see Mäki 2009).

A certain picture in the economic world, and defended by economist Jeffrey Sachs, sees the scientist as the apothecary in front of the chest of drawers containing his *materia medica*[5]. According to the metaphor, the economist is like the apothecary when she chooses a certain medical substance, or a mixture of them, in order to cure a specific condition. The economist has at her disposal a number of mathematical models, has evidence from experiments, is aware of statistical analyses, and perhaps also possesses historical knowledge of the type of problem she is faced with. While it is unlikely that a single item of her knowledge — a mathematical one, or a historical one — will provide a cure for all "economic illnesses", a capable economist will use her judgment to select those items that are needed for a specific cure from her economic chest of drawers.

Beyond the metaphor, Jeffrey Sachs reckons that an economist's expertise is limited, and that in order to resolve economic problems in the real world one needs to resort to a variety of theories (items of knowledge, the drawers in the metaphor). Sachs himself embodies the figure of the practitioner economist, involved in advising several governments in Europe and Latin America, during their transition from communist to market-based economies.

Is the picture of these apothecary economists, like Sachs, or Anders Åslund — Åslund and his role as advisor to the Russian government is described in Angner (2006) — an adequate one for the resolution of economic problems, the formulation of economic plans, the construction of economic tools and so on? Or should we rely on *committees*, rather than individuals, like the Monetary Policy Committee of the *Bank of England* (see Budd 1998)? The advocacy of the power of groups, crowds, and in the cases discussed in this thesis especially *teams*, over that of individuals has spurred from various scientific fields in recent decades and will be analyzed in the second half of this thesis.

While groups take over the role of individuals, as for example committees take over individual experts (or *should* take over, as theorists of group power suggest), a question is left as to what type of responsibility these groups should have. With power, goes the old saw, comes responsibility, but the statement is not always true in practice, as the status of individual anonymity behind the group can hide guilt and blur responsibilities. It seems necessary then to have at least *some* account of group responsibility, if one is to defend the idea of groups as providers of economic consensus.

The third and final question, then, is related to the responsibility of

[5]I owe this picture to my supervisor, Pieter Ruys. He was presented with the concept of the economist as apothecary by famous economist Jeffrey Sachs, who in personal conversation was defending such idea of the role of an economist in the field of real economies and concrete economic problem.

groups.

Query 3. What features should belong to a group seeking consensus, for example in economics, in order for it to be also held accountable for the actions it performs?

Far from developing a theory of corporate responsibility, the final section of the thesis will start from a prominent contemporary account, Pettit and List's account on group agents and group responsibility, and develop it by first asking whether the account is sound, unsound, or needs revisions and corrections.

To conclude this section, following are the summaries of the individual chapters.

In chapter 2 I introduce the Lehrer-Wagner theory of consensus and its interpretations. I present the mathematics of the model and illustrate how it fits possible interpretations. The core of the analysis will be on how to defend the Lehrer-Wagner model for consensus as a realistic or a normative model. In order to provide a comparison, I will present two similar mathematical models, which are part of the same theory of consensus. While the Lehrer-Wagner model is commonly defended as an aggregation model (Lehrer and Wagner 1981), I will argue, instead, that the model can be best defended as an updating model. This will be the main conclusion of chapter 2.

Provided that the Lehrer-Wagner model can be defended as an updating model, the question is whether it is rational to update one's beliefs in the light of disagreement (cf. *Query 1*, above). Chapter 3 deals with that problem. As mentioned in this introduction, probably the two main answers to the problem of disagreement come from Aumann's theorem and the Lehrer-Wagner model, the former in a Bayesian framework and the latter in a linear updating one. In chapter 3 I will discuss both approaches to disagreement. While some remarks and disclaimers will be in place, I will conclude that disagreement does not constitute evidence on which to update, and therefore does not justify a conciliatory view on disagreement.

Chapter 4 contains a mostly programmatic discussion of how the formation of consensus can be investigated with formal tools such as the Lehrer-Wagner model. Drawing from the conclusions in chapter 3, the model need not be considered a truly consensual model, but can be taken as an aggregation one, similar to a sophisticated voting mechanism[6]. In fact, it

[6]Many years after the publication of their work (Lehrer and Wagner 1981), Wagner

can be taken as a voting mechanism that conveys a great deal more information at the group level than, for example, majority voting does. Chapter 4 investigates the Lehrer-Wagner model as an aggregation model, and provides a strategy for one of the unsolved problems in the original formulation of the Lehrer-Wagner model, namely the assignment of weights.

In the second half of the thesis, the focus moves from the abstract treatment of the problem of disagreement in an epistemological context, to the problem of disagreement and consensus formation in the sciences and in particular in economics. In chapter 5, I present the topic of consensus as treated by philosophers of science, and present a number of reasons why consensus is a desirable outcome in a scientific debate. I then discuss the problem of the origin of consensus, the desideratum of *rational consensus*, and illustrate a standard example of the application of the canons of scientific rationality in the physical sciences. The final section of chapter 5 will analyze a number of phenomena typical of the economic and social world, which make the standard scientific method of consensus formation partly inapplicable in economics.

Chapter 6 continues the previous chapter along the same lines. The question now is to try to provide a *positive* account of consensus formation in economics. After reviewing some alternative sources of knowledge in economics, other than modeling and testing, I discuss the problem of expertise and argue that economic experts are the primary and also irreducible source of economic knowledge and economic consensus. The meaning of that irreducibility will be discussed when explaining the role of experience and the concept of tacit knowledge in economics. In the final sections of the chapter, I will provide a brief and schematic illustration of the biases and shortcomings related to expert judgment, and present two models for directing expert judgment. The thesis defended at the conclusion of chapters 5 and 6 will be that, while the role of expertise in economics is not reducible to the application of the rational methods of modeling and testing, there are methods for reducing biases in expert judgment that can be applied to economic theorizing.

In chapter 7 I will discuss the problem of responsibility in small committees, for example the committees of experts referred to in the previous two chapters. While the topic is not new, and a number of theories of corporate responsibility exist, the chapter will start the discussion from Pettit and List's theory of group agency and responsibility. Pettit (2007)

---

was indeed convinced of this fact, that the model is a particularly advanced aggregation mechanism, rather than an updating model. I thank Carl Wagner for sharing his views on the matter in personal conversation.

argues that group agency implies group responsibility. I will discuss the three conditions the authors give for group responsibility and argue that there are counterexamples to Pettit's thesis. Finally, I will add two additional desiderata to Pettit and List's conditions, and argue why the addition is important for avoiding the type of counterexamples I previously illustrated.

Chapter 2

The Lehrer-Wagner model

2.1 Consensus and compromise models

The practice of using the term ‘consensus’ in different and often incompatible ways is reflected by the variety of studies that deal with the problem of consensus. From a purely definitional point of view, achieving a consensus in a group means finding a general agreement or identity of judgment among a number of initially different opinions. By contrast, a compromise can be defined as the decision to *settle on* a statement (or set of statements) even though the members of the group have not, internally, come to fully endorse a unique subjective judgment identical (or similar enough) to that of all other members of the group.

It is important to point out that the meaning of agreement, in this context, is restricted to the context of belief, whereas, in general, agreement can refer to actions as well. For example, we can agree *to do* something because we have reached a consensus or a compromise on what course of action to take, as a group. On the other hand, if we agree *on* something (e.g. whether the arguments in this chapter are cogent or not) we therefore have a consensus, not a compromise.

I can then define here a consensus as an *internal convergence of beliefs among the members of a group*, whereas a compromise is an *external (imposed or chosen) agreement to accept a belief as the belief that is endorsed by the group.* The term ‘convergence’ refers here to the *outcome* of the process of deliberation, aggregation, or any other suitable process that can lead to consensus. More on the different processes in section 2.1.1. The term ‘internal’ indicates that the justification for the consensual belief is “accessible by the mind”, in conformance with the common philosophical distinction between internalism and externalism in epistemology.

The foregoing definitions will guide the following discussion of different *consensus models*, which, albeit the fact they all use the same term 'consensus' in the "name tag", they in fact serve very different purposes and produce very different results.

2.1.1 Two families of models

A number of linguistic or logic-based models[1] are called 'consensus models' in the sense that they use a logical or mathematical function in order to extract a unique value from a set. In other words, these models produce convergence of logical or mathematical values, and are essentially aggregation algorithms (see Xu and Da 2003), to wit, methods for aggregating beliefs or information. Normally, the goal of aggregation functions is to satisfy a number of properties[2] selected on the basis of a specific desideratum (or set of desiderata), for instance, to have a *democratic* social choice function. Formal properties are the main criteria of evaluation of linguistic or logic-based models, even though the properties themselves might be justified on independent grounds, such as ethical, practical, or other reasons[3].

However, linguistic models are consensual only insofar as they produce an artificial form of agreement — as convergence of mathematical or logical values — by means of algorithmic methods. From the point of view of linguistic or logic-based models, consensus is a purely logical phenomenon, that is, *identity* of values from an originally diverse set. The rationality of such models is often grounded on the overall plausibility of the procedure (e.g. as said before, satisfaction of properties), but no mention is normally made about the convergence of beliefs (the agreement) implied in the notion of consensus as defined above. In other words, the agents involved in these types of models need not be *belief-bearers*, and there is no requirement that consensus be produced as the convergence of individual *beliefs*, in the internalist epistemological and psychological sense given in section 2.1.

A second branch of studies on consensus goes under the name of "consensus decision making", and comprises a number of "models" (more often only sets of institutionalized practices) that can be applied in group de-

[1] For some representative examples, see Herrera-Viedma et al. (1995) and Herrera et al. (1997).

[2] Any *aggregation* function has the principal role of aggregating information, beliefs, or other values; this is what separates them from other types of functions. Functions that aggregate, however, are compared and evaluated among each other on the basis of their capacity to satisfy a number of (desired) formal properties.

[3] Pioneer of this type of work was Kenneth Arrow, whose *impossibility theorem* set the ground for future Social Choice Theory research and its ramifications (see Arrow 1963).

liberation in order to reach consensual decisions. These are, for example, medical consensus models (widely used by the U.S. medical community, see Solomon (2007)), consensus models in legal theory, and political and economic consensus models like the Dutch *Polder Model* (see Schreuder 2001; Plantenga 2002). The term 'consensus', in this context, is often used more as a metaphor, or as an ideal that the models in question should approximate. In fact, consensus need not mean, in the case of these models, a full convergence of logical or mathematical values: A supermajoritarian decision, for example, can be considered a consensus, even though evidently a full convergence of values might not have been reached by the whole voting group.

Consensus models in this second family produce genuine consensus in the sense that they take into account the convergence of beliefs among the bearers of beliefs, the agents of the group. However, they normally fail to provide a concrete and precise method for achieving complete or even partial convergence, or the method is so demanding that it can only lead to consensus in very special cases. For example, unanimity voting produces consensus in the sense that it both provides an algorithm for full convergence and is based on the psychological convergence of the beliefs of the agents in the group, yet it produces consensus so rarely that it can hardly be considered a valid alternative to voting or algorithmic procedures in a great many practical situations.

Similarly, the Polder Model is an institutional procedure by which different parties involved in a decision are made to sit down at the same deliberative table and discuss in order to reach a decision that satisfies and is suitable to all or at least most parties involved. The procedure is called consensual because it stresses the communitarian aspect of deliberation, but it does not guarantee that the full convergence of values will be reached, nor that such convergence will be a consensus rather than a compromise.

2.1.2 The Lehrer-Wagner model

The Lehrer-Wagner model is meant to be both a deliberative model and an algorithmic one. It is deliberative in the sense that it produces a consensus of opinions among a group of rationally deliberating individuals, and it is algorithmic because it guarantees the production of consensus under a wide range of conditions; specifically, according to Lehrer (1976) and Lehrer and Wagner (1981), under a *minimally rational* set of conditions[4].

[4]The *mathematical* and *rational* conditions for consensus will be discussed, respectively, in sections 2.2.1 and 2.2.2.

The model has the goal of producing a real consensus, rather than just a compromise: All the dissenting members of a group, if rational, will agree with the aggregate value, which is a function of their original individual values. The Lehrer-Wagner shows an impossibility result similar to Aumann's *impossibility of disagreement* theorem (see Aumann 1976): Rational agents who recognize the nature of their dispute cannot fail to agree, at least given the minimal set of conditions under which the model converges, on the numerical value produced by the model.

The first class of models discussed in section 2.1.1 was meant to produce convergence of mathematical or logical values, no matter what these values represent. The second class, discussed in the same section, was meant to promote a certain degree of agreement among belief bearers (epistemic, political, economic agents), no matter whether such agreement is complete or only partial, truly consensual or perhaps the result of some bargaining process. The Lehrer-Wagner model tries to link the two tasks, by presenting a mathematical and algorithmic method for producing convergence of views among belief bearers.

In the next section I will present the details of the model, that is, its mathematical structure and its interpretations and possible applications. The rest of this chapter will analyze the Lehrer-Wagner model, together with two related ones, and analyze whether they can be truly considered algorithmic models for consensus.

2.2 Outline of the model

Imagine a relatively small committee of international scientists, who are asked to estimate a suitable and feasible CO_2 emission cut for international emission cut enforcement. The committee may be composed of people possessing diverging scientific knowledge, different agendas and different degrees of commitment to the necessity of finding a consensus. Alternatively, consider a company's board of directors, whose members are asked to assess the performance of the company's CEO and of some key managers, in light of a recent investment on a product and its market performance.

Let us assume that the goal of both committees is to agree on an official position, that is, to find the consensual stance of the board. The question the Lehrer-Wagner model addresses is whether such ideal committees can rationally end their sessions in disagreement. The answer from Lehrer and Wagner (1981) is that they cannot. Whether such analysis is correct will be the problem of this chapter, after having introduced the Lehrer-Wagner model in more details in the remainder of this section.

2.2.1 The mathematical model[5]

The Lehrer-Wagner model produces a consensus by means of iterated weighted averaging of the beliefs (expressed, for example, in form of probability assignments) that the members (agents) of a group hold on the issue under deliberation.

In the forthcoming paragraphs I will illustrate the model step by step. In the first stage the agents in the model (the committee members) assign a certain measure m_{ij} to themselves and to all others, where $m \in [0,1]$ and where i is the agent assigning the measure, and j is the agent receiving it. These measures form a $N \times N$ matrix W, with entries w_{ij}, where N denotes the size of the group and each row W_{i*} is normalized, that is, $w_{ij} = \frac{m_{ij}}{\sum_{k=1}^{N} m_{ik}}$. The matrix W is called the "matrix of weights" of the Lehrer-Wagner model and is exemplified below.

$$W = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1N} \\ w_{21} & w_{22} & \dots & w_{2N} \\ \dots & \dots & \dots & \dots \\ w_{N1} & w_{N2} & \dots & w_{NN} \end{pmatrix} \tag{2.1}$$

In the second step, agents provide their judgment p on the subject matter on which the group is deliberating. These judgments form a column of numbers (for instance, probabilities) P, with entries p_i, as exemplified below.

$$P = \begin{pmatrix} p_1 \\ p_2 \\ \dots \\ p_N \end{pmatrix} \tag{2.2}$$

W and P make up the initial information-set, that is the situation in which all members of a group have assigned a certain measure to their fellows (the interpretation of this will be provided in subsection 2.2.2) and have expressed their belief on the subject matter. Normally, the entries in P will differ from each other, denoting the fact that a consensus has yet to be

[5] An early exposition of Lehrer's theory of consensus can be found in Lehrer (1976), a presentation of the mathematical model appeared in Wagner (1978). For two previous expositions of the underlying idea, and formal theory of consensus, see French (1956) and DeGroot (1974). The complete exposition of the Lehrer-Wagner theory of consensus, the mathematical model, and an extension of it, are in Lehrer and Wagner (1981). Most of the discussion in this chapter will be based on this latter work. In the chapter I will only present the basic version of the model, in appendix A I will briefly present also the extended version.

reached. When the entries in P will be equal to each other (or approximately so, if we pragmatically agree on a certain degree of approximation) then the model will be said to have reached consensus, meaning that all members of the group are holding the same opinion on the subject matter.

In the following I will explain how the mathematical model goes from a state of dissensus to one of consensus. By multiplying the matrix of weights k times and then by the column of probabilities P, that is W^kP, a theorem shows that, under certain conditions, the values of the obtaining column P_C will be equal to each other as the powers k of W rise.

$$\text{(Column of consensual probabilities)} \quad P_C = W^k P \quad \text{for} \quad k \to \infty \tag{2.3}$$

The conditions for convergence are that the weights in each row of W be normalized (as explained above), and that there be a "chain of respect". The concept of chain of respect is clearly metaphorical, but it is formalized by Lehrer and Wagner, and plays an important role in Lehrer and Wagner's model of consensus (Lehrer and Wagner 1981, 129-133, see also Theorem 7.4). The authors explain the concept as follows: "Convergence towards positive consensual weights results from iterated aggregation if there is a chain of positive respect from each member of the group to every other member of the group, and at least one member assigns positive weight to himself. We call this communication of respect." (Lehrer and Wagner 1981, 27)

Lehrer and Wagner give the mathematical conditions for convergence in the second half of their book: *the formal foundations of rational consensus* (Lehrer and Wagner 1981, part two). For our purposes, it is sufficient to say that if a normalized matrix is *reducible*, then it does not converge to consensual weights, and thus does not produce a consensus in the group of agents. A matrix is reducible if its entries can be split into two distinct matrices without changing the order of the entries[6]. For illustration matrix M is reducible to matrices X and Y (below).

$$M = \begin{pmatrix} .7 & .3 & 0 & 0 \\ .2 & .8 & 0 & 0 \\ 0 & 0 & .4 & .6 \\ 0 & 0 & .1 & .9 \end{pmatrix} \; ; \qquad X = \begin{pmatrix} .7 & .3 \\ .2 & .8 \end{pmatrix} \; ; \qquad Y = \begin{pmatrix} .4 & .6 \\ .1 & .9 \end{pmatrix}$$

In the example above, the matrix M in fact converges to two different

[6]For a formal notion of reducibility see Meyer (2000, pp. 209, 671) or Royle and Weisstein (2010).

values, which are equal to the convergence values of respectively matrices X and Y. X converges to the consensual weights .4 and .6; Y converges to the consensual weights .14 and .86.[7] The *non*-consensual matrix M^* would then look like this.

$$M^* = \begin{pmatrix} .4 & .6 & 0 & 0 \\ .4 & .6 & 0 & 0 \\ 0 & 0 & .14 & .86 \\ 0 & 0 & .14 & .86 \end{pmatrix}$$

In the Lehrer-Wagner model, if the conditions for convergence are satisfied, the matrix of weights converges to a consensual matrix with identical rows, and the product $W^k \cdot P$ (for $k \to \infty$) yields the consensual column P_C. Each line l of P_C is the updated opinion of agent i on the subject under deliberation and all the values in P_C are identical, denoting the fact that a consensus has been reached.

The proof of convergence of the model is omitted here, although the precise mathematical formulation of the conditions for convergence must be attributed to Perron-Frobenius in the Perron-Frobenius Theorem. Other versions of the theorem and proofs for convergence can be found in Lehrer and Wagner (1981, pp. 129-133) and in Jackson and Golub (2007). The conditions for convergence have to do mostly with the requirement of normalization and the composition of the matrix of weights. What weights represent is not straightforward and it will be the main topic of the next section.

2.2.2 Interpretations

Some of the elements of the model are of fairly straightforward interpretation. This is the case for the term 'P', which contains the opinions of the members of the group on the issue that is under deliberation. If the members cast their opinion in terms of a probability value, then P will contain probabilities. If otherwise — for example if the opinion is expressed in terms of a quantity, or of discrete values — P will contain the appropriate numerical values that express the information provided by the members of the committee.

As explained in section 2.2, the Lehrer-Wagner model can only handle problems that are representable in a mathematical form. The limitation derives, in part, from the necessity of having the opinion of each member expressed in the form of a numerical value, typically, but not necessarily, a probabilistic value. In a standard case, representable in the Lehrer-Wagner

[7] The calculation was obtained with 100 iterations, that is X^{100} and Y^{100}.

model, the members of a committee will express their belief as a probability value. Alternatively, however, the column can contain integers (e.g. if the problem is to find a consensus about a quantity, rather than a probability), or also an array of preference sets (e.g. when the problem is to find a consensual ordering). A convergence theorem (see above, section 2.2.1) guarantees that the rows in the column (or the array) will converge to a unique value, or a unique ordering. Most of the discussion in this and other chapters will be on the use of the model in its common interpretation, that is, with probabilities.

The term 'k' in 2.3 indicates the number of rounds that the model takes to reach convergence. The requirement that k tend to infinity is purely theoretical, because in all practical problems the matrix of weights reaches convergence in a finite number of steps, provided that the other mathematical conditions are satisfied. In practice then, k needs to be "large enough" for the matrix to converge, in order for the model to reach a consensus.

In context of practical deliberation, one can imagine k to represent the number of times a committee goes through alternative phases of deliberation and updating (the members, after deliberation, change or update their beliefs). A worry with this interpretation is the fact that phase after phase, at least in the simple model (not in the extended version however, see appendix A), the matrix of weights remains unchanged, this indicating that the members update their opinions only on the subject matter, not on the expertise of the other committee members. In itself, the interpretation of the term 'k' is also straightforward, even though it might be contested whether the term makes sense in a context of practical deliberation. This problem will be treated in the forthcoming section 2.4.

More difficult than the interpretation of P is the interpretation of the matrix W, that is, the interpretation of the weights (ws), as normalized *measures* that agents in the model assign to each other. Lehrer and Wagner (1981) discuss possible derivations of the weights from four different cases in which the model could be applied, namely a *census problem*, an *estimation with minimal variance*, a *weather forecasting* problem and a problem of *subjective estimation.* A thorough discussion on the interpretation of weights, in the remainder of this section, will be essential for understanding what type of "consensus" the model yields, that is, the rationality of the consensual results from the model, a problem which will be focal in section 2.4.

Without giving here the details of cases 1 and 2, which can easily be found in Lehrer and Wagner (1981, pp. 138, 139), it is to be noticed that both of them make use of the model as a mechanical aggregation procedure, where what is aggregated is objectively obtained information (e.g information from instruments). Clearly when the subject matter is consensus, we tend to think of that in terms of agreement among different opinions. If, instead, the agreement sought is rather among measuring instruments 'consensus' is simply a term for denoting convergence of measures. In turn, the *rationality* of such consensus can only be defined as the *appropriateness* of using a specific aggregation function, rather than another one, for example in virtue of a number of properties that it guarantees. For this reason, when the model is used as an aggregation function, it can only be said to be consensual in the sense in which linguistic-based models (see the beginning of section 2.1) are consensual, that is, as externally motivated aggregation procedures.

Similar remarks can be made for case 3, the forecasting example: Imagine a forecaster who is trying to estimate the probability of rain tomorrow and is assigned a "verification score", given by $F = ((f_1 - O_1)^2 + \cdots + (f_N - O_N)^2)/N$ (Lehrer and Wagner (1981, 140), taken from Sanders (1963)). The verification score is a measure of past performance in forecasting; f_i is the forecaster's judgment (forecast probability) on a particular past event, and O_i takes values 1 if the event occurred, and 0 if it didn't. The Lehrer-Wagner model applies to this scenario in the following way:

> This verification [F] score can [...] be computed for a sequence of consensual probability forecasts as well, and Sanders offers empirical evidence that a consensus of probabilities based on even simple arithmetic averaging can attain a verification score better than that of any individual contributing to the consensus. While Sanders did not investigate weighted averaging it is clear that such a refinement is possible. (Lehrer and Wagner 1981, 140)

If the model is taken as a tool of obtaining a more accurate forecast, then its rationality lies in the fact that the property it has is to "produce better forecasts". In this case, there need not be a convergence of the beliefs of the forecasters in order to justify the "use" of the model, but, as in the previous two examples, the consensus produced is only figuratively such, that is, as convergence of information. By using the model as a forecasting tool, agents need not rationally converge to the belief that the model yields as *their own* rational belief.

In this interpretation then, the model cannot be taken as an impossibility-of-disagreement statement: Forecasters may still be disagreeing as to which forecast is the correct one, even though the model is shown, a posteriori, to be a better predictor of the weather. There is however, a different possible formulation of the forecasting scenario, one which presents the problem of forecasting as convergence of beliefs. This latter case brings us to the case that Lehrer and Wagner give as an example of possible interpretation of weights, viz. the subjective weights example (Lehrer and Wagner 1981, 140).

> When a decision problem involves neither highly structured estimation subject to a prior analysis of weighting schemes, as in the examples 1 and 2 above, nor a statistical record of past performance, as in the preceding example, then the choice of weights becomes a subjective enterprise. (*idem*)

Suppose a group of forecasters are asked to estimate the probability of a certain meteorological event. They gather and assign to each other a certain measure that represents the value of accuracy that a certain forecaster thinks her colleague has. So, m_{ij} is the measure of accuracy that forecaster i assigns to forecaster j. Forecasters do not always agree on how accurate a certain colleague of theirs is, and so the rows in the matrix of weights (resulting from normalized measures) need not be alike.

The question now, expressed in the terms of Lehrer (1976), is whether the forecasters can disagree after recognizing their present situation. Granted that the forecasters accept to give at least some minimal weight to at least some of the other forecasters, and that no subgroups arise[8], the forecasters should accept to aggregate their opinion with that of the other forecasters to whom they have assigned part of the weight. The reason for this is that assigning a weight to a fellow forecaster amounts to accepting that my opinion on the subject matter be a function of my own initially expressed opinion and the opinion of (some of) the other forecasters.

To see this, consider that in the aggregation process, my updated opinion on the probability, for instance, of rain is "$w_{me,me} \cdot p_{me} + w_{me,forecaster1} \cdot p_{forecaster1} + \cdots + w_{me,forecasterN} \cdot p_{forecasterN}$". In other words, my updated opinion on the probability of rain is the weight I assign to myself times my probability forecast, plus the weight I assign to *forecaster1* times her probability forecast, and so on for all the weights I assign to my fellow forecasters and their relative probability forecasts.

Aggregation, in this case, is the acceptance that my belief as to what the correct forecast is be a function of the forecasts of all (or some of) the members in the group, including myself. It remains to be seen, at this point, how the mathematical function in the Lehrer-Wagner model differs from an averaging or a compromising procedure. In other words, it remains to be explained why convergence of the Lehrer-Wagner model to a unique

[8] In which case the matrix of weights may be reducible and fail to converge, see the conditions of convergence in subsection 2.2.1.

mathematical value is a consensus and not just another form of averaging. Section 2.4 will take up that problem.

2.2.3 Scope of the model

Before proceeding to the evaluation of the Lehrer-Wagner model and its *rationality*, in this subsection I will make some preliminary considerations about the model's domain of application. In the first place, the model makes sense only for relatively small groups of agents. The reason is that agents in the model are required to assess each other's competence, or to provide a value or "trust" for each other, at least under a number of important interpretations of the matrix of weights[9]. Assessing a large number of fellow group or committee members, however, may be considered impractical or even unfeasible for large groups such as nations, large associations, large committees of stockholders, etc. For this reason, the model is not a substitute for electoral systems, as it requires a capacity of mutual assessment that is not realizable for any large body of voters.

Secondly, the model requires the problem to be expressible in a mathematical form. Clearly not all problems are such that they can be treated mathematically, even though they may be perfectly clear in their formulation. For example, particularly complex geopolitical decisions, or management issues, might be clearly expressible yet fail to be properly formalizable. What extension a hypothetical Palestinian state next to Israel should have, which management practices a certain company should adopt in a market competition, and so on, are hard-to-formalize problems not because of some inherent vagueness in the formulation of the problem, but because of the number and nature of the sub-issues they involve. Similarly, problems involving ethical sub-issues are often hard to find a mathematical formulation — see for example the problem of three children and a flute in Sen (2009, 12-14).

Due to its algorithmic nature, the Lehrer-Wagner model can produce consensus (or convergence) only on issues that can be expressed in a numerical form, whereas some of the informal procedures, mentioned above (see section 2.1.1) are aimed at producing, or at least promoting, consensus on a wider range of possible issues on which the parties disagree. The limited scope of the Lehrer-Wagner, nonetheless, is an important one. Economic and scientific problems often are reducible to the assessment of a certain value or group of values, and even numerous social and political issues are often largely dependent on the assessment of some quantifiable variables.

[9]Nonetheless, as I will explain in section 2.3, some extensions or specifications of the model make it apt for larger groups, as they make the assignment of weights not dependent of personal assignment but on objectively detectable measures.

The first example of a possible application of the Lehrer-Wagner model, which opened section 2.2, was a simplified version of what was happening at the 2009 United Nations Climate Change Conference in Copenhagen[10]. There, much of the dispute was on the exact CO_2-emission-cut goal, in percentage value. We can imagine a great number of more or less similar examples: scientific committees for the evaluation of a new drug, city councils evaluating urban planning criteria (see section 2.3.1), etc.

Additionally, it is important to stress the fact that agents in the model are an abstract category. In the appropriate context, the model's agents can be, for instance, groups such as parties in a political context. As an example, in the Polder model (mentioned in section 2.1), the parties (agents) involved are three: the government, the employees and the employers. The model allows for collectivities to be agents in the model, so that convergence can be sought among groups, rather than among single individuals.

It is fair to say that the limitations of the Lehrer-Wagner model, in fact, point it in direction of the precise range of problems it can deal with, rather than leaving it to handle the problems of consensus in its full generality. It is in that range of possible applications that the model will be evaluated and discussed as a model for consensus.

Before proceeding to the evaluation of the rationality of the Lehrer-Wagner model, it will be useful to introduce two similar models, which are, in part, extensions of the Lehrer-Wagner model itself. The models presented in the next sections are a consensus model for environmental management introduced in Regan, Colyvan and Markovchick-Nicholls (2006), and the Bounded Confidence model (Hegselmann and Krause 2002). The expositions of the two will both clarify some of the issues about the meaning and methods for weight-assignment, discussed in section 2.2.2, and provide some examples of applications of the Lehrer-Wagner model.

2.3 A family of consensus models

It should be noted that what I have so far called a 'model' is in fact a general framework for consensus production that can take a number of different

[10]In this and the following examples it is clear that a major role is played by political and economic interests. While such interests, it may be argued, put those examples outside the scope of the Lehrer-Wagner model, there are also important scientific elements in them, which can be dealt with by the "standard" interpretation of the Lehrer-Wagner model as a model for scientific consensus. It should also be pointed out that the model need not be taken only as a model for (exclusively) scientific consensus, even though that is the assumption in this thesis. A version of the model as a consensus model for non-factual matters is pursued in Martini, Sprenger and Colyvan (2011).

specifications. The terminological choice of this section may be confusing in the sense that what are here called a 'model' are in fact specifications of what has been previously also called 'model'. For simplicity I refer to the environmental management model, the Bounded Confidence model, and also to the possible further specifications in the latter, all as models. The group forms a family of what can be called "lower level models", where the Lehrer-Wagner theory of consensus and its related mathematical model are the overarching framework.

2.3.1 An environmental management model

The first model presented in this section (see Regan, Colyvan and Markovchick-Nicholls 2006) is in fact a development of the basic mathematical model presented in Lehrer and Wagner (1981), applied in the context of environmental management. The task that Regan, Colyvan and Markovchick-Nicholls (2006) analyzes is the formulation of a list of criteria to be applied for the selection of urban open spaces for a Californian environmental conservation project. The practical problem is to determine a consensus weight for each of a number of proposed environmental criteria; the consensus weight attached to a certain criterion will determine its position in the multicriteria decision tree that the commission is asked to produce.

Examples of criteria are "reduces environmental risk" and "provides recreational opportunities and benefits" and "contributes to biodiversity". These criteria are of very different nature and respond to very different rationales. Moreover, they will be prioritized differently depending on the interests of those who are considering them. A consensual ranking is clearly a very welcome result for any committee trying to come up with a decision tree for the selection of urban open spaces that accommodates most parties.

The problem is a scientific one, insofar as the criteria are not purely subjective[11], but rely on objective and scientific evaluations. Nonetheless, due to the complexity of the problem, such ranking of criteria cannot be derived directly from scientific models, but rather needs to be agreed on by a group of experts. The sought agreement is influenced by stakeholder's interests and personal preferences, and the variables that enter the evaluation of a criterion over another are not easily codified and evaluated by mechanical methods like computer models.

For the reasons stated, a group of experts gathered in deliberation seems to be the best solution to a *fair* as well as an *efficient* ranking. The set-up

[11] For instance, one of the criteria is "improves quality of urban system"; this assessment depends on state-of-the-matter assessments, not only on individual tastes.

of the deliberation group that Regan, Colyvan and Markovchick-Nicholls (2006) has in mind, operates very similarly to the Lehrer-Wagner model with the important difference that the entries in the matrix of weights W do not derive directly from measures of trust assigned by each agent to all others. Instead, weights are derived from the difference in opinions among the members, that is, from the difference in the assigned values in column P of the model[12].

$$w_{ik} = \frac{1 - |p_i - p_k|}{\sum_{k=1}^{N} 1 - |p_i - p_k|} \tag{2.4}$$

Equation 2.4 states that w_{ik} is a function of the distance between p_i and p_k "where i refers to the individual who is assigning the weights, k refers to the individual being assigned a weight and N is the number of group members" (Regan, Colyvan and Markovchick-Nicholls 2006, 172). The basic idea is that agent i will be willing to give more weight to agent k, if agent k's opinion is closer to hers and the closer k's opinion is, the more weight agent i will give her.

The reasons behind the choice of function 2.4 are summarized in a number of desiderata (Regan, Colyvan and Markovchick-Nicholls 2006, 172). In short, the desiderata state that each agent should receive the highest weight from herself, that higher weights are given to individuals with similar values of p and vice versa, and that weights are normalized for each row W_{i*}. Clearly 2.4 meets all the desiderata and is a simple function of immediate use[13].

The specific motivation for choosing to derive weights from a measure of "distance in opinion", rather than a direct measure of trust of accuracy assignment, as originally thought in Lehrer and Wagner (1981), is that:

> The consensus convergence model described above [the Lehrer-Wagner model] requires each individual in the group to assess all other group members and then assign a weight to each member according to their degree of respect for or agreement with that member's expertise or views on the issue at hand. For the urban open space MCDM case study, *this approach is infeasible for a*

[12] In the example given by Regan, Colyvan and Markovchick-Nicholls (2006) what appear in the column are not probabilities but weights to be assigned to a specific criteria of environmental assessment.

[13] Exemples of a derivation of weights from a partial data set from the Californian Conservation Management Project is given in Regan, Colyvan and Markovchick-Nicholls (2006, 172).

> *number of reasons.* (Regan, Colyvan and Markovchick-Nicholls 2006, 172. Italics added.)

The reasons referred to are summarized in three points:

- Complexity of the task. It is not stated explicitly but it is inferable from the paper that the deliberating group in the case study was too large to consider feasible the task of gathering weights for all agents.

- Manipulability of the method[14]. Agents, in the Lehrer-Wagner model, are supposed to give honest assignments of weights, "if, on the contrary, the weights represent an egoistic attempt to manipulate social decision making, then it is unacceptable to use those weights though they were a disinterested summary of information." (Lehrer and Wagner 1981, 74). That weights are not given disinterestedly, however, is a possibility in any realistic setting like the one that Regan, Colyvan and Markovchick-Nicholls (2006) are investigating. Thus the need to derive the weights from some other measure.

- Quantification of elements that seem inherently non-quantifiable: "While most people would agree that they have different degrees of respect for, or agreement with, other group members' positions, translating that to a numerical value is a non-trivial task. Furthermore, group members may feel reluctant to explicitly quantify degrees of respect, as it could lead to rifts and ill feelings within the group. This is an undesirable outcome when the purpose of the exercise is to reach consensus.

The opportunity of deriving the weights from something different than direct assignments of a measure of trust seems thus motivated; nevertheless, such a move imports a complication in the original (and general) formulation of the model. The question that should be asked is whether it is rational to make weights be derived from opinion distances, or whether they should be derived from some other measurable variable. The problem will be treated in section 2.4. In the next subsection I will present the Bounded Confidence model.

[14]The concept of strategy-proofness is also at times used, especially in social choice theory.

2.3.2 The Bounded Confidence model

Hegselmann and Krause (2002) introduce a model[15] based on the idea that people neither completely share others' ideas nor ignore them altogether; rather, they will "take into account the opinions of others to a certain extent in forming [their] own opinions" (Hegselmann and Krause 2002, 2). The extent to which an agent in the model will share her opinion with other agents is determined by how many other agents that person is willing to share her opinion with. In turn, whom exactly, among all other agents, that agent is willing to share her opinion with, is determined by *how close* those agents opinions are to hers.

An example will illustrate the procedure. Suppose we take the case of the 2009 United Nations Climate Change Conference in Copenhagen mentioned in section 2.2. Members of the deliberation group have to agree on a certain reduction emission target (say between 5% and 40%). Suppose member A declares a target of 10%; according to Hegselmann and Krause, it is reasonable to assume that A will be willing to aggregate her opinion only with those other members whose opinion lays, for example, in the 2% distance interval from her own opinion. That is, if agent B declares a 8.5% target, then agent A will accept to aggregate with agent B; if, however, agent B declares a 5% target, then A will refuse to aggregate with her. Clearly the interval can be picked *ad hoc*, and the authors investigate under which intervals the model produces consensus or, alternatively, *opinion fragmentation* or *opinion polarization*.

Mathematically, the "neighborhood" that a certain agent is willing to aggregate with is given by a *confidence set*. Given the opinion profile $\{x_i\}_{i\in I}$ and the confidence level ϵ_i, for agent i, the *confidence set* for this agent is defined by: $I(x,\epsilon_i) = \{j \in I : |x_i - x_j| \leq \epsilon_i\}$, which is based on the absolute difference in opinions between agents (Hegselmann and Krause 2002, 382).

Once the confidence set is defined, the opinion (x) of each individual i, at time $t+1$, will be:

$$x_i(t+1) = \frac{1}{|I(x,\epsilon_i)|} \sum_{j\in I(x,\epsilon_i)} x_j(t) \quad \text{for} \quad t = 0,1,2\ldots \tag{2.5}$$

The convergence properties of this model depend on how large the interval ϵ is and the authors of the model analyze the formation of consensus by

[15]Properly speaking the Bounded Confidence model is not an extension of the mathematical model presented in section 2.2.1, but rather of an extension of that model (outlined in Lehrer and Wagner (1981, 53-72)) which allows for changing matrices at each step of the iteration process. For economy of space, the main idea of the extended model is presented in appendix A.

means of computer simulations. I leave the issue of the conditions for the formation of consensus in Bounded Confidence to the interested reader. What is relevant for the purposes of this chapter and the investigation of the Lehrer-Wagner model is what is highlighted in an article extending the exploration of Bounded Confidence (Hegselmann and Krause 2005). In this article, the authors investigate the model by using average measures alternative to the arithmetic average used in the original model; for instance the authors investigate convergence by means of *geometric*, *harmonic*, *power* and *random means*.

The declared scope of the extension carried out in Hegselmann and Krause (2005) is to investigate the convergence properties of 2.5 obtained by using different means. "The simulations [. . .] indicate that for all the different means there is a stable pattern of opinions after finitely many time-steps." A proof of this and a "stability theorem" is given to show that the opinion pattern stabilizes after finitely many time-steps of aggregation. This does not mean that convergence occurs always, as it depends on the choice of ϵ, but rather that after a certain time the initial group of opinions has either converged, polarized into two values, or fragmented in a number of different values (opinions).

Besides this investigation, perhaps of rather technical interest, what the extension of the Bounded Confidence model shows is that there need not be a single form of averaging in order to achieve opinion stabilization, and, what is relevant for the Lehrer-Wagner model, for convergence of opinions.

For the purposes of this chapter it is important to notice that the extension of the Bounded Confidence model challenges the rationality of Lehrer-Wagner model by showing that the aggregation function used in the original model is not unique. Lehrer and Wagner (1981) are silent on whether their model is a rational one among many, or whether the model is uniquely rational, but the issue is important if one wants to claim, as Lehrer (1976) does, that rational agents who understand the implications of their situation of disagreement *are rationally required to update* like agents in the Lehrer-Wagner model do. More on this point will be said in the next section.

2.4 The meaning of rational consensus

There can be, and there are, several ways in which the procedure a model describes (e.g. how to change one's belief in a case of disagreement) can be defined as *rational*; consequently, there are several ways in which a model can be said to be rationally justified or unjustified. Most models for aggregating or updating beliefs or information consist of a number of steps. For instance, the Lehrer-Wagner model has at least three steps towards reaching a consensus: 1) the association of each agent with all other agents in the model via the assignment of a certain measure; 2) the evaluation of the problem in a probabilistic (or at least formal) way; and 3) the iteration and convergence process to consensus. Defense of a model, however, may fail to provide a story that justifies each one of its steps, but rather just gives an overarching justification of the whole. The first question that arises, thus, is whether it is enough to justify a model as a whole, or whether justification is necessary for each of its parts.

Moreover, justification of a certain model may give a number of positive reasons as to why the procedure it describes is rational, yet fail to provide a reason as to why *only* such a procedure is the rational one. In other words, a model may be justified without necessarily being uniquely justified. For instance, a certain pooling algorithm may be shown to be the only one that satisfies a number of desired properties. Similarly, one may have reasons for claiming that a certain consensus model is the only justified consensus model; this would amount to showing that the procedure it describes is *uniquely rational*. Unique-type rationality can be a very strong desideratum, one that perhaps we do not want to ask from a model, but it is a desideratum that must be taken into account when considering the topic of rationality. If a procedure is not uniquely rational, then there should be a number of reasons as to why it is a valid (or better) alternative to the other ones, if there are any, suggested in the relevant literature.

In general for this section, the question to be addressed is what makes a certain model a model of *rational consensus*, rather than a simple pooling algorithm or deliberative method. The aforementioned subdivision of the problem provides us with a useful taxonomy[16] in the analysis of the rationality of the Lehrer-Wagner model that will follow.

Table 2.1: Taxonomy for evaluating the type of justification for a model

	In whole	In part
Is the model justified?	yes/no	yes/no
Is the model uniquely justified?	yes/no	yes/no

Table 2.1 is a simple graphical checklist for the type of analysis that

[16]Part of the material for this taxonomy was taken from Martini, Sprenger and Colyvan (2011).

should be carried out when assessing a model for consensus. Clearly, different interpretations as to what exactly is required from a model for consensus will lead to different answers as to whether, for example, the Lehrer-Wagner model represents a rational procedure for consensus or not. In the following, the definition of consensus given at the beginning of section 2.1 will lead the investigation. For now I will leave open the issue as to whether whole or part-for-part justification is required, the same will hold for the other elements in table 6.1. I will return to that taxonomy at the end of this section.

Lehrer and Wagner justify their model *as a whole*, even though some justification is provided for the individual parts as well; moreover, they do not consider whether the consensus procedure their model describes is the only rational one. In the following I will argue that the explicit defense of the model (as provided in Lehrer (1976) and Lehrer and Wagner (1981)) is not sound.

Keith Lehrer provides a "consistency argument" (Lehrer and Wagner 1981, 22)[17] in order to defend the claim that the result is a consensual one. The argument claims that once agents accept to enter into deliberation via the Lehrer-Wagner model, they accept that their judgment be a function of two pieces of information: 1) the mutual assessment of their degrees of expertise and 2) each member's individual (and independent) judgement on the subject matter under deliberation. In particular, what is required as a minimal condition for convergence is that all agents give at least *some* weight to the other agents[18], as refusing to do so would amount to pure dogmatism. Lehrer states the argument as follows.

> Actual disagreement among experts must result either from an incomplete exchange of information, individual dogmatism, or a failure to grasp the mathematical implications of their initial stage. What is impossible is that the members of some community of inquiry should grasp the mathematical implications of their initial state and yet disagree.
>
> (Lehrer 1976, 331)

And again:

[17]See also Nurmi (1985, 123), about the problem of consistency. Nurmi argues that the Lehrer-Wagner model is not consistent, in the sense of consistency specific of social choice theory.

[18]The mathematics of this condition was given in section 2.2.1. To recall, the matrix of weights of the Lehrer-Wagner must not be *reducible*. Informally the condition is that there must be a "chain of respect" among agents (see Lehrer and Wagner 1981, 98).

> [...] there is a consistency argument in favor of such aggregation [the aggregation of, for instance, my own opinion with the other group members' opinions]. If a person refuses to aggregate, though he does assign positive weight to other members, he is acting as though he assigned a weight of one to himself and a weight of zero to every other member of the group.
>
> ...
>
> One justification for aggregation is consistency, since refusing to aggregate is equivalent to assigning everyone else a weight of zero and aggregating.

(Lehrer and Wagner 1981, 43)

The line of defense that Lehrer and Wagner adopt is at least questionable. They claim that agents are rationally required to aggregate, because they have already accepted to deliberate with the model. To them, this means that agents have accepted to give weights to one another and, when that is the case, they are expected to aggregate.

A first observation is that one could refuse to endorse this type of deliberation procedure in the first place and request an independent justification for why it is rational to accept the model in the first place. In other words, while agents already in the model may be needed to accept its mathematical implications, a person (the member of a committee) may not agree with the procedure described by the model in the first place.

That critique is not conclusive, and a counter-argument is at hand; it is true that agents who do not accept the model in the first place, need not be considered irrational if they don't aggregate, nonetheless in Lehrer (1976) it is claimed that the model is a realistic representation of a deliberation procedure and that agents who refuse to aggregate are *equivalent* to dogmatic agents in real-life terms. If the model is realistic, then independent justification for why agents should aggregate is not needed, as claiming that agents do not aggregate amounts to saying that agents are dogmatic; if that is not the case, on the other hand, Lehrer needs to provide an independent justification of the aggregation process[19].

In sum, the best possible formulation of the standard argument in defense of the model is that its mathematical structure captures *a typical* situation

[19]It might also be argued that for the Lehrer-Wagner model to be realistic it should also include the trade-off between the "cost", incurred by individual members, of agreeing with an opinion that does not fully represent their beliefs, and the costs involved with the possibility of not reaching a consensus at all.

of disagreement and refusing to change one's opinion would be equivalent, in mathematical terms, to assigning a null weight to all other members and full weight to oneself. This situation, which could prevent the model from reaching a consensual value, is one of pure dogmatism.

There are at least two remarks to be made about the argument by consistency. The first remark is that the argument relies on the assumption that the model captures a typical situation of disagreement, in other words, that the model is a realistic representation. The second observation is that it is not so clear from the literature on disagreement that one should not (or should *never*) be dogmatic.

I briefly address the two concerns without attempting to give a definitive solution.

1) Thesis: "The model is not a realistic representation of a situation of disagreement." While it is hard to state exactly what it takes for a model to be *realistic* or *non-realistic*[20], there are at least some intuitive reasons to claim that the model is not realistic. Firstly, it may (and often does) take a very large number of iterations of the matrix or weights (W) in order to reach a consensus; but in a realistic scenario, a group in deliberation is unlikely to undergo several iterations (at least as many as the mathematical model normally requires) to reach a consensus. A second worry is that the simple version of the model uses the same matrix W at each step of the iteration process. This would not be a problem for an averaging procedure, but if the model needs to be a realistic one, it is not clear why deliberating agents would use the same information over and over again.

It is true that Lehrer and Wagner provide an extended version of the model [21] in which, at each step, new information is gathered and the matrix of weights changes accordingly; nonetheless the extended version has the disadvantage that the conditions for convergence are stronger and, mostly, that the information gathered at each step is not at all compatible with a realistic approach: In the extended model the weights gathered for the second step of the iteration, w_{ij}^2, are assessments that each agent gives on the expertise of all members of the group; this time the expertise in question is not expertise in judging the subject matter, but rather in judging their fellows. In other words, at step 2 weights represent the *experts' expertise on judging other experts*. Mutatis mutandis, at step 3 weights represent *expertise on judging expertise on judging expertise*. So on for the successive

[20]It could be argued that a model, exactly because it is a model, is never a *realistic* representation of its object. For the purposes of this thesis it is not necessary to decide the question of which properties make a model a "realistic" one. For a discussion see Frigg and Hartmann (2006).

[21]See Lehrer and Wagner (1981, Chapter 4: *The extended model*).

steps. The model quickly becomes cognitively absurd, as no one would expect agents to be able to represent *n-th* degrees of expertise and to judge their fellows on such *n-th* degrees.

2) I address now the second concern, the thesis that "dogmatism need not be a sufficient condition for irrationality". Lehrer (1976) claims that giving a null weight to all other group members and full weight to oneself is dogmatic and therefore irrational, but the 'therefore' is not obvious and needs justification. For instance, recently Kelly has argued for a form of "epistemic egoism without apology" Kelly (2005, 192). In brief, a situation of disagreement between epistemic peers does not commit one to revise one's belief. This situation coincides with the case of dogmatism that Lehrer advises against, and whereas authors such as Thomas Kelly and Adam Elga provide reasons pro or against dogmatism or, call it by another name, "epistemic egoism", Lehrer does not have arguments either pro or against, exception made for his *moral* stance against dogmatism. A moral stance, however, is not enough to *rationally* justify aggregation.

It should be clear by now that the classical line of defense of the model is not sound. For this reason, in the following section I will provide a defense that does not need the assumption that the model is a realistic representation of how groups deliberate, nor the assumption that goes from the rejection of dogmatism to a rational requirement of belief-updating. The defense I will provide eludes the need for independent justification of the aggregation procedure.

With reference to table 6.1 (above), the advantage of the argument I will present in section 2.5 is that it provides a justification of the model at each of its steps, rather than a holistic one. It is reasonable to claim that such justification carries across each step and in turn justifies the model as a whole.

2.5 Lehrer-Wagner as an updating model

How can we defend the rationality of the Lehrer-Wagner model without appealing to the realisticness[22] of the model or rejection of dogmatism? This line of defense starts by observing that in the Lehrer-Wagner model agents are not merely aggregating, that is, their judgments are not simply being merged with the other agents' judgments. On the contrary, agents are updating their belief in the light of disagreement from the other agents.

[22]The neologism 'realisticness' is here used in the sense of Mäki (1998), who suggests to use it when one is to refer to the property (or a set of properties) — e.g. "the property of being realistic".

Since it is rational (non-dogmatic) to give a positive weight to at least some of the other agents in the group, our belief will turn out to be a function of our own judgment and other agents' judgments taken as *internal* evidence. This is because in the scenario we are considering there are two sources of evidence, external evidence (facts on which agents base their independent assessment) and internal evidence, that is disagreement in the group.

Granted that disagreement counts as evidence, and that rejection of dogmatism calls for updating, in the light of new evidence agents should also accept their final judgment to be a function of all the weights they assign to the other agents multiplied by relative assessments the agents give on the subject matter. This amounts to take, according to Lehrer and Wagner (1981), at each step k of the process, an updated arithmetic average. For instance, in a group of three agents, at step 1, *agent* 1 will update her judgment in the following way[23]: $w_{11} \cdot p_1^0 + w_{12} \cdot p_2^0 + w_{13} \cdot p_3^0$, mutatis mutandis for the other agents. Call the result from the previous step p_1^1 (the opinion of *agent* 1 at step 1 of the iteration process). At step 2 some disagreement may still be present; in the light of this, agent 1 will again update her judgement as before, but this time with the new set of updated assessments reached by all agents at the end of step 1: The new updating function for *agent 1* will be $w_{11} \cdot p_1^1 + w_{12} \cdot p_2^1 + w_{13} \cdot p_3^1$, call this p_1^2. So on and so forth for the successive steps. The mathematical properties of Markov chains guarantee that at some finite step of this process $p_1^k = p_2^k = p_3^k$.

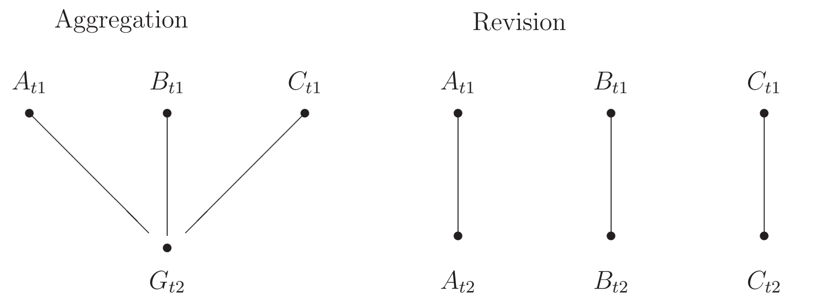
To restate the point, we can assume that in the Lehrer-Wagner model agents are not aggregating, rather updating. The fact that their updating process is such that it converges to a consensus is only a matter of the chosen updating procedure. The important result from using this method is that the final result is not a compromise. Imagine the following scenario: We are to decide on the probability of rain tomorrow such that the forecast will be out on tonight's news. I claim the probability is .7, my colleague Tom claims that it is .85 and my other colleague Isabelle states .4. According to much of the literature on forecasting (e.g. see Armstrong (2001)) we should give, as a forecast, the unweighted average (for instance, the arithmetic average .65). This is what we *agree to broadcast*, not what *we agree as to being the right forecast*. In the Lehrer-Wagner model, the situation is different, since we are all updating our judgment in a rationally justified manner and coming to hold the same belief, namely that the probability of rain is, for example, .58[24].

[23]Call p_i^0 agent i's subject matter assessment at the beginning of the deliberation process, viz. step 0.

[24]The value .58 is derived from the following set of unnormalized weights for the group in question: $w_{1*} = \{1, .1, .8\}$; $w_{2*} = \{.5, 1, .9\}$; $w_{3*} = \{.5, .3, 1\}$, which reflects the fact

The difference is remarked in Bradley (2006), where the author uses the terms "aggregation problem" and "revision problem". As Bradley states it, the Lehrer-Wagner model "purports to show that rational revision must lead to a consensus." (Bradley 2006, 147). In short, in the model agents update according to a specific mathematical procedure which, thanks to the convergence properties of Markov chain processes, makes their individual updates converge to a unique value.

Figure 2.1: Aggregation and Revision



These observations can be made explicit by use of figure 2.1. In figure 2.1, the graph labeled 'aggregation' shows a process in which three agents' beliefs [25] are merged, through an appropriate function, into a unique belief G_{t2}, which need not necessarily be endorsed by any of the agents, but is *endorsed* by the group[26]. The graph labeled 'revision' shows a process in which three agents' beliefs are revised (in light of new evidence) and transformed into the three newly endorsed beliefs A_{t2}, B_{t2}, C_{t2}, with no further specification as to whether they coincide with (or differ from) one another.

that agents are quite self-confident, that *agent 2* is not considered a very good weather forecaster, that *agent 3* is considered a very good expert and that both *agent 2* and *agent 3* do not quite have a strong opinion in either sense on *agent 1.* This example shows that the *LW* model requires more information than the information required for a compromise (e.g in majority voting), namely, information on the mutual opinions among experts in the group.

[25] The agents are labeled A, B and C, tx indicates the time.

[26] If talking of groups that can endorse beliefs in this context seems inappropriate, the same conclusions can be reached by saying that G_{t1} is the belief *associated with* the group.

In the Lehrer-Wagner model, the two processes illustrated above are combined as illustrated in figure 2.2.

Figure 2.2: Consensus in the Lehrer-Wagner Model

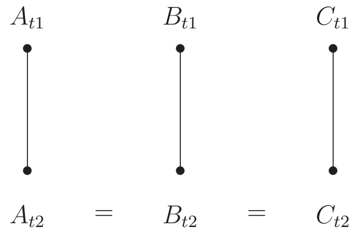


Figure 2.2 shows a process in which three agents' beliefs are revised (in light of new evidence) and transformed into the three newly endorsed beliefs A_{t2}, B_{t2}, C_{t2}. Moreover, in this case, the three new beliefs coincide.

In conclusion to this chapter, it turns out that the Lehrer-Wagner model does not belong to either of the two families presented in section 2.1.1, because it is neither a purely aggregating procedure, nor a general framework for consensual deliberation. One can conclude from the consideration made in this section that the best defense of the Lehrer-Wagner model is as an updating one. The peculiarity of the particular updating procedure used is that it leads to convergence of views under specific mathematical conditions.

It must be noticed, however, that the justification of the Lehrer-Wagner model as an updating procedure that leads to convergence, does not solve the problem of uniqueness. On one hand, the model in its original formulation is quite general as to the admittance of different ways of deriving the weights that are used for updating. On the other hand, in the discussion of the Bounded Confidence model (section 2.3) I explained that there are several possible averaging procedures allowed, as Hegselmann and Krause (2005) shows. A more complete account on the rationality of the model should explain why a specific updating function is rational, or if its "just as rational" as some other one, or else why it is "more rational" than others.

In the next chapter, I will consider two specific updating methods, in the context of the epistemology of disagreement: the Bayesian updating method and the linear method. I will investigate the question of whether either of the two can be used to resolve a situation of disagreement, and consequently lead to the formation of a consensus.

Chapter 3

Resolving epistemic disagreement

3.1 Epistemology of disagreement

One of the fundamental questions for the epistemology of disagreement is "what should you do when you discover that someone [an epistemic peer[1]] firmly disagrees with you on some claim P?" (Frances 2010). The three main answers, offered in the literature, go under the names of *steadfast position*, *precautionary stance*, and *conciliatory stance*. They prescribe, respectively, to "hold on to your beliefs", "suspend your beliefs", and "update your beliefs", when you are disagreeing with someone whom you deem to be as good as you at making judgments on the issue object of the dispute. The three viewpoints just mentioned provide an essential taxonomy of the problem of disagreement.

It is quite uncontroversial that there is one way only to hold a steadfast position. Imagine I were told "whenever you face a situation of disagreement on the truth of P with another epistemic agent, your epistemic attitude on P should remain unchanged". It would be meaningless for me to ask "*how* should my epistemic attitude remain unchanged?". There is also one way only of suspending one's own beliefs, as prescribed by the precautionary stance, although it should be noted that different practical directives could be

[1]Elga postulates the following condition for "epistemic peerhood": "[. . .] you count your friend as an epistemic peer with respect to an about-to-be-judged claim if and only if you think that, conditional the two of you disagreeing about the claim, the two of you are equally likely to be mistaken." (Elga 2007, footnote 21). Despite arguments to the contrary (see King 2011), and for the sake of the analysis in this chapter, I will assume that it makes sense to postulate the existence of epistemic peers, or at least an imperfect version of Elga's idealized condition of epistemic peerhood.

given, for the purposes of action theory, in order to address decision making impasses that may arise from a precautionary stance on disagreement.

An interesting position to hold on disagreement is the conciliatory stance, because the directive "update your beliefs on the matter of a disagreement with an epistemic peer" prompts the further question "*how* should I update my beliefs?". In the first place, a distinction must be drawn between cases of disagreement that admit of continuous values, and cases that only admit of binary values. In the binary case, the only way to change one's beliefs, in a situation of peer disagreement, is to move *all the way* towards the position of our epistemic peer. Following is an example. Suppose I learned that you believe that extraterrestrial life exists. Assuming I initially believed that extraterrestrial life does not exist, the only way for me to be conciliatory is to change my belief from "I don't believe in aliens" to "I believe in aliens"[2].

If, in the same exact situation, you also adopted a conciliatory spirit, then we would end up in a circle in which we will never resolve our disagreement, unless given other further instructions on how to get out of the deadlock. While binary cases might need a special treatment, if the conciliatory stance is to be defended, those are normally not the standard cases analyzed in the disagreement literature, which instead focuses on situations where the matter of disagreement is on continuous or probabilistic values (see Lehrer 1976; Aumann 1976; Hegselmann and Krause 2002; Elga 2007; Christensen 2007, 2009; Kelly 2010).

A conciliatory account on the problem of peer disagreement, on the one hand, will have to provide a rationale for the general claim that we should respond to disagreement by moving closer to our adversary's position. On the other hand, it should also, possibly, provide some concrete instruction on *how* close we should move, that is, which updating function we should use, once we discover we are in a situation of disagreement with an epistemic peer. The former task has been at the center of much epistemological discussion on disagreement. The latter however (namely the problem of which specific way of updating in the light of disagreement is "best"), despite some exceptions[3] has been mostly neglected.

In the following sections I will analyze in more detail the conciliatory position, and the *equal weight view* (EWV, henceforth) in particular. EWV

[2] Of course I am assuming here that the belief in question, namely that extraterrestrial living beings, or aliens, exist, can only take two values, that is 0 and 1.

[3] Recently, the task of providing a rationale to some updating procedures has been carried out in, Shogenji (2007) and Jehle and Fitelson (2009). Shogenji shows that the proportional weight view is incompatible either with Bayesianism or with probability theory, while Jehle and Fitelson undertake the task of analyzing different versions of the so-called Equal-Weight-View.

is at times explicated in Bayesian terms (Elga 2007), the details of which I will present in section 3.2. I will argue that a Bayesian treatment of disagreement (exemplified with Elga's EWV) is irrational in the sense that it provides only a subjective, and in fact arbitrary, procedure for updating on disagreement. Conciliatory positions, however, can also be explicated in terms of linear updating (see Lehrer 1976; Lehrer and Wagner 1981; Hegselmann and Krause 2002, 2005), which will be discussed in section 3.4. In that section I will show that linear approaches are also irrational, albeit this time in the sense of *anti-rational*, and therefore also fail to provide a solution to the peer disagreement problem.

3.2 Bayesian treatment of disagreement

Any position mandating a rational agent to update one's beliefs in the light of disagreement has to hold true the claim that disagreement constitutes evidence on which to update. While it can be ignored, for now, the problem of what specific type of evidence disagreement is, it is in the light of new evidence that we gain new beliefs, and justifiably change our current ones. For that reason, if disagreement has to have a special role in our epistemology and our process of belief formation, it is because it constitutes evidence on the basis of which we should update.

Assuming that disagreement constitutes evidence on which to update, perhaps the most intuitive way to deal with it is by means of Bayesian conditionalization. Bayesian theory prescribes how to rationally update one's degrees of beliefs in the light of new evidence and, despite its critics, remains at the moment the most advanced theory for treating evidence. Indeed, one of the most discussed views on disagreement (EWV), is at times presented in Bayesian terms:

> **Equal weight view.** Upon finding out that an advisor disagrees, your probability that you are right should equal your prior conditional probability that you would be right. Prior to what? Prior to your thinking through the disputed issue, and finding out what the advisor thinks of it. Conditional on what? On whatever you have learned about the circumstances of the disagreement.

Elga (2007, 490)

Elga (2007) provides a number of very convincing reasons as to why EWV is a rational stance on the problem of peer disagreement. The salient point of his argument is the observation that in a situation of disagreement it is true that "[epistemic peers] are equally likely to be correct." (Elga 2007, 487). Sosa summarizes the point as follows:

> Several philosophers have converged on the view that the proper response to disagreement among ostensible epistemic peers is for each to give equal weight to the opponent's view. We are told *that it would be unreasonable to downgrade the opponent based simply on the substance of the disagreement.*

(Sosa 2010, 5)

For the sake of the argument, I will assume that Elga is correct in stating that, in a situation of disagreement, epistemic peers are rational in thinking that they are equally likely to be correct. Nonetheless, I will argue that Elga's EWV falls short of providing a satisfactory answer to a further question, mentioned above, that a conciliatory stance on disagreement prompts; to wit, "*how* should we update our belief on the subject matter my epistemic peer and I are disagreeing about?" It is by failing to address that question that Elga's view, and in fact any Bayesian interpretation of "updating in the light of disagreement", is irrational, where the sense of 'irrational' will be made clear in the foregoing paragraphs.

Let us put in symbolism Elga's EWV. Call a's degree of belief in H '$P_a(H)$', and b's degree of belief in H '$P_b(H)$'. Let us assume that a's and b's degrees of belief differ, and that $P_b(H) = y$; it is irrelevant, for now, to know what the value of $P_a(H)$ is. Elga says, at this point, that rational agents should conditionalize on the evidence gained from the "circumstances of disagreement". But what does a learn, exactly, about the "circumstances of disagreement"?

The most obvious thing she learns is that $P_b(H) = y$, that is, she learns what b thinks about H. The first attempt to conditionalize, then, should be made starting from this piece of information, in the following way: Upon finding out what b thinks of H, a's posteriors should be equal to $P_a(H|Disagreement)$, more specifically to formula (3.1).

$$P_a(H|P_b(H) = y) \tag{3.1}$$

(3.1) expresses the simplest interpretation of the commandment "conditionalize upon learning about the circumstances of disagreement." The problems with that interpretation arise when we develop the terms of (3.1), using Bayes' theorem, as shown in (3.2):

$$\frac{P_a(P_b(H) = y|H) \cdot P_a(H)}{P_a(P_b(H) = y)} \tag{3.2}$$

In (3.2), the denominator and the left-hand side of the numerator are items of information that cannot be straightforwardly obtained in a situation of disagreement. Consider the following. The left-hand side of the numerator reads '$P_a(P_b(H) = y|H)$' and the denominator reads '$P_a(P_b(H) = y)$'. Let us ask "what is the question I should be in a position to answer in order to obtain the information required in the left-hand side of the numerator, and in the denominator?"

First, let us look at the numerator. I assume I am an epistemic agent (a); in order to know the value of $P_a(P_b(H) = y|H)$ I should be able to answer the question "given that I know H, what can I say about the probability that b will assign probability y to H?". But, knowing that H is the case, how am I entitled to form an opinion about b's priors for H? The question is not trivial and I will show in the following that answering the question implicates a large number of additional assumptions on the situation of disagreement.

Suppose I knew that H is quite an obvious and widely accepted theory; I thus think your priors for H will likely be the same as mine, provided our cognitive abilities are similar. Instead, now suppose I knew that you detest everything that, you think, only *seems* obvious; perhaps you would think H is *too* obvious to be true. If that is the case, knowing H, I should think your opinion on the truth of H would likely differ greatly from mine. On what principles, then, should I form an opinion on your priors for H, given only my knowledge that H is the case?

It is evident in this case that the expression $P_a(P_b(H) = y|H)$ is a fully subjective value; in other words, there is no systematic way of evaluating such expression. Even more, however, the example above shows that slightly different initial situations, one in which I think b is a "conformist", and one in which I think b is an "nonconformist", are likely to lead me to give completely different values to the same expression. It is easy to imagine a wide array of possible initial assumptions and variations in the picture which will make the values of that same expression change randomly and without apparent conformity to rational rules.

Equally worrisome is the case for the denominator, $P_a(P_b(H) = y)$, the details of which are very similar in nature to those for the previous case. On what rational grounds can I form priors about your priors? Suppose I know nothing about you, other than the fact, assumed throughout this chapter, that we are epistemic peers; how can I have a justified claim about the fact that your priors for H will be, say, y? It should be clear that very few situations of disagreement, perhaps none, will reveal the type of information required for Bayes' formula to be implemented in the case of disagreement. Furthermore, it is arguable that there cannot be rational guidance for obtaining that type of information, given that apparently irrelevant variations in the initial assumption can change the value that formula (3.1) will take.

To summarize the points just made, it seems that at least the simple interpretation of Elga's EWV does not leave us with an algorithmic solution for responding to disagreement, but rather with a formula that only a very idealized agent could use, one able to calculate the posteriors in the light of disagreement. That is, in itself, not yet a reason for claiming that EWV is anti-rational. The view, however, is not rational in the sense that it leaves the rational agent facing disagreement with a number of *arbitrary* rules for updating, as it was shown in the examples above. To reiterate, Elga's EWV is not irrational in the sense that it clashes with some specific principle of rationality. But using Bayesian updating in response to disagreement leaves us with an arbitrary procedure, and Elga's EWV is in that sense *not rationally motivated*.

One could argue that the problem is not with the use of Bayesian theory *per se*, but rather with the interpretation of conditional probability by means of the so called *ratio analysis*. Hájek has brought the fact to attention in a number of cases where the so called ratio analysis of probability does not give us the right answer, even though we have a very clear intuitive notion of what the conditional probability for those cases are (see Hájek 2003a,b).

Hájek suggests an alternative: "I suggest that we reverse the traditional direction of analysis: Regard conditional probability to be the primitive notion, and unconditional probability as the derivative notion." (Hájek 2003a, 315). While it is an interesting research project to see whether the notion of conditional probability could be taken as a primitive in the case of peer disagreement, the task cannot be undertaken in the space of this chapter and will have to be left for future work.

To conclude this section, I have not claimed that Bayesian updating is in principle the wrong approach to disagreement, but that if the desideratum is to have a finer-grained theory on how to respond to disagreement, and especially one that can be rationally motivated, the foregoing arguments suggest that Bayesian theory is at its present stage of analysis not a rational way for updating on disagreement. Lack of such a finer-grained theory results in the proposal of an arbitrary strategy for responding to disagreement.

3.3 The Equal Weight View reformulated

Elga himself seems to recognize that there may be issues with his own initial formulation of EWV in Bayesian terms, and, still maintaining the Bayesian framework, gives a partial reformulation of the original definition as follows:

> "[*as before*] Conditional on what? On whatever you have learned about the circumstances of how you and your advisor have evaluated the claim."
>
> Elga (2007, 490: footnote 26).

What Elga seems to be doing with this reformulation is to interpret the "circumstances of disagreement" in the previous formulation as the "circumstances of how you and your advisor have evaluated the claim." The latter interpretation, however, is at risk of being too loose.

Imagine the following classical example of a situation of disagreement, the split-the-bill case (see Christensen 2009). You and I are at a restaurant and, at the end of our dinner, each of us calculates the bill independently. I claim that the bill is $43, and you say that it is $45. One of us must be wrong although we do not know who. Does disagreement tell us anything about the way in which each of us has evaluated the claim "the bill is $\$X$"? It does not. All that the situation of disagreement tells us is that one of us *must be* wrong (clearly, we *can*, in principle, both be wrong). Imagine the following similar situation: I calculate the bill with an abacus, I then calculate it again with pen and paper. The two results do not match. Clearly I must have made a mistake in at least one of the two calculations. The mismatch, however, does not tell me anything about the correctness or wrongness of either.

What should I do then in a situation of disagreement? I should go over my calculation (my reasoning) again. This much seems straightforward. However, if this is the implication of EWV, as reformulated in Elga's footnote 26, then disagreement does not play any special role in my epistemology, at least not a role in any way different from all other situations in which I might become aware of a mismatch between two results which should be by hypothesis equal, had I not made any mistake.

The question that epistemology of disagreement investigates is what special role, if any, disagreement has in our state of belief towards a certain proposition, or theory, H. If I should update my beliefs in the light of disagreement it is because disagreement itself constitutes evidence, and not because I gain new *factual* evidence in the process of rechecking my reasoning. If I find out that I am disagreeing with someone, the least I can do, provided I give at least some credit to the other person, is to check my reasoning once more. But this is not to say that disagreement is evidence, at least not evidence on which I would update, whereas it is, by pure logic, evidence of the fact that one of us is mistaken.

That disagreement is evidence, however, is implied by the conciliatory position when claiming that, once we have taken into consideration all the available evidence, redone the calculations over and over again, and checked our reasoning enough many times, if we are still disagreeing, then we should move our belief towards the other party's belief.

It should be clear that the second interpretation of the EWV cannot answer the problem of disagreement because it does not treat disagreement as evidence, or at least not the type of evidence on which one can update. It remains to be seen if there are other ways to update one's beliefs in the light of disagreement. The next section will be dedicated to a possible alternative: the linear updating interpretation of the conciliatory stance.

3.4 Disagreement and linear updating

Not all belief updating needs to be treated with Bayes' formula for calculating the posteriors, although it is a different question whether all belief updating should be at least compatible with the rules of Bayesian reasoning[4]. The problem, however, will not be discussed in this chapter. It is a fact that linear updating functions are used in a number of conciliatory models, like the Lehrer-Wagner model, the Bounded-Confidence model, most weighted average models, and also in the often mentioned split-the-difference view, at times conflated with the EWV but not necessarily equivalent.

Linear updating is a form of updating one's beliefs by means of a *linear function*. The question I will address here is whether linear updating is a suitable form of updating in the light of disagreement. For simplicity, I will consider only the split-the-difference view, although the foregoing arguments hold for all other forms of linear updating. The split-the-difference view states that if you and I disagree on a certain issue, then we should take the arithmetic average of our beliefs, so to converge to the middle point. The view is clearly a good algorithm for updating because, given a case of disagreement, we know exactly by how much we should scale each of the disagreeing parties' judgments down, that is by $\frac{\sum_{j=1}^{n} O_j}{n}$, where n is the

[4]Bradley shows that Bayesianism is incompatible at least with some forms of linear updating (see Bradley 2006, 151: Proposition 3).

number of members in the disagreeing group and O is their probability judgment. For that reason, a conciliatory view that uses linear updating does not run into the problems of arbitrariness highlighted in section 3.2 for Bayesian updating. But is linear updating a rational answer to the problem of peer disagreement? In the following I will argue that it is not.

To recall, the problem of peer disagreement is "what rational epistemic attitude should we take, when we disagree with an epistemic peer?" Consider the following scenario. You and your friend Sara are planning to do some bird-watching at the Isle of May (in Scotland) this coming weekend. You have been there a few times and you know that the trip is not worthwhile in case of bad weather. You and Sara both happen to be weather forecasters; you think that the probability of rain for the weekend is .8, whereas Sara thinks otherwise. Moreover, you and Sara think that, both of you being epistemic peers, you should "split the difference", that is, take an arithmetic average of yours and Sara's judgments.

Why should you split the difference? "It is rational", is the answer from the conciliatory split-the-difference view. In fact, you are not just holding a "conciliatory spirit" in the discussion with Sara about what to do over the weekend. You are not worried about keeping your friendship with Sara going; that would not be an epistemic reason for updating in the light of disagreement. Rather, you think that your disagreement with Sara constitutes evidence for the fact that you and Sara are equally likely to be wrong in your respective forecasts.

> [...] you count your friend as an epistemic peer with respect to an about-to-be-judged claim if and only if you think that, conditional the two of you disagreeing about the claim, the two of you are equally likely to be mistaken.
>
> Elga (2007, footnote 21)

If that is the case, as it should be if the split-the-difference view is rational, claiming that you should split the difference means that you agree with the statement "I think that the probability of rain is .8, but I might be wrong". Exactly how wrong? "I might be wrong with probability .5". The question I will address in this section is whether the belief, or state of belief, 'I think that the probability of rain is .8, but I might be wrong with probability .5' is rational. At least two cases must be distinguished, namely whether the probabilities in question are objective or subjective, as follows.

Suppose your belief 'the probability of rain is .5' is your subjective degree of belief. Being a professional forecaster, you have a number of mathematical models, charts, and weather maps available, plus a great deal of background knowledge, all of which makes you a reliable expert on the topic of weather prediction. At the end of the day, however, you know that tomorrow it will either rain or not rain, and you know that the probability you are assigning to the event RAIN is only an expression of your limited knowledge as an expert. This picture is perfectly compatible with, and in support of, the idea that the expression "the probability of rain tomorrow at the Isle of May is .8" is about subjective probabilities.

If that is the case, what does it mean to add to your sentence the clause "but I might be wrong"? Of course you might be wrong, the probability you expressed is a function of your limited knowledge on the event 'it will rain tomorrow at the Isle of May', which, it was assumed in this context for the sake of the argument, is an event in a deterministic world. Provided you knew all about weather patterns and so on, you would either *know that it will rain tomorrow*, or *know that it will not*. But assuming that the probability you express is subjective, then what sense does it make to add the clause "but I might be wrong"? In particular, what sense does it make to add the clause "but I might be wrong with probability .5"?

It is clear that adding the clause "I might be wrong" to an expression containing subjective probabilities is redundant, thus the idea of updating in the light of disagreement finds no ground in the context of subjective probabilities. In such context, disagreement does not constitute any new evidence in addition to that which one already possesses. For that reason, scaling down one's initial opinion is not a rational procedure as it is equivalent to double-counting: I know I may be wrong, but disagreement, not even with epistemic peers, should not prompt me to revise my belief further.

Let us now imagine the second possibility, that the expression "tomorrow it will rain on the Isle of May with probability .8" is not an expression of your subjective probability for the event RAIN, but rather about an objective probability. Whatever objective probabilities may be exactly, let us assume that when by giving your forecast (probability of rain = .8), you are giving an estimate of the objective probability of the event RAIN. That probability may be indicated, for instance, by the mathematical models you are using for computing the probability of rain.

Now suppose your friend Sara shares her forecast with you, she uses different forecasting models, and her models indicate a different estimate of the objective probability for the event RAIN. If that is the case, you are faced with a situation in which the probability of the event RAIN is a certain value x according to you, and a different value y, according to Sara. It is rational to say that both yours and Sara's values constitute *information*, which should indicate what the objective probability of rain is. Let us call such value 'α'[5].

Now suppose that both you and Sara have read Scott Armstrong's *Principles of Forecasting*, and learned that combining information is in many cases more likely to give the most accurate forecast. Is it then rational for you to update your own judgment and merge it with Sara's? In this case there seems to be nothing against the rationale of that procedure, since you are reporting on information from your forecasting model, which of course could be subject to further revision in the light of additional information. But given the scenario just presented, one can hardly claim that there was disagreement at all between the epistemic agents (the epistemic peers) in the first place.

In fact, in the scenario just presented, you and Sara had a number of beliefs related to the event RAIN. You believed that whatever the probability of rain was, it was an objective probability with value α. You believed that the most accurate forecast was some adequate combination of all the available information pointing in the direction of the *true* value of α. And finally, you believed that the information available to you was .8, whereas Sara believed that the information in her possession was, let us assume, .9.

It is hard to see how that situation was a situation of disagreement. Imagine you and Sara decide to take a straight average of your forecasts. The value of your combined forecasts (your consensual forecast) is then '$.8 \cdot .5 + .9 \cdot .5$'. Does this mean that you have somehow changed your beliefs for event RAIN from .8 to 8.5? You have not. You never believed that "the probability of rain is .8" because what you really believed was the fact that your forecasting model gave the indication '.8' about the *objective* probability of rain α.

3.5 Conclusion

It is easy to see that the arguments in the previous section apply not only to the split-the-difference view, but to all other forms of linear updating, such as the Lehrer-Wagner model or other forms of weighted average.

It is time now to summarize the conclusions reached so far. The conciliatory stance on disagreement implies that disagreement constitutes evidence on which rational epistemic agents should update their beliefs. Sometimes this idea is presented in Bayesian terms. Section 3.2 showed that, if updating

[5]The argument becomes slightly more complicated if we allow for vague objective probabilities, the problem cannot be addressed in the space of this chapter, even though the conclusions do not vary significantly.

is based only on the information regarding the disagreement itself, then the prescriptions the conciliatory view gives become arbitrary, and in that sense irrational. On the other hand, if updating is done in the light of the process by which the agent goes over her reasoning once again, then disagreement does not seem to have any special role in the updating process. In particular, disagreement does not constitute evidence, as the conciliatory view would require.

Bayesianism, however, was not the only option. In section 3.4 I argued that linear updating is also not a rational response to disagreement, when what is being updated are subjective probabilities. If instead what is being updated are objective probabilities, over which one does not have a specific truth-implicating belief, then the case is not one of genuine disagreement.

At this point, a question may arise as to what exactly the foregoing arguments imply. In particular, I have not argued for the positive claim that "one should not update her beliefs, and should not move her belief in the direction of her peer's belief, in a situation of factual disagreement." What I have argued is that the prescriptions coming from the conciliatory stance, which seems, *prima facie*, rational and justified, run into problems when interpreted in either Bayesian or linear terms. Clearly the possibility remains that other forms of updating in the light of disagreement be viable, and that both linear and Bayesian updating are simply inadequate theories for capturing our rationality when it comes to the problem of disagreement.

While the latter possibility cannot be excluded on a-priori grounds, it should be noted that there are strong independent reasons for accepting the validity of Bayesian and linear updating as a way of capturing human rationality, and assuming that rational response to disagreement is simply a special case of human rationality would grant that the two approaches should be no exception when it comes to the problem of disagreement.

To conclude, while leaving the possibility open for further options as to how to update in the light of disagreement, the arguments so far presented provide strong support to the thesis that the present conciliatory stances on disagreement are irrational, in the two senses of "rationality" that were made explicit in this chapter.

Chapter 4

Consensus and networks

4.1 The status of the Lehrer-Wagner model

The conclusion from chapters 2 and 3 was that the Lehrer-Wagner model cannot be taken as a genuine model for consensus formation, at least not in the sense of consensus that was given in section 2.1. As an aggregation model, the Leher-Wagner model cannot be defended as a consensual one (section 2.5), and as an an updating model, on the other hand, it is not a rational aggregating function (section 3.4).

Nonetheless, it was said early in chapter 2 (section 2.1), that it is a common linguistic use to call some aggregation models 'consensus models', even though they cannot be said to produce a true consensus of beliefs among the members of, for example, a deliberating committee. Similarly, the Lehrer-Wagner model should be considered a particularly refined aggregation model or voting mechanism[1]. The Lehrer-Wagner does not only take into considerations the opinions of the deliberating members on the issue under deliberation, like most voting functions, for example, do, but it also uses the information contained in the opinion that members have towards their fellow deliberators.

Given the presence of a matrix of weights in the Lehrer-Wagner model, and given that a certain interpretation of such matrix considers weights as the opinions that members have towards other members of the deliberating group, this chapter undertakes the task of studying the process through which disagreement is "resolved" in the Lehrer-Wagner model, when different *opinion structures* are present in the deliberating group.

From now on in this thesis, the Lehrer-Wagner model will be referred

[1]Many years after the presentation of the model, one of its authors shared a similar opinion — cf. footnote 5, in the introduction to this thesis.

to as a 'consensus model' even though I have argued against the truly consensual nature of it in chapter 2 and 3. The use, maintained in this and the next chapters, is one of pure convenience, given that the customary use of the terminology in the literature does not distinguish between the epistemic concepts of *consensus* and *compromise* that were instead pointed out in section 2.1.

The following sections are organized as follows. In section 4.2 I discuss one of the principal unresolved problems associated with the Lehrer-Wagner model as an aggregation model, to wit, the problem of how members in a committee should assign weights to each other. In sections 4.3 and 4.4 I make a proposal for a strategy for assigning weights. In section 4.5 I provide some examples of noteworthy networks, and in section 4.6 I motivate the proposal from the descriptive and normative points of view. In section 4.7 I draw some conclusions to this chapter.

4.2 Weight assignment in the Lehrer-Wagner model

The fact that agents in the Lehrer-Wagner model assign weights to each other is very important for the consensual nature of the results from convergence, insofar as weights represent trust or confidence among a group's members (see Lehrer 1976; Lehrer and Wagner 1981). However, how agents are to assign weights to one another remains one of the major unsolved problems in the work of both Lehrer and Wagner.

Lehrer and Wagner (1981) do not give a strategy or algorithm for assigning weights, although four different examples of possible assignments are given. The first three examples involve some mechanical procedures for aggregating information and, in those cases, weights are not meant to represent trust or confidence in other agents' judgments, but rather some more or less objective quantity (Lehrer and Wagner 1981, 138-140). For that reason the first three cases are omitted here. In their fourth example, on the other hand, the authors take the weights to be dependent on subjective assignments.

> When a decision problem involves neither highly structured estimation subject to a prior analysis of weighting schemes, as in the examples 1 and 2 above, nor a statistical record of past performance, as in the preceding example, then the choice of weights becomes a subjective enterprise. (Lehrer and Wagner 1981, 140)

The idea of a subjective assignment of weights is what drives consensus in the model, according to Lehrer (1976). However, making the assignment of weights a "subjective enterprise" raises a number of problems that have been pointed out in the subsequent literature on consensus.

One of the main critical stances on subjective assignment of weights is in Regan, Colyvan and Markovchick-Nicholls (2006). That paper takes the Lehrer-Wagner model as a practical option for consensus seeking in medium-sized committees. The authors present a case study in which a panel of experts are to formulate a list of criteria, which will in turn be used for selecting urban open spaces for a Californian environmental conservation project.

The problem that the committee in Regan, Colyvan and Markovchick-Nicholls (2006) has to deal with cannot be resolved by a purely scientific and objective analysis, because the list of criteria involves both ethical principles and complex multi-disciplinary evaluations, which make it impossible for one to rely entirely on "hard science". But the decision making process is not entirely subjective either, since important elements of the evaluation need to be assessed on the basis of specific expert knowledge and scientific data. Moreover, the agreement will most likely be influenced by stakeholders' interests and personal preferences.

In such context, it is argued in Regan, Colyvan and Markovchick-Nicholls (2006), the Leher-Wagner model could provide a useful framework in order for the committee to achieve a consensual resolution; in particular, a consensus that takes into consideration not only the opinions of each expert, but also the weights associated with their opinions. In other words, the Lehrer-Wagner model would take into consideration also the degree of trust, or respect, that the committee members have towards each other. As the authors stress, however, asking experts to subjectively assign weights to each other would be open to a number of both theoretical and practical problems.

In the first place, it would be a very impractical task to ask each member of the committee to assign a weight to all of his or her fellows. Secondly, members may conceal their true agendas in order to manipulate the results[2]. Furthermore, Regan, Colyvan and Markovchick-Nicholls (2006) provides a third argument against subjective assignment of weights; the argument is reported in full below, as it is difficult to summarize.

> Third, and most important, the assignment of a numerical value on a person's degree of respect for each of the other members in the group is abstract and provocative. While most people

[2]Nurmi shows that the Leher-Wagner model, when weights are assigned subjectively, is manipulable (see Nurmi 1985, 15: *Proposition 1*).

> would agree that they have different degrees of respect for, or agreement with, other group members' positions, translating that to a numerical value is non- trivial. Furthermore, group members may feel reluctant to explicitly quantify degrees of respect for other group members, or reveal their true weight of respect, as it could lead to rifts and ill feeling within the group. (Regan, Colyvan and Markovchick-Nicholls 2006, 172)

For these reasons, the authors propose a method for assigning weights based on the relative distance of two agents' opinions. In other words, the weight w_{ij} that agent i assigns to agent j will be a function of the distance between agent i's and agent j's opinions[3].

A similar suggestion has also been proposed in Hegselmann and Krause (2002). There, the authors present the *Bounded Confidence model*, a model for consensus similar to the Lehrer-Wagner, and suggest that agents should aggregate their opinion on the subject matter under consideration only with those other agents whose opinion is at a certain distance ε from their own. In other words, the admitted weights are only 0 and 1. Agent i will assign weight 1 to agent j if and only if agent j's opinion lies within the "confidence interval" '$\pm\ \varepsilon$' from hers, otherwise, she will give agent j weight 0. In that way, all *normalized positive weights* in each row $\mathbf{W}_{i*}$ of the matrix $\mathbf{W}$ will be the same.

Both the proposals in Regan, Colyvan and Markovchick-Nicholls (2006) and Hegselmann and Krause (2002) have a number of advantages; in particular, they solve the problems I highlighted above about a subjective assignment of weights. With a distance-based mechanism, weights would no longer be assigned subjectively, at least in part avoiding manipulability of the assignments. Moreover, weights would be derived directly from the information about the agents' opinions, thus providing an economical, easily quantifiable, and "sentiment-free" measurement.

There are, however, also drawbacks to that proposal. For example, imagine a case in which the decision on which the group is seeking agreement is, by nature, highly polarizing (e.g. people have very high personal stakes in the matter that is the object of the decision). The goal is to obtain a solution that is not only the win of a majority, but a function of all the opinions of the members in the group.

Adopting either of the two aforementioned solutions would not achieve the goal of promoting agreement, rather the opposite effect of polarizing the two groups. In particular, an extreme type of manipulation is still possible

[3]Details on the function and derivation of weights are left to the interested reader (see Regan, Colyvan and Markovchick-Nicholls 2006, 172).

with the proposal in Regan, Colyvan and Markovchick-Nicholls (2006): Members of the two subgroups may conceal their *true* opinion and decide to give an extreme opinion (1 or 0), knowing that it is opposite to the opinion of the other subgroup. If the other subgroup does the same, the resulting matrix will be reducible to two sub-matrices, which will independently converge to two different values.

The Bounded Confidence model has similar problems. As Hegselmann and Krause (2002) show, in a important number of cases (if ε is "small enough"), the model converges to two or more independent opinions, meaning that the original opinions of the group's members stabilize on multiple non-communicating convergent paths (see Hegselmann and Krause 2002, 10-20). In other words, as the *confidence interval* (ε)[4] decreases, the members of the group will tend to stick to their own opinion instead of moving closer and closer to the others.

While the drawbacks of assigning distance-dependent weights are not sufficient to defeat either of the proposals, there seems to be a question to be asked about the rationality of assuming that agent a will assign a higher weight to b, if b's opinion is closer to her own, as both Regan, Colyvan and Markovchick-Nicholls (2006) and Hegselmann and Krause (2002) assume. The assumption is rational insofar as it is rational to expect that people will tend to converge towards those positions that are closer to theirs[5].

While that *can possibly* be the case, one can think of many scenarios where it *needs not to*. For instance, one can imagine that a mother would put much respect or trust in her son, even if she did not agree with her son's opinions. Similarly, the president of a nation would give a high weight to the president of another nation with which she had strong economic interests, even when, taken out of context, her opinions would differ greatly from those of that president. The possible cases are many but these examples should be sufficient to clarify the point being made here, namely that distance-based assignment of weights is rational only when some assumptions about the deliberating committee are made.

In light of the aforementioned problems, the methods in Hegselmann and Krause (2002) and Regan, Colyvan and Markovchick-Nicholls (2006) should

[4] "Confidence interval" is used here in the sense of Hegselmann and Krause (2002), not to be confused with the homonymous concept used in statistics.

[5] So far, I have taken both methods in Regan, Colyvan and Markovchick-Nicholls (2006) and Hegselmann and Krause (2002) to be normative in character. Whereas the choice is not problematic for the former, it is arguable whether the latter should be taken as a normative model, at least in the authors' intentions. In principle, however, there seems to be no reason for prohibiting that the *bounded confidence* model be taken normatively, regardless of the original authors' intentions.

be taken as one of a number of possible solutions, each of which has advantages as well as disadvantages. It is in the light of those considerations that in the next section I will present an alternative method for assigning weights, one that seems particularly fruitful, for example, for counterbalancing the dynamics that tend to make a group split (polarize) on very sensitive issues.

4.3 Social influence and networks

The method for the assignment of weights suggested in this chapter takes its rationality from the observation that, in real life, groups do not come in the idealized form that is often assumed by most consensus models. Lehrer and Wagner (1981) assume that agents give *honest* assignments of weights.

> [...] if, on the contrary, the weights represent an egoistic attempt to manipulate social decisions making, then it is unacceptable to use those weights though they were a disinterested summary of information." (Lehrer and Wagner 1981, 74).

However, honest assignments cannot be taken as a realistic assumption. A member of a group may tend towards the opinion of other members not necessarily because those opinions are similar to hers, but also for a number of other possible reasons, e.g. political or economic interest, kinship, etc. In other words, in several concrete scenarios, people are not, so to speak, "on equal grounds", but rather "networked", to wit, organized in a structure (a network) in which the degree of connectedness of different agents varies depending on that agent's position in the network.

Theorists have recently drawn attention to the phenomenon of network formation in initially homogenous groups, and on how a network can affect the flow of information within a group[6]. In general, a network is a structure of connected elements (e.g. the agents in a model) that can be represented mathematically with a *graph*. A graph is an ordered pair $P = (N, E)$ which includes a set N of *nodes*, or *vertices*, and a set E of *edges* (see Weisstein 2011). The nodes of a graph, as used in this chapter, represent the agents in a consensus model, and the edges represent the existing connections among agents.

The literature on networks has evolved especially in economics and sociology. DeMarzo, Vayanos and Zwiebel (2003) address the problem of how persuasion biases can be more effective depending on how well-connected

[6]For a recent comprehensive treatment of networks in economics and sociology see Jackson (2008).

an agent is in the group. Biases, persuasion, and the structure of a network can arise from disparate situations — a city council's members may be linked more or less strongly by political, economic and even family-related interests. Similarly, the members of an environmental panel may be linked by reasons of national interest, ideology or a number of other factors.

It is in the interest of a modeler who wants to find an optimal solution for how agents should assign weights to each other, to maximize or minimize the effects of a network structure among the agents within a deliberating group. If, for example, a panel is composed of members (agents) whose opinions are known to tend towards the opinion of a known node (agent) in the network, an appropriate schema of weights would reduce the *influence* of the *central node* on the other nodes[7].

In the following sections, I will first discuss the existing literature on assigning weights on the basis of the structure of a group (section 4.4), and then formulate a proposal for assigning weights that is based on the idea of maximizing or minimizing biases and the effects of a network in a group (section 4.5).

4.4 Deriving weights from network structures

The idea of deriving weights from a network structure was first suggested by French. French (1956) analyses the convergence properties of different networks and provides a number of theorems for convergence of different "networked groups". The theory in French (1956) is only sketched out, and the proofs are presented informally, but Golub and Jackson (2007) address the same question — which networks allow a deliberative group to reach consensus — and provide necessary and sufficient conditions for convergence of the opinion of the members in a networked group.

Golub and Jackson (2007) propose a convergence model in which the flow of information in the group is conveyed through the existent network, even though their approach does not allow agents to give different weights to one another when they are equidistant in the network. Similarly, the idea in French (1956) was to sketch a mathematical theory of social influence, where influence ("power" in French's words) is represented by the capacity of one node to "exert influence" on another node (French 1956, 182-183).

Both French (1956) and Golub and Jackson (2007), however, deviate

[7]This case is exemplified in figure 4.2 — see section 4.5.2 — and will be treated in that section.

from the intentions of the Lehrer-Wagner model in that they have an essentially descriptive approach. They assume that there is a certain network of connections and influences among members of a group, and that an edge between two elements, a and b, of that group will affect in a specific way the dependence relation between the opinions of a and b. The question they answer is "what we can say about how a consensus, if any, will evolve, when such dependence relations are present in a group?".

In this chapter I suggest two main variations from the aforementioned literature. In the first place, the method suggested in the foregoing sections allows for more flexibility in the assignment of weights, that is, weights are not fixed for all links but are: a) dependent on the total distance (in number of links) between agents, and b) some variations are allowed to occur in the attribution of weights even when agents' mutual positions are equidistant.

The second and most important variation from a derivation of weights based on the network structure of a group is that here the idea of a *normative* model for consensus formation, as it was described in Lehrer and Wagner (1981), is maintained. The idea, that is, will be to use the network structure in order to maximize or minimize, according to the intentions of the modeler, the structure of influence relations in a group.

4.5 Network-dependent weights

4.5.1 A balanced network

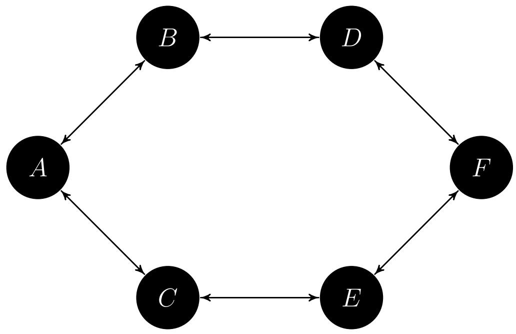
The idea behind a derivation of weights from the underlying network structure of the deliberating group is that there are scenarios in which agents assign weights based on their preferences for, or biases towards, other agents.

Imagine the following scenario in which the group is composed of diplomats from different countries. For simplicity I will assume that each agent (each diplomat) has exactly two *proximate neighbors*[8]. Agent A is neighbor with agents B and C, B will then be neighbor with A (the relation of being neighbor is symmetric) and with D, C on the other hand is neighbor with A and E. Again, for simplicity, I assume that the group is small and that it is closed, that is, there is a member F that has E and D as neighbors. Figure 4.1 represents such a group.

For example, neighboring countries may have common interests: trade, military security, environmental safety, etc. As the distance between two

[8]An agent is a "proximate neighbor" with another agent if there is an edge that connects them without passing through any other agent.

Figure 4.1: A six-node network



countries increases, however, those factors will most likely play less and less a role in their preferences towards one another.

From figure 4.1, a list of instructions for the derivation of weights can be formulated, as illustrated in table 4.1. Table 4.1 gives a simple set of rules for the derivation of weights based on the relative distance (in number of nodes) from agent to agent. For example, if $a = 1; b = 0.8; c = 0.5; d = 0.3; e = 0$, then $w_{ad} \in (0.5, 0.3)$, because agent A is two links away from agent D. According to the schema, an agent will give herself a higher weight, her proximate neighbor a slightly lower weight, her next neighbor an even lower weight, and so on. The schema can be reformulated for any number of nodes in a ring-shaped network.

Hartmann, Martini and Sprenger (2009) analyze the case of a ring-shaped network, like the one in figure 4.1, in order to provide a formal definition of epistemic peers and investigate the dynamics of consensus formation. Initially, the paper assumes a stricter schema for the assignment of weights, such that if agent A is distant x nodes from agent B, then there is only one possible weight that w_{ab} and w_{ba} can take. In other words, with reference to table 1, each " $\ldots \in (a, b) \ldots$" [or (b, c), etc.] is substituted by "$\ldots = \alpha \ldots$" [or β, etc.], where α, β, etc. are fixed values between 0 and 1.

[9]The items listed below should be read as follows: If and only if the number of edges between x and y is 0, then the weight that x gives to y is a value in the open interval (a, b). Similarly for the other cases (e.g. the number of edges between x and y is 1). The idea is that as the number of edges (that is, the distance) between vertices (agents) in the network increases, the weights decrease. Similar tables can be written, at the discretion of the modeler, in order to make weights increase, decrease or remain constant depending on the number of edges between two vertices.

Table 4.1: Weight-derivation for a six node graph[9]

1. $w_{yx} \in (a, b)$	$\iff$ # edges between x and y is 0 (case for $x = y$)
2. $w_{yx} \in (b, c)$	$\iff$ # edges between x and y is 1 (where $a \leq b$)
3. $w_{yx} \in (c, d)$	$\iff$ # edges between x and y is 2 (where $b \leq c$)
4. $w_{yx} \in (d, e)$	$\iff$ # edges between x and y is 3 (where $c \leq d$)

Hartmann, Martini and Sprenger have shown (see Hartmann, Martini and Sprenger 2009, 116: *Theorem 1*) that the consensual results deriving from the model are equal to the arithmetic average of the values in the column **P** (see equation 2.2 in section 2.2.1.) of the deliberators' opinions. In the same paper, it was shown that also by relaxing the schema for the assignment of weights (as from table 4.1), the results are robust; that is, the consensual value will be approximately equal ($\approx$) to the arithmetic average of the column **P**.

What those results mean is that whenever the group forms a symmetric network, like the one of figure 4.1 (or any isomorphic extension of that network with more agents), the different weights assigned by agents to other agents balance each other out and the situation is equivalent to that of a group where agents assign no weights to one another. The analysis, so far, is still at the descriptive level, and in Hartmann, Martini and Sprenger (2009) it was used in order to justify an "equal weight view" among epistemic peers (see Kelly 2005; Elga 2007).

From the normative point of view, however, if the choice of the modeler is to derive weights from a network structure like the one presented in this section, the implications are that whenever such structure of relations, or *power structure* (as in French 1956), is present, then the role of matrix of weights in the Lehrer-Wagner model is null or almost null. In other words, if the goal of the modeler is to reduce the influence of opinion that members exert towards each other in the group represented in figure 4.1 (or any similar group), then we can be sure that the members' influences will simply cancel each other out in such group.

4.5.2 Other networks

The example provided in section 4.5.1 is only one of numerous possible network formations that can in principle be studied in order to provide some normative guidelines for the modeler of consensual opinion formation. A thorough analysis of the dynamics of consensus formation with the Lehrer-Wagner model has yet to be carried out and is beyond the scope of this chapter. The following remarks will serve as an illustration of the many theoretical possibilities that such analysis can disclose.

The ring-shaped network presented in the previous section was an example of an extremely regular network, in which all the weights even each other out. It could, as stated before, represent a group of countries each of which has a bias towards its neighbors, and the finding there was that when that is the case, the global effect of all the biases involved does not influence the results of the consensus.

In other networks, the same effect may not occur, and the biases might influence the formation of consensus. Indeed French shows that there can be cases where “[a member of the group] will influence the others but no one will influence him.” (French 1956, 189); when that is the case, in the limit all members will converge to that member’s opinion. For instance, there may be groups in which one of the agents plays the role of the leader, and to whom all other agents assign a high degree of respect or confidence (a high weight). An example of this is what Elga calls ‘gurus’, that is “people to whom we should defer entirely” (Elga 2007, 478). Identifying the presence of a guru in a group might have important epistemic implications.

The epistemic value of the presence of a guru was one of the points left open in Hartmann, Martini and Sprenger (2009): “[...] we did not address the question of whether the leader bias is beneficial or not — this depends on the leader’s factual competence and honesty. The Lehrer-Wagner model is silent on these questions.” (Hartmann, Martini and Sprenger 2009, 120).

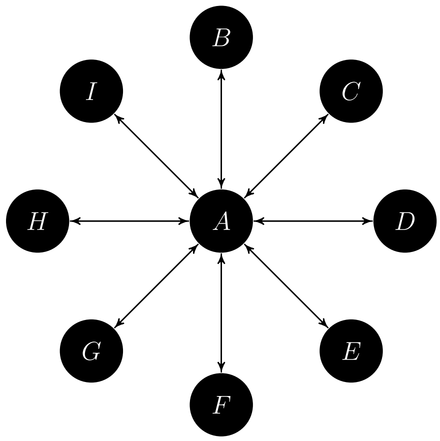
The Lehrer-Wagner model, when weights are assigned subjectively, is not sensitive to the presence of a leader in the group. However, in a star-like network (see figure 4.2), the consensus will be biased towards the opinion of the leader[10].

Deriving weights from a given star-shaped network would, on one hand, allow us to represent a real case of consensus formation in case a leader should be present. On the other hand, from the normative point of view, it would allow us to reduce or maximize the leader’s influence on the other members. The presence of a leader can be regarded as a positive or a negative effect on the group, depending on whether the opinion of the leader in question is considered “good” (e.g. truth conducive) or “bad” (e.g. misleading).

With the scheme for deriving weights exemplified in table 4.1, if properly modified, it is possible to balance the weights that agents assign to one another so that the opinion of the leader will be weighted higher or lower, depending on which of the two the modeler thinks of as the appropriate

[10] For an explanation of this see Hartmann, Martini and Sprenger (2009, 120: *Theorem 3*)

Figure 4.2: Nine-node star-shaped network



strategy. For example, if we deem the leader to be a negative influence on the group, the schema for the assignment of weights could be built along the lines of table 4.2 below.

Table 4.2: Weight-derivation for a nine-node graph

1. $w_{yx} \in (a, b) \iff \#$ edges between x and y is 0 [case for $x = y$]
2. $w_{yx} \in (b, c) \iff \#$ edges between x and y is 2 (where $a \leqslant b$)
3. $w_{yx} \in (d, e) \iff \#$ edges between x and y is 1 (where $d \ll c$)

The derivation of weights from table 4.2 causes the influence of a leader, which would normally sway the results of the consensus, to be diminished. Agents will still be giving more *confidence* (or preference) to the leader, due to the preference structure of the network, but their opinions towards the leader, will be automatically scaled down. The idea is that the consensus should be a measure of an agent's independent opinion on the matter under consideration, so that factors affecting that independence, if possible to detect, should be minimized.

4.6 Justifying network-dependent weights

4.6.1 Normative justification

At the end of section 4.2 it was said that the method for deriving weights suggested in this chapter should be taken as one of many possible methods, each of which will have advantages as well as drawbacks. In this section I will enumerate a number of reasons for justifying weight derivation on the basis of the underlying network structure. While the focus of this chapter is on the normative advantages, as in the spirit of the original formulation of the Lehrer-Wagner model, I will provide some reasons for a descriptive take on the method as well.

A normative interpretation of the Lehrer-Wagner model was the one endorsed in Regan, Colyvan and Markovchick-Nicholls (2006). According to that paper, weights should be assigned as a function of the distance between agents' opinions. This is rational, in those cases in which agents assign higher weights to agents with opinions similar to their own; for instance, we may think that musicians may assign higher weight to musicians with musical tastes similar to their own.

That scenario, however, need not always be the case. As I explained in section 4.5, we can imagine cases in which we give a higher weight to someone who is "close" to us, or whom we view as the leader, no matter what her opinion on the matter under deliberation is. In all these cases the problem is whether such *preference*, the influence of the network on the group's decision, is epistemically advantageous or not.

The main reason for normatively adopting a network-based derivation of weights is that it is in principle possible to exploit the epistemic advantages of a certain network or minimize its influence, depending on whether the particular network is positively or negatively affecting deliberation. That there are biases in groups and small committees — for example the panel of experts that are the object of the case study in Regan, Colyvan and Markovchick-Nicholls (2006) — is an assumption that does not need many arguments in its defense. If that is the case, however, the opinions of the members will move in one way or another according to the forces and biases that are present in the group. So far, this is the descriptive account of networks and social groups studied in French (1956).

Once we get to know what type of network is in place, however, the problem is whether we would like to maximize or minimize its effects. With the method suggested in Lehrer and Wagner (1981) — subjective assignment of weights — the effects of a network will be maximally evident, since agents will assign weights following their own biases completely. As it was argued in section 4.3, it is unrealistic to assume, like Lehrer and Wagner (1981) do, that weights represent an "honest assignment" of respect or trust from member to member.

On the other hand, the proposals in Hegselmann and Krause (2002) and Regan, Colyvan and Markovchick-Nicholls (2006), while partially avoiding the shortcomings of a fully subjective assignment, base their rationality on an assumption that holds only in special cases, namely, when it is in fact the case that I give a higher weight to those agents whose opinion is closer to my own.

The proposal in this chapter, instead, was to assign weights on the basis of the network structure of a group, in order to exploit or reduce its effect. Clearly in this case the results from convergence will be maximally manipulable, not by the agents in the model anymore, but rather by the modeler herself. While the exposition of the principles according to which one should want to minimize or maximize the effects of a network cannot find a place in this chapter, it is clear that a modeler should support its strategy with sound principles from psychology and decision theory in order to formulate an appropriate schema for assigning weights similar to those given above in tables 4.1 and 4.2.

4.6.2 Decriptive justification

Some interpretations take the Lehrer-Wagner model as a descriptive model of consensus formation[11]. In the second part of this section I will justify a network-depended assignment of weights from a descriptive point of view. A descriptive interpretation of the Lehrer-Wagner model claims that the model is a representation, however idealized, of how consensus is formed, rather than a deliberating method that groups should use in order to achieve agreement.

Agents often seek consensus by deliberating and putting their opinions together, and trying to come up with a "group opinion". But the process is not one of pure amalgamation, as some pooling algorithms may imply, rather it is, normally, an iterative process, in which agents come closer and closer to each other's opinions until a single one emerges as the consensual one. Moreover, it is reasonable to assume that agents will tend to assign (perhaps

[11]Indeed early versions of consensus models that use the properties of convergent Markov chains make reference to DeGroot (1974), who takes the model to be descriptive in character. In fact, if the Lehrer-Wagner model is taken as an "impossibility of disagreement" result, as Lehrer (1976) does, it is necessary to take the model to be descriptively accurate, and not only rational from the normative view point. This point cannot be developed further in the space of this chapter.

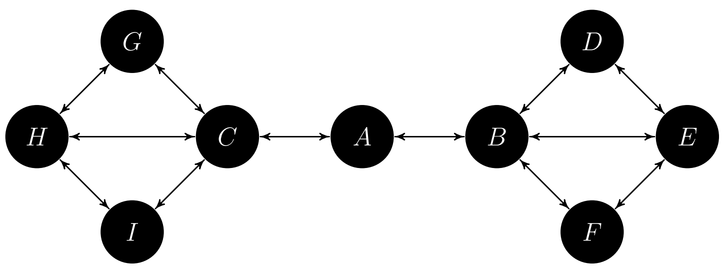
unconsciously) some *trust* or *confidence* to other agents. The higher the degree of trust I assign to agent i, the more i's opinion will influence my own opinion while moving towards the sought consensus.

How the mathematics of the Lehrer-Wagner model applies to the situation just depicted is fairly straightforward. The iterative nature of the deliberating process is captured by the subsequent rounds of the model towards convergence, and the measure of trust is represented by the weights that agents assign to each other. In the chosen model, where the opinions of some influence my own opinion to an extent greater than the opinion of others, at each round my opinion will be driven mostly by those agents to whom I have given higher weights — this is the idea that my opinion at each round will be affected mostly by those agents that I deem more trustworthy.

An interesting extension of the idea of describing consensus formation through the Lehrer-Wagner models is to apply the theory of networks to its system of weight assignment. Some networks will, in concrete cases, affect the formation of consensus, and it is in principle possible to study the effects of the network on the consensus that is produced.

A very straightforward example will clarify the latter point. Figure 4.3, represents a group in which two subgroups are present, groups (B, D, E, F) and (C, G, H, I), and the two are linked by a "mediator" (A).

Figure 4.3: Double ring network with a mediator



In this case it is clear the mediator plays a central role in the formation of a consensus, as it is the only one that shares a certain measure of trust or confidence with the two subgroups. Indeed, were A to be deleted from figure 4.3, together with the two edges that link it with B and C, then the two subgroups would not converge to a consensus, but to two independent opinions, one driven by the communication of trust in subgroup (B, D, E, F), and the other driven by communication of trust in subgroup (C, G, H, I).

As French (1956) saw, one can study the way a group converges to one consensus, or to more and distinct opinions, by looking at the distribution of weights in the matrix **W** characteristic of a specific group. That matrix, in turn, will likely be dependent on the network of the group, thus providing a descriptive justification for a network-dependent assignment of weights.

4.7 Conclusion

This chapter started off by considering a well known model for consensus formation, the Lehrer-Wagner model. The use of a formal model for the analysis of consensus formation, as well as the development of a normative theory of consensus, is a very fruitful approach for the study of consensus in social epistemology.

The model developed by Lehrer and Wagner, however, besides providing such a useful tool of investigation, also opens a number of problems, among which the problem of how a group of agents should assign weights to one another, or how to represent the distribution of weights (trust, confidence) that is reasonable to assume takes place in a group that is seeking consensus.

The paper does not claim that there is a one-fits-all solution for determining weights; for example, the methods suggested in Regan, Colyvan and Markovchick-Nicholls (2006) and Hegselmann and Krause (2002) seem to be fit for certain situations but unfit for others.

The proposal of this chapter was to take some ideas from French (1956) and, more recently, Golub and Jackson (2007), on how to derive weights from the network structure of a group. Those ideas were extended to a normative theory of weight assignment, where the analysis of a network is used in order to minimize or maximize its effects in the group, according to the modeler's judgment. The approach is particularly fruitful in those cases in which we wish to obtain a *rational* and *unbiased* consensus in networked groups. The examples provided in this chapter constitute only a small fraction of a potentially large number of possible applications and further lines of investigation.

Chapter 5

Consensus in economics PART 1

5.1 Disagreement and consensus in science

In the previous chapters I analyzed the topics of disagreement and consensus from the point of view of epistemology. In that context, disagreement and consensus were treated from a purely abstract point of view. It did not matter, for example, what the context of disagreement was, whether on scientific matters, on everyday issues (like in the split-the-bill case — cf. section 3.2), or other situations. All that was assumed in chapters 2-4, was that disagreement is on *factual* matters.

The topics of disagreement and consensus, however, are specific of the contexts they appear in. For example, disagreement in science is almost never truly *irreducible*. Real world epistemic agents are hardly ever *epistemic peers* in the sense in which epistemology defines them. Evidence is normally incomplete, and is not equally available to all agents. For those and probably several other reasons, the treatment of "real-world" disagreement should be context-specific. That is the case for disagreement and consensus in science, the subject of the second part of this thesis in chapters 5, 6, and 7.

The forthcoming sections will introduce the problem of consensus and disagreement in science, and specifically in the economic sciences. As it was the case for chapter 4 (cf. section 4.1), the meaning of 'consensus' should be taken more liberally than in the first two chapters. In this context, consensus, intended as convergence of views, can arise for a variety of reasons, and it would be too demanding to require a perfect convergence of individual beliefs in order to be able to speak of "scientific consensus".

The aim of this chapter is twofold. On one hand I want to motivate the search for consensus in economics. While concurring with much philosophical literature on the need for dissenting opinions, theories and methodologies in science, I intend to highlight here some reasons as to why a certain degree of consensus is desirable in economics.

The second goal is to confront the myth of *rational consensus* in economics. While admitting that much of the discourse in this chapter could be applied to other social sciences, the type of examples and the specific matters discussed mandate caution when applying the same arguments to other sciences. It is indeed arguable that the extent to which one can expect *rational consensus* to arise in a specific science will depend on the advancement of that science, its methodology, its object of investigation and so on. When that is the case, then, one cannot hope to provide one-size-fits-all solutions.

5.1.1 Consensus: rational causes and social causes

In this section I will outline some of the factors which may or may not justify a want for either consensus or disagreement in science. Should consensus be a goal for science? Or should we prefer to let disagreement thrive among scientists and theories alike? Much of the answers to those questions will depend on the stated goals of scientific communities, which in turn will depend on the preferences and desires of the scientists that are members of those communities. It seems possible, however, to give at least some normative indications of why certain goals should be favored over others, as it has been done repeatedly in the philosophy of science literature.

On the one hand, Objectivism and Progressivism[1] are inclined to see disagreement as a pathological situation in science; that is, as the indication of an imperfect state of the science itself. Those positions take from Kuhn the idea that disagreement is symptomatic of a “pre-paradigm” phase (the stage of immaturity of a science) (Kuhn 1970). In other words, according to objectivism and progressivism, disagreement is present because the science has not established a paradigm on which to build cumulative progress. Some progressivists or objectivists also make the additional assumption that successive levels of consensus constitute closer and closer approximations to the *best formulation* of a science[2].

On the other end, philosophers like Mill, Feyerabend, and Solomon defend the idea that disagreement in science is not only uneliminable, but all the more desirable. Any specific consensus which forms in science is likely

[1]By Progressivism it is meant here the attitude, typical of positivism, of seeing the course of science as a cumulative one, in which better and better theories are produced as the science evolves. Kuhn uses the more colorful term “Whig historian [of science]” to refer to a progressivist view of the development of science (see Kuhn 2000, 54 and 282).

[2]For an essential survey of some of the assumptions of Objectivism and Progressivism see Solomon (2001, chapter 1 and chapter 6)

to conceal at least some hidden and perhaps partly arbitrary assumptions of that science. To the contrary, disagreement will often highlight those assumptions, and be a motor for the progress and development of science. A typical claim made by supporters of this stance on the disagreement-or-consensus debate, is that a consensus is hardly ever based on purely rational grounds. Agreement, in other words, is not justified solely by the methodological principles of a science, but is often more likely to be the product of external factors, and is therefore an undesirable side effect of the *scientific community*, rather than the science itself. Agreement can arise due to social dynamics (e.g. political pressure, psychological conformity, etc.), or the more or less benevolent rhetorical work of scientists trying to convince other scientists. Whereas some consider the scientist's self interest to be a feature conducive of scientific progress (see Kitcher 1990), social dynamics and rhetorical arguments can also be detrimental to the goals of science (see Russell 2009).

It is arguable whether anyone, in the philosophy of science literature, holds either of the two extreme views presented above, yet it is clear that there is a stark contrast between advocates of the *rational causes* for consensus, and those who see much of scientific consensus as the product of *social* and/or *political causes* (Solomon 2001, 5). A pluralist attitude towards the problem would most likely admit that both positions are supported by cogent arguments, none of which, taken in isolation, is sufficient for rejecting the claims of the opponent. In a pluralist attitude, in the following sections I will limit the scope of my own arguments to providing some positive claims as to why consensus is desirable in science, and to show how such arguments are in principle compatible with the classical arguments provided by supporters of the "disagreement-first" stance in science.

In section 5.1.2 I outline some of the arguments traditionally presented in favor of disagreement, while in section 5.1.3 I provide three reasons for wanting at least a certain degree of consensus in science. Such list should not be taken as exhaustive of the possible reasons for wanting consensus in science; the selection is rather on a number of reasons that fit the economic sciences in particular. Furthermore, I will explain why the reasons I give for wanting consensus in science are not in conflict with those views, presented in section 5.1.2, which champion disagreement in its stead.

5.1.2 The value of disagreement

The desirability of disagreement in science is often attributed to the interactional value of dissensus and variety of views and opinions. A "healthy dose" of disagreement can only be desirable in science, if science is intended as an interactional discipline. Even the most cumulative disciplines such as mathematics undergo, from time to time, major makeovers. Progression is not always smooth and linear, it can take steps, jumps, and u-turns, and it is often the case that disagreement within a discipline, or subfield of it, is the motor allowing and promoting such progression.

> "Conceptual heterogeneity is necessary for the continued development of a science if science is construed as a selection process. Some commentators on science praise "pluralism." Let a thousand theories bloom." (Hull 1988, 521), see also Solomon (1994, 328)[3].

"Let a thousand theories bloom" can be taken as a political, pragmatic, or epistemic stance. From the political perspective, tolerating dissent can be seen as a *just* practice in science. Giving room for debate among people and programs who deviate from the predominant paradigm is a principle of democracy, where some degree of ability to make one's voice heard is granted to all. A pluralistic stance can be motivated by pragmatic reasons too: Variety in research topics is likely to please more scientists, or it may be practically impossible to reach a consensus that eliminates competing theories, and so on. The limits to the acceptance of dissent in the scientific field are often procedural ones: The decision on what are "acceptable scientific statements" is normally based on adherence to the dicta of the *accepted method* in that specific science[4].

But allowing disagreement in science, according to some, is not only *just*, or practically necessary, but also conducive of better science; this is an epistemic reason for pluralism in science. The epistemic role of disagreement has been the subject of chapter 2, although one thing is to outline, in abstract terms, the role that disagreement has in the process of holding or withdrawing one's belief; a different task is to explain, at the macro level, what disagreement has to do with the success and validity of entire theories and paradigms in science. At the macro level the epistemic value of disagreement is often attributed to the positive effects of collaboration and judgment aggregation. More on those two aspects will be said later on in this and the next chapter, but for now I will focus in the the rest of this section on some specific defenses of the goal of disagreement.

[3]The metaphor of the thousand flowers is attributed to Mao Zedong as the motto of the *Hundred Flowers Campaign*.

[4]The statement can be confirmed by a quick look at the stated requirements for acceptance of most scientific journals. Rejections from publication are often based on grounds of "form", rather than ones of "content".

According to Solomon, a widespread consensus[5] (no pun intended) among philosophers of science and historians of science is that, in the ongoing scientific debate "dissent is seen as the stage of competition, consensus the stage when there is a winner of the competition." (Solomon 1994, 99). Even though we might recognize that a plurality of views might benefit some secondary goals of scientific investigation (such as, for instance, pleasing the scientists' personal research interests), in the end it seems that what counts is that intellectual conflict be resolved, with the formulation of a theory, or the gain of some "piece of knowledge", regarding that part of reality that the science in question investigates.

Even the epistemic virtue of disagreement, in the light of the former considerations, can thus be seen as a transitory state. Disagreement, plurality of views, collaboration between scientists with different methodologies and different opinions, is only conducive of more qualitative science; but the end point is the final product, a theory or, more generally, a piece of knowledge, stripped of objections and contradictions. Not all scholars, including Solomon, agree with the aforementioned consensus. Solomon attributes to Longino the view that disagreement is not just a "means to finding "the best theory"", but also the "appropriate result of inquiry" in science (Solomon 1994).

An eminent proponent of such view is John Stuart Mill. In chapter 2 of *On Liberty*, Mill presents a number of arguments that are representative of the views that Longino and Solomon defend: Consensus is not the only, and perhaps not even the principal, end-point, or ultimate goal, of scientific investigation. The reasons for wanting disagreement in science are summarized by Mill as follows.

> First, if any opinion is compelled to silence, that opinion may, for aught we can certainly know, be true. To deny this is to assume our own infallibility.
>
> Secondly, though the silenced opinion be an error, it may, and very commonly does, contain a portion of truth; and since the general or prevailing opinion on any subject is rarely or never the whole truth, it is only by the collision of adverse opinions that the remainder of the truth has any chance of being supplied.
>
> Thirdly, even if the received opinion be not only true, but the whole truth, unless it is suffered to be, and actually is, vigorously and earnestly contested, it will, by most of those who receive it, be held in the manner of a prejudice, with little comprehension or

[5] Solomon calls it the "consensus on consensus" (Solomon 1994, 98)

> feeling of its rational grounds. And not only this, but, fourthly, the meaning of the doctrine itself will be in danger of being lost, or enfeebled, and deprived of its vital effect on the character and conduct: the dogma becoming a mere formal profession, inefficacious for good, but cumbering the ground, and preventing the growth of any real and heartfelt conviction, from reason or personal experience.
>
> (Mill 1859, 128)

It is not the goal of this chapter to analyze the reasons Mill adduces for wanting disagreement and rejecting consensus. Prima facie, they all seem to be valid reasons for claiming that disagreement is no less an end-point of science than consensus. The first reason appeals to the fact that even the strongest consensus might turn out to be false, hence the need for dissenters of even the strongest opinions. The second reason states that consensuses do not always contain "all the truth" on a certain subject, hence that disagreement is needed in oder to reveal those parts of the consensus which are not true. The third reason is that it is epistemically stronger, according to Mill, to possess an opinion and at the same time know why the deniers of that opinion are wrong, rather that to possess an opinion simpliciter. While in the domain of logic having a proof of a certain statement is enough to guarantee that its negation is false, scientific, social, political, and in general most opinions, are not known with the same strength with which logical statements are. The fourth reason is mostly a variation on the third one, and for the purposes of this chapter, equivalent.

Leaving Mill aside for the moment, I will go back to his theses on consensus in section 5.1.4, and show that they are compatible with at least a number or equally good reasons for wanting consensus. But first, in the next section, I will present the vindication for consensus.

5.1.3 The value of consensus

Any application of theories and models in science to concrete problems, at least for those sciences that have such direct application, requires the formulation of a coherent set of statement on which the practitioners are to act. In other words, reasoned action requires a coherent principle of action, a consensus. This is the pragmatic reason for wanting consensus in science.

It should be noted here that the term 'consensus' is used liberally in this context; a consensus can be the product of bargaining, voting, aggregation and so on, and takes the sense of "converging view", or "converging numerical value" (for example for probabilistic beliefs). To recall, in chapter 2 consensus was opposed to compromise, where in the latter the *group view* converges, even though the individuals may still retain their personal beliefs on the subject matter. Nevertheless, in this context it seems unreasonably demanding to require the consensus needed for action to retain the same meaning it had in chapter 2. All that is needed for reasoned action is the elimination of conflicting views, or contradictory statements, and the goal can be achieved by a diversity of means, such as voting or deliberation.

The second motive for wanting a consensus in science is an epistemic one, and has to do with some recent trends in social epistemology. Traditional Cartesian epistemology attributes beliefs, knowledge, rationality and other epistemic concepts to the individual alone; it is, in that sense, an individualistic epistemology (see Solomon 1994; Goldman 2009a,b). Recent discussions in social epistemology however — for an example, see List and Pettit (2011) — suggest at least a partial redefinition of the concept of epistemic agent, such that it includes certain types of groups and collectivities which function and should be considered agents to all intents and purposes. If talk about group agents seems strange in the context of science, consider how often scientific theories, models, and views are attributed to research centers, universities, or other institutions. It is not uncommon linguistic usage to claim that 'such and such research center has discovered a new protein [...]', or that 'such and such University has cut the budget for basic research by [...] percent'.

The idea of taking groups as epistemic agents, however, is not only meant to accommodate a linguistic use; some advocate that it is also rationally motivated, and can be pragmatically advantageous. Pettit and List argue that institutions trying to perpetuate themselves through time are by all means better seen as agents, rather than simple aggregates of individuals[6]. If such agents are to possess coherent sets of beliefs, aggregation is necessary in order to form a consensus in the set of (possibly) diverse individual opinions that make up the group. The need for consensus, therefore, becomes equivalent to a rational person's epistemic requirement of having a coherent set of beliefs.

Finally, a third reason for wanting consensus in science is motivated by a reflection on Kuhn's work. Kuhn divides the timeline of a science and its development into periods of "normality" (Kuhn 1970, Chapter III), and periods of "crisis" and consequent change in paradigms (Kuhn 1970, Chapters X-XII). One of the characteristics of the period of normality is the fact that the majority of the scientific community works under a given paradigm. What this means is that there is relatively little disagreement about the validity of the assumptions of that specific paradigm and, as a

[6]For an introduction to the problem see Pettit (2004); List and Pettit (2005, 2011).

consequence, there are relatively few resources dedicated to the pursuit of other paradigms, even if there can still be plenty of disagreement *within* a given paradigm. To give an example, the method of mathematical modeling and econometric testing is, according to many, a paradigm in contemporary economics, although this does not mean that there is no debate as to which models are acceptable, which tests are valid, and so on.

In light of Kuhn's subdivisions of sciences in periods of normality and periods of crisis, one must notice that the consensus that exists when a paradigm is left unchallenged allows for a better allocation of resources. Assuming that a given paradigm is the best advancement of a science, up to a specific time in history, then selective allocation of resources in periods of normality prevents energies from going to support less efficient paradigms. To use the same example as above, the current paradigm in economics allows fewer resources to go into the development of historical approaches to economics which, according to the paradigm, would deliver less benefit to the science itself. This does not mean that the historical paradigm should be ignored. Historians of economics develop their paradigm in parallel to the mainstream one, only with fewer resources.

Probably the major objection to this Khunian argument for consensus, comes from the observation that differential allocation of resources can bring the science into a lock-in or path-dependence, where deviation from the paradigm becomes harder and harder, because it implies a vast reallocation of resources. A lock-in is not a negative phenomenon in itself, but it is when the path (the paradigm) in question is an inefficient one, albeit it being still consented upon by the community of researchers. Arguably there was widespread consensus around the geocentric model of the universe, and one could argue that resistance to the new evidence brought by the telescope was at least in part due to such lock-in effect.

On the other hand, equal allocation of resources is not always the best alternative, and is in fact the least desirable when a clear "by-far best paradigm" can be identified. For example, it would be wasteful to share a government's resources equally among, say, western (or 'evidence-based') medicine, homeopathy and acupuncture, if the latter two could be shown to contribute to the population's health to a much lesser extent than western medicine. It is also for this third reason then, that a certain degree of conformity (consensus) in science is desirable.

5.1.4 Compatibility of disagreement with consensus

While it is perfectly fine to claim that both the advocates of disagreement and those of consensus have good reasons for defending their claims, if one wants to be liberal on the subject, one should at least show that the claims on one side are compatible with those made on the other side. In this section I argue that the two positions outlined in sections 5.1.2 and 5.1.3 are indeed compatible. The claim that science should pursue *only* variety of opinions would be just as narrow-minded as the claim that science should pursue the elimination of all disagreement, and strive for consensus alone.

To see how the reasons provided in section 5.1.3 are compatible with Mill's theses one should start by noticing that Mill's theses are concerned with the epistemology of science in the first place. If by 'science' we mean solely the formulation of theories and models of the world, and of methodologies for explaining and possibly predicting the world, then it is clear that disagreement might have more epistemic benefit than consensus. If science were taken to be abstract investigation only, the first reason (section 5.1.3) would not apply, the second reason would be mostly a perhaps pedantic subtlety, and only the third reason would have some bearing on the disagreement-consensus debate. One needs to balance the allocation of resources to the most promising line of investigation, with the need of keeping research in an open-minded framework, as Mill suggests. But, as Mill rightly argues, in an epistemic context consensus is never the end-point of science.

Nevertheless, economics is not a purely abstract or theoretical science. The idea of this chapter and the following chapters is to take economics as a science concerned primarily with economies (that is economic phenomena, as they appear in different contexts), and economic policies. Clearly the expectations from economics become stronger, when one views the science as being concerned with economies and policy making, but it is hard to deny that economics gains its reputation and value from the very fact that it is, by many, deemed capable of dealing with concrete economic problems. Not just a science that only *explains* economic phenomena, but one that *predicts* and *manipulates* the economic environment as well. This type of science is concerned with what Reiss (2008) calls 'social capacities'.[7].

In that sense, economics is a "toolbox science", where investigation and manipulation go hand in hand. For policy making, however, the epistemological problems Mill mentions should be put in the background. Lack of clear

[7]In Reiss (2008) knowledge about *social capacities* is "knowledge about causal relationships that are *stable across suitable ranges of circumstances*". Of course, a number of renowned methodologists think that economics can *not* achieve knowledge of social capacities, nor of course the ability to predict and manipulate economic reality (see Reiss 2008, chapter 9 - Social Capacities). In this and the other chapters I side with Reiss's optimism about the possibility of achieving social capacities in economics, and chapter 6 will suggest part of a methodology for achieving that goal.

and explicit guidelines (which have to be agreed on), verifiable predictions (on which there need to be some consensus), and so on do not serve the interest of an applied science, even a pluralistic one. Different methodologies and theories can be admitted at the theoretical level, but contrasts and disagreements need to be eliminated if the goal is to evaluate a piece of economic knowledge against the results of policy making.

In conclusion to this section, one can see that the reasons for wanting consensus, outlined in section 5.1.3, apply mostly to economics as an applied science, whereas Mill's theses apply mostly to economics as a theoretical enterprise, and the fact that the distinction between the two is often blurred does not help the debate, at least in the field of economic applications. The reason is that any science which aims at giving policy advice (or technological advice) needs to evaluate its methodology against the results from policy making. In order to be open to evaluation, however, a methodology needs to be at least in principle identifiable, and identifiability quite clearly requires at least a certain degree of consensus.

The debate whether it is more important to have consensus or disagreement in science, and in particular in economics, is probably not going to be easily settled. In the preceding sections, however, I hope to have been able to at least show that, on one hand, consensus is *one of* the desiderata in economics, and secondly, that such a desideratum is not in conflict with Mill's (and his successors') thesis on the importance of disagreement in science. In the foregoing sections I will take up the problem of *how* consensus is formed in economic sciences.

5.2 The normative question

5.2.1 Stating the normative question

Accepted that consensus is one of the desiderata in economics, the second task of this chapter is to see what *ought to* drive the formation of such consensus. Here one needs to distinguish between a factual question about consensus formation in economics, and a normative one. The former asks how a consensus is, as a matter of fact, driven by this or that factor, political or methodological, rhetorical or evidential, and so on. The latter question, on the other hand, has as its ultimate goal a *normative theory of consensus formation in economics*. Such theory would try to delineate what criteria should be used, in order to evaluate the truth or falsity of a certain economic matter (a single economic statement, a group of statements, or a whole theory), on which economic consensus has formed or ought to form. The normative question that will guide the rest of this and the next chapters can then be formulated as follows:

What criteria should regulate the acceptability or non acceptability of a specific economic consensus as a piece of *economic knowledge*?

Some remarks are in place here. Firstly, a criterion should identify economic statements that can be turned into policy advice, at least to a certain extent. Clearly some statements in economics, like in science in general, can be underdetermined with respect to the world, because they are too general, or vague. Such statements can still be considered economic knowledge, and even be economically relevant, but may not be suitable for any policy recommendation. If that is the case, evaluating the acceptability of a given consensus may be pointless, since, in the spirit of Mill, we could argue that any purely theoretical stand benefits more from the presence of disagreement than the existence of a consensus. An exception may be made insofar as a specific theoretical consensus might positively *influence* a direction of research. Even when that is the case, any consensus criterion should still keep as its goals the formation of consensus on social capacities.

Secondly, the notion of knowledge used in the formulation above should be taken as a fallibilist notion. In other words economic statements can always turn out to be false, but in so far as we can justify them to the best of our abilities, they can justifiably become part of our stock of "economic knowledge", on which we act and promote policies. The epistemological standards for 'knowledge' are often stronger: Knowledge is *justified true belief*, where the concept of 'justification' varies across different epistemological stances. However, it does not seem proper of a pragmatic approach, as is the one undertaken in this chapter, to suggest a methodology that can help answer economic questions once and for all. All that the normative question I am considering asks is for a "best answer". It will become clear in the following sections what is meant by that; for now, suffice it to say that a method should be allowed to be fallible, at least when what one has in mind is "applied theory", as it is the case here.

Thirdly, it is clear that assuming a realist perspective, as the question above does, is not an innocent assumption. Consensus in science, according to a realist, would ideally converge towards the *truth* (with or without capital 'T' (see Mäki 1995; McCloskey 1995)), or the *right* theories. An instrumentalist, on the other hand, may defend the thesis that consensus must ideally converge towards those theories or models that are most successful, or which produce the best results. The problem is open: Any philosophical account able to uniquely identify a certain goal, will want a consensus that forms in the direction of that goal. The space of this chapter does not allow for further investigation into this matter. And, for the moment being, I will assume a realist stance.

A final issue with this formulation of the normative question is the fact that, it should be noted, the question of how consensus should form, from a normative point of view, in economics is parallel in important respects to the question of how we should gain knowledge in economics. To motivate this claim, suppose that a consensus forms among economists, around the problem of the relation between the quantity of money and the price level. The example is loosely taken from Reiss (2008, 168). Imagine it is possible to agree on the truth of a specific correlation, complete with conditions under which the correlation breaks down and so on, between the two variables, such that, so claims the consensus, by manipulating the former we can control the latter. In the words of Reiss (2008) we have come to agree on the truth of a “social capacity”.

The truth of this social capacity is, for now, only agreed on; the question remains of whether such consensus is correct, whether the correlation is in fact found to hold in the economy. In fact the problem is equivalent to establishing whether such newly formed consensus is knowledge or not. In the first place, is it a true belief, the one shared among the economists in this imaginary example of consensus? This question pertains to the problem of verification: whether we can verify, or test, the statement of the correlation. In the second place, is the statement justified, and *how* is it justified? The latter question pertains to the problem of the methodology, which will be the focus of the following sections.

There are several ways, in the sciences and in economics, to justify a given consensus as a rational one or not, at least as many ways as there are methodologies.

> [...] there are a plurality of methods for gathering evidence: indeed observation with the naked senses; instrument enhanced scientific observation; statistical methods (such as data-reduction and analysis techniques, index numbers, regression, ANOVA etc.); mathematical modeling; computer based methods such as simulations; laboratory experiments, thought experiments and the analysis of natural experiments; testimony and expert judgment.
>
> Reiss (2008, 3)

While Reiss, above, is speaking about methods for gathering *evidence*, the discourse is equivalent to the problem of consensus and knowledge. There are several ways to justify a consensus as rational and several ways to gather knowledge of economic objects and phenomena.

In the following section, I will focus on that part of this methodology that is preferred by at least a large part of the contemporary world, the scientific and empirical method, which comprises a number of subfields and is primarily focused on modeling and experiment, hypothesizing and testing.

5.3 Rational consensus formation in science

Solomon talks about "rational causes for consensus" when referring to the position according to which "overwhelming evidence" is the main motor for consensus in science (Solomon 1994, 5). A certain version of *scientism* holds that science, defined in broad strokes as the strict adherence to the method of hypothesis formulation and testing, gains authority by the *rules of evidence*, evidence that is present in the natural and social world. What exactly counts as evidence cannot find place for discussion here; in general, evidence is recognized as one of the pillars of the Scientific Method, together with the use of mathematical (or at any rate formal) tools for the formulation of hypotheses.

Salmon writes "[...] the use of scientific methods is believed by many people to be the best way to obtain genuine (though not infallible) knowledge about the world. Despite this admiration for science and its methods, it is not easy to say exactly what science or the scientific method is [...] let us assume that science involves such features as laws, testability, prediction and explanation [...]" (Salmon 1999, 405). Notwithstanding the difficulties with the formulation of a standard definition of scientific method — in fact, there is no such standard definition — I will loosely define here the scientific method as that method of investigating natural or social phenomena by formulation of hypotheses, derivation of predictions, and verification of the occurrence of those predictions. This is known as the *hypothetico-deductive model of science*[8].

While empirical confirmation of a theoretical hypothesis is not possible, as it is tantamount to the logical fallacy of affirming the consequent, on the other hand a negative tests can *falsify* it (cf. Karl Popper's *falsificationism*). Consequently, hypotheses that have resisted testing over and over again are

[8]The model was first formulated by William Whewell in the 19th century, although its core principles were established much earlier in the works of Galileo Galilei and Isaac Newton.

preliminarily accepted as good ones, until new evidence comes to reject them. In the philosophical literature this process is called *confirmation* of scientific hypotheses. The theory of scientific confirmation has been widely extended since its early formulations in the 1950s, in order to accommodate logical problems with the initial formulations, and also to capture probabilistic phenomena; a thorough review of the literature is in Earman and Salmon (1999).

Despite the many yet-to-be-settled philosophical debates, the hypothetico-deductive model and the theory of scientific confirmation have had a strong appeal in economics. In fact, their principles are substantially similar to the ones proposed by Friedman for the methodology of the economic sciences, in his famous and much-influential article *The Methodology of Positive Economics* (Friedman 1953). There, Friedman claims that the core of economic methodology is the formulation of hypotheses, the derivation of informative predictions, and the testing of the former by verification of the occurrence of the latter. "Only factual evidence can show whether [a theory] is "right" or "wrong" or, better, tentatively "accepted" as valid or "rejected." (Friedman 1953, 8)

The hypothetico-deductive model, as well as Friedman's methodology of positive economics, have been extensively criticized, especially as descriptive tools. It must be conceded that the "rational consensus", or "scientific consensus", approach is too simple to account for concrete scientific practice. It is however more difficult to assess the approach, when taken as a normative one. To many, scientific criteria are those by which consensus should be evaluated, and in that sense they apply to the context of justification, not the context of discovery.

It would be out of place, for the purposes of this thesis, to aim at giving a definitive answer to the problem of whether the hypothetico-deductive model, or some variant of it, are adequate for a normative assessment of scientific theories. For that reason, in the following sections I will only provide some examples, where it can be argued that the ideal of the scientific method works at its best (section 5.4), in order to contrast them with others, taken from the field of economics, where instead the ideal of the scientific method runs into an impasse (section 5.5).

5.4 An example: celestial navigation

The example I discuss in this section comes from physics. While one may argue that any example from physics bears little relevance to phenomena in the socio-economic world[9], the goal here is to present an example that can clearly and easily fit the view of rational consensus presented in the previous section. While there are some (relatively extreme) positions, which would argue that also the example presented below would not fit the rational consensus view, I will deal with that type of criticism in section 5.4.1.

The example chosen here is the theory of Celestial Navigation. "Celestial navigation is a technique for determining one's geographic position by the observation of identified stars, identified planets, the Sun, and the Moon. Its basic principles are a combination of rudimentary astronomical knowledge and spherical trigonometry." (van Allen 2004, 1418) The technique relies heavily on the model of celestial mechanics, in particular the calculation of the position of 57 so called *navigational stars*.

Figure 5.1: A schematic representation of the theory of celestial navigation.

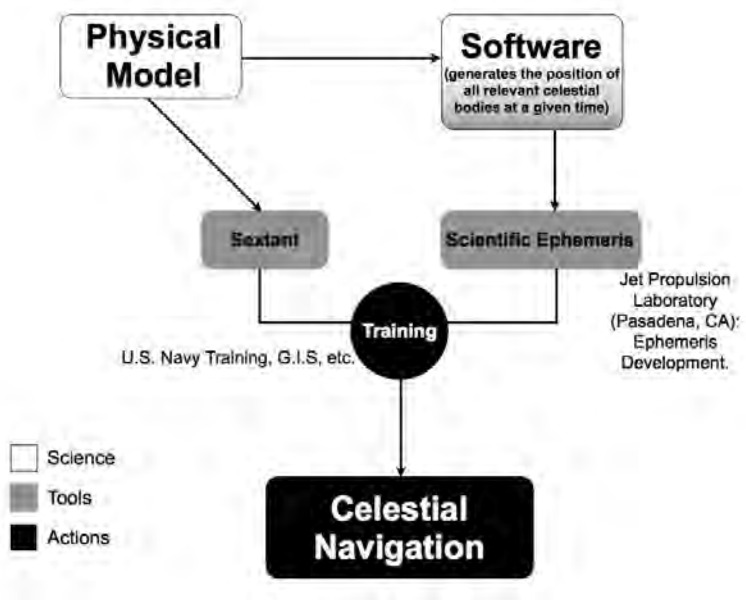


Figure 5.1 illustrates the main theoretical and practical components of celestial navigation. At the top level is the physical model of celestial

[9]That, however, has not been the dominant view in much of 20^{th} century economics (Lo 2004).

mechanics, plus the development of software for handling the data. The physical model is a piece of the underlying science, whereas the software is in part a product of science and in part of engineering, although the limits between the two may be blurred. Scientific ephemerides and the sextant are essentially tools for the application of the underlying science (e.g. trigonometry). Finally, at the bottom, is the practice of celestial navigation itself; it requires training in the use of the tools used to apply the underlying science.

Celestial navigation may not be considered a science in the proper sense; it is in fact a technique made possible by the application of different sciences. If one abstracts from the strictly practical issues of the technique, however, it is possible to isolate the scientific principles and theories that make navigation possible. On those, a large degree of consensus has formed, allegedly on mostly scientific and rational grounds. Celestial navigation cannot be considered explanatory of any specific natural phenomena; nonetheless, it focuses on prediction, which is often one of the principal focuses of the scientific method (cf. Friedman's instrumentalist theses in Friedman (1953)).

The schematic representation in figure 5.1 should be sufficient to give an idea of how the elements of the model of celestial navigation are combined. The schema applies to celestial navigation at sea, and, with the proper adjustments (omitted here), to navigation in outer space.

Celestial navigation allows ship crews to calculate their position on the surface of the earth. Models of celestial mechanics are used in order to compute ephemerides, tables with long lists of the positions of the relevant stars at any given time. Using a sextant and a chronometer navigators can then determine their latitude and longitude on the earth's surface. Learning celestial navigation requires precision and long training, or sophisticated computer programs in the case of automated vehicles (see Sigel and Wettergreen 2007); in that sense, it is not a "simple" science, nonetheless, it can be considered, for the purposes of navigation, an exact science.

Stars can be computed as points moving along invariant trajectories with known speed; disturbing factors are mostly known, errors can be calculated, and increases in the sophistication of the tools allow for greater navigational accuracy. Moreover, the physical system itself, the relative motion of the starts in the sky, is simple and stable[10].

[10]Stability must be considered a relative term in this context. From a reference frame of billions of years the universe may not be considered a stable system. It is arguable that astronomy could not compute the development of the universe in millions or billions of years, since unresolved issues such as the presence and quantity of dark matter and dark energy are a major determinant on the possible scenario of evolution of the universe. Despite all that, the time frames considered in the example provided in this section are of

Salmon's characterization of the scientific method (cf. section 5.3) mentioned four factors that can all be found in the theory of celestial navigation, that is, laws, testability, prediction and explanation. Celestial navigation involves *laws*, the laws of mechanics, which in the time frame of interest are exact up to approximation. In other words, small errors in calculation produce small deviations between the predicted values and the actual values. The model provides simple *predictions* (e.g. the position of a vessel on a map), which can be as accurate as desired, conditional on the available technology. Those predictions are *testable*, they concern observable objects — stars are points, the horizon is an imaginary line — which in turn can be used to verify the predictions of the theory. Moreover, the theory underlying celestial navigation is *explanatory*: With mathematical principles (spherical trigonometry and elementary arithmetic), and very general and tested assumptions on the relative positions of the objects in question (position of stars to earth, etc.), we can explain why such and such observations on the celestial sphere indicate my position on the surface of the earth.

In short, the theory of celestial navigation seems to fit well to the ideal of the Scientific Method. In other words, if asked about the theoretical justification of the theory, the *rational* answer would be to provide the inquirer with the principles and computations involved with the theory, and present the evidence available for *confirmation* of the theory. Possibly, we would want to compare the theory, or specific bits of it, with alternative ones, and therewith settle questions such as "what are the advantages of relying on an atomic clock rather than on a pocket chronometer?", "how safely can we navigate a vessel in the dangerous Great Barrier Reef?", and so on.

5.4.1 Socio-historical influences on celestial navigation

The question in this section is whether evidence and computation are the *only* principles at work, when we *agree* on the truth of some statement concerning the theory of celestial navigation? Even defending the weaker point that the hypothetico-deductive model is only normatively valid, in the case of celestial navigation, comes under attack from the ranks of the Social Studies of Science.

The concept of horizon and the method of computation of one's distance from the horizon, had to undergo several mathematical refinements through

such a small magnitude that the system of celestial mechanics can be considered extremely stable for the purposes of celestial navigation.

the centuries. Moreover, calculation of the latitude was for a long time a major unresolved issue.

> To determine the longitude Λ of the observer, a knowledge of the simultaneous longitudes of the respective substellar points is essential. In other words, it is essential to know the *absolute time* (for example, GMT, the mean solar time at Greenwich). The historical challenge of developing an accurate chronometer for maritime use rests on this simple fact.
>
> (van Allen 2004, 1421)[11]

According to the sociologist of science, then, the difficulty is to reconcile the claim that only the scientific method, in its ideal form, should be at work in the evaluation of the theory of celestial navigation, with the observations that the formulation of the concepts involved in the theory have been, and in all likelihood will be in the future, subject to historical development, not at all immune from sociological and political factors.

The answer is that there is no need for reconciliation. It is undeniable, save for denying the existence of historical documents, that the concepts involved in the theory evolved through time and where subject to political, social, and contingent events. The same holds for the mathematical theory involved in the computation (spherical trigonometry) and also, in modern times, for the development of the technologies involved in the observations and computations. However, this is not to say that those factors are at work nowadays, in the precise moment in which a navigator out at sea is trying to locate her vessel on a map.

To qualify the latter statement, note that such navigator, had the aforementioned historical and social events taken a different course, would possibly now be in a different position with respect to her capacity of navigating at sea with a certain degree of precision. For instance, had the British Parliament never promulgated the *Longitude Act* in 1707, perhaps the development of the chronometer (necessary for determining the longitude) would have been delayed greatly. It is thus possible to envisage a chain of events, for which we would be now in a very different position with respect to our capacity of navigating at sea, due to that possible "missed" act in 1707. Nevertheless, the fact that we can now navigate at sea, or in space missions, is only "counterfactually dependent", on those sociological and historical

[11] For the historical account of the "challenge" that Van Allen refers to, see Sobel (1995).

factors, just as much as my last statement was possible because David Lewis provided an analysis of counterfactual reasoning about 40 years ago.

To make the previous point stronger, I will draw an analogy from the theory of knowledge regarding the Kantian definition of the *a priori*. Kant defines *a priori* knowledge as that type of "[...] cognition independent of all experience and even of all impressions of the senses." (Kant 1787, 136 - B2) If the statement were given too narrow an interpretation, then Kant would be saying that *a priori* knowledge relies entirely on innate ideas, present in the mind even before the person has ever seen the light. With some intermediate steps, that would lead to the paradox of someone who has been locked up and fed in a sealed dark chamber since their birth, but who could still reason of mathematics or logic.

Without pretension of historical accuracy, it must however be reckoned that clearly the meaning of Kant's statement should not be taken to imply that paradox. A possible interpretation of the apparent absurdity is that *a priori* knowledge is independent of all experience, in the sense that it cannot be shown false by appeal to experience. In other words, we need experience in order to acquire a language, the capacity to reason, etc., but our presently exercised ability to reason a priori is not dependent, for its validity, on experience and sensations.

The analogy applies to the case of celestial navigation — sociological, historical, and political factors have shaped the development and formulation of the theory through the centuries, but we cannot appeal to any of those factors today as what justifies the fact that we rationally rely on that theory, and not on others, for navigation.

The discourse so far may have given the impression that there is a stark separation between theories (such as celestial navigation) which can be justified by reason, and other theories whose justification, or better, acceptance in the relevant community, has to be found in socio-historical factors. It is certainly imaginable a scenario in which the historical developments of human society had deviated greatly from the course they in fact took, or in which the social constructs would be extremely different from the ones in fact present today. In those scenarios, we could thereby certainly imagine that the theory of celestial navigation would nowadays not be justified in the eyes of the scientific community, and, for example, that some other theory would be; perhaps one which relies on entirely different principles, and absolutely incompatible with ours.

It does not seem to be the task of philosophy of science to account for such scenarios. Possible alternative scenarios, where the theory of celestial mechanics would not have come to be consensually believed, are so distant from the actual one that it is reasonable to consider the theory justified by *scientific reasons* alone. Therefore, the consensus formed around the theory derives from what Solomon (1994) refers to as "rational causes for consensus" (cf. section 5.3, above).

5.4.2 More examples

Is the example of celestial navigation representative of all scientific phenomena and the scientific theories around them? The theory of celestial navigation relies on one of the oldest mathematical theories (trigonometry), on a branch of physics that has enjoyed hundreds of years of astronomical observations, as well as a physical system that can be considered rock-stable for the time frame of interest. Because of those and possibly other reasons, the example cannot be considered representative; in fact, it is quite an atypical situation in science to have such favorable conditions for the investigation of a phenomenon.

However, the example is not an outlier either. Metrology, the science of measurement, has undergone important historical challenges and is far from uncontroversial, but its hypotheses and the physical systems it works with can be compared to the case of celestial navigation. Similar considerations can be made for stoichiometry, the science that deals with quantitative relationships between reactants and products in a chemical reaction, and even for some branches of medical diagnostics: those that develop software for aiding physicians diagnose patients on the basis of statistical relations among symptoms, physical preconditions, past medical history, etc.

In general it must be reckoned that the example of celestial mechanics, while it is not *typical* in science, cannot be regarded as an outlier either. Science is filled with cases of phenomena that can be studied by the methods of hypothesizing and testing with high levels of precision and little influence from social, historical and contingent factors. So, what is the case for economic theory? Repeatedly throughout the history of economic methodology economics has been ranked among the so-called "inexact sciences", even though it would be more correct to speak of exact and inexact theories, as inexactness and exactness can, and often do, cohabit the same particular science (see Helmer and Rescher 1959).

But what the meaning of inexact is, is not at all clear. Hausman (1992), for example, lists a number of reasons for attributing inexactness to a science, which do not discriminate between the economic sciences and others, for example physics, which would instead be considered exact. More on the latter point will be said at the end of this chapter in section 5.6.1. The point of the following sections, however, is to try to discriminate between economic theories and other theories that belong to that class identified above as the "exact class". Is there any difference between a typical economic theory (or model), and the theory of celestial navigation, or a certain chemical theory of stoichiometry?

The next section will highlight some of the differences between economic theories and models, and those scientific theories and models that can legitimately go under the "exact class". I will highlight some of the limitations imposed to the method of evidence and testing in economics. The list of arguments that follows should not be taken as exhaustive, nor should its division into three sections be taken as reflecting some objective division of the problem space, but rather as serving the goal of simplicity in exposition.

5.5 Economic methodology under scrutiny

The problem addressed in this section is how consensus should form, be evaluated, and eventually accepted in the field of economics. In other words, should consensus form around the principles of scientific modeling and evidence, or some other principle or set of principles? It was assumed, in sections 5.2.1 and 5.3, that the rationalist's answer to the problem on consensus is acceptable from a normative viewpoint if also metaphysical realism and a correspondence theory of truth are assumed.

One should reckon, however, that pragmatic considerations would require that the methods of modeling and testing also be available and successfully applicable to the field in question. That, I argued, is the case in at least some fields of physics, chemistry, medicine and so on. The success of modeling and testing that natural sciences have seen in the past few centuries has arguably pushed the so called "social sciences" into acquiring a similar methodology. To restate the point in Salmon (1999) "Clearly, the label "scientific" carries a certain cachet in our society. Scientists are respected, and their work is heavily funded by public and private agencies. More importantly, the use of scientific methods is believed by many people to be the best way to obtain genuine (though not infallible) knowledge about the world." (Salmon 1999, 405).

Despite Salmon's observation, and what I have assumed about the validity and opportunity of the scientific method, it is a much debated issue whether the scientific method (or, as Salmon, uses the term, 'methods'), is successfully applicable to all — or, if a part, which part — of economic problems. In the following four sections I will give some examples of many of the issues involved in using a methodology typical of the natural sciences, to problems in the economic world. The analysis will focus on a number of mostly pragmatic issues, rather than theoretical problems.

One important theoretical issue, for example, is the nature of human will in relation to economic forecasting. The problem dates back to John Stuart Mill; the argument against the possibility of performing mathematical calculations of human factors is at times used against the possibility of using identical, or even similar, methodologies in the natural and the social sciences. This and similar other problems will not be discussed here. Instead, I will focus on problems related to the practices of economists as they investigate and try to understand economic phenomena at work in the economic system.

5.5.1 Instability and unaccounted-for factors

The problem of the stability of a system was mentioned in section 5.4; the model of celestial mechanics, it was said, can be considered stable for time frames of some few thousands of years or even longer, allowing for exact calculations up to approximation. As a working definition, a system is stable within a timeframe T, if the same laws operate at every time t of T. For instance, imagine a biological system, such as "Rabbit-World", and a population model for rabbits. The system the model represents can be considered stable if the same biological laws (or trends) operate within the timeframe T considered by the model. If rabbits in Rabbit-World, at some point within T, undergo a behavioural change that speeds up their reproductive habits, the system's stability will have been broken.

Problems for a model can also come from factors occurring outside the predictive powers of the model. As a working definition, an unaccounted-for factor is a factor in the system that the model or theory has not included in its formal structure. It should be noted that a break in the stability of a system, that is, a change in the underlying forces (laws of nature), may also be considered an unaccounted-for occurrence, but for pragmatic reasons it seems legitimate to limit the domain of stability to forces and laws, and confine the domain of unaccounted-for factors to changes in the initial conditions, external influences, etc. For example, imagine that in the area initially postulated by the population model for rabbits, a certain number of rabbits were introduced from without. The model did not account for that type of population increase; moreover, it would not be convenient to add such a variable either, thus complicating the mathematics of the model, unless the introduction of individuals from without the population were constant, or could be at least given a certain probability.

A "normal" population model would account for increases in rabbit population that are dependent on reproductive rates, mortality rates, seasonal changes, etc. But unless there were constant and predictable introductions of rabbits from without into the environment, it would not be standard to introduce such phenomenon as a variable in the population model for rabbits.

In the example of celestial navigation, the system in question can be considered stable. The apparent motions of stars in the celestial sphere are subject to constant laws — similarly for the bodies in the solar system. Moreover, unaccounted-for factors can often be neglected; even when they cannot be ignored, their effects are visible in small increments, not as dramatic changes in the laws of the system. In a cosmological model, examples of such factors could be the formation of a black hole in a certain region of space, or on a smaller scale, a previously unknown asteroid perturbing a planet's orbit. But such events often show their effects in much longer time-frames than the ones of interest for an economic system.

The problem of stability and unaccounted-for factors is a major cause of imprecision, if not randomness, in several areas of economics. At least two types of phenomena are identifiable in the economic world. On the one hand, some events are known to happen but it is extremely hard to predict them with meaningful accuracy. These types of events are, for instance, runs to banks. We know that runs to banks occur: they can be triggered by news, irrational behavior, or other causes. Runs to banks are disruptive events, they show their effects in short periods, and are highly unpredictable, by scientific methods at least. The sudden appearance and rapid development of such factors in the economy afford little or no time for fixtures in the models, thus leaving inflexible mathematical methods unable to give reliable predictions.

On the other hand, some of the factors that go unaccounted for by mathematical or statistical models are simply unknown. Examples of that latter type are phenomena like the Internet and its effects on communication, or the effects of *containerization* (a mode of freight transport standardized in the 1970s) on global trade. Such selection is merely illustrative, but two prominent nontechnical works have been written on the subject: Malcom Gladwell's *Outliers*, and Nassim Nicholas Taleb's *The Black Swan* (see Taleb 2007; Gladwell 2008). The important aspect of these phenomena, which occur in the social as well as in the natural world, is that they are very frequent in economics.

In the economic world, changes in the underlying laws (for example the speed at which information spreads), and occurrence of an unpredictable event (a run to a bank) are a constant threat for a forecaster. When that is the case, the process of modeling, theorizing, and testing can be painfully slow to account for the developments of the environment. Moreover, a model's exactness and rigidity is more often an impediment rather than a virtue, when the occurrence of unforeseen events requires a change in the initial assumptions and conditions. Such changes are often very costly, in terms of changes in the specific methodology, the reformulation of models, adjustments in how the data is collected and so on. Consideration on the costs can cause institutional failures if adjusting the tools in order to have accurate predictions is taken to be more costly than accepting the price of a wrong prediction.

5.5.2 Openness

Openness to exploitation is the second phenomenon surveyed in this section. The fact that economic systems can be exploited is often attributed to irrationality. For example, people holding probabilistically inconsistent sets of beliefs, according to Bayesian theory, are subject to so called Dutch-Bookies, sets of bets that will seem rational to them, but which will make them lose money consistently. Irrationality, nonetheless, is not the only source of openness in economics. Openness to exploitation, that is, to profit making, can come from flaws in the economic system itself, or from the legislative structure that is superimposed on the economic system.

Charles Ponzi started off the infamous *Ponzi scheme* by taking advantage of a flaw in the system regulating international reply coupons for postage stamps (Dunn 2004). Initially, Ponzi was not acting in breach of the law, but simply taking advantage of different international rates in order to buy international reply coupons in a country at a certain price, and exchange them in another country for stamps of a value higher than the price paid for the coupons. Similarly, recent reports in the New York Times claim that U.S. banks are shifting speculative activities from proprietary trading, limited by the so called Volker Rule currently implemented in the United States, to client-related business. In other words, banks would be allowed to engage in risky speculations, intendedly forbidden by the legislation, under a different name (Schwartz and Dash 2010). In this second case, more than a flaw in the system itself, the possibility of exploitation seems to derive from a gap in the legislative system, if banks are indeed allowed to operate as before by simply changing the name of their activities.

On the one hand, legislation seems to open the system to loopholes and gaps which will be exploited by those who are smart enough to figure out a way to do so. If that is the case, it seems that economic liberalism had an argument against regulatory practices in economics. That is only one side of the coin though. Indeed, regulatory practices are meant to protect individuals from the irrationality trap (mentioned at the beginning of this chapter), which also opens economic systems to exploitation.

How the exploitation of economic systems affects predictions in economics, can be explained with another example. Before the 1990s only big savers could access interest-bearing financial instruments (treasury bonds or corporate bonds), but by the late 1990s most small savers could open interest-bearing checking and savings accounts. Clearly, the precise configuration of the former financial instruments will have to guarantee an equilibrium between risk and yield of those financial instruments. But banks can try to surpass their competitors by exploiting the loopholes of regulations, in order to offer their financial instruments to a wider range of customers. Once that happens, the equilibrium that was calculated at the beginning will no longer be the real equilibrium, because the number and type of customers have changed.

In general, a fully formalized system would be desirable, as the more formalized a system is, the less easy it is to exploit its gaps (see Suppes 1968). Ideally, one would like to have a fully axiomatized system of economic laws, similar to axiomatizations in mathematical or theoretical physics. According to Suppes (1968), the advantages of formalization are many, among which are *explicitness*, *standardization*, *objectivity*, *self-contained assumptions*. For instance, explicitness would guarantee that the terms of the theory have the same meaning for all the users, a fact that is not always the case in economics, where several of the terms are fuzzy or vague. A self-contained set of assumptions, on the other hand, "is a way of setting off from the forest of implicit assumptions and the surrounding thickets of confusion, the ground that is required for the theory being considered." Suppes (1968, 655)[12]

While admitting that Suppes is right in posing formalization as a desideratum, in science, it would be a highly ideal normative stance to request such level of formalization, for economics, as the one Suppes envisages. The viewpoint taken in this chapter was certainly to analyze a *normative* theory of consensus formation in economics, while still maintaining some considerations of minimal feasibility in the normative requirements. Feasibility does not seem a requirement in Suppes' assumptions, when he puts formalization as a desideratum in science, without discriminating among *sciences*. Perhaps it is not a case that Suppes considers mostly examples from physics, with one main exception (an example from linguistics) which is also a highly

[12] It is tempting to give reason to McCloskey's consideration on the persuasive power of rhetoric, when, in order to justify why a self-contained set of assumption is a good thing, Suppes talks about the "*forest* of implicit assumptions", the "surrounding *thickets of confusion*", and the "*ground* that is required for the theory being considered." (Suppes 1968, 655 - italics added).

disputable one in the context of his paper.

5.5.3 Observables and variables

The third and last source of problems for the application of the scientific method to at least some economic problems is the definition of the variables, and identification of the observables. Looking back at the example of celestial navigation, we can try to identify the observables and variables there. The main observables for celestial navigation are lines (the horizon) and points (celestial bodies), and perhaps a few others. The variables are angles and distances.

In the other examples provided in section 5.4.2, the identification of observables, and the related measurements becomes slightly more complicated. Medical diagnostics relies on the reading of symptoms (the observables), a process that can involve quite subjective factors; on the other hand measurements of the variables (e.g. temperature, pressure, electrical activity of the brain, etc) often rely on tested tools and statistical methods. While there clearly is ample room for error, the system investigated in medicine is a natural one, and at least under normal conditions statistical methods succeed in making errors and biases cancel each other out.

What is the case for economics? In general, measurements in economics are a particularly complicated matter[13]. Not mainly because of the difficulty of calculating what is identified as the quantity to be calculated, but rather because of the difficulty in agreeing on what such quantities *should* represent and what they *should* leave out. The rate of inflation and GDP are very important factors in several macroeconomic models, thus their measurement is a central task in economic theory. In the remainder of this section I will briefly discuss those two examples. Both examples are mainly illustrative.

I start from the rate of inflation, which measures how prices increase in time.[14] Since a list of the price increases of each and every single commercial commodity would be both intractable and uninformative, the price increase is given as an average. Moreover, since averaging among all commodities would also be intractable and unnecessary (most people probably do not care much about the price increase in polarizing lenses for SLR cameras), the price increase is given as an average of the price increases for *some* commodities. Commodities that are included in the calculation of the inflation rate make

[13]For a comprehensive discussion of some issues related to measurement in economics see Boumans (2007).

[14]Reiss provides a thorough discussion on the formulation of the Consumer Price Index, problem that is directly related to the calculation of the rate of inflation (see Reiss 2008, chapters 2-4).

up the market basket (or commodity bundle). The problem then is to decide which commodities should go into the market basket. The simple answer is "the important ones", and it is not far from the "true" answer.

The problem is not merely a technical one. The answer to the question above depends on what one wants the inflation rate to represent. Because the inflation rate represents the rate at which our money looses value, at the end of the day inflation rate represents how much poorer we become as possessors of money. Clearly, however, that depends on what we buy, if the price of diamond rings goes up tenfold, and about 0.0001 % of the population buys diamond rings, the population has not really become considerably poorer, but only that section of the population which buys diamond rings (1 in a million). On the other hand, if the price of wheat or rice increases, it is a much larger part of the population that becomes poorer.

What goods should go into the market basket, then, does not seem a matter of which techniques give the "best answer", because there is no unique best answer. Any prescription as to which goods should and should not be included, which ones should be added and which ones should be eliminated, will depend on what one wants to represent and what one wants to leave out. In the early 1990s internet provider fees did not exist; less than twenty years later they now represent a significant monthly fee for about 77% of the population of North America, 58% of Europe and 80% of Australia[15]. There is no programmatic or principled reason for adding or eliminating items from the basket, and whereas some cases like bread and internet fees are clear-cut, most of them are not.

In fact, the problem of calculating the inflation rate is handled mostly by a complex number of statistical methods that should, in principle, cancel out biases[16]. Different price indexes aim at making sure that different sets of goods are represented, price averages are weighted according to good's consumption rates, market baskets are kept up to date with changes in the marketplace and seasonally adjusted to reflect possible seasonal trends. The complexity of dealing with inflation measurement is already a proof of the impossibility to fully "rationalize" the theory of inflation. Necessity of adjustment in the basket, for example, is by definition a process that needs to be motivated by considerations that go beyond the domain of economic analysis into the social, sociological, or even political spheres.

The second example mentioned at the beginning of the section was the calculation of GDP. Whereas calculation of price indexes and inflation is

[15]Source: http://www.internetworldstats.com/ Retrieved August 29, 2010.

[16]For an example of these methods, I refer to the *BLS Handbook of Methods* (Bureau of Labor Statistics 2010).

fairly unproblematic, calculation of GDP has been under the critic's spotlight for a long time. Criticisms are cast from all fronts (see Hershey 1995); for example, a GDP index should be able to record quality increases, it should avoid double counting; for instance, the cost of destroying a block of flats and building a hotel instead both go into standard GDP counting. For a survey of the many biases present in the calculation of most GDP indicators I refer to Roubini (1998); however, what several critiques fail to point out is the nature of the problem of calculating GDP.

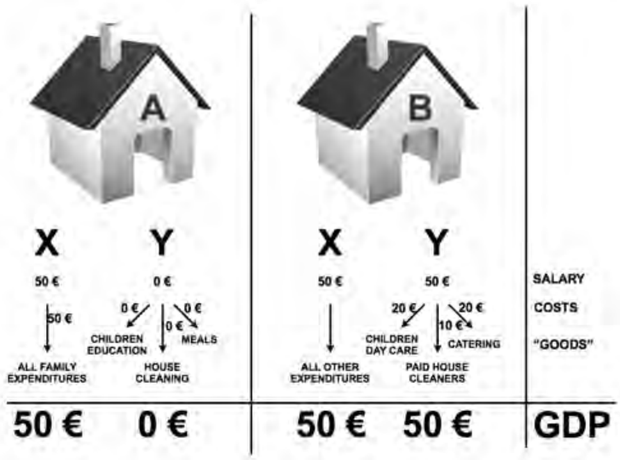
One of the major controversies about GDP indexes is that they leave out any non-monetary transactions in the economy. Virtually all household activities are systematically left out from GDP calculations.

> The general definition of the production boundary may then be restricted by functional considerations. In the SNA (and in the U.S. accounts), certain household activities — such as housework, do-it-yourself projects and care of family members— are excluded, partly because by nature these activities tend to be self-contained and have limited impact on the rest of the economy and because their inclusion would affect the usefulness of the accounts for long-standing analytical purposes, such as business cycle analysis.
>
> (Bureau of Economic Analysis 2009)

The statement that "these activities tend to [. . .] have limited impact on the rest of the economy" is quite clearly misguided. A toy-example should illustrate why the impact of "private" activities is not negligible.

Two households (A and B) are made up by two individuals each (X and Y). In both households, X earns 50 euros. In household A, Y earns no income, and takes care of raising the children, cooking meals, and cleaning the house. In household B, Y has a paid job and earns 50 euros; here, Y does not have time to take care of the children and other household activities, and therefore pays for children day care, meals catering, and house cleaners. Households A and B, intuitively, produce about the same amount of total wealth, even discounting for the extra amount of quality that household B may gain from outsourcing. For the purposes of GDP calculation, however, household B produces twice as much as household A, an amount that does not seem counterbalanced by the (possible) benefits of outsourcing. If the foregoing example were to be used in the context of two very culturally different societies, assumption that is realistically plausible, then "traditional" society would have a very large number of A-type households, whereas the

Figure 5.2: An elementary example of GDP calculation.



"progressive" society would have predominantly B-type households. The bias of the calculated GDPs has a large, quite definitely not a "limited", impact.

The effects of the examples above are accrued by considering other traditionally "private" activities, such as nail polishing, lawn mowing, house redecorating, and so on. Those activities also contribute to a country's GDP only when they involve a monetary exchange. One does not need to go as far as comparing third world countries with industrialized nations, in order to see that the GDP of a country with a fully developed service sector will be much higher than one with a smaller service sector but equivalent *real* provision of services.

It is worthwhile analyzing a little more the point that the impact of such unaccounted factors is limited in an economy. What that means is that factors such as the private education a father imparts to his child does not have a *monetary* value; in other words, it cannot be substituted for another good of the same monetary value. On the other hand, a teacher's salary does have a value because, at least under ideal conditions, a family can choose the schools they send their children to, and perhaps even the teachers who are dedicated to their children. Whereas monetary transactions can buy education in a market, there is no market of the services offered in the privacy of a household by its members (see Bureau of Economic Analysis

(2009, chapter 1) and Australian Bureau of Statistics (2000, chapter 1)).

Monetary value, in this context, is at odds with *social value*, where in the previous example the latter is the value that is associated with the father's activities within the household. However, some definitions describe an economy as "the wealth and resources of a country or region, esp. in terms of the production and consumption of goods and services" (New Oxford American Dictionary 2005). If one accepts that definition, it is clear that transactions between husband and wife, parents and children, a neighborhood council and the elderly of that neighborhood, are all interactions and exchanges that bring wealth and resources into the economy, even though they are not exchangeable, nor can they be purchased in a market.

A further problem, is that the definition of an economic measure often times also defines an economic *target*. Most countries' economic agendas, whether implicitly or explicitly, will contain a GDP-growth target. But any switch from a private transaction to an equivalent public one will result in economic growth, as defined in Bureau of Economic Analysis (2009), without guarantee that there ever was any concrete production of a *new* (or a better) service. In this sense, then economic indicators are not just objective looks into a certain state of the economic world, but can at the same time define a certain end or goal.

The idea of creating objective and measurable indicators is arguably an important one, because it allows assessments and ameliorations when necessary. Nonetheless, the previous discussion should have highlighted the fact that GDP indicators are not synonymous with "wealth", not even just material wealth. Wealth, however, is what we ultimately desire, at least with respect to our material needs.

In short, it seems that calculation of GDP, as it is done nowadays by most national economic bureaux, is a source of misrepresentation of the true state of the economy as well as, potentially, a booster of change, in good or bad, in the distribution and nature of human economic but also social activities. Analyses on the issue of GDP calculation have been done repeatedly and even prominent economists such as Amartya Sen have suggested alternative indicators. Nonetheless, the task of finding an indicator, or indicators, alternative to GDP is essentially a human task, thus subject to the evolution of the human psychology. What an economic measure is not, is something that can be developed with sole reference to economic theories, and corrected and refined by testing or experimenting, as instead several of the measures we use for the natural world can.

5.6 On the epistemology of the inexact sciences

The issues listed in section 5.5 are certainly known in economics, even though, for one reason or the other, they are often presented and analyzed in isolation. For instance, on the issue of the limits of GDP accounting an enormous amount of literature has been written, which has produced suggestions for several alternative economic (or welfare) indicators, for example the *human development index*, and even the *gross national happiness* indicator.

Similarly, the problem of formulating meaningful systems for predicting stock market changes has been treated in several instances. The problem is analogous to the one described in McCloskey (1992): An economist cannot be "smart" and informative at the same time; to wit, she cannot make a forecast which will yield her a net gain, and at the same time make her forecast public. In the same way, a model that predicts the value of stocks will have to make some assumptions about the behavior of buyers and sellers. The behavior, however, is affected by the predictions of the model, rendering its initial assumptions invalid. The case is often made for arguing against the possibility of meaningful prediction and/or policy advice in the field of economics (see Reiss 2008, 166).

The point of the foregoing sections was to gather those and other issues side by side, in order to show, as a whole, where the ideal of scientific methodology fails to yield a consensus, when applied to the economic world. The problem has been put before as a problem of "inexactness" of the social sciences. Inexactness is the reason why the simple application of modeling and testing fails to yield a consensus on statements about the world. This indeed is possibly one of the principal issues with the so-called inexact sciences, that they fail to draw a consensus on many, if not most, statements about the portion of the world they attempt to describe and manipulate. But what exactly is inexactness? The concluding sections of this chapter will try to cast some light on the issue of inexactness in economics.

The idea of inexactness, in economics dates back to at least John Stuart Mill, who divided the sciences into those that are capable of accurate predictions and those that are not [17]. Inexactness, at least in the social sciences, was attributed by Mill to the fact that social sciences have to deal with the behavior of individuals. Due to the complex psychology of human behavior, and perhaps also to the contribution of the free will, it is thus not possible to detect the same exact regularities that we can observe, for example, in some areas of physics.

[17]Salmon reports on Mill in Salmon (1999).

Much later, Helmer and Rescher returned to the idea of inexactness when describing the problem of predictability in some areas of research. To them (Helmer and Rescher 1959), the idea of inexactness is not related to the problem of free will or human behavior, but rather to the complexity of the environment, and the difficulty of formalization of the problem space in the social sciences. Because Helmer and Rescher's idea will be in the background of much of the following chapter, their theory will be discussed in more detail later on.

More recently, Hausman applied the concept of inexactness to economics in describing how economic models make use of *ceteris paribus* clauses, and how economic laws are derived from pragmatically selected generalization in what he calls the *inexact deductive method* (Hausman 1992). The next section will analyze Hausman's conception of inexactness and argue that it does not capture the whole scope of inexactness in economics.

5.6.1 A comparison with Hausman

Hausman (1992) takes from Mill the observation that economics is an *inexact* science, and elaborates on that. Mill's thesis, according to Hausman, can be interpreted in four ways:

> There are at least four ways in which one might attempt to analyze inexactness or the notion of a "tendency." These are not mutually exclusive, and some may be combined with one another:
>
> 1. Inexact laws are approximate. They are true within some margin of error.
> 2. Inexact laws are probabilistic or statistical. Instead of stating how human beings always behave, economic laws state how they usually behave.
> 3. Inexact laws make counterfactual assertions about how things *would be* in the absence of interferences.
> 4. Inexact laws are qualified with vague *ceteris paribus* clauses.

(Hausman 1992, 128 - italics in the original)

Hausman is fairly liberal about the four possible interpretations, although he provides some moderate criticism of the first three and defends a version of (4) (Hausman 1992, 128-151). The problem with the account Hausmann provides is that all of the conditions (1 to 4) seem to be typical of about any of the sciences, even in many of the most *exact* formulations of specific theories we can come up with. On the other hand, the "inexactness" pointed out in sections 5.5.1 to 5.5.3 above, seems to be of a very different nature from the one Hausman talks about. The following will justify the two points.

Let us go back once again to the example of celestial navigation. In section 5.4, we may ask, are the laws regulating the system exact or inexact[18]? If by that we mean the laws in themselves, as inscribed in the "book of nature", they must certainly be exact, but if that is the case (excluding arguments from the nature of free will or similar ones) also the laws that regulate human behaviour are exact. Therefore, that cannot be what Hausman means.

What he means, instead, is probably the laws as formulated in scientific practice: "the apparent retrograde motion of celestial bodies on the celestial sphere follows the trajectory α with velocity β", or "the moon appears every Monday July 3rd at 00:00 hours with x degrees, etc.". These laws are: 1) approximate, they are true only within a certain margin of error; 2) statistical, in the sense that the data points on which they are based are always handled by statistical methods; 3) make counterfactual claims: If there were no interferences, the calculations would not need to be revised and adjusted every so often; 4) qualified with (perhaps not vague) *ceteris paribus* clauses: If secular motions were not in play, if our instruments were a great deal more precise, etc.

Hausman's idea of inexactness seems to be grounded on his description of inexact laws. It is reasonable indeed to assume that a theory or model containing inexact laws are also themselves inexact, and that a science that contains mostly inexact theories or models is an inexact science. If the characterization of inexactness is the one provided by Hausman then probably all sciences are inexact, at least all applied sciences. In particular, also those parts of science described in section 5.4, which seemed to be good candidates for the application of the *rational method* of consensus formation, in virtue of their fairly clear adherence to the method of modeling and testing, should be tagged as 'inexact'.

The examples provided in section 5.4 were meant to provide items of comparison against which to have a look at the economic sciences. Economics, in virtue of the reasons provided in sections 5.5.1 to 5.5.3 above, is inexact in a way that is not reducible to any, or any combination of, the four reasons provided in Hausman (1992). It remains to be argued, in the following sections, which source of inexactness is the one that characterizes economics

[18]It is important to notice that the right question of comparison with Hausman is whether the laws, not the science, are exact or inexact. In the end however, also Hausman wants to claim that it is economics itself, as a science, that is inexact, not just its laws.

in specific, and what possible alternatives to the strict adherence to the scientific method are at hand, if any.

5.6.2 The nature of inexactness in economics

In section 5.6, a difference was mentioned between Mill's reasons for calling the social sciences 'inexact', and the reasons advanced in Helmer and Rescher (1959). Helmer and Rescher attribute the property of a science to be exact or inexact to the availability of a formal characterization of the phenomena that science investigates. In line with that idea, it is no longer possible to characterize a science as either exact or inexact, because such attribution is dependent on the specific matter of investigation, so that the same science can be either exact or inexact depending on the specific phenomenon it analyses.

> We speak of an "exact science" if this reasoning process is formalized in the sense that the terms used are exactly defined and reasoning takes place by formal logico- mathematical derivation of the hypothesis (the statement of the fact to be explained or predicted) from the evidence (the body of knowledge accepted by virtue of being highly confirmed by observation).
>
> [...]
>
> In an inexact science, conversely, reasoning is informal; in particular, some of the terminology may, without actually impeding communication, exhibit some inherent vagueness, and reasoning may at least in part rely on reference to intuitively perceived facts or implications.

(Helmer and Rescher 1959, 25,26)

The difference between Mill's and Helmer and Rescher's interpretations is that in the latter no science is precluded from making exact predictions, or giving exact explanations, so long as there is a formal (and faithful) representation of the phenomenon investigated. It is not clear from the text whether Helmer and Rescher claim that the representation of the phenomenon must preserve as much richness and complexity of the phenomenon itself as possible, or whether all that matters is that such representation *give the correct predictions*[19].

[19] This is the position defended in Friedman (1953).

Whether we only need prediction, or also want our formal representations to be faithful representations of economic complexity, it is clear that the complexity of the economic world makes both harder to achieve, in comparison with other (especially natural) sciences. However there are two types of critique to economic formalism, which should not be confounded. One could argue against the possibility to adequately capture economic phenomena, due to the fact that the phenomena are set in an environment that is too complex to be captured by formal means.

The criticism, however, is tempered by at least two considerations. On one hand computer models are now able to handle much more complexity that it has ever been possible to handle in the history of mathematical modeling in economics. On the other, the criticism would have no effect against Friedman's instrumentalist approach, where all that matters are predictions, and not how realistic the assumptions are.

It is in the context of prediction and policy making then, that the problem of complexity bears. How can we form a valuable consensus on the social capacities of economic sciences, that is, on policy applications, from the relatively simple representations of a very complex economic reality? The conclusion from the foregoing analysis was that the system of economics is too complex for the method of hypothesizing and testing to be sufficient for producing economic knowledge.

5.7 Conclusion

When applied to the study of scientific phenomena that can be considered exact for the intended purposes (e.g. explaining the relative positions of two heavenly bodies at time t, calculating a chemical reaction's yield, etc.), the scientific method is the best available reference point for evaluating the "goodness" of statements about the world. I avoid intentionally here to talk about the *truth* of those statements, to avoid metaphysical quarrels. All that one can infer from the discussion about the celestial navigation model (section 5.1) and similar models, is that by adopting a minimal realist perspective, the evidence produced by the method of modeling and testing ought to drive the formation of consensus, since that evidence seems to be the best candidate around which agreement can be sought.

When the matter of investigation, however, is not such that it can be easily formalized, or whose formalizations cannot be easily interpreted, the distance between the statements produced in the model world and the statements about the world is such that the scientific method does not seem to produce a measure of justification for the latter on the basis of the former.

The question then arises as to what, if anything, should be the criteria of consensus formation in economics. In order to try answer the question, one needs to look at alternative sources of knowledge (in the form of statements about the world) in economics. The next chapter will start off with a look at possible alternative sources of knowledge in economics.

Chapter 6

Consensus in economics PART 2

6.1 Sources of knowledge in economics

Knowledge comes from a variety of sources, and in epistemology it is common to divide them between those which provide justified beliefs (knowledge), and those which only warrant unjustified, or partially justified, beliefs (opinions or credences). Justification itself does not come from a single source; the literature broadly identifies *direct sources* of justification (e.g. sensory experience, introspection, etc.), and *indirect sources* of justification (e.g. testimony).

The scientific method of theorizing, modeling and testing seems to be the best source of knowledge for several disciplines (cf. section 5.3), at least from a normative perspective. Nonetheless, it was shown in section 5.5, that a number of phenomena that seem to rationally pertain to the domain of economics (e.g. welfare, market equilibrium), cannot be easily (or at all) handled by the scientific method. The goal of that section was to provide some examples, in which the application of the scientific method is either inadequate or logically unsound, for contingent or necessary reasons.

The scientific method is not the only source of knowledge, neither in the social nor in the natural sciences. Geology applies the scientific method of theorizing and testing extensively to its objects of investigation, although most pioneering geological knowledge was acquired by experience, and also at the present time field methods in geology are still an essential source of knowledge about geological phenomena. To say that the field geologist does not possess the Science of geology, but only "practical knowledge" of geology, is to miss the point that she possesses *empirically justified statements* about the objects of geological investigation. Those statements are obtained from *experience* (more on experience later), not from theory.

Other sources of knowledge are historical, for example in political sciences or international relations. The idea is that, by recognizing patterns in a succession of events in the pasts, we can reliably infer to the occurrence of similar events in the future. The notion is not alien from scientific methodology, being simply an application of the principle of empirical induction, commonly applied to experimental or statistical sciences. The major difference between the historical method of induction and the scientific one is that, in the macro field of historical occurrences, the inductive process is not as easily formalizable and quantifiable as in scientific experimentation. That, however, does not seem to imply that we cannot have reliable knowledge of political, historical, or cultural phenomena.

Like in most other sciences, in economics knowledge comes from sources alternative to the ones allowed by the scientific method. The major problem is that there seems to be very little consensus as to the principles that should guide the production of economic knowledge, in particular, that type of knowledge that enables policy making. According to McCloskey and others, the situation is dire (see Reiss 2008, xvii).

> Economists don't believe one another. [...] I think it's worse in economics than in what we English speakers call 'science'. And I know it's worse than in historical science. Historians don't believe everything they read in the library. But they expect, rightly, to be able to rely on sheer factual assertions by their colleagues and to have some confidence in their interpretations, if signs of haste or of party passion are absent. [...] I could claim that in economics we have nothing like this degree of scientific agreement.
>
> (McCloskey 2003, quoted from Reiss (2008))

If McCloskey is right, however, anyone looking at economics as a toolbox science (cf. section 5.1.4), should be even more worried. After all, policy makers are definitely more inclined to rely on economists than on historians. But is it possible for economics to ground their consensus formation on anything else other than the hypothetico-deductive model and Friedman's methodological approach? (cf. section 5.3).

The next two sections will discuss that point. The next section will discuss experimental methods, while section 6.1.2 will discuss the historical method in economics. It must be noted that, while experiments are by all means considered an essential part of the scientific method, in this case they are discussed independently because that is not the case in the economic sciences[1]. When speaking about the *economic* scientific method, then, it would be more correct to speak about mathematical/computer modeling and econometric testing. The latter, at least, is the predominant interpretation given to Friedman's idea of *positive economics*.

6.1.1 Experiments in economics

According to some scholars, in recent decades experimental economics has gained a new and independent epistemological status, by supporting economic claims independently of the presence of a preexisting mathematical or statistical model for those claims[2]. Sugden (2005, 2008) defends the idea that experimental economics is nowadays evolving into "systematic inductive enquiry" (Sugden 2008, 623), from the initial stages in which it was instead only meant to test the assumptions of existing theories.

> "The first sustained programs of experimental economics — those in which Kahneman and Smith worked — were presented as tests of core components of existing theories.
>
> (Sugden 2008, 623)

Kahneman and Tversky's early experiments were meant to test the hypothesis of rational choice, widely assumed in economic models and theories of the time, and, to a certain extent, of the present too. In general, the goal of early experiments was to test the validity of assumptions in order to go back to the theory and, if necessary, modify it. According to Sugden, later experiments were taken to be independent confirmation of "economic laws" or "trends" (similar to some experimental laws of chemistry — e.g. Fick's laws of diffusion, or Gay-Lussac's law for gases), that were *induced* from those experiments even before specific theoretical hypotheses had been formulated.

Sugden reports that "for most of the twentieth century [. . .] economics was generally understood, by both practitioners and methodologists, to be a nonexperimental science." (Sugden 2008, 621) From the point of view of the scientific method, as applied in many areas of the natural sciences, economic methodology was perhaps a "crippled" science, lacking an experimental method. Assuming the truth of Sugden's stance, the question then arises

[1]For a perspective on how the experimental method relates mainstream economics see Sugden (2005, 2008).

[2]For a general introduction to experimental economics, in relation to epistemology and philosophy of science, see Guala (2005).

whether economics, with the addition of an experimental face, can provide more reliable knowledge of economic phenomena.

There are at least two reasons why, despite the improvements that experiments can carry to economic methodology, the Scientific Method still remains of limited applicability to economic phenomena.

The first reason is the following. Clearly experiments can aid economic methodology in finding regularities and, possibly, "economic laws". From those, we can hope to build models and theories whose assumptions will likely be more realistic, and less dependent on *ceteris paribus* clauses. Nonetheless, that might not be enough to resolve some of the problems mentioned in sections 5.5.1, 5.5.2 and 5.5.3. For example, where knowledge of the model, if made available in the system, is going to affect the predictions of the model itself, or where the conditions of an economic system change very rapidly, more realistic assumptions will not help. Moreover, the problem of measurement could possibly even be amplified by experimental economics, highlighting questions like what exactly is being measured in experiments (see Ross 2007).

The second reason has to do with the problem of devising and performing experiments in the macro scale. The scale of these hypothetical macro-experiments is such that "experimental macro-economics" might never be feasible under the current state of technology. However, two remarks are in place here: On the one hand, improvements in microeconomics might bring about unexpected development in macroeconomics. In fact, if one sees microeconomics as the foundations and starting point of macro, clearly it should be the case that improvements in micro carry over to macro.

On the other hand, there already is a way of running "experiments" in macro scenarios, and that is by means of computer simulations. Although at the moment our machines' computing power to run macro-scale *realistic* economic experiments is nowhere close the levels that would be necessary, it is possible to envisage futuristic scenarios in which such computing power is achieved. Because of that, the argument that economic macro experiments cannot be performed by means of computers seems to be based on contingent factors.

6.1.2 Historical investigation in economics

An old Latin adage, attributed to Cicero, claims that *historia magistra vitae est* (history is life's teacher). According to some scholars, indeed, history is not only recollection of the past, but also a guide to future actions. As said before, epistemologically the idea is a simple application of induction: By looking at the past and finding patterns of recurrence, I can hope that the system in which I should take a decision now has not changed, in the relevant aspects, from the past, and that the same causes in the past will yield same effects in the future.

In practice, the concrete everyday investigations of historians into the past and the logical application of the principle of induction are far apart, making the direct application of history to "future economics" highly dependent on interpretation. A well-known, though perhaps overworn example is at hand: the Great Depression of the 1930s in the United States of America. The economic events of the period that goes under the name Great Depression have been, and still are, a heated topic of discussion. Scholars, both historians and theoreticians, still debate over the causes of the economic events of the late 1920s and 1930s.

The major debates were not exactly over what caused the initial Wall Street Crash of 1929 — dramatic market fluctuations are known to happen and are often attributed to random processes — but rather as to what made the initial situation persist for over a decade. It would be beyond the point of this section to recall the threads of the various debates among the schools of thought relating respectively to Friedrich Hayek, John Maynard Keynes (and neo-Keynsians), and Milton Friedman, but what is certain is that the debate about the Great Depression is far from settled[3].

What one can conclude from a glance at the course of economic history, and the literature written on it, is that history is quite definitely not the place where to look for economic consensus. That however, is a descriptive point; the question remains as to whether it is the place where one *should* look. The foregoing comments should not be taken as a particularly developed theory of the relation between economic history and knowledge of the economic world as it functions and develops. However, it is arguable that economic history is in no place to be a normative basis for consensus anymore that economic theorizing is.

The reason for that is that "historical induction" is at least as controversial as other forms of economic analysis. On the one hand, as it is the case in the historical reconstruction of the Great Depression, it is possible to reconstruct the historical events in a way that will support one or another of a range of policy applications. On the other hand, it is always possible to argue that a certain historical period differs in significant aspects from the present one, thus possibly invalidating any inference from the former to the latter. While neither of these are conclusive arguments against the possibility

[3]To give the idea of the extent of the debate nowadays, it should be enough to highlight the fact that on the causes and possible lessons of the Great Depression there have been written a dedicated Wikipedia article and a volume of the popular *for Dummies*® series (Wikipedia 2010; Wiegand 2009).

of correct historical inferences from the past to the present, the arguments do show that any such inference is likely to raise reasonable disagreement from other historians.

The conclusion from section 5.5 and the subsequent sections, was that statements about the economy and economic systems are underdetermined, in the sense that often models and theories fail to provide adequate support and justification for specific economic predictions or explanations. Similarly, one can easily argue that historical investigations also tend to only vaguely justify economic predictions or explanations. Whereas the complete denial of any value to the study of history, for the purposes of economic applications, is probably both theoretically unjustified and of little practical use, one must reckon that there is no easy way to go from historical observations to predictive statements and policy prescriptions.

6.1.3 Methodological liberalism

It is time to draw some preliminary conclusions about sources of knowledge in economics, and the problem of consensus seeking on those theoretical issues that have a direct application in the economic world (e.g. forecasting). Ariel Rubinstein, discussing the practical role of economic modeling, resorts to poetry:

> As in the case of a good fable, a good model can have an enormous influence on the real world, not by providing advice or by predicting the future, but rather by influencing culture. Yes, I do think we are simply the tellers of fables, but is that not wonderful?

(Rubinstein 2006, 882)

Rubinstein defends the idea that models are not that much different from fables, both in their characteristic structure (idealized worlds, unrealistic claims, etc.), and in the function they perform: they influence the culture, and, arguably, the minds of policy makers, managers, CEOs, bankers, etc. It is far from clear that most economists would agree with Rubinstein, and that their (at least intended) goal, when building models or constructing experiments, is to tell fables, or less literarily, influence culture. It is undeniable that science influences culture (see Russell 2009, Part V), but that is often taken to be a secondary effect of science, not its primary purpose.

If economic theorizing were simply story-telling, there would be little debate as to why econometrics is superior to economic history, or why experiments are more accurate than careful conceptual investigation; in fact there would be no debate because books of history tell wonderful stories just as much as models, and reports of experiments influence culture just as much as mathematical models. If the situation were really as Rubinstein portrays it, then all methods would be on a par with all others, at least from a normative point of view. Such a situation would give little room for distinguishing, a priori, good economics from bad economics. The resulting situation would be one of methodological equality: no method can, a priori, be claimed superior to another. This is the sort of situation that McCloskey seems to advocate, or perhaps the methodological anarchism Paul Feyerabend defends (Feyerabend 1988, 1999).

Clearly there may be other methods for discriminating between good economics and bad economics. It is assumed here that one should be able to at least say something a priori on the quality of a method from an epistemic point of view, without resorting to considerations of style, utility, ethics and so on.

Without rejecting, in principle, the idea of methodological liberalism (McCloskey), or the stronger claims of methodological anarchism (Feyreabend), the next sections will try to provide an alternative analysis to the problem of how knowledge should be tackled in economics and what principles would ideally guide the formation of economic consensus.

What is left for the space of this section is to highlight the fact that in a situation in which the principal sources of economic knowledge underdetermine claims about economic phenomena, one can still ask the question of what type of justification such claims may have by looking at other, perhaps less obvious, sources. Accepting the fact that in economics there does not seem to be a way to univocally justify an official method over another, in the following sections I will investigate those sources of economic knowledge that are most proximate to the recipient of economic consensus, be these public or private institutions. Such sources are institutionally called "economic experts", and are those who, with any previous background information (historical, theoretical, practical, etc.), have a say in the public or private arena on the theoretical matters underlying economic decision-making.

6.2 Experts in economics

It was argued, so far, that no one of the methods mentioned (modeling and econometric testing, modeling and experiments, historical investigations, etc.) is individually sufficient for the economic toolbox that a science wishing to apply its results needs[4]. At the same time it is hard to deny that there is such a toolbox in any political administration, government agency, banking institution, or company. Even more, one can argue that there are *several* toolboxes, all partly different from one another, and perhaps also partly or wholly incompatible with one another. The latter are descriptive statements, but the fact that a consensus needs to form, for a policy-making action to be taken, should be granted. The question then remains as to how such consensus forms, and how it should form.

The claim defended in this section is that there is a source of knowledge that is often overlooked, and that is *expertise*. Experts, I will argue, are themselves sources of economic knowledge, and irreducible to any of the previously mentioned sources or combinations of those. But before delving into the subject of experts and expertise, there is a common misunderstanding about expertise which must be dealt with.

That experts are a source of knowledge is a trivial thesis, at least in the following sense: Experts constitute a large part of the communicative channels through which whatever type of knowledge is gained, by scientific or other methods, is passed on to the decision-makers. This function of experts should not be minimized[5], but the role of an expert as a communicator is essentially different from her role as an irreducible source of knowledge in a specific discipline. We can exemplify the role of an expert as a communicator by sticking to the example of celestial navigation (cf. section 5.4).

In the example of celestial navigation, it is likely that the practitioner (the sailor, pilot, etc.) will not know the details of the science behind celestial navigation. Practical knowledge about navigating by use of celestial bodies, indeed, comes from training or navigation manuals. It is probably then up to the scientific expert to *communicate* the science, that is, to reduce it to practical operational principles. Despite the possible dangers involved in such operation, this does not impinge on the science of celestial navigation itself, as it was described in section 5.4.

Moreover, it is arguable that in those fields where it was said that "exact science" is possible, it is relatively easy to double check the information conveyed by the expert to the layman, with other experts or on textbooks. This is not true of the situation in which experts are direct (and not only indirect, as in the case of the communicator's role) sources of knowledge. The meaning of the expression 'direct sources of knowledge' as referred to experts, will be explained in the remainder of this section.

[4]It was assumed in the introductory sections of chapter 5 that economics does have aspirations to be (also) a toolbox science.

[5]For an introduction and a brief survey of the topic see Russell (2009).

6.2.1 Experts and tacit knowledge

The goal of this section is to make explicit the sense in which experts can be thought of as sources of knowledge beyond their role as communicators. In that light, experts are the depository of a form of knowledge that has been called with various names in the literature: 'personal knowledge' (Polanyi 1958), 'background knowledge' (or 'implicit knowledge', as opposed to explicit knowledge) (Helmer and Rescher 1959), but mostly, and currently, 'tacit knowledge' (Collins 2010).

The definition of tacit knowledge is problematic because different authors elucidate the concept in different ways; some define it as an actual state of knowledge (knowledge that *is* tacit), some as a modal concept (knowledge that *can only be* tacit). For working purposes, I will give Collins's characterization, without attributing it a definitional value: "Tacit knowledge is knowledge that is not explicated." (Collins 2010, 1). Also Helmer and Rescher (1959) write on the relation between tacit knowledge and experts.

> "The expert has at his ready disposal a large store of (mostly inarticulated) background [read 'tacit'] knowledge and a refined sensitivity to its relevance, through the intuitive application of which he is often able to produce trustworthy personal probabilities regarding hypotheses in his area of expertness."
>
> (Helmer and Rescher 1959, 38)

Experts make use of tacit knowledge in most natural and social sciences. A doctor diagnosing a patient might rely on decision aid software (see section 5.4.2), although almost invariably there will be a great deal of personal judgment involved in the diagnosis. Similarly, conservation biologists make large use of their knowledge of the field, besides their theoretical science, when giving advice like forecasting the expected success of a particular conservation program.

That there is tacit knowledge, and that it constitutes a large part of one's "professional" knowledge (viz. knowledge of the investigated environment), is a well-known fact in the field of business planning and forecasting. Explicitation and transmission of tacit knowledge is the leading concept of Ikujiro Nonaka's business theory (Nonaka 1991; Stillwell 2003). Nonaka claims that most of the potential in a business enterprise is the ability to render the large batch of knowledge present in a company's human capital explicit and usable for the company's development.

The discussion of knowledge in the business literature tends to blur important distinctions that a science-based approach would like to draw between *how-to* knowledge (or skills), factual knowledge (information), or theoretical (predictive or explanatory) knowledge. For example, Gourlay's definition of tacit knowledge includes values and emotions: "Tacit knowledge is a non-linguistic, non-numerical form of knowledge that is highly personal and context specific and deeply rooted in individual experiences, ideas, values and emotions." (Gourlay 2002, 2). Whereas one must recognize the importance of values and emotions in a decision making context, it is debatable whether they play a decisive epistemic role in the creation of scientific knowledge in the form of explanation and prediction.

Here, the focus is on the epistemic value of tacit knowledge, that is, on its role in forecasting and explaining scientific (and in particular economic) phenomena.

> "A source of characteristic examples of the predictive use of expert judgment is provided by the field of diagnostics, especially medical diagnostics. [...] the use of background information, in a way that is not systematized but depends entirely on the exercise of informal expert judgment, may appropriately lead to predictive conclusions in the face of prima facie evidence which points in the opposite direction."
>
> (Helmer and Rescher 1959, 40)

In the above quotation, the "prima facie evidence" is evidence that can be managed by a medical decision aid system, a software with a number of inbuilt functions and a database, which mechanically formulates a diagnosis on the basis of the inputs (symptoms, etc.) given by the doctor or nurse in the standard pre-clinical visit procedure.

Cooke reports on the use of experts for assessing risk in the engineering sector.

> "As in the nuclear sector, expert opinion entered the aerospace sector because of the desire to assess safety. In particular, managers and politicians needed to assess the risks associated with rare or unobserved catastrophic events. The likelihood of such events could obviously not be assessed via the traditional scientific method of repeated independent experiments."
>
> (Cooke 1991, 19)

I will come back to the problem of *using* expert opinion in the best possible way later. For now, the question I want to address is what makes expert opinion valuable and not reducible to a simple combination of textbook data.

6.2.2 What is experience?

Expertise, and its characteristic tacit knowledge, can be a product of historical, scientific, mathematical, or conceptual investigation, but what seems peculiar to it is that it is almost invariably highly dependent on experience.

Knowledge deriving from experience, like tacit knowledge in general, is most of the times imperfect, vague, and hard-to-describe. Moreover, assessments based on experience are also hard to replicate in a controlled scenario (Cooke 1991, 18), a factor that makes the scientific mind suspicious of any claim based on experience.

If I were asked the time by a passerby, looked at my wristwatch, and uttered "it's 3 o'clock in the afternoon", the passerby would likely walk away confident, to a certain extent, that it is in fact 3 o'clock in the afternoon. If, in the same situation, I were to raise my head to the sun, check the surroundings, and utter "it's 3 o'clock", the passerby would likely walk away just as uncertain of the time as she was before our encounter. Nonetheless, some people can quite reliably know the time if, for example, they have acquired a sensitivity towards the relation between the time of the day and the conditions of their surrounding environment.

If I step outside of my house, which stands opposite to a school, and see a line of cars parked all along the curbside, but not much movement of students in the street, I can very reliably assert that it is five to ten minutes to one o'clock, when the students are about to come out of school, and their parents are waiting for them outside. If, instead, I see no cars lined up, but only a small number of students standing in front of the school and chatting, I can reliably infer that it must be five to fifteen minutes past the hour, when most of the students that came by car left, and those few who remain are waiting for the school bus.

With more and more experience, more and more exposure to the same scenarios from day to day, and perhaps a richer number of cues that can aid my estimations, I can probably become fairly accurate and determining the hour even when conditions are slightly changed from the *standard* picture described above.

With rich enough scenarios and refined sensitivity, experience can be a very reliable source of knowledge. Besides, such type of knowledge is flexible, contrary to mechanical or "automated" knowledge. Regardless, experiential knowledge cannot be verified, other than by "proxy" indicators, past performance, trust, etc., and can easily be faked, so that the passer-by will likely prefer to put her trust on a watch than on my refined sensitivity for telling the time.

A person's reliability in telling the time does not come from computing, for the above example, the number of cars, parents, students, and their relative positions in the street. More commonly, such knowledge is arrived at unconsciously by the background work of the brain when we interact (through the senses) with the environment around us.

The example of time telling, as are most illustrative examples, is simplistic. Psychologists, however, have dedicated plenty of work to the role of experience in problem solving tasks (see Ericsson 2001; Simon 2001; Fischhoff 2001). Experience seems to be one of the factors that affect expertise, intended as performance at problem solving abilities. Experience determines *skills* (e.g. playing musical instruments), as well as *analytical abilities* (e.g. chess playing, forecasting), although the correlation between experience and expertise is very domain-specific and assumes both positive and negative values[6].

Expertise and experience certainly have a role in the economic sciences too. One of the standard roles attributed to economics is *explanation*. What is irreducible of the expert is the fact that she has experience and thus tacit knowledge of economic matters. Whereas knowledge derived from the application of the scientific or the historical methods can at least partially be communicated and made explicit in textbook style writings, knowledge gained from experience is harder to be put to use.

6.3 Tacit knowledge in groups

So far the concept of tacit knowledge has been outlined mainly as a personal one. Tacit knowledge resides in experts in the form of experience, implicit information, or other forms. However, there is also a sense of tacit knowledge that applies to groups of epistemic agents.

The basic idea is simple. Imagine a group of two agents a and b, each of which is in possession of a certain item of knowledge, respectively I_a and I_b. Since there is not any standard partition of knowledge in its fundamental elements, it is possible that item I_G, also an item of knowledge, be a set G of the items a and b. Trivially, then, the group of two agents will not know I_G as long as a and b do not communicate their individual items of knowledge to each other and, by that means, the group.

As an example, imagine the following scenario: Jack and Sally are heading

[6]Two important experiments testing the calibration scores of business students forecasting stock prices and earnings showed an "inverted" expertise effect, where more experienced subjects (graduate business students) did worse at forecasting than the less experienced ones (undergraduate students) (see Yates and McDaniel 1991; Staël von Holstein 1972).

to the river bank for a picnic. While, Sally is unaware of where Jack, who is driving, is going, Jack knows the exact location. Sally has heard the news in the morning and knows that the river has overflown during the night, so that any vehicle directed towards the flooded area is bound to be caught in the muddy terrain. Together, Sally and Jack know that they are going to get stuck in the flooded area, but as long as they do not communicate their individual information to each other, that item of knowledge — and arguably a valuable one — only remains in the group as tacit knowledge.

A 2004 article in the *New York Times* reports the words of Samuel R. Berger, national security adviser to President Clinton: "We've learned since 9/11 that not only did we not know what we didn't know, but the F.B.I. didn't know what it did know." (Shenon 2004) The reference is to the fact that, due to the inefficient communicative structure of the Federal Bureau of Investigation, while the community, as a whole, had knowledge about the risk of terrorists attack, on the other hand such knowledge was not explicit, and thus unknown to the group intended as an operational entity. Knowledge that remains implicit, even though it is present, makes the group unable to operate in the light of that knowledge.

It is evident then that tacit knowledge exists both at the individual and at the group level, albeit with different characteristics and for different reasons. Consensus, it should be noted, is a composite of the individual convergent (or identical) opinions of a group, but it can also be the information, as a primitive composite, that is present at the group level, and which is then transmitted down to the members of the group. It should be clear then that tacit knowledge is important for the formation of consensus both in its individual and group forms.

One final remark should be made about the nature of expert knowledge, both at the individual and at the group level. In the examples above, experts and the group possessed "factual knowledge"; in philosophical terms, they possessed some evidence of facts. But this is not the only type of knowledge that experts possess. In the context of economics, there are at least two other types, which are of primary importance for policy making.

The first one is knowledge of the future, to wit, forecasts. Experts can predict the future of events related to their field of expertise. Formal methods (statistical models, computer-based methods, etc.), which are normally used for forecasting, can and should be substituted by "human predictors" when the conditions require it. In that way, experts become forecasting devices, when formal tools are not available or not suitable for a specific problem. A full analysis of experts as forecasting devices is found in Helmer and Rescher (1959) and Cooke (1991). More on that will be discussed later on this chapter, in the discussion of the Delphi method (cf. section 6.6.1).

The second type of knowledge related to expertise is *how-to* knowledge. How-to knowledge is not necessarily related to manual or physical ability, but it refers to the complex set of skills that are required in problem solving situations. Indeed, part of the task that economic sciences are faced with, when their primary concern is with policy making, is that of problem solving.

Problem solving is in some respects more complex than scientific understanding and forecasting, insofar as it involves the ability to identify the problem, identify the tools that can resolve it, and use the tools available to resolve it. How-to knowledge is here the ability to think creatively, to identify solutions for relatively new and non-standard scenarios, and so on. In this sense, economics turns into economic engineering, a process where explanation and prediction are only of secondary importance, while the primary aim is the achievement of specific goals with given means.

An example of the way economic engineering is achieved is given in Guala (2001). In general, auction theory and the licensing of broadcasting licenses by several governments, is often taken to be an example of the successful use of economic knowledge for policy making (see also Reiss 2008, chapter 5). The policies that were implemented in actioning broadcasting-spectrum licenses involve a very complex number of skills, and a sensitivity to the specificity of the problem, for which experts are said to have a special type of "practical" knowledge. A large part of tacit knowledge (see section 6.2.1), is 'how-to' knowledge, and plays an important role in economic applications.

6.4 Some preliminary conclusions

Given the status of theoretical knowledge in economics, most of the situations of consensus in the sciences must be a direct product of *economic experts*. These will be people formed in the scientific, historical, and business studies but probably also trained in the economic field, and experienced in economic decision making for businesses, governments, economic organizations and other institutions. Descriptively, this seems to be the best depiction of the situation in the economic sciences.

However, since the emphasis of this thesis is on the normative question (cf. section 5.2), the foregoing considerations will focus on assessing whether such situation is optimal, that is, whether reliance on expertise is desirable in economics. Economic experts clearly have advantages over textbook economics, if textbook economics is transmitted uninterpreted to the policy makers or to whomever it is to be transmitted to. It seems that in economics the situation should resemble more the case of, say, a geopolitical advisor

— who communicates to the policy-maker by interpreting her theory and making extensive use of her tacit knowledge — than to the case of the hypothetical instructor of celestial navigation mentioned in section 6.2.

The interpretative layer, however, and the use of personal non-explicated knowledge, makes the judgment of the expert subject to important biases and flaws. Such biases are important insofar as they both undermine the justificatory status of expert knowledge (hence the skepticism towards experts expressed in the literature), and they produce knowledge that is either, incorrect, unjustified, or in general, “bad” knowledge.

The following section will present the main problems associated with expert knowledge, both from an individual and a collective viewpoint.

6.5 The drawbacks of expert elicitation

It should be noted that the great majority of the literature on the methods for using expert judgment tends to argue, by empirical testing, that such methods are superior to informal deliberation and elicitation. However, little is said as to the various problems that make the use of expert judgment and elicitation problematic. An exception is in Cooke (1991), who provides a survey of the major biases and some mathematical problems in expert elicitation.

This section is an attempt to categorize and put in a schematic form the various issues that affect the use of expert elicitation. Even a brief survey of the many biases and other problems that affect experts and their judgment, as individuals and in groups, would require covering two very large literatures: on the topic of individual biases, the psychological literature on cognitive biases, and the one on logical and probabilistic fallacies of reasoning. On the topic of group biases, the literature on judgment aggregation, with related impossibility theorems, and the psychological literature on biases that arise specifically at the group level.

Realistically it would be close to impossible to handle even a small section of such literature. For that reason, this section will cover the problem of the disadvantages of expert elicitation from a very schematic point of view. A table will give an essential representation of the various problems associated with eliciting expert’s opinion, and some essential references. Each of the four points in the table with be briefly explained individually and some preliminary conclusions will be drawn.

Table 6.1: Taxonomy of the drawbacks related to "expert judgment"

	Logics & Mathematics	Psychology
Individual Level	ELICITATION	INDIVIDUAL BIASES
Group Level	AGGREGATION	SOCIAL BIASES

6.5.1 Elicitation

Elicitation is the process of asking an expert her opinion on a certain subject matter. The problem seems trivial. If I want to know whether a certain statement is true or false, I can just ask the expert for a yes/no assessment. Oftentimes, however, the information obtained from a yes/no statement is not very valuable.

In the engineering sector, for instance, expert elicitation is often used for risk assessment (Cooke 1991, 19, *quoted in section 6.2.1*). Risk is assessed as a probabilistic value; therefore the expert, asked to assess the risk that associates a certain scenario with a certain danger, should provide a value between 0 and 1. This form of elicitation is called 'point estimation'.

Point estimation, however, does not provide information on the uncertainty associated with a certain assessment; that is, it provides only one degree of information (Cooke 1991, 46). If a mechanical engineer is asked about the risk of a particular machine component to break beyond the point of functionality, she might give an *informed estimate*. Were I to be asked the same, I would probably give a very uninformative estimate. Informativeness and non-informativeness is second level information on my assessment, that is, a measure of how confident I am that my assessment be correct. Clearly, knowing how confident an expert is about her assessment is relevant information for the decision maker.

Elicitation is a formal problem because the formal treatment of yes/no answers, point estimates, or interval estimates is different and different measures do not have a clear psychological interpretation. A simple example is the problem of comparing different individual scales. We can assume that it is clear to most assessors what the meaning of 'risk 0' and 'risk 1' are. What lays in the middle is not as straightforward. Expert A's scale for risk might be linear, while Expert B's could be logarithmic; if that is the case, then different estimates will correspond to the same piece of subjective information. It is difficult to know a-priori what piece of information they correspond to, although psychological scaling seems to be the most reliable method (Cooke 1991, chapter 14).

6.5.2 Individual biases

Even assuming that individual elicitation, as a mathematical task, is somehow a solvable problem, the issue would still remain of individual psychological biases which affect human judgment. The landmark reference for this subject is in the works of Amos Tversky and Daniel Kahneman; notorious biases are 'framing', 'hindsight', 'confirmation', and several other ways in which human judgment consistently deviates from the principles of logic or probability theory. Classified cognitive biases are known because they have been tested in controlled settings and are normally replicable.

More in general, individual biases are psychological predispositions that influence one's judgment when certain cues are given. Biases need not be normatively undesirable. A bias to truth, were such a bias to exist, would certainly be epistemically desirable. Expert A's bias towards expert B, that is, A's tendency to follow B's opinion, is normatively desirable if B's opinion tends to be consistently better (e.g. closer to the truth) than A's. Unfortunately, in most contexts in which expert judgment and tacit knowledge are involved it is impossible to tell a-priori whose opinion is likely to be the best one (Armstrong 2001, 421-422).

To conclude this brief section on individual biases, it should be noted that cognitive biases and biases affecting behavior[7] are, in principle, a red flag on the good epistemic status of the judgment that is elicited.

6.5.3 Aggregation

Section 6.3 motivated the study of expert elicitation from the viewpoint of groups rather than single individuals. Cooke writes "Most decision-making bodies are not individuals but groups." (Cooke 1991, 171). The statement is true not only for decision making bodies, but also for advising committees, expert panels, etc. Moreover, whereas Cooke only makes a factual statement, the claim that most decision bodies *should be* groups rather than individuals is also true, in a great many contexts[8].

[7]These are political, emotional, or other types of biases, which have not been mentioned here because their study is not as developed (and replicable) as the research on cognitive biases. Nonetheless, in principle they are equally affecting the epistemic value of one's judgments, and a proper handbook of expert elicitation should at least give mention of them and possibly give guidelines on how to avoide such biases.

[8]There is no space here to provide a thorough defense of this statement. Group judgment and the aggregation of judgments, in many scientific and social contexts, is considered to be epistemically superior to individual judgment. The branch of philosophy that goes under the name of 'social epistemology' provides a number of justifications for that thesis. For a philosophical analysis see Goldman (1999).

Group judgment, however, almost invariably requires some artificial method of consensus formation. The meaning of consensus here needs not be as narrow as it was taken to be in the initial chapters of this work (2, 3, 4), where consensus was analyzed from a purely epistemological point of view. There, consensus meant convergence of individual beliefs to a common view, and the prospects of having a formal model capable of producing *mathematical consensus* seemed destined to fail. In the context of policy advising, decision making and, in general, from the viewpoint of philosophy of science, consensus has normally a broader meaning. Consensus can be, and normally is, reached as a compromise, or an aggregation, of reported beliefs. Because the problems related to reaching a compromise belong to the next section, I will start from aggregation below.

Aggregation is the mathematical or logical merging of initially different mathematical or logical values. Values can be numerical estimates (e.g. life expectancy, in number of years), probabilities (e.g. risk), binary values (e.g. yes/no for qualitative assessment), and possibly other values. Judgment aggregation gives rise to a number of problems that stem from various impossibility theorems. The pioneering work on the literature has been Arrow's work in social choice theory. His famous theorem is on the impossibility of formulating a social preference ordering that satisfies a number of seemingly innocuous principles.

Arrow's work gave life to a whole literature of impossibility results. Saari (2008) gives a compendium and a general mathematical theory of the major impossibility results. List, in a number of papers, has extended the results of Arrow's and proven new ones; List and Pettit (2011) provides a good compendium of the major results in collaboration with different authors. Finally Xu and Da (2003) gives a broad comprehensive overview of aggregating operators, without impossibility results.

The conclusions that can be drawn from the literature on aggregating judgment is that there are almost always trade-offs to make between conditions when aggregating expert judgment. In relation to their impossibility result for the aggregation of propositional attitudes, List and Pettit (2011) take a constructive attitude.

> How should we interpret this impossibility theorem? [...] our result can be taken to show that, if a group seeks to form intentional attitudes, it must relax at least one of the four conditions [the passage refers to the four conditions, analogous to Arrow's conditions, from which the impossibility is derived].
>
> (List and Pettit 2011, 45)

Which trade-offs to make, however, is hardly ever a purely technical choice, other than for convenience. Whether to sacrifice, for example, *collective rationality* or *universal domain* (see List and Pettit 2011, 49), may depend on the composition of the group that is required to aggregate. A particularly cohesive group, one where major differences in individual opinions are not expected, may accept the drawback of having incomplete group attitudes by taking a supermajority rule of two-thirds or above (see List and Pettit 2011, 53). Whether such group is functional, and avoids the impasse of not achieving a majority, is an evaluation that can only be made a posteriori in the operational context.

6.5.4 Group biases

Often times deliberating groups do not apply an aggregating procedure directly. Deliberative methods can themselves bring the opinions of the group closer to one another through simple discussion, feedback, and evaluation. In fact, a convergence effect is observed in the Delphi deliberative procedure (more on this in section 6.6.1), where iterations of problem assessment and provision of feedback to the members leads, in some cases, to a certain degree of convergence.

In general, a deliberative group, in light of the necessity of forming a group opinion, might often simply settle on a compromise solution without necessarily requiring aggregation[9]. A compromise is normally the result of psychological processes[10], obtained as exchange of opinions and more or less slow settlement and elimination of the individual differences.

Suppose one (e.g. a prime minister) is to assess a number of factors that will affect the inflation rate in order to stabilize the rate within a specific interval. The prime minister will gather a committee around a table, ask them to exchange opinions, and come to a feasible number of policy advices. Despite that seeming the most intuitive procedure, hardly ever is the strategy successful in order to obtain valuable expert opinion.

[9]It should be noted that aggregation is not always the best solution and in some cases it leads to absurdities. Imagine a group that is trying to decide whether to cut public sector wages, thus losing the public sector section of the electorate, or increase corporate taxes, loosing the support of some corporations instead. A compromise, half wage-cut and half-tax increase, might loose both electorate and corporate support. Even though in this case one can blame the problem on the fact that one should not "*aggregate* apples and oranges", it is possible to formulate a similar argument with one type of entity alone. A similar example was given in the introduction of this thesis, the example of King Solomon and the two women.

[10]Negotiation can also be seen as a psychological process.

In the first place, a committee of such sort is very unlikely to even be able to get to an orderly and fair discussion without a moderator. The moderator (or chair) is the first form of superimposition of an “elicitation mechanism” over the committee. As trivial as its role may seem, the moderator should prevent the group from running into social dynamics that will likely lead it to low performance. For instance, the most aggressive individual will likely take over the discussion floor imposing her views on the rest of the group. The often-cited adage “repeat a lie a thousand times and it will become a truth”, at times attributed to renowned Nazi or Communist dictators, stands as a voucher for the idea, taken from the psychological literature, that it is not the right opinion that tends to be believed by the masses, but the one that is repeated most times and most fervently.

Biases that arise at the group level are at times called social dynamics, that is, forces which are a product of the group as a whole, and affect the individuals composing it. Social dynamics are tendencies of individuals to act in ways that are conditioned by the forces of the group.

6.5.5 Relying on experts

As the previous section should have indicated, reliance on expert judgment is problematic, especially when the distance between the theory and the practice is large. The more precise and operational the science is, the less the room for personal judgment will be, when moving from the abstract domain of the theory to its concrete application. But whenever room for personal judgment is large, biases of different types and judgment errors will likely be numerous.

Trout (2009) warns against the use of so-called “experts” in a number of situations where simple statistical methods are shown by psychological research to perform better. The decisions of probation officers evaluating parole candidates would better be substituted by evaluations based on the felons’ past record (Trout 2009, 163-168). Similarly, argues Trout, student interviews are worse at selecting students for admission than other methods that consider only past academic record (Trout 2009, 163-168).

McCloskey (1992) asks the question why professed economic experts are not themselves the richest men, all other things being equal. The question is rhetorical, and the paper is meant to show the absurdity of the concept of a forecast that is both profitably informative and public[11]. Yet, the effect is to cast doubt on the professed “expertise” of economists as *predictors*.

[11] If a forecast is public, the public has already exploited its information for their profit, thus making it unprofitable.

To throw even more doubt on the efficacy of economists as predictors are two papers (Staël von Holstein 1972; Yates and McDaniel 1991), showing that expertise seems to be negatively correlated with stock market predictions. Yates and McDaniel (1991) confirms and extends the results in Staël von Holstein (1972): "[...] subject's probability forecasts of stock prices where shown to be surprisingly inaccurate" and "[...] although we might expect that greater experience will lead to demonstrably greater accuracy, it instead simply resulted in more useless variation in judgments" (Yates and McDaniel 1991, 75). Although, as the authors admit, the results are far from being generalizable, the aforementioned papers cast doubt about whether the expertise of stock prices forecasters is real or only professed.

Angner applies the studies on the effect of overconfidence to a historical case, showing how professed individual experts in government advising positions "are likely to fall prey of significant overconfidence" (Angner 2006, 19). By means of conclusion, the author states "I have assumed throughout that we are on the whole better off relying on serious economic analysis in public decision making. My point is that to make the best of the situation, we need to be aware of the limitations of expert advice and try to anticipate diverse negative consequences." (Angner 2006, 20)

The model of expertise that Angner presents is a simple one: There is one individual, trained as an economist, who advises directly the Russian government on which economic policies it should adopt. Even though it is impossible to know exactly how the chain of power was working, it is arguable that the advisor to a government was consulted in briefings and perhaps written memoranda. The scenario seems open to the influence of all the problems described in section 6.5, with the only exception of aggregation problems[12].

In the next section two models for expert advising will be presented. The major difference between these models and the folk models of expert advising is that the latter lack almost any formal structure and take the number of individuals in the advising structure to be only a contingent issue. The next section will present a number of methods for deliberation and expert elicitation where the attempt is to control the type and quality of interaction between experts, in order to avoid at least some of the issues involved with expert judgment.

---

[12]Note that group biases can arise not only in advising committees, which would not be the case for the example that Angner (2006) gives, but also from the interaction between advisors and policy makers.

6.6 The advantages of expert elicitation

Before presenting the methods, it should be said that there are at least some facts about expertise and expert judgment that have been quite extensively investigated: a) group judgment is epistemically superior to individual judgment, and b) structured judgment is epistemically superior to unstructured judgment.

a) *Groups outperform individuals.* The first fact has been investigated in several respects. The first pronouncement came in the form of a mathematical theorem, the *Condorcet Jury Theorem*, showing that if some initial conditions are met, a group as a whole is better at *tracking the truth* than its individuals in isolation. The theorem has successively been extended and its assumptions relaxed. Empirical literature has also shown that experts tend to perform better than individuals at problem solving and forecasting tasks (Rowe and Wright 2001; Armstrong 2001b; Aspinall 2010; Stewart 2001).

The literature on the *Nominal Group Technique* (more on this later) has shown that groups provide a better exploration of the problem space, enhancing creativity (Delbecq and Van de Ven 1971), and producing more input in the form of available information (Nonaka 1991). What should not be assumed, when developing a formal tool with the idea of capturing a natural (or social, in the case of economics) phenomenon, is that the formulation of the problem is both easy and innocent with respect to the conclusions that can be derived from that tool. Groups have a better chance at formulating the problems more accurately because they will likely possess more information than the single individual on the issue in question.

b) *Structured judgment outperforms unstructured judgment.* The principle that one ought to structure the judgment process of a group seems to go against what was stated before about the superiority of tacit knowledge over mechanical methods in fields where formalization of the problem is hard to achieve. The confusion, however, arises from a misinterpretation of what it means to structure a judgment process.

The existence of a chairman in a discussion session clearly limits the freedom of expression of the participants, but, ideally, only insofar as such freedom goes against the principle that all information should be made available. Similarly, the requirement of anonymity in the Delphi Procedure, limits the interaction of experts by not allowing them to recognize each other other than by means of the expressed judgment. In other words, in Delphi, an expert is identified with its judgment, not with its physical, behavioral, interactional features. While one could object to such limitations by arguing that such aspects are an essential part of the human dimension of interactions, it turns out, from the research briefly reviewed in section

6.5, that such interaction is in many cases conducive of "bad judgment", for example, if the social attitudes of the expert made her otherwise sound judgment look unworthy.

Structuring judgment then means, on one hand, limiting the effect of those factors that are irrelevant to the judgment itself and potentially conducive of negative social dynamics, such as weighting someone's judgment by factors (e.g. personality) that have nothing to do with the judgment itself. On the other hand, it means to facilitate the creation of those social dynamics that are instead conducive of better judgment. For example, constraining judgment into being a product of groups rather than individuals, or facilitating the provision of feedback from the group level back to the individuals.

It is clear that the procedure of directing expertise in the formation of judgment can only partly be decided on a-priori grounds. Some procedures will fit different purposes better than others, but such assessment is often made a-posteriori.

The foregoing sections will review two of the methods for constraining expert judgment, the *Delphi Method* and the *Nominal Group Technique*. Each review is divided in three parts: 1) I describe the procedure in its essential features; 2) I give a brief history of its formulation, uses and goals; 3) I list the positive contribution of the procedure to the formulation of group judgment.

6.6.1 The Delphi project

In the Delphi Method the administrator of the procedure formulates a problem for which she wishes to obtain forecasts or subjective judgment. The formulation should allow for either yes/no answers or a probability value in the form of point estimates or interval estimates. The problem is administered to a number of selected experts[13], with the request to provide their judgment.

Experts submit their reports (and in some cases the *reasons* for their personal judgments) to the administrator, who can choose, depending on the situation, to formulate a summary of the first round of reports. Either the summary, or the individual reports, are fed back to each of the experts with the request to consider their judgment again and resubmit their (possibly revised) judgment. The process is iterated as many times as it is feasible (iteration is time consuming and costly), or until the process stabilizes (more

[13]Selection of experts can present problems and biases too, for a treatment of the issue see Collins (2004).

iterations do not produce significant change in the judgment).

The Delphi Method was developed in the 1950s[14] and 1960s at the RAND corporation, mostly by the work of Dalkey and Helmer (Dalkey and Helmer 1963; Dalkey 1969). Extensive reviews of the method were published in Dalkey et al. (1972) and Linstone and Turoff (1975). Recent evaluations and commentaries are in Rowe and Wright (1999), Rowe and Wright (2001) and Rowe, Wright and McColl (2004).

The rationale and philosophical foundation for the use of Delphi was outlined in Helmer and Rescher (1959). There, the authors motivate the use of experts in situations in which a formal analysis is not yet possible. The sciences for which formalization of problems is particularly hard are classified as "inexact sciences" (see section 5.6 and 5.6.2), even though Helmer and Rescher deny that there is a qualitative distinction between exact and inexact sciences. Instead, their claim is that in all sciences there some are degrees of inexactness as well as domains that resist formalization more than others[15].

> In an inexact science [...] reasoning is informal; in particular, some of the terminology may, without actually impeding communication, exhibit some inherent vagueness, and reasoning may at least in part rely on reference to intuitively perceived facts or implications. Again, an inexact science rarely uses mathematical notation or employs attributes capable of exact measurement, and as a rule does not make its predictions with great precision and exactitude.
>
> (Helmer and Rescher 1959, 25-26)

The Delphi method makes use of expert opinion in those domains in which the formalization of the problem would be either impossible or misrepresentative of the problem (e.g. too simple to lead to robust conclusions about the object of investigation). The major advantages of Delphi are listed below.

- **Aggregation.** Delphi performs better than single individuals by simple aggregation of the individual judgments. "Delphi effectiveness over comparative procedures, at least in terms of judgmental accuracy, has generally been demonstrated [...]. Rowe and Wright [...] found

[14] Because the original work was done in the context of military decision making and evaluation — the project was originally developed for the US Air Force — very early presentations of the Delphi Method were protected by secrecy.

[15] A similar picture is justified by the observations made in sections 5.3 and 5.5 of this work.

that Delphi groups outperformed 'statistical' groups (which involve the aggregation of the judgments of noninteracting individuals) in 12 studies, underperformed these in two, and 'tied' in two others, while Delphi outperformed standard interacting groups in five studies, underperformed in one, and 'tied' in two." (Rowe, Wright and McColl 2004, 378) The application of Delphi, as well as that of several other methods, seems to validate the first fact that was mentioned in section 6.6, that is, the first step towards better judgment is the passage from individuals to groups.

- **Anonymity.** In the Delphi method there is no personal interaction among experts, or more precisely, no direct physical interaction. Surveys and elicitation are done via mail (in the original implementations of the method) or computer interaction (now becoming the standard method for Delphi interaction (see García-Magariño et al. 2008, 2010)). Anonymity and lack of personal interaction eliminates the several psychological biases that can arise from direct group interaction. Some bias can still occur with the provision of feedback, especially when feedback is provided on the reasons that have been given to support an expert's opinion in previous rounds.

- **Feedback.** Feedback is meant to provide experts with more information from round to round, in the hope of improving their accuracy. Experimental studies show that, in weather forecasting and engineering applications, feedback can improve an expert's calibration (Cooke 1991, 26,27, 72-76). Unfortunately, in the context of stock market prediction, calibration seems to produce no improvements and experts perform worse than non-experts. An analysis of the role of feedback in Delphi is in Rowe, Wright and McColl (2004).

- **Convergence.** Iterations of the rounds of elicitation in Delphi produce partial convergence of the views of the experts involved. This is often attributed to the effects of feedback. Convergence, together with the evidence that Delphi outperforms 'statistical groups' (see first point), points towards the conclusion that the convergence is towards the truth. In directly interacting groups, on the other hand, convergence is normally unrelated to the truth of the matter (herd behavior).

.

This list concludes the brief survey on the Delphi method. The motivations for the use of Delphi as part of the methodological apparatus of economics will be motivated in section 6.7, after a review of the *Nominal Group Technique* in the coming section.

6.6.2 Nominal Group technique

The main goal of the Nominal Group technique is not forecasting, but rather problem exploration and program management. It was argued in the sections of chapter 5.5, that one of the issues in economics is the formulation of the problem itself. Correct formulation of problems is, in science, as important as their solution, inasmuch as the conclusions derived from an overly simplified problem will likely bear little resemblance with the *real* problem.

A Nominal Group is put together, as in Delphi, by an administrator in order to explore a certain problem or to manage a certain project (the two are not exclusive). The technique divides the process of expert elicitation into *phases*, in which the experts are called to give their contribution, and in minutely specified steps for each round. The complexity of the process, and the detailed description of each step, would make their exposition here pedantic, and the interested reader can refer to Delbecq and Van de Ven (1971) for a detailed description.

In its original formulation the Nominal Group technique has five phases: *Problem Exploration*, *Knowledge Exploration*, *Priority Development*, *Program Development*, and *Program Evaluation*. The last three steps can be considered of primary interest for the business sector, whereas the first two phases are of important theoretical interests and will therefore be the focus in this section.

Problem Exploration involves a number of steps in which the experts are asked to provide feedback on the understanding of the problem itself, rather than provide solutions to previously formulated problems: "The organizational representative [. . .] then indicates that the purpose of the meeting is to understand the *problems*, not to explore solutions." (Delbecq and Van de Ven 1971, 470) The process involves both personal assessment (individuals write down their judgments on 3" x 5" cards) and interaction (individuals read aloud what they have written), but the interaction is strictly mediated by the administrator.

Once the problem has been explored and the goals have been established, the phase is over; the administrator, taking from the pool of Problem Explorers and from other pools of experts, then forms a second group, smaller than the first, in order to elicit knowledge about possible solutions of the problems selected in the first phase. *Knowledge Exploration* takes the same modalities of the previous phase, but is aimed at making the tacit knowledge present in the second group explicit.

It was said in section 6.2, that a large, yet mostly unexploited in the economic context, source of knowledge of economic phenomena, correlations and solution possibly resides in the form of tacit knowledge, gained from the diverse practices of economists as historians, mathematicians, modelers, experimenters and so on. That section, however, posed the problem of how to make such knowledge "working knowledge", that is, how to make it explicit. Nonaka (1991) poses the same problem, but seems to focus on the problem of eliciting *how-to* knowledge in large companies and organizations. Moreover, Nonaka only gives general indications for that task, without giving any particular procedural guideline.

Contrary to that, Delbecq and Van de Ven (1971) provide a precise methodology for letting experts express their subjective judgment on the problem under discussion. As said before, the Nominal Group technique is not meant for forecasting; the although limited interaction of experts in a small group would in all likelihood trigger those biases that were thought to bring the group to underperformance in the Delphi studies (e.g. dependence of the judgments, tendency to follow the strongest opinion, etc.). The main use of the technique is in "(a) identifying strategic problems and (b) developing appropriate and innovative programs to solve them" (Delbecq and Van de Ven 1971, 467).

> This research indicates that [freely] interacting groups produce a smaller number of problem dimensions, fewer high quality suggestions, and a smaller number of different kinds of solutions than groups in which members were constrained from interaction during the generation of critical problem variables.
>
> (Delbecq and Van de Ven 1971, 772)

The Nominal Group technique was initially developed by Delbecq and Van de Ven (Delbecq and Van de Ven 1971). Subsequent papers developed the technique and compared it to similar methods, including Delphi (Van De Ven and Delbecq 1974). Since then, the technique has had little theoretical development but has been, and still is, widely applied in many practical contexts whenever the assessment of the problem and the identification of solutions is a particularly complex task.

The advantages of the technique are listed below (the list is not exhaustive).

- **Group process structure.** The full justification of the superiority of a structured group over brainstorming or free interaction is given in Van De Ven and Delbecq (1971). Among the several advantages reported, are the following facts: A nominal group "avoids the dominance of group output by strong personality types" (one of the biases mentioned in section 6.5), "avoids evaluation of elaborating comments while problem dimensions are being generated", and several others (Van De Ven and Delbecq 1971, 206,207). One can notice that the advantages that Delbecq and Van de Ven report are mostly in the psychological dimension. Nonetheless, such advantages would reflect on the epistemic dimension too, if the psychological biases are conducive of partial or incomplete judgment.

- **Alternation.** The Nominal Group Technique separates the periods of personal problem and solution exploration from the periods of sharing. According to the authors, this gives the opportunity to all the members to reflect privately without disturbance such as immediate objections and comments (see previous point). The alternation of the two steps, in both the first and the second phase, can be seen as a direct feedback process similar to the Delphi one, with the only difference being that in this case the process is not anonymous.

- **Social Interaction.** A number of positive reasons for using Nominal Groups is given that makes reference to the positive *social* advantages of the technique. For example, it allows exchange between different types of expertise and even between the experts and the "clients", viz. those who are the receivers of benefits from the project under analysis. Or, also, it gives the administrators the opportunity to select the appropriate expertise at different stages of the process. Because these advantages belong to a dimension that is not discussed in this work, to wit, the proper decision-making aspect of economic sciences, where the interactions are not among economists alone but between economic sciences and the political or social sphere, these advantages should not be considered as epistemic advantages *per se*. These aspects of the technique, on the other hand, are important if the goal is to address the decision-making process in relation to the use of expertise, as for instance in (Collins 2004).

6.7 Conclusion

It was argued, in this and the previous chapter, that the *rational model* of consensus formation in economics suffers from a number of drawbacks due to the fact that the system investigated by economic sciences is a rapidly adapting system, and one where the object of analysis can respond to information coming from the analysis itself and thereby defy its predictions. The debate on whether even adaptive systems are to a certain extent predictable by the means of the scientific method, modeling and testing, is far from settled. More than trying to provide a definite answer to that point, these two chapters were meant to suggest that the sources of economic knowledge are larger than the scientific method implies.

While those conclusions are not new in the literature on economic methodology, the problem remains to provide a valid alternative method that is able to take into account all the different sources that are available to economics for understanding and manipulating the economic environment. In this chapter, it was suggested that besides modeling economic phenomena themselves, the goal of economic methodology should be to model those who were identified as the primary sources of *applicable* economic knowledge and around which economic consensus forms. Those are the *economic experts*.

On the one hand, reliance on the scientific method has the advantage of precision, replicability, communicability and perhaps other advantages as well, but that comes at the cost of flexibility, as a model is a mathematical construct that does not respond to information that was not included in the model at the time of its formulation. As it happens, however, that information is essential for understanding the volatile nature of many economic phenomena.

On the other hand, relying on expertise has the advantage of flexibility, often however at the expense of those advantages that modeling and testing offered. The first idea of section 6.6.1 was to provide an initial framework for *modeling expertise*. The reason for modeling expertise is to avoid the shortcomings of expert judgment, especially in the context of committees of experts. The second idea of this chapter was to suggest a shift from the concept of individual expertise to that of collective expertise. While the justification of the power of groups over individuals was mostly assumed throughout the chapters, the idea that a single economic expert can manage the variety of tasks that require him or her in the real world is at best illusory.

The picture given in the introduction to this thesis, the one of the economist as the apothecary drawing from his stock of medical substances, each apt for a specific illness, should be replaced by a normatively more adequate representation. Economists are mostly trained in a specific subfield of their science, almost always a highly specialized area and with a very specific methodology; that is true for all areas of economics, from econometrics to history of economics.

What the policy maker requires from economists, however, is to treat and solve problems that go well beyond the limits of an expert's area of competence. Clearly an economist is likely to have been able, by her own intelligence, to gain knowledge of the mechanism of the economic world beyond the specificity of her own domain of specialization, but it is unrealistic to think that she has been able to gain all the understanding and competence that is necessary for the task she is assigned by the policy-maker.

Figure 6.1: Pictorial representation of an expert's domains of competence.

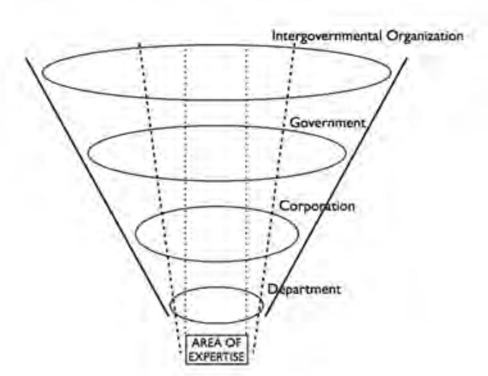


Figure 6.1 represents the idea just expressed. The training of an economist grants her some competence (dotted line in the figure) which runs across different social organizations (e.g. a corporation or a government), but which will probably not cover the full extent of that organization's need. It is likely that an expert will have been able, through her own cognitive resources, to gain a decent understanding of phenomena outside her own direct field of competence (dashed line in the figure). It is however virtually impossible that an expert's competence can reach the whole domain required at higher and higher levels of social organization (continuous line in the figure).

That latter domain, on the other hand, can easily be covered by the combination of the competence of more than one expert, even though at a price. The price is that of coordination, group biases, and the other shortcoming of *collective expertise*. The idea in this chapter was to suggest some strategies of expert elicitation for groups of economic experts. While some of these strategies have been extensively applied to other domains, like engineering and policy-making, the contribution of chapter 6 was to justify the adoption of such strategies for the field of *positive economics*, that is, for the development of that *economic toolbox* that policy makers, managers and other institutional figures need in their decision making processes.

Chapter 7

Responsibility incorporated

7.1 Introduction

The previous two chapters of this thesis concluded by supporting the idea that economic methodology should be guided by the concept of modeling expertise, in addition to the traditional scientific idea of modeling phenomena. The second idea in section 6.7 was that expertise is better exploited when it comes as *collective expertise* than when it comes as individual expertise. It must be reckoned, however, that especially in the field of applied economics, where the science has direct consequences for the policy-making and, in turn, for the wellness of those affected by the adopted policies, modeling experts goes hand in hand with the ethical principles that the science, or better, the scientist, is expected to follow.

It is a fairly accepted standard to regulate the behavior of scientists, or at least those scientists who work in particularly sensitive domains, by adopting moral standards in the form of ethical codes or even special legislative provisions that the community enforces on its members. The classical example of such code is the *Hippocratic oath* for physicians, but the use of ethical codes has spread to other areas of science as well, for example in genetics. While in the field of business ethical codes have been studied and experimented for some decades now, in economics the idea is relatively new and in part inspired by the recent financial crisis of 2007-2010 (Epstein and Carrick-Hagenbarth 2010).

When groups are involved, however, the problem of responsibility can become complicated because the concept of *group responsibility* or *collective responsibility* is still a controversial one. Nonetheless, according to some scholars, any discussion about groups, group deliberation, and group action leads to the question of responsibility or culpability from the collective viewpoint. The idea that groups can act in concert to provide analysis and evaluation of economic matters, as individuals do, makes them resemble individual agents so closely that the problem of responsibility for these "unusual agents" needs to be addressed.

The traditional idea of an agent is of something which is able to deliberate, make a decision, and act accordingly; in virtue of that, an agent is also open to blame for the actions taken or the consequences of such actions. The question that arises when talking about group agents is then whether groups are also subject to blame, when the consequences of their actions are blameworthy, or deserving of praise otherwise. The need to settle this question, according to Pettit (2007), derives from the fact that *group agency entails group responsibility.*

The conditional 'group agency $\supset$ group responsibility' implies that if a group can deliberate, make a decision, and act accordingly — that is, to most accounts, be considered akin to an agent — then it must be the case that that group is also liable or praiseworthy for the consequences of its actions. In this final chapter I will analyze Pettit and List's account of group agency, group responsibility, and the conditions for both. I will argue that their account falls short of providing sufficient conditions for responsibility, given the recognition of a group as an agent. In addition, I will suggest an extension of their account, which makes use of the same tools for group deliberation that were presented at the end of chapter 6.

7.2 Committees and moral responsibility

The concept of responsibility and accountability has become almost a technical term as applied more and more frequently to corporate bodies. Some argue that corporate bodies such as governments, corporations, decision-making and even advising committees should be held responsible for their faults, and, correspondingly, deemed praiseworthy for their merits. Among the several accounts that have been given on the topic, some argue that corporate responsibility is a practical necessity (Coleman 1990), in order to prevent corporate powers from doing harm in pursuit of their interest and pressure them into producing welfare.

Others instead, argue that the idea of corporate responsibility is not justified on pragmatic reasons, but rather that it comes from the analysis of the design of the group agents themselves, which shows them to be akin to individual agents, and thus open to the same (or similar) moral standards.

> No incorporated agency, without incorporated responsibility.

> [...] This, in a slogan, is the line I defend in this article.[1]
>
> (Pettit 2007, 172)

According to this stance, individual agents are analogous to collective ones, such that the same criteria for holding the former responsible apply to the latter as well. Pettit (2004) and Pettit (2007) defend the argument that corporate agents should be considered autonomous agents in their own right, not reducible to the individuals composing them. Pettit (2007) and List and Pettit (2011) develop on the same idea, identifying a set of three features, found both in individual and in corporate agents, which justify an attribution of moral responsibility. The aforementioned texts will be the main references throughout this chapter. I will assume Pettit and List's stance on corporate agency and corporate responsibility, and develop on their premises. Henceforth, I will refer to their account as the *entailment account*, since, according to it, it is true that '$agency \supset responsibility$'[2].

In the following sections I will argue that Pettit and List's account on corporate responsibility does not fully capture the extent of conditions that seem necessary to attribute moral responsibility to a group agent. In other words, the slogan reported above does not hold in general: There can be incorporated agency without incorporated responsibility, when the incorporated agents are "ill-designed"; more on the meaning of that will be said in the next sections. In addition to the three conditions that the entailment account requires, I make the additional point that one needs to say more about the design of corporate agents, and in particular about their deliberative structure, in order to attribute them responsibility. I will draw from the technical literature on expert deliberation in order to defend the thesis that some *deliberative procedures* are more apt than others for satisfying the conditions for moral responsibility given in Pettit (2007) and List and Pettit (2011).

7.3 Philip Pettit: Group agency and group responsibility

The term 'collective responsibility', or the equivalent 'corporate responsibility', refers to the attribution of moral blame and moral responsibility

[1]For more references where a similar position is defended, see Pettit (2007, 172), footnote 3.

[2]Pettit's "slogan" (above) gives the contrapositive of my statement, that is '$\sim responsibility \supset \sim agency$', the latter implies my '$agency \supset responsibility$'; the proof is trivial.

to so-called group agents or corporate agents. Corporate agents are collectivities of individuals that are so formed and designed as to possess the characteristics that are normally attributed to "personal agents" (e.g. human beings).

List and Pettit (2011) list a number of salient features which belong to agents. An agent i) "has representational states that depict how things are in the environment", ii) "has motivational states that specify how it requires things to be in the environment", and iii) "has the capacity to process its representational and motivational states, leading it to intervene in the environment whenever that environment fails to match a motivating specification." (List and Pettit 2011, 8)

In other words, an agent can form internal states of how the world is configured, has a certain "plan" for how the world should be, and, when the plan and the state of the world do not match, is able to conjure strategies for manipulating the world so to make it match with the plan. According to Pettit and List, those are minimal conditions for agency. In the same work, as well as in Pettit (2007), the authors give three conditions under which an agent can be considered morally responsible for an action.

Pettit writes that an account of moral responsibility reflects the doctrine of the Christian catechisms, whereby a deed constitutes serious sin if and only if there has been "grave matter [. . .], full knowledge of the guilt, and full consent of the will." (Pettit 2007, 174) The full explication of the three conditions is given in the following[3]:

1. **Value relevance.** The group is an autonomous agent that faces a significant choice between doing something good or bad or right or wrong.
2. **Value judgment.** The group has the understanding and the access to evidence required for making judgments about the relative value of such options.
3. **Value sensitivity.** The group has the control required for being able to choose between the options on the basis of its judgments about their respective value.

(Pettit 2007, 175)

[3]It must be noted that the concept of responsibility comes as *responsibility for* and *responsibility to*. We are responsible *for something*, but responsibility is also always *directed to* something or most often someone, which can be a person, a group of people, an institution and so on. In this chapter, as well as in Pettit and List's work, the concept of responsibility is only analyzed in terms of *responsibility for*.

7.4 A caveat on Pettit's account

Among the conditions that Pettit (2007) and List and Pettit (2011) list, in order for a corporate agent to be held responsible, condition (2) is particularly important, since it has to do not only with the group's possession of a specific characteristic, namely whether this "constitutes an autonomous agent" (first condition), but rather with the precise way in which that characteristic is implemented in the structure of the group. The following considerations will clarify this point.

To iterate, condition (2) claims that a group is fit to be held responsible if it "has the understanding and the access to evidence required for making judgments about the relative value of [the moral option it faces]" (Pettit 2007, 177). A little further in the text, Pettit expands on that condition, explaining how the group might gain understanding of the situation it faces and access the evidence that will guide its moral decision. Understanding and gathering of evidence, according to Pettit, might be done for example by means of "a vote in the committee of the whole, a vote in an authorized subgroup, or the determination of an appointed official", in general by means of some voting mechanism: "by means of a vote or something of the kind." (Pettit 2007, 187)

Because the mental states attributed to the group agents are dependent on those of the individual agents composing it, the faculty of *understanding* and *access to evidence* must be attributed in the first place to individuals. In most groups individuals will most likely have access to different collections of evidence and almost certainly the understanding of that evidence, be it completely or just partially shared, will differ among them. This situation is a typical one and, with a jargon that has become technical in the philosophical literature, the group is said to be starting deliberation from a *situation of disagreement*.

Solving disagreement therefore becomes the task the group is faced with, when it is to assess a certain issue and to form the beliefs of the group as a whole. If a group's beliefs were simply the summation of individual beliefs, then, under the initial conditions, viz. condition of disagreement, the group would hold inconsistent sets of beliefs. Persistence of disagreement would halt most functions of the group as an agent; for example, the capacity to act according to a rational and coherent strategy. Pettit suggests that deliberation and voting, or some similar aggregation procedure, can resolve initial disagreement and thereby provide the initial requirements for the fulfillment of condition (2).

While disagreement can in principle be resolved by deliberation and voting, it is not clear that this alone can satisfy the requirements of condition

(2). In the following sections I will provide some examples where deliberation and voting are not enough to make the epistemic group agent in question also a morally responsible agent. In particular, some ill-designed group structures, even though they may fulfill the conditions for being a group agent, may well fail to satisfy the responsibility condition as given by Pettit and List.

7.5 An ideal example

Consider the following situation: A decision maker seeks recourse to an especially appointed committee to provide evaluation and analysis of the facts and problems concerning the subject matter that is object of deliberation. Such committee is diversely composed, with experts from the various disciplines involved. The committee produces research, discusses, and is aware of the fact that a final and coherent message must in the end be delivered to the policy maker.

It is expectable that, at the beginning of its works, the committee will be divided on several sub-issues pertaining to the matter that is under deliberation; at least some of those issues will likely be logically interconnected so that the conclusion (or conclusions) will depend on the premises or some subsets of them[4]. However, in this example, the group does not use a decision structure on "one level" only, as it is the case in most of Pettit's deliberative examples where the members deliberate and vote communally, and the conclusion is reported directly as the outcome of the group deliberation.

Instead, in the imaginary committee considered here, the structure of the group is multilayered, the bottom levels reporting to the ones further up in the pyramid of power, and the top reporting to the policy maker. The group can be thus said to be a "hierarchical" one, or "multilayered" (the terms will be used interchangeably here). Such composition of the group is intuitively a realistic one (more on this in section 7.7), and the first question to answer is whether a so formed committee could fulfill the conditions for agency. It will be useful here to recall the conditions for agency given in List and Pettit (2011), whereby an agent should posses all of the following three features.

- **First feature.** [An agent] has representational states that depict how things are in the environment.

[4] As in Pettit's *Deliberative Dilemma* cases, the opinion of the group can thus end up being significantly independent from the opinion of its members (see Pettit 2004); in the same text, see also the justification for the thesis that the attitudes (beliefs, preferences, etc.) of groups supervene on the mere aggregate of the attitudes of their members.

- **Second feature.** [An agent] has motivational states that specify how it requires things to be in the environment.
- **Third feature.** [An agent] has the capacity to process its representational and motivational states, leading it to intervene in the environment whenever that environment fails to match a motivating specification.

(List and Pettit 2011, 8)

In order to see whether a group with a hierarchical structure, as the one introduced in this section, can be an epistemic agent of the type Pettit and List have in mind, it will be useful to think of what difference such structure could possibly make, for the fulfillment of the three features, from a non-hierarchical one.

In the first place, a hierarchical group should be able to have representational states and motivational states just as a the deliberative groups depicted in List and Pettit (2011). For that, it should be noted that the representations of a layered group are just the same as those of Pettit and List's groups: they are grounded on the individual psychological representations, and possibly transmitted from agent to agent across the group. As in Pettit and List, there is no ontological claim here as to what exactly such representations may be (see List and Pettit 2011, 10); all definitions are functional ones, that is, any object that can *serve as* a representation is considered as one.

Similar remarks can be made for the presence of motivational states, although a caveat is due here. It may be argued that, in a hierarchical structure, the bottom layers of a group have different motivational states from those of the top layers because at different levels of a hierarchy the motivations guiding action might change.

To respond to that, it should be noted that the problem does not concern multilayered structures alone: A certain subset of individuals, in one of the groups that are exemplary in Pettit and List, could have motivations which differ from the ones of another subset of that same group. If that were the case, one could argue whether the agent has coherent motivational states at all, and therefore whether it can be taken as an agent in the first place. But the problem is not specific of the type of hierarchical groups I introduced above. Therefore, provided that one can identify a group's motivational states, the particular structure (hierarchical or not) should not make a difference.

The third feature of group agents (see above), ought to guarantee that groups can act in the environment surrounding them in order to match their representations with their goals (motivational states). A hierarchical group might function differently from other types of groups, as the "executive power" is not shared homogeneously. As in the comments to the second feature, however, the same is true for groups where the composition is not multilayered, and this particular feature should not be problematic for the fulfillment of the agency conditions.

The foregoing comments were not meant to provide a definite statement on the possibility for hierarchical groups to perform as agents. The claim that no hierarchical structures can be agents would have to find an essential feature of such structures that is incompatible with agents in the sense of Pettit (2004), Pettit (2007), and List and Pettit (2011). The burden of proof in that case is on the opposer of the view defended here. The comments above were instead meant to show that there are no *obvious* features that make a hierarchical structure incompatible with group agency. That should have served the purpose. In the next sections I will proceed with the argument stated at the beginning of this section, that some hierarchical structures, though meeting the demands for agency, might fail to meet the demands for moral responsibility.

7.6 The responsibility requirement

The pyramidal structure of the group described in section 7.5 implies that the message informing the policy maker (the transmission of which can be considered the *action* that the group, as an agent, performs) is delivered to the policy makers by the top layer alone. While it is not possible to claim that the message formed at the top is independent of the rest of the group, because it is the *product* of the group agent as a whole, I will argue that a multilayered group does not meet condition (2) for moral responsibility.

The hierarchical structure described above is ill-designated because the group does *not* have "understanding and access to the evidence required for making judgments"[5] (Pettit 2007, 175). To be more precise, access to evidence is different at each layer, unless there is a mechanism of transmission for evidence across the layers. This latter condition, however, is not guaranteed by either a voting or a deliberation procedure, as I will argue in the rest of this section and in the following ones.

In general, while the hierarchical group described above satisfies the conditions for agency, its composition and chain-of-command do not make it

[5]It can be argued that the group does not have "the control required for being able to choose between the options" (Pettit 2007, 175) either, but this problem will not be addressed here.

able to function as a *responsible* agent, that is, one which evaluates evidence, forms judgments, and chooses among a number of options. Whereas in groups with one layer it is easier to argue that the evidence has been transmitted in the deliberative phase, the options have been laid out in front of the members, and the group has thus acted in communion of its members, that is not the case for the type of groups discussed in this and the previous sections.

To stress the previous point, it seems possible to imagine cases in which a certain judgment or action is formed *as the product of the group* (and specifically what we would consider a group agent), and yet the structure of that group is such that the ultimate responsibility for its judgment or action cannot be ascribed to the group as a whole. Suppose, for instance, that the group is compartmentalized (as many real world groups and institutions are), and that different layers of the group have different agendas. If so, the group would still fulfill the conditions for agency without fulfilling those for responsibility.

The reason for that is that in a compartmentalized group, only the agendas of the top layers are the basis on which decisions are taken (or policy suggested); and that is because the understanding of the goals of the group, the ability to access all evidence, and the control on the information transmitted, all belong only to the top layer (or layers). Nonetheless, the group can still be called a group agent because it fulfills the conditions for group agency outlined in section 7.3 (conditions i, ii, iii).

The consideration from the fictional scenario described in this section suggest that not all decision making configurations make a group fit to be held responsible, even when intuitively we would claim that the group is an autonomous agent. Those considerations, however, are at this stage still hypothetical. Evaluation of whether a group satisfies any of the conditions above cannot be abstracted from the empirical analysis of the specific group in question. One can advocate the *possible* existence of a group which satisfies conditions for agency and fails to meet conditions for responsibility, but while this might be a *logical* problem for Pettit and List's account, it is not at the logical level that such argument is an interesting one. For those reasons, the next section will give some concrete historical examples where it is possible to see the difference between group agents that can and cannot be also held morally responsible for their actions.

7.7 Historical examples

In this section, without pretense of historical accuracy, I will give two concrete scenarios where the considerations of the foregoing sections find application. The first example is one of an ill-designed group structure, where, despite the group being an agent according to Pettit and List's conditions for group agency, it cannot be claimed that the group was responsible for the decision taken as a consequence of its deliberation. The second example will serve as a contrast to the first one, and will present a case in which the group agent in question fulfills both the conditions for agency and those for moral responsibility.

7.7.1 The decision to use the atomic bomb

The decision to use atomic bombs on the cities of Hiroshima and Nagasaki, as the very final act of WW2, was preceded by a long and intricate evaluation & decision process by the so called *Interim Committee*. The committee was especially set up to analyze the costs and opportunities of using nuclear weapons against Japan, and to formulate advice on the issue directly to president Truman. A look at one of the major historical sources on that decision process (see Alperovitz 1996) shows that the decision group was not homogenous. In particular, there were at least three layers that clearly influenced Truman's final decision: a group of four scientists from the *Manhattan Project* (not officially members of the Interim Committee but providers of advice to it), the Interim Committee itself, and a member of the committee, James F. Byrnes, who had direct influence and access to Truman's inner circle of advisors.

The structure of the interim committee was a hierarchical one. The reasons why Byrnes was not at the same level of the other members of the committee are thoroughly explored in Alperovitz (1996), where it is argued that any piece of advice coming from the committee was in large part influenced by its most influential member, who had direct access to the president. The analysis of the decision making process that led to the use of the atomic bomb, points to the conclusion that such decision was not the product of one man (which in that case would have likely been Byrnes himself), because no such decision could have been taken without the intense discussion and reporting from the scientific advisors and the other members of the committee. The group, in this sense, can be considered an agent because it was operating as one, fulfilling conditions (1) to (3) in section 7.2.

Nonetheless, it seems clear that the final opinion had been largely influenced by Byrnes' agenda, and the agenda of the political group it represented (distinct from the Interim Committee), which had independent reasons, other than the termination of the conflict with Japan, for using atomic bombs on Hiroshima and Nagasaki. Alperovitz (1996) attributes the 'historical responsibility' to Byrnes and the political agenda he represented for the information provided to Truman, which led to the final decision on the use of the bomb, even though the support of information and analysis that led to the decision was the product of the Interim Committee as a group agent, in Pettit and List's sense.

Alperovitz's analysis, like any historical analysis, is in principle open to objections and criticisms, but that does not undermine the claims made in this section. *If* the analysis in Alperovitz (1996) is correct, *then* the case is one in which a group (the Interim Committee) is an agent, yet the responsibility of the actions it produced (namely, the formulation and transmission of information to the president) cannot truly be attributed to the group as a whole.

The Interim Committee, while fulfilling the conditions for agency, was not able, as a group, to access all evidence. In particular, it could not evaluate the evidence on the basis of the agenda that only Byrnes (among the components of the group) had in mind[6]. This is because there was no sharing of information from the top layers down, whereas there clearly was a lot of communication from the bottom layers up.

Most members of the Interim Committee, then, even though they were in fact acting as a group together with Byrnes, the "prominent" member, could not be deemed responsible for selecting and filtering the information that was passed on to the decision makers, and ultimately the president. It is in that sense that the Interim Committee, while on one hand acting as a group agent, is not responsible for the decision taken in the case of the atomic bomb.

The next example will serve as a contrast to the example of the Interim Committee. The claim I will defend is that in the next example the decision process was structurally different from the case analyzed above, and was such that the moral responsibility of the action taken should be rightfully placed on the group as a whole.

7.7.2 The UN Security Council and Resolution 1441

In 2001 intense pressure from United States and the George W. Bush administration opened the case for a possible military intervention in Iraq,

[6]Byrnes's agenda, according to Alperovitz (1996), included the display of extraordinary force to the world and especially the Russians, as the signs of a possible incumbent East-West conflict were already in sight before the end of the war.

in order to eliminate Saddam Hussein's regime. The case was brought to the UN Security Council, backed by the Bush administration which built a case against the regime, claiming it was developing WMDs (weapons of mass destruction).

Regardless of the truthfulness of those allegations, it is important to understand how the case was built, and what the decision making process involved. The case against Iraq culminated with the unanimous support of the UN Security Council to *Resolution 1441*, which gave the Iraqi regime an ultimatum for compliance with previous resolutions mandating inspections of Iraqi military sites and disposal of all WMDs. The aftermath of Resolution 1441 is known, as well as the allegations that the USA and the UK overrode the terms of the resolution, waging war on Iraq upon its alleged non-compliance with the terms of the ultimatum.

In this case, the question I want to highlight is not about the responsibility for the conflict itself, but rather whether the Council was responsible for passing the resolution, which gave the US and the UK an opportunity for waging war under the UN umbrella. Like in the case of the decision to use the atomic bomb, the group that was delegated to discuss and vote on the resolution seems to fulfill the conditions for qualifying as an autonomous agent. The UN Security Council is a highly structured group with precise decision making procedures, it deliberates and votes on issues whose logical structure may well open the group to deliberative dilemmas of the type that Pettit uses to make his case for the autonomy of certain group agents.

Unlike the Interim Committee, however, the Council is not organized in a multi-layered structure. Concerning Resolution 1441, all the Council's members had access to the evidence, and the deliberation process and final resolution were conducted "communally", that is, with all the members accessing the same evidence, and deliberating over the same official agendas. To be sure, this is not to say that the evidence presented by the US was "good evidence". This is also not to say that there were not hidden agendas guiding the individual votes by each member of the council. But clearly the deliberation, voting and decision making processes for passing Resolution 1441 was transversal and transparent. If the works of the Council were made in light of wrong or constructed evidence, or guided by hidden agendas, that would not be a problem for the claim that the responsibility of Resolution 1441 falls, historically, on the council itself and not on some or only one of its members.

To be even more precise, the case may be compared to a jury deliberating and voting over a legal case. The evidence may be tainted and the accusations may be ill-motivated, but the jury is responsible for the verdict itself, at least according to the conditions for responsibility formulated by Pettit and List. The responsibility for tainting evidence, withholding it, or similar morally condemnable actions, should be independent from the responsibility for the decision itself. In the UN Resolution 1441 case, given the evidence and the deliberation that occurred, it was the Security Council that took the decision and that should be responsible for it. On the other hand, for the Interim Committee case, had the committee had a homogeneous and transparent structure, one can easily argue from the analysis in Alperovitz (1996) that the conclusions it would have arrived at, and which were transmitted to Truman, would have been different.

7.7.3 Preliminary conclusions

To conclude section 7.7, it is evident that the claim according to which there can be no incorporated agency without incorporated responsibility (see Pettit 2007, 172) is in need of specification. The claim has a rationale, but perhaps only in those cases in which the group agent gains autonomy in virtue of specific features that make it fulfill conditions (2) in "a proper way", in addition to the fulfillment of condition (1) and (3). In particular, transmission of information and transversal access to evidence across the group are required.

In List and Pettit (2011), the authors envisage the possibility of group agents formation "without joint intention": "Individuals may combine into group agents in two very different ways, depending on whether or not joint intention is involved." (see List and Pettit 2011, 24). The claim, motivated with examples in this section, is that whenever a group forms without joint intention, and this may be due, like in the case of the Interim Committee, to the lack of a transparent decision procedure, then the group may fulfill condition (1) on group agency yet fail to fulfill condition (2).

Despite the above criticisms, I think that the conditions for moral responsibility in Pettit (2007) and List and Pettit (2011) should not be rejected altogether, nor should it be rejected the claim that incorporated agency implies incorporated responsibility. The conditions seem justified by strong independent reasons which make it hard to reject them as a whole[7], and the latter claim seems to capture a notion that is both intuitive and pragmatically useful. It is intuitive because it seems natural to link agency with responsibility, and it is pragmatically useful in virtue of the considerations the two authors make, about the *opportunity* of considering agents

[7] The details are given in full in the two texts referred to throughout this work (Pettit 2007; List and Pettit 2011).

responsible for their actions[8].

For the reasons just stated, the *entailment account* (see section 7.2) should not to be scrapped. In the following sections, I will illustrate how the problem of the composition of the group, and the possibility that such group may not fulfill all conditions for moral responsibility, can be resolved leaving substantially intact all of Pettit's conditions, but adding some important appendices to condition (2).

7.8 Decision making in group agents

The question to be answered in this and the foregoing sections is which deliberation procedures are more apt to meet the demands that condition (2) imposes on groups, in order for them to be open to moral blame or praise. To recall, the main desiderata arising from that condition are the following:

a) A group agent should be able to understand value judgments about options (Pettit 2007, 185).

b) A group agent should be able to access evidence on the relative value of the options it faces in a certain choice (Pettit 2007, 185).

7.8.1 Deliberation, voting and condition (2)

It should be clear from the examples given in the previous sections that some group structures are not fit to meet the demands of condition (2). The problem, however, runs deeper than those two examples. We can ask whether an institution that uses, for example, majority voting is able to fulfill those demands. Or else, we can ask whether a group that deliberates with an open discussion on the issues it faces, before voting, is so able.

Both voting and deliberation are open to failure, with respect to the fulfillment of condition (2). In voting, evidence is analyzed privately and only the conclusions drawn from a private assessment of the problem are expressed in the vote. That, however, seems to contrast with the natural interpretation of point (b) at the beginning of this section; if evidence is a purely private matter, it is not the group that is "able to assess evidence on the relative . . . ", as point (b) requires. The natural interpretation of

[8] Such reasons are explained as reasons of *deterrence* and reasons of *development* (see Pettit 2007, 175-176). More precisely, agents (group agents included) are and *ought to* be held responsible both in order to prevent them from acting regardless of the consequences of their actions, and in order to educate them to the principles of "self-awareness" and "self-regulation".

condition (b) seems to be that the group be able to assess evidence *as a group*, albeit perhaps only imperfectly.

Moreover, it does not seem that with voting alone the group as such is able to understand the value judgments about the options it faces. As before, if such understanding is a purely private matter, it is not the group itself that is in possession of such understanding. Some individuals of the group may have in mind a number of moral options and understand them in a certain (private) way. Others might have access to other options, or the same options but with a different understanding of them. The preparation of the agenda itself, on which the members of the group are called to vote, seems to be an essential part of the fulfillment of desiderata (a) and (b) above.

Similar considerations can be made for open discussion and deliberation. Part of the problems with deliberation were already presented in the discussion of the Interim Committee example above. In this context, I should add that, in general, the model of unstructured deliberation is open to a large number of biases and group dynamics which can in principle, and in practice, make the group unable to satisfy conditions (a) and (b) above.

Free deliberation does not guarantee that the evidence is shared equally by the members of the group, nor that an effort is made to make sure that the members are aligned on the understanding of such options. While complete sharing and alignment might seem a very ideal situation, more can be done in structured deliberation towards the fulfillment of those goals. The contrast here is between free and unstructured deliberation and voting, and a number of mechanisms that committees can put in place to structure the deliberation and voting process. The next section will be devoted to explaining the idea of *structured deliberation*.

7.8.2 Structured deliberation and condition (2)

In this section I will present a framework for deliberation and voting which, although perhaps still imperfectly, goes towards the direction of fulfilling condition (2) and its sub-conditions (a) and (b). The underlying idea is to regulate the processes of deliberation and subsequent voting in such a way that, in the first place the members of the group are given the opportunity to align themselves on a common understanding of the values at stake, and secondly, that the initial information is shared among the members.

The following considerations are taken from a vast literature on expert deliberation and group management. In the 1960s and 70s two prominent methods were developed in order to resolve some of the issues that committees and expert groups ran into, when open and free discussion and subsequent voting was adopted. Those methods are the *Delphi* and the *Nominal Group Technique* (NGT, henceforth). The problems they tried to tackle were mostly group biases[9] arising from the dynamics that free and unstructured group discussion generated.

The goal of structured deliberation is to organize the process with which members of a group discuss the items in their agenda, and come to a consensual stance that will be the group's stance.

- **Feature 1.** Structured deliberative methods allow the members an independent assessment of the evidence, free (at least ideally) from biases and group dynamics.

Structured methods ought to maximize the contribution of individuals to the group analysis and evaluation of the items in the agenda. For example, giving members time to analyze the issues in private should avoid that the influence of group pressure take over the individuals psychologies. The risk of that happening is recognized in Pettit (2004), where the author wants to exclude the possibility that the group take over the individual agents in the context of "how natural and institutional persons relate to one another within the psychology of a given member."(Pettit 2004, 189) According to Pettit, a model by which the institution takes over the natural person (the individual agents) "is clearly crazy, suggesting that persons take over psychologies in the way demons are said to assume possession of souls."(Pettit 2004, 189).

Nonetheless, the possibility is not excluded from being realized, if unstructured discussion is allowed in the group, with no constraints, for example, on how much influence a specific member is allowed to exert on the group, or how personal evaluation and public disclosure should be separated in time, and so on. Conversely, structured discussion makes sure that the contribution of the individual is not overshadowed by the influence of the group or its most fervent members. Epistemically, the combination of independent assessments in a committee is normally valued more highly than when the judgments are dependent on one another, or conditioned by peer pressure. Similarly, a biased deliberation structure will put the weight of responsibility more on one individual (or a number of them) rather than the group as a whole.

[9]Here 'biases' should not be understood in the narrow sense in which the literature on decision making understands it; according to that literature, biases are mostly failures in judgment (see Kahneman 1982; Cooke 1991). The term 'bias' is used here in a less technical sense, for example, if a certain piece of evidence is available to only a member of the group, the group is biased because its judgment is based on incomplete information, and thus epistemically weaker than if the group had access to more information.

- **Feature 2.** Structured deliberative methods allow the evidence produced at the individual level to be transmitted across all the members of the group.

In unstructured deliberation, and when the group is not homogeneous (as in the Interim Committee case), information is not guaranteed to pass from member to member, and across different layers of the group. As discussed in section 7.8.1, voting is a clear example of that phenomenon, where all that is passed to the group are the result of the vote, not the reasons, which are intuitively the main carrier of evidential force for the group. Similarly, in hierarchical groups, deliberation can be filtered in such a way that it only selectively passes up to the top or to whichever subsection of the group is designated to take action.

While the requirement to "filter" information is a necessary one when the total amount of information is too large, and cannot possibly pass through every individual member of the group, the *mode* in which such filtering is done bears on the consequent evaluation of the moral responsibilities that should be attached to the group. If the information is filtered on the basis of reasons and agendas external to the those of the group (as it was the case of James F. Byrnes and the Interim Committee), the group, according to what was said in section 7.6, cannot be held responsible for the use that is made of that information. To recall, this is because the group has not fulfilled condition (b).

In order to avoid the aforementioned problems, structured deliberation as in Delphi and NGT promotes the exchange and transmission of information. In addition, independent institutions (e.g. moderators) are put in place when filtering is necessary. For example, in Delphi the moderator can, if she sees fit, decide not to transmit in full the reasons provided by individual members in support of their assessment, and provide a summary instead. In NGT, where the group can be split into layers to form a hierarchy, it is nonetheless made sure by the structure of the deliberative process that information is transmitted across the layers (in the technical terminology of NGT called "phases" (Delbecq and Van de Ven 1971, 469[10])).

7.9 Conclusion

It is clear that the conditions provided in the previous section still run the risk of producing ill-designed deliberative groups, for which, while the deliberation can be said to be a product of the group, the responsibility for

[10]The passages from phase to phase are explained throughout the paper.

the actions taken cannot be attributed to the group as a whole. In other words, it would be too bold to claim that the issues related to the attribution of corporate responsibility are resolved entirely by structured deliberation, in addition to Pettit and List's conditions for moral responsibility. Like condition (2), structured deliberation can also fall short of providing sufficient requirements to avoid the type of situations exemplified by the Interim Committee example. The attempt in this chapter, however, was not to provide sufficient conditions, but rather a number of conditions that are necessary in order to hold a group agent accountable for its actions.

Extensions of the methods presented in this chapter are highly desirable. In the first place, better specifications of how structured deliberation is to be carried out are partly available in the literature, and partly in need of extension. Due to the largely empirical nature of such deliberative prescriptions, different specifications will serve different purposes and fit a particular situation better than others. This work is mostly concerned with the *technical* development of deliberative methods, based on principles and observations from the psychological, philosophical and decision making literature.

Secondly, it is clear that a full specification of how a group should deliberate will also be dependent on political and social factors. This however, seems to be a task for a different approach to the problem, which is not the one that was taken in this chapter. The Delphi method and the Nominal Group Technique were originally developed for technology forecasting and policy or industrial planning, but their potential goes beyond the original intentions. Such potential has in great part yet to be explored. Of Delphi, Eto (2003) writes that "it is also suitable if there is the (political) attempt to involve many persons in [decision making] processes" (Cuhls 2003, 97, quoted from Eto (2003)) and it is clear that similar potentials exist for NGT as well. It is the task of further exploratory work to develop those and other techniques in order to make them suitable for the different contexts in which group deliberation is needed.

Chapter 8

Conclusion

This thesis started off by posing three queries about disagreement in small committees, about the process of resolving such disagreement and reaching a consensus, and about the assessment of responsibility for the work of such committees.

Query 1 dealt with the problem of whether we can reach a consensus, or at least rationally change our beliefs and move closer to the disagreeing parties' opinions, by recognizing a situation of disagreement and its *rational implications*. The conclusion to that question was negative. Both Bayesian and linear methods are not *rational* methods for updating on disagreement because they do not provide an acceptable procedure for updating in the light of disagreement.

A more tentative conclusion is that disagreement does not constitute evidence in itself, as instead part of the literature seems to claim. As stated in chapter 3, the fact that neither linear nor Bayesian methods provide a rational procedure for changing one's beliefs in a situation of disagreement gives quite some ground for the claim that disagreement is not evidence, even though other methods for updating on disagreement could be, in principle, available.

The arguments given in the first half of the thesis bear on much of the literature on disagreement and consensus. In particular, if the arguments are correct, they show that disagreement is not resolved into a consensus by the formal methods presented in the first part of the thesis. More often, one must consider such formal solutions as compromises, that is as bargaining solutions, where we agree *to do* something (taking a statement for true), without agreeing on what the content of the compromise is. While reaching a consensus would constitute a net improvement of the group over the initial situation of disagreement, I argued here that such improvements do not occur as frequently as several consensus models advocate.

To conclude about chapters 2 and 3, I argued that resolving resilient disagreement cannot result in a consensus, at least not in the internalist sense of 'consensus' used in those chapters. An independent reason for such claim is the intuition that all I come to know when I discover I am disagreeing with someone on a factual matter is that either I or the other person must be wrong. That alone, however, does not tell me anything about the possible *reasons* as to why I may be wrong. If the disagreement is *irreducible*, that is, it cannot be resolved by further analysis of the problem in question, then rationality requires one to hold on to her own beliefs.

The idea of consensus in science does not imply the fact that all the scientists have internalized and agreed upon the truth of the statements that make up a certain consensus. In this sense, what is called a "scientific consensus" in the literature can be the product of compromise, negotiation, and only under special circumstances a truly consensual resolution.

Addressing *query 2*, I investigated the issue of how consensus, now in the more liberal sense of "convergence" postulated in chapter 4, arises in a specific context; for example the context of science, or a specific science. As claimed in the introduction, here the choice had to fall, in part arbitrarily, on some subject, and in chapters 5 and 6 the choice was to zoom in on the field of economics. Specifically, *query 2* was a normative question, and the second part of the thesis investigates the problem of how we should evaluate the acceptability of a certain consensus in economics.

Chapters 5 and 6 served the purpose of justifying reliance on economic experts over reliance on the scientific ideal of a "science of economics" as expressed in Friedman's concept of "positive economics". While I did not challenge the point that there should, and can, be a science of economics, in Friedman's sense, I also argued that positive economics, as modeling and testing of hypotheses, does not exhaust the domain of theoretical contribution to understanding economic problems and manipulating the economic world.

On the one hand, there are cases in which positive economics simply fails to give the right answer to an economic problem, for example because the formulation of a certain model influences the environment itself that the model purports to study. On the other hand, there are situations in which it is in principle possible to build models of a certain phenomenon, but the speed at which the system changes is so fast that resorting to expertise can be practically more convenient, as experts and tacit knowledge tend to respond more quickly to changes in the environment.

Chapters 5 and 6 defended the idea that expertise is still superior to the utilization of mechanical methods, such as statistical analysis and modeling, at least for those problems illustrated in sections 5.5.1, 5.5.2, and 5.5.3 of chapter 5. That conclusion, however, raised a number of issues related to the shortcomings (biases) of expert judgment. Presence of biases and, in the case of groups of experts, problems of aggregation, are the major critiques normally brought against advocates of expertise over mechanical methods. For that reason, it was argued that, when defending the use of expertise, one needs to say something about the possible resolution of biases and mistakes of judgment.

Indeed, in chapter 6 I put forth the idea of "modeling expertise", by means of the Delphi Method and the Nominal Group Technique. While those two methods are only a partial solution to the problems affecting expertise (in particular small committees of experts), they are a step forward into thinking about the various issues that affect expert judgment, on which economic consensus is (and should) be based. The natural extension of those two proposals is a normative theory of consensus formation in economics, one that takes into account a possibly broader range of problems, and provides a number of rational decision-making procedures with which economic experts can collectively develop the "economic toolbox" that economics should furnish policy makers with.

Pretending to provide a short answer to *query 2* is unrealistic, due to the complexity of the problems that the query raises. Nonetheless, the goal of chapters 5 and 6 was to indicate that any answer to that query will have to look not only into the scientific method narrowly intended — application of theory and verification, or modeling and testing — but more broadly into other possible sources of evidence, such as expert judgment, and how to bring those sources into a methodology of economics. A possible direction as to how to bring expert judgment into such a methodology was sketched, albeit certainly only imperfectly, at the end of chapter 6.

It was observed, at the very beginning of the final chapter, that the idea of taking expertise and expert judgment into a methodology of economics brings up the problem of responsibility of the experts themselves. While the issue of individual expert responsibility is well covered in the literature, the same cannot be said for the concept of collective expert responsibility. But if one is to defend the use of expert committees in economics, one should also say something about the responsibility that is attached to the actions of such committees.

The latter, at least, is the position taken in List and Pettit (2011), who argue that any group possessing the features of a group agent should also be held accountable for the actions it takes as an agent. In the final part of the thesis, I accepted List and Pettit's account of group agency, but argued that their account falls short of providing sufficient reasons for group responsibility. Furthermore, I extended their account by means of those same methods (Delphi and Nominal Groups) suggested in the preceding chapter in order to model expertise.

Appendix A

Lehrer-Wagner: the extended model

In this appendix I present an extension of the Lehrer-Wagner model presented in chapter 2; the extended version is in Lehrer and Wagner (1981, chapter 4). The difference between the version of the model given in chapter 2 and the extended version is that in the latter the matrix of weights is not identical for all iterations of W^k. Instead, the extended model is allowed to contain different matrices at each step of the iteration process. Consensus, in this model, obtains when function A.1 converges.

$$P_C = W_1 \cdot W_2 \cdot \dots \cdot W_n \cdot P \tag{A.1}$$

In A.1, the number n is the number of the nth matrix that is needed for convergence, and at each step W_α, new weights w_{ij} are gathered among the agents of the model. Clearly, the conditions for convergence of the extended model are much stronger than for the original version; details are omitted here and the interested reader can find them in Lehrer and Wagner (1981, chapter 8: *Convergence to Consensus: the Extended Model*).

The underlying motivation for using different matrices at each step of the iteration process is that the original model with a constant matrix W, runs the risk of appearing redundant because it uses the same information (the measure of respect, or accuracy) contained in the weights w_{ij}s, over and over again. Under most conditions, the Lehrer-Wagner model takes more than one step to reach convergence, and at all steps the same matrix of weights is used in the basic version of the model.

The extended version of the model, on the other hand, allows agents to update the weights they assign to the other agents in the deliberating group. The modification is meant to account for the idea that in the course of a deliberation process, while agents update their opinion at different stages of such process, group members may change their opinion on their fellows and wish to change the weights they assigned them.

Bibliography

Angner, Erik. 2006. 'Economists As Experts: Overconfidence in Theory and Practice.' *Journal of Economic Methodology,* 13(1):1-24.

Alperovitz, Gar. 1996. *The Decision to Use the Atomic Bomb.* New York: Vintage Books.

Armstrong, J. Scott, ed. 2001. *Principles of Forecasting: a Handbook For Researchers and Practitioners*. Norwell, MA: Kluwer Academic Publishers.

Armstrong, J. Scott. 2001. 'Combining Forecasts.' In *Principles of Forecasting: a Handbook For Researchers and Practitioners*. Ed. J. Scott Armstrong. Norwell, MA: Kluwer Academic Publishers (2001).

Arrow, Kenneth J. 1963 [1951]. *Social Choice and Individual Values*. New York: Wiley.

Aspinall, Willy. 2010. 'A Route to More Tractable Expert Advice.' *Nature*, Vol. 463:21.

Aumann, J. Robert. 1976. 'Agreeing to Disagree.' *The Annals of Statistics* 4(6):1236-1239.

Australian Bureau of Statistics. 2000. *Australian National Accounts: Concepts, Sources and Methods.* Accessed August 30, 2010 http://www.abs.gov.au/AusStats/ABS@.nsf/MF/5216.0.

Becker, Gary S. and Kevin M. Murphy. 1992. 'The Division of Labor, Coordination Costs, and Knowledge.' *The Quarterly Journal of Economics*107(4):1137-1160.

Bishop, Michael A. 2004. *Epistemology and the Psychology of Human Judgment.* New York: Oxford University Press.

Blanchard, Olivier. 2009. 'The State of Macro.' *Annual Review of Economics* 1:209-228.

Bonanno, Giacomo and Klaus Nehring. 1997. ‘Agreeing to Disagree: a Survey.’ *SSRN Working Paper* 97(18), doi: http://ideas.repec.org/p/cda/wpaper/97-18.html.

Boumans, Marcel. 2007. *Measurement in Economics: a Handbook.* London: Academic Press.

Bradley, Richard. 2006. ‘Taking Advantage of Difference in Opinion.’ *Episteme* 3(3):141-155.

Bradley, Richard. 2007. ‘Reaching a Consensus.’ *Social Choice and Welfare* 29(4):609-632.

Brown, Bernice and Olaf Helmer. 1964. ‘Improving the Reliability of Estimates Obtained from a Consensus of Experts.’ *Project RAND P-2986*, doi: http://www.rand.org/pubs/papers/2008/P2986.pdf.

Budd, Alan. 1998. ‘The Role and Operations of the Bank of England Monetary Policy Committee.’ *The Economic Journal* 108(451):1783-1794.

Bureau of Economic Analysis. 2009. *NIPA Handbook: Concepts and Methods of the U.S. National Income and Product Accounts.* Accessed August 29, 2010 http://www.bea.gov/national/pdf/NIPAhandbookch1-4.pdf.

Bureau of Labor Statistics - Office of Publications and Special Studies. 2010. *BLS Handbook of Methods*, Chapter 17: Consumer Price Index. Accessed August 29, 2010 http://www.bls.gov/opub/hom/homtoc.htm.

Carlson, Elof Axel. 2004. *Mendel’s Legacy: The Origin of Classical Genetics.* Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Cato Institute. 2009. ‘With All Due Respect Mr. President, That Is Not True.’ Advertisement page on *The New York Times,* January 28, 2009.

Chiesa, Arturo and Raffaele Chiesa. 1990. ‘A Mathematical Method of Obtaining an Astronomical Vessel Position.’ *Journal of Navigation* 43:125-129.

Christensen, David. 2007. ‘Epistemology of Disagreement: the Good News.’ *Philosophical Review* 116(2):187-217.

Christensen, David. 2009. ‘Disagreement As Evidence: The Epistemology of Controversy.’ *Philosophy Compass* 4:1-12.

Christensen, David. 2009. 'Disagreement, Question-Begging and Epistemic Self-Criticism.' *Philosophers' Imprint* (forthcoming). Accessed July 29, 2011 http://www.brown.edu/Departments/Philosophy/onlinepapers /christensen/Conciliationism.pdf.

Coleman, James. 1990. *Foundations of Social Theory*. Cambridge: Belknap.

Collins, Harry M. and Robert Evans. 2002. 'The Third Wave of Science Studies: Studies of Expertise and Experience.' *Social Studies of Science* 32:235-296.

Collins, Harry. 2004. 'Interactional Expertise As a Third Kind of Knowledge.' *Phenomenology and the Cognitive Sciences* 3:125-143.

Collins, Harry. 2010. *Tacit and Explicit Knowledge*. Chicago: University Of Chicago Press.

Cooke, Roger M. 1991. *Experts in Uncertainty*. Oxford: Oxford University Press.

Cuhls Kerstin. 2003. 'Delphi Method.' Technical Report *Fraunhofer Institute For Systems and Innovation Research*. Accessed July 27, 2011 http://www.unido.org/fileadmin/import/16959_DelphiMethod.pdf.

Dalkey, Norman and Olaf Helmer. 1963. 'An Experimental Application of the Delphi Method to the Use of Experts.' *Management Science* 9(3):458-467.

Dalkey, Norman. 1969. 'An Experimental Study of Group Opinion: the Delphi Method.' *Futures* 1(5):408-426.

Dalkey, Norman, Daniel L. Rourke, Ralph Lewis and David Snyder. 1972. *Studies in the Quality of Life: Delphi and Decision-Making.* Lexington, MA: Lexington Books.

DeGroot, Morris. 1974. 'Reaching a Consensus.' *Journal of the American Statistical Association* 69:118-121.

Delbecq, André and Andrew Van de Ven. 1971. 'A Group Process Model For Problem Identification and Program Planning.' *Journal of Applied Behavioral Science* 7(4):466-492.

DeMarzo, Peter M., Dimistri Vayanos and Jeffrey Zwiebel. 2003. 'Persuasion Bias, Social Influence and Unidimensional Opinions.' *The Quarterly Journal of Economics* August: 909-968.

Dietrich, Franz and Christian List. 2010. 'The Aggregation of Propositional Attitudes: Towards a General Theory.' In *Oxford Studies in Epistemology - Vol. 3*. Ed. Tamar Szabo Gendler. Oxford: Oxford University Press (2010).

Dunn, Donald. 2004. *Ponzi: The Incredible True Story of the King of Financial Cons.* New York: Broadway.

Earman, John and Wesley C. Salmon. 1999. 'The Confirmation of Scientific Hypotheses.' In *Introduction to the Philosophy of Science.* Ed. M. H. Salmon et al. Indianapolis: Hackett Publishing Company (1999).

Elga, Adam. 2007. 'Reflection and Disagreement.' *Noûs* 41(3):478-502.

Elster, Jon. 1998. 'Introduction.' In *Deliberative Democracy.* Ed. Jon Elster. Cambridge: Cambridge University Press (1998).

Epstein, Gerald and Jessica Carrick-Hagenbarth. 2010. 'Financial Economists, Financial Interests and Dark Corners of the Meltdown: It's Time to Set Ethical Standards For the Economics Profession.' *Working Paper Series* Political Economy Research Institute - University of Massachusetts Amherst. Number 239, doi: http://ideas.repec.org/p/uma/periwp/wp239_revised.html.

Ericsson, Anders. 2001. 'Expertise'. In *The MIT Encyclopedia of the Cognitive Sciences (MITECS)* Eds. Robert A. Wilson and Frank C. Keil. Cambridge, MA: The MIT Press (2001).

Eto, Hajime. 2003. 'The Suitability of Technology Forecasting/Foresight Methods For Decision Systems and Strategy. A Japanese View.' *Technological Forecasting and Social Change* 70:231-249.

Everett, Theodore J. 2001. 'The Rationality of Science and the Rationality of Faith.' *The Journal of Philosophy* 98(1):19-42.

Faust, David. 1984. *The Limits of Scientific Reasoning*. University Of Minnesota Press.

Feldman, Richard. 2007. 'Reasonable Religious Disagreements'. In *Philosophers without Gods: Meditations on Atheism and the Secular Life.* Ed. Louise M. Antony. Oxford: Oxford University Press (2007).

Feyerabend, Paul K. 1988 [1975]. *Against Method: Outline of an Anarchistic Theory of Knowledge*. (Rev. ed.) New York and London: Verso.

Feyerabend, Paul K. 1999. *Paul K. Feyerabend: Knowledge, Science and Relativism — Philosophical papers, Vol.3.* Edited by John Preston. Cambridge: Cambridge University Press.

Fisher, R. A. 1936. 'Has Mendel's Work Been Rediscovered?' *Annals of Science* 1:115-137.

Fischhoff, Baruch. 2001. 'Judgment Heuristics'. In *The MIT Encyclopedia of the Cognitive Sciences (MITECS)*. Eds. Robert A. Wilson and Frank C. Keil. Cambridge, MA: The MIT Press (2001).

Frances, B. 2010. 'Disagreement.' In *Routledge Companion Epistemology*. Eds. D. Pritchard and S. Bernecker. London: Routledge (2010).

French, John R. P. Jr. 1956. 'A Formal Theory of Social Power.' *Psychological Review* 63(3):181-194.

Friedman, Milton. 1953. *Essays in Positive Economics*. Chicago: University of Chicago Press.

Frigg, Roman and Stephan Hartmann. 2006. 'Models in Science.' *The Stanford Encyclopedia of Philosophy (Summer 2009 Edition)*. Ed. Edward N. Zalta. Accessed July 22, 2011 http://plato.stanford.edu/archives/sum2009/entries/models-science/.

García-Magariño, Iván, Jorge J. Gómez-Sanz and José R. Pérez-Agüera. 2008. 'A Multi-Agent Based Implementation of a Delphi Process (Short Paper).' In *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008).* Eds. Padgham, Parkes, Müller and Parsons. May 12-16, 2008, Estoril, Portugal. Accessed July 17, 2011 http://grasia.fdi.ucm.es/jose/igarciamagarino-Delphi-ShortPaper.pdf.

García-Magariño, Iván, Jorge J. Gómez-Sanz and José R. Pérez-Agüera. 2010. 'A Complete-Computerised Delphi Process with a Multi-Agent System.' (Unpublished manuscript.) Accessed July 10, 2011 http://grasia.fdi.ucm.es/main/myfiles/7-promas08-delphi_0.pdf.

Genest, Christian and James V. Zidek. 1986. 'Combining Probability Distributions: a Critique and an Annotated Bibliography.' *Statistical Science* 1(1):114-135.

Gladwell, Malcolm. 2008. *Outliers.* New York: Little, Brown and Company.

Goldman, Alvin. 1999. *Knowledge in a Social World.* Oxford: Oxford University Press.

Goldman, Alvin. 2010. "Why Social Epistemology Is Real Epistemology." In *Social Epistemology* Eds. A. Haddock, A. Millar, and D. Pritchard. Oxford University Press (2010).

Goldman, Alvin. 2009. 'Systems-Oriented Social Epistemology.' (Working Paper.) Accessed July 12, 2011. http://fas-philosophy.rutgers.edu/goldman /Systems-Oriented%20Social%20Epistemology.pdf.

Goldman, Alvin. 2009. 'Social Epistemology: Theory and Applications.' *Royal Institute of Philosophy Supplement.* Vol. 64.

Golub, Benjamin and Matthew O. Jackson. 2007. 'Naïve Learning in Social Networks: Convergence, Influence, and the Wisdom of Crowds.' *Working Papers Series FEEM Working Paper.* No. 64. Accessed July 17, 2011 http://www.adis.u-psud.fr/docs/ADIS_JACKSON_GOLUB_19062007.pdf.

Goodin, Robert E. 2001. 'Consensus Interruptus.' *The Journal of Ethics* 5:121-131.

Gordon, Theodore J. 1994. 'The Delphi Method.' *AC/UNU Millenium Project. Futures Research Methodology.* Accessed July 23, 2011 http://www.millennium-project.org/FRMv3_0/04-Delphi.pdf.

Gourlay, Stephen. 2002. 'Tacit Knowledge, Tacit Knowing or Behaving?' *Paper of the European Conference on Organizational Knowledge.* Accessed July 11, 2011 http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.103.7950&rep =rep1&type=pdf.

Guala, Francesco. 2001. 'Building Economic Machines: the FCC Auctions.' *Studies in History and Philosophy of Science.* 32(3):453-477.

Guala, Francesco. 2005. *The Methodology of Experimental Economics.* Cambridge: Cambridge University Press.

Hájek, Alan. 2003. 'What Conditional Probability Could not Be.' *Synthese* 137:273-323.

Hájek, Alan. 2003. 'Conditional Probability is the Very Guide of Life.' In *Probability is the Very Guide of Life: The Philosophical Uses of Chance.* Eds. Henry Jr. Kyburg and Mariam Thalos. Chicago: Open Court Publishing (2003).

Hausman, Daniel M. 1992. *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press.

Hausman, Daniel M. 2008. 'Philosophy of Economics.' *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition).* Ed. Edward N. Zalta. Accessed July 1, 2011 http://plato.stanford.edu/archives/fall2008/entries/economics/.

Hartmann, Stephan, Carlo Martini and Jan Sprenger. 2009. 'Consensual Decision-Making Among Epistemic Peers.' *Episteme* 6(2):110-129.

Hegselmannn, Rainer and Ulrich Krause. 2002. 'Opinion Dynamics and Bounded Confidence: Models, Analysis and Simulation.' *Journal of Artificial Societies and Social Simulation* 5(3):1-33.

Hegselmannn, Rainer and Ulrich Krause. 2005. 'Opinion Dynamics Driven by Different Ways of Averaging.' *Computational Economics* 25:381-405.

Helmer, Olaf and Nicholas Rescher. 1959. 'On the Epistemology of the Inexact Sciences.' *Management Science (pre-1986)* 6(1):25-52.

Hempel, Carl G. and Paul Oppenheim. 1948. 'Studies in the Logic of Explanation.' *Philosophy of Science* 15(2):135-175.

Herrera-Viedma, Enrique, S. Alonso, Francisco Chiclana and Francisco Herrera. 1995. 'Basis For a Consensus Model in Group Decision Making With Linguistic Preferences.' *Proc. of the 3th European Congress on Fuzzy and Intelligent Technologies*. Accessed July 24, 2011 http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.163.2528.

Herrera, Francisco, Enrique Herrera-Viedma, andJosé Luis Verdegay. 1997. 'A Rational Consensus Model in Group Decision Making Using Linguistic Assessments.' *Fuzzy Sets and System* 88:31-49.

Hershey, Robert D. 1995. 'Counting the Wealth of Nations; G.D.P.'s Accuracy Is Under Attack From All Sides.' *The New York Times*, December 19, 1995.

Hsu, Chia-Chien and Brian A. Sandford. 2007. 'The Delphi Technique: Making Sense of Consensus.' *Practical Assessment, Research & Evaluation* 12(10):1-8.

Hull, David. 1988. *Science As a Process*. Chicago: University of Chicago Press.

Jackson, Matthew O. 2008. *Social and Economic Networks*. Princeton: Princeton University Press.

Matthew O. Jackson and Benjamin Golub. 2007. 'Naïve Learning in Social Networks: Convergence, Influence and Wisdom of Crowds.' *Working Papers 64, Fondazione Eni Enrico Mattei*. Accessed July 21, 2011 http://www.bepress.com/feem/paper124.

Jehle, David and Branden Fitelson. 2010. 'What is the "Equal Weight View"?' *Episteme* 6: 280-293. doi: 10.3366/E1742360009000719, ISSN 1742-3600.

Kahneman, Daniel. 1982. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge (UK): Cambridge University Press.

Kant, Immanuel. 1998 [1787] *Critique of Pure Reason.* Cambridge: Cambridge University Press.

Kelly, Thomas. 2005. 'The Epistemic Significance of Disagreement.' In *Oxford Studies in Epistemology — Vol 1.* Eds. John Hawthorne and Tamar Gendler Szabo. Oxford: Oxford University Press (2010).

Kelly, Thomas. 2010. 'Peer Disagreement and Higher Order Evidence.' In *Disagreement*. Eds. Richard Feldman and Ted Warfield. Oxford: Oxford University Press (2010).

King, Nathan L. 2011. 'Disagreement: What's the Problem? or A Good Peer is Hard to Find.' *Philosophy and Phenomenological Research* Early View. Accessed July 13, 2011. doi: 10.1111/j.1933-1592.2010.00441.x.

Kitcher, Philip. 1990. 'The Division of Cognitive Labor.' *The Journal of Philosophy* 87(1):5-22.

Kuhn, Thomas. 1970 [1962] *The Structure of Scientific Revolutions.* (Second edition.) Chicago: University of Chicago Press.

Kuhn, Thomas. 2000. *The Road Since Structure.* Chicago: University of Chicago Press.

Lagueux, Maurice. 2008. 'Are We Witnessing a Revolution in Methodology of Economics? About Don Ross's Recent Book on Microexplanation.' *The Erasmus Journal for Philosophy and Economics* 1(1):24-55.

Lam, Barry. 2007. 'On the Rationality of Belief-Invariance in Light of Peer Disagreement.' (Unpublished manuscript.) Accessed July 22, 2011http://faculty.vassar.edu/balam/RationalityofBeliefInvariance.pdf.

Lehrer, Keith. 1976. 'When Rational Disagreement is Impossible.' *Noûs* 10(3):327-332.

Lehrer, Keith. 2001. 'Individualism, Communitarianism and Consensus.' *The Journal of Ethics* 5:105-120.

Lehrer, Keith. 2001. 'The Rationality of Dissensus: A Reply to Goodin.' *The Journal of Ethics* 5:133-137.

Lehrer, Keith and Carl Wagner. 1981. *Rational Consensus in Science and Society.* Dordrecht: Reidel.

Lewis, David 1980. 'A Subjectivist's Guide to Objective Chance.' In *Studies in Inductive Logic and Probability — Vol. 2*. Ed. R. C. Jeffrey. Berkeley: University of California Press (1980).

Linstone, Harold A. and Murray Turoff, eds. 1975. *The Delphi Method: Techniques and Applications.* Reading, MA: Addison-Wesleu Publishing Company.

List, Christian and Robert E. Goodin. 2001. 'Epistemic Democracy: Generalizing the Condorcet Jury Theorem.' *Journal of Political Philosophy* 9:277-306.

List, Christian and Philip Pettit. 2002. 'Aggregating Sets of Judgments: An Impossibility Result.' *Economics and Philosophy* 18:89-110.

List, Christian and Philip Pettit. 2005. 'On the Many as One.' *Philosophy and Public Affairs* 33:377-90.

List, Christian and Pettit, Pettit. 2011. *Group Agents: The Possibility, Design, and Status of Corporate Agents.* Oxford: Oxford University Press.

Lo, Andrew W. 2004. 'The Adaptive Markets Hypothesis: Market Efficiency From an Evolutionary Perspective.' *The Journal of Portfolio Management* 30th Anniversary Issue, 15-29.

Mäki, Uskali. 1988. 'Realism, Economics, and Rhetoric: A Rejoinder to McCloskey.' *Economics and Philosophy* 4:167-169.

Mäki, Uskali. 1995. 'Diagnosing McCloskey.' *Journal of Economic Literature* 33(3):1300-1318.

Mäki, Uskali. 1998. 'Realisticness.' In *The Handbook of Economic Methodology.* Eds. Davis, John B., D. Wade Hands, and Uskali Mäki. Cheltenham: Edward Elgar (1998).

Mäki, Uskali, ed. 2009. *The Methodology of Positive Economics - Reflections on the Milton Friedman Legacy.* Cambridge: Cambridge University Press.

Martini, Carlo, Mark Colyvan and Jan Sprenger. 2011. 'Resolving Disagreement Through Mutual Respect.' (Under review).

McCloskey, Donald N. 1998 [1985]. *The Rhetoric of Economics.* Madison: University of Wisconsin Press.

McCloskey, Donald N. 1992. ' "If You're So Smart..." '. In *Praxiologies and the Philosophy of Economics.* Eds. Wojciech Gasparski and Marek Mlicki. New Brunswick, NY: Transaction Publishers (1992).

McCloskey, Donald N. 1995. 'Modern Epistemology Against Analytic Philosophy: A Reply to Mäki.' *Journal of Economic Literature* 33(3):1319-1323.

McCloskey, Deirdre. 2003. *How to Be Human (Though an Economist).* Ann Arbor, MI: University of Michigan Press.

Meyer, Carl D. 2000. *Matrix Analysis and Applied Linear Algebra.* Philadelphia: Society for Industrial and Applied Mathematics (SIAM).

Mill, John Stuart. 1859 [2008]. 'On Liberty.' In *On Liberty and Other Essays.* Ed. Mary Warnock. Oxford: Oxford University Press (2008).

New Oxford American Dictionary. 2nd ed. Oxford: Oxford University Press, 2005.

Nonaka, Ikujiro. 1991. 'The Knowledge Creating Company.' *Harvard Business Review* 69(6):96-104.

Nurmi, Hannu. 1983. 'Past Masters and Their Modern Followers.' In *Exploring the Basis of Politics.* Ed. S. Hänninen. Tampere: Finnpublishers.

Nurmi, Hannu. 1985. 'Some Properties of the Lehrer-Wagner Method for Reaching Rational Consensus.' *Synthese* 62:13-24.

Pettit, Philip. 2004. 'Groups With Minds of Their Own.' In *Socializing Metaphysics.* Ed. Frederick Schmitt. New York, Rowman and Littlefield (2004).

Pettit, Philip. 2007. 'Responsibility Incorporated.' *Ethics* 117:171-201.

Plantenga, Janneke. 2002. 'Combining Work and Care in the Polder Model: an Assessment of the Dutch Part-Time Strategy.' *Critical Social Policy* 22:53-71.

Polanyi, Michael. 1958. *Personal Knowledge: Towards a Post-Critical Philosophy.* Chicago: University of Chicago Press.

Quinton, Anthony. 1975. 'Social Objects.' *Proceedings of the Aristotelian Society* 75:17.

Reiss, Julian. 2008. *Error in Economics: Towards a More Evidence-Based Methodology.* New York: Routledge.

Regan, Helen M., Mark Colyvan, Lisa Markovchick-Nicholls. 2006. 'A Formal Model for Consensus and Negotiation in Environmental Management.' *Journal of Environmental Management* 80:167-176.

Ross, Don. 2007. 'The Economic and Evolutionary Basis of Selves.' In *Distributed Cognition and the Will: Individual Volition and Social Context.* Eds. Don Ross et al. Cambridge, MA: MIT Press (2007).

Royle, Gordon and Weisstein, Eric W. 2010. 'Reducible Matrix' *MathWorld - A Wolfram Web Resource.* Accessed July 22, 2011. http://mathworld.wolfram.com/ReducibleMatrix.html.

Roubini, Nouriel. 1998. 'Output and Inflation: Are We Mismeasuring Them?' (Unpublished manuscript.) Accessed July 12, 2011 http://pages.stern.nyu.edu/ nroubini/MEASURE.HTM.

Rowe, Gene and George Wright. 1999. 'The Delphi Technique As a Forecasting Tool: Issues and Analysis.' *International Journal of Forecasting* 15:353-375.

Rowe, Gene and George Wright. 2001. 'Expert Opinions in Forecasting: The Role of the Delphi Technique.' In *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Ed. J. Scott Armstrong. Norwell, MA: Kluwer Academic Publishers.

Rowe, Gene, George Wright and Andy McColl. 2004. 'Judgment Change During Delphi-Like Procedures: The Role of Majority Influence, Expertise, and Confidence.' *Technological Forecasting and Social Change* 72(4):377-399.

Rubinstein, Ariel. 2006. 'Dilemmas of an Economic Theorist.' *Econometrica* 74(4):865-883.

Russell, Nicholas. 2009. *Communicating Science.* Cambridge: Cambridge University Press.

Saari, Donald G. 2008. *Disposing Dictators, Demystifying Voting Paradoxes.* Cambridge: Cambridge University Press.

Schwartz, Nelson and Eric Dash. 2010. 'Despite Reform, Banks Have Room for Risky Deals.' *The New York Times*, August 25, 2010.

Salmon, M.H. 1999. 'Philosophy of the Social Sciences.' In *Introduction to the Philosophy of Science.* Eds. M. H. Salmon et al. Indianapolis: Hackett Publishing Company (1999).

Sanders, Fredereick. 1963. 'On Subjective Probability Forecasting.' *Journal of Applied Meteorology* 2:191-201.

Schreuder, Yda. 2001. 'The Polder Model in Dutch Economic and Environmental Planning.' *Bulletin of Science Technology Society* 21(237).

Schelling, Thomas. 1971. 'Dynamic Models of Segregation.' *Journal of Mathematical Sociology* 1:143-186.

Sen, Amartya. 2009. *The Idea of Justice.* Cambridge, MA: Harvard University Press.

Shenon, Philip. 2004. '9/11 Panel Plans Hard Questions About the F.B.I. and Justice Dept.' *The New York Times*, April 6, 2004.

Shogenji, T. 2007. 'A Conundrum in Bayesian Epistemology of Disagreement.' (Unpublished Manuscript) Accessed July 23, 2011 http://www.fitelson.org/few/few_07/shogenji.pdf.

Sigel, Deborah A. and David Wettergreen. 2007. 'Star Tracker Celestial Localization System for a Lunar Rover.' *Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*. San Diego, CA, USA, Oct 29 - Nov 2, 2007.

Simon, Herbert A. 2001. 'Problem Solving'. In *The MIT Encyclopedia of the Cognitive Sciences (MITECS).* Eds. Robert A. Wilson and Frank C. Keil. Cambridge, MA: The MIT Press (2001).

Sobel, Dava. 1995. *Longitude.* New York: Walker.

Solomon, Miriam. 1994. 'Social Empiricism.' *Noûs* 28(3):325-343

Solomon, Miriam. 2001. *Social Empiricism*. Cambridge, MA: MIT Press.

Solomon, Miriam. 2007. 'The Social Epistemology of NIH Consensus Conferences.' In *Establishing Medical Reality: Essays in the Metaphysics and Epistemology of Biomedical Science.*. Eds. Harold Kinkaid and Jennifer McKitrick. Dordrecht: Springer (2007).

Sosa, Ernest. 2010. 'The Epistemology of Disagreement.' (Unpublished manuscript) Accessed July 22, 2011 http://philpapers.org/autosense.pl?searchStr=Ernest%20Sosa.

Staël von Holstein, Carl-Axel S. 1972. 'Probabilistic Forecasting: An Experiment Related to the Stock Market.' *Organizational Behavior and Human Performance* 8:139-158.

Steele, Katie. 2006. 'The Precautionary Principle: a New Approach to Public Decision-Making?' *Law, Probability and Risk* 5: 19-31.

Stewart, Thomas. 2001. 'Improving Reliability of Judgmental Forecasts.' In *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Ed. J. Scott Armstrong. Norwell, MA: Kluwer Academic Publishers (2001).

Stillwell, William D. 2003. 'Tacit Knowledge and the Work of Ikujiro Nonaka: Adaptions of Polanyi in a Business Context.' *Tradition & Discovery* 30(1):19-22.

Sugden, Robert. 2000. 'Credible Worlds: the Status of Theoretical Models in Economics.' *Journal of Economic Methodology* 7(1):1-31.

Sugden, Robert. 2005. 'Experiments As Exhibits and Experiments As Tests.' *Journal of Economic Methodology* 12(2):291-302.

Sugden, Robert. 2008. 'The Changing Relationship Between Theory and Experiment in Economics.' *Philosophy of Science* 75:621-632.

Suppes, Patrick. 1968. 'The Desirability of Formalization in Science.' *The Journal of Philosophy* 65(20):651-664.

System of National Accounts 2008. European Communities, International Monetary Fund, Organization for Economic Co-operation and Development, United Nations and World Bank. (2008) Accessed August 30, 2010 http://unstats.un.org/unsd/nationalaccount/SNA2008.pdf

Taleb, Nassim Nicholas. 2007. *The Black Swan: The Impact of the Highly Improbable*. New York: Random House and Penguin.

Trout, J.D. 2009. *The Empathy Gap*. New York: Viking/Penguin.

van Aaken, Anne, Christian List and Christoph Luetge, eds. 2004. *Deliberation and Decision: Economics, Constitutional Theory and Deliberative Democracy*. London: Ashgate.

van Allen, James A. 2004. 'Basic Principles of Celestial Navigation.' *American Journal of Physics* 72(11):1418-1424.

van Fraassen, Bas C. 1984. 'Belief and the Will.' *The Journal of Philosophy*, May 1984.

VandeVen, Andrew and André Delbecq. 1971. 'Nominal Versus Interacting Group Processes for Committee Decision-Making Effectiveness.' *The Academy of Management Journal* 14(2):203-212.

VandeVen, Andrew and André Delbecq. 1974. 'The Effectiveness of Nominal, Delphi, and Interacting Group Decision Making Processes.' *The Academy of Management Journal* 17(4):605-621.

Wagner, Carl. 1978. 'Consensus Through Respect: a Model of Rational Group Decision-Making.' *Philosophical Studies* 34:335-349.

Weatherson, Brian. 2007. 'Disagreeing about Disagreement.' (Unpublished Manuscript.) Accessed July 27, 2011 http://brian.weatherson.org/DaD.pdf.

Weisstein, Eric W. 2011. 'Graph.' *MathWorld — A Wolfram Web Resource.* Accessed July 23, 2011 http://mathworld.wolfram.com/Graph.html.

Wiegand, Steve. 2009. *Lessons from the Great Depression For Dummies®*. Indianapolis, Wiley Publishing.

Wikipedia Contributors, "Causes of the Great Depression" *Wikipedia, The Free Encyclopedia,* Accessed December 20, 2010 http://en.wikipedia.org/w/index.php?oldid=407093381

Xu Z. S., Q. L. Da. 2003. 'An Overview of Operators for Aggregating Information.' *International Journal of Intelligent Systems* 18:953-969.

Yates, Frank J. and Linda McDaniel. 1991. 'Probability Forecasting of Stock Prices and Earnings: The Hazards of Nascent Expertise.' *Organizational Behavior and Human Decision Processes* 49:60-79.

Young, H. P. 1988. 'Condorcet's Theory of Voting.' *American Political Science Review* 82:1231-1244.

Summary

Situations of disagreement are a very common occurrence, possibly even the norm, in most types of human interaction. At the same time humans spend a great deal of energy seeking to eliminate disagreement and reach a consensus. From the most private kinds of interactions, e.g. a group of friends planning to go to the movies, to the most difficult scenarios, like global diplomacy, consensus is looked for among all classes and occupations: politicians, physicians, businessmen, and also scientists.

This Ph.D. thesis deals with the problem of disagreement and consensus from three different perspectives. Chapters 2, 3 and 4 analyze consensus and disagreement from a formal perspective. Starting from one of the best-known models for rational consensus, the Lehrer-Wagner model, I provide some extensions and specifications of the model, arguing, however, that the model provides only a general framework for aggregation, rather than the normative notion of rational consensus originally sought by its proposers. Furthermore, I argue that the formal methods for resolving epistemic disagreement, thus including the Lehrer-Wagner model, are inadequate for answering the standard question of how it is that two (or more) epistemic peers can resolve their disagreement on a specific (and quantifiable) subject matter.

In chapters 5 and 6 I look at the phenomena of disagreement and consensus as occurring in the field of scientific investigation, in particular in economics. In that context, I argue that consensus and disagreement have a lot to do with the dynamics of expert judgment and deliberation occurring in a specific scientific field. I further argue that the subject matter of economics does not guaranteed the type of objectivity that is often attributed to its method and that is often assumed to be present in other sciences, particularly in the natural sciences. For that reason, the dynamics of expert judgement and deliberation are all the more important for the quality of the output of economics sciences, which for the purposes of this thesis is assumed to consist, mostly, of policy-applicable knowledge. I investigate the literature on expert judgment and deliberation and suggest a number of possible directions for future research into how a thorough discussions of the problems related to expertise can and should enter the field of economic methodology.

In the last section of the thesis, it is assumed that the formation of a specific scientific consensus bears its fruits to society, whether in a positive or negative way. When that happens, however, the promotors of consensus become what in ethics are called “moral agents”, that is subjects open to praise or blame and, in general, responsibility. In chapter 7 I discuss the problem of responsibility in small committees, for example the committees

of experts referred to in the previous two chapters. While the topic is not new, and a number of theories of corporate responsibility exist, the chapter begins with a discussion of Pettit and List's theory of group agency and responsibility. Pettit (2007) argues that group agency implies group responsibility. I discuss the three conditions the authors give for group responsibility and argue that there are counterexamples to Pettit's thesis. Finally, I add two additional desiderata to Pettit and List's conditions, and argue why the addition is important for avoiding the type of counterexamples I previously illustrated.