This is a preprint of a paper published in the Journal of Political Philosophy 19(1): 1-27, 2011.
For the official version, see DOI: 10.1111/j.1467-9760.2010.00369.x

Group communication and the transformation of judgments: an impossibility result*

Christian List, LSE

March 2007 / April 2009

Abstract

While a large social-choice-theoretic literature discusses the aggregation of individual judgments into collective ones, there is much less formal work on the transformation of judgments in group communication. I develop a model of judgment transformation and prove a baseline impossibility theorem: Any judgment transformation function satisfying some initially plausible conditions is the identity function, under which no opinion change occurs. I identify escape routes from this impossibility and argue that the kind of group communication envisaged by deliberative democrats must be 'holistic': It must focus on webs of connected propositions, not on one proposition at a time, which echoes the Duhem-Quine 'holism thesis' on scientific theory testing. My approach provides a map of the logical space in which different possible group communication processes are located.

*This paper was presented at the 8th Augustus De Morgan Workshop, King's College, London, 11/2006, the 2nd Indian Conference on Logic and Its Relationship with Other Disciplines, IIT-Bombay, 1/2007, the Dutch Social Choice Colloquium, Tilburg, 6/2007, and seminars at LSE, 5/2007, ANU, 9/2007, Bristol, 10/2007, Edinburgh, 1/2008, Geneva, 3/2008, Juan March Institute, Madrid, 4/2008, Warwick, 5/2008, IHPST, Paris, 7/2008, Bogazici Univ., Istanbul, 12/2008, and the Catholic Univ. of Leuven, 12/2008. I thank the participants, the anonymous referees and Nick Baigent, Richard Bradley, Franz Dietrich, Wulf Gaertner, Robert Goodin, Mathias Koenig-Archibugi, David Makinson, Kai Spiekermann, Katie Steele, Laura Valentini and Leif Wenar for comments.

1 Introduction

Aggregation and deliberation are often contrasted as two very different approaches to collective decision-making. While aggregation is the merging of conflicting individual opinions into a social outcome, deliberation involves the discussion of these opinions and their possible transformation by the individuals deliberating. Jon Elster summarizes the difference between the two approaches as follows:

> 'The core of the [deliberative approach] ... is that rather than aggregating or filtering preferences, the political system should be set up with a view to changing them by public debate and confrontation ... [T]here would [then] not be any need for an aggregation mechanism, since a rational discussion would tend to produce unanimous preferences.'[1]

The contrast between the two approaches is probably overstated. More plausibly, they are complementary, not contradictory. In many real-world collective decisions, aggregation is preceded by some form of group communication – in the best case, by the kind of reasoned deliberation envisaged by deliberative democrats.[2] Nonetheless, social choice theory, our best formal theory of collective decision-making, has focused mostly on aggregation and said little about pre-decision communication. Game theorists have recently given more attention to communication, investigating for example the incentives for and against truth-telling in deliberative settings,[3] but we still lack a social-choice-theoretic model of the transformation of opinion under various forms of group communication. The aim of this paper is to contribute to filling this gap in the literature.

[1] See Elster (1986, p. 112). On deliberative democracy, see, e.g., Cohen (1989), Dryzek (1990, 2000), Fishkin (1991), Gutman and Thompson (1996), Bohman and Rehg (1997).

[2] E.g., Miller (1992), Knight and Johnson (1994), Dryzek and List (2003).

[3] E.g., Austen-Smith and Feddersen (2006), Calvert (2006), Landa and Meirowitz (2006), Hafer and Landa (forthcoming).

I model opinions as judgments – acceptance or rejection – on certain propositions, drawing on the theory of judgment aggregation.[4] The propositions may be logically interconnected, so that the judgments on some propositions constrain those that can rationally be held on others. This way of modelling opinions is very general: As illustrated below, it can represent not only beliefs but also preferences. While the theory of judgment aggregation focuses on *judgment aggregation functions*, I here analyze what I call *judgment transformation functions*. A judgment transformation function maps each admissible profile of individual sets of judgments on the given propositions not to a collective set of judgments on them – as an aggregation function does – but to another, possibly revised profile of individual sets of judgments. The input profile represents the individuals' pre-communication judgments, the output profile their post-communication judgments. The process may or may not lead to consensus. The concept of a judgment transformation function is very flexible, with different such functions representing very different communication processes. Some may satisfy conditions of good democratic deliberation, while others may capture indoctrination or the blind mimicking of some charismatic leader.

Using the new model, I prove a baseline impossibility theorem. When the propositions under consideration are logically connected with each other, any judgment transformation function satisfying some initially plausible conditions must be maximally conservative: It must be the identity function, under which nobody ever changes his or her judgment on anything. The

[4] Inspired by the 'doctrinal' and 'discursive' paradoxes (Kornhauser and Sager 1986, Pettit 2001), judgment aggregation was formalized by List and Pettit (2002, 2004), combining Arrow's (1951/1963) axiomatic approach to social choice theory with a logical representation of propositions. Further results and model extensions were provided by List (2003, 2004), Pauly and van Hees (2006), Dietrich (2006, 2007), Nehring and Puppe (2008), van Hees (2007), Dietrich and List (2007a,b,c, 2008), Dokow and Holzman (forthcoming, 2006) and Pigozzi (2006). Judgment aggregation theory is closely related to abstract aggregation theory, e.g., Wilson (1975), Rubinstein and Fishburn (1986), Nehring and Puppe (2002), and to the theory of belief merging (Konieczny and Pino Pérez 2002).

theorem's conditions thus imply the strongest version of what Gerry Mackie has called the 'unchanging minds hypothesis':

> '[P]ublic deliberation on a pending item seldom seems to change anyone's mind.'[5]

The conditions, informally stated, are the following:

(1) Any profile of rational individual judgment sets is admissible as input to the communication.

(2) The output of the communication is also a profile of rational individual judgment sets.

(3) If there is unanimity on *every* proposition before communication (not just on a single proposition), this is preserved after communication.

(4) The individuals do not always ignore their pre-communication judgments in forming their post-communication judgments.

(5) The communication focuses on one proposition at a time, which in turn can be shown to be necessary for protecting the communicative process against strategic manipulability.[6]

Since only a degenerate communicative process without any opinion change satisfies these five conditions together, which would rule out effective group deliberation as envisaged by deliberative democrats, I consider relaxing some of them. The significance of the new theorem, I suggest, lies not in establishing the impossibility of deliberative democracy, but rather in showing which conditions can and cannot be met if group communication is to be effective. If any one of the five conditions is dropped, the theorem's negative conclusion no longer follows. So the result provides a map of the logical

[5] See Mackie (2006, p. 279).

[6] Below I distinguish a weaker and a stronger version of this condition.

space of possible communication processes and thereby shows us what, from a 'bird's eye' perspective, the functional relations between the inputs and outputs of effective deliberative processes can and cannot look like.

What, then, are the most plausible escape routes from the impossibility? I argue that, except in special cases, the first four conditions are hard to give up, but the fifth – the focus on a single proposition at a time – is a plausible candidate for relaxation. Thus effective group communication as envisaged by deliberative democrats requires some kind of 'holism': The objects of judgment transformation cannot generally be single propositions in isolation, but must be larger 'webs' of interconnected propositions. This echoes the Duhem-Quine thesis on holism in science, according to which one cannot empirically test a single proposition in isolation, but only in conjunction with a larger web of related propositions.[7]

The present conclusion refines Mackie's suggestion that the network structure of opinions affects whether or not deliberation can change minds: '[D]ue to the network, the effects of deliberative persuasion are typically latent, indirect, delayed, or disguised.'[8] My result shows that the unchanging minds hypothesis is true when communication is restricted to one proposition at a time,[9] but false when communication is sufficiently holistic. The price of this holism is strategic manipulability of the communication process, by providing incentives for strategic misrepresentation of individual judgments.

It may be tempting to think that the holistic property of judgment transformation is just a trivial consequence of the presence of logical connections between propositions. To see that this is not the case, notice that rational opinion change without holism is entirely possible if one of the other four conditions of the theorem is suitably relaxed. Even a process as *non*-holistic

[7] See Quine (1951).

[8] Mackie (2006, p. 279) says: 'The network structure of attitudes explains why the unchanging minds hypothesis seems to be true, and why it is false.'

[9] In the presence of logical connections between propositions.

as deference to the majority on each proposition, for instance, can lead to perfectly rational post-communication judgments if the theorem's first condition – specifying the domain of admissible inputs to the communication – is appropriately adjusted, as shown below. The theorem's conclusion is a genuinely joint implication of its five conditions.

After a discussion of the impossibility result, I give some examples of feasible judgment transformation functions that do allow opinion change, distinguishing between those that generate consensus and those that generate something less than consensus. Among the latter are transformation functions that offer a new theoretical explanation of the much discussed phenomenon of deliberation-induced 'meta-agreement'.[10] This phenomenon is relevant to democratic decision-making because it helps to avoid some of the notorious paradoxes of aggregation. Thus the paper contributes new positive results in addition to its new theoretical model and impossibility theorem. I conclude with a brief discussion of how the present work is related to game-theoretic works on communication.[11]

[10] The concept of 'meta-agreement' was introduced in List (2002). See also note 2 and empirical evidence in List, Luskin, Fishkin and McLean (2000/2006).

[11] Important related works include Lehrer and Wagner's (1981) model of rational consensus, which can be seen as a probabilistic analogue of the judgment transformation model (where opinions are represented not by binary judgments, but by subjective probability assignments), and the theory of conciliation and consensus in belief merging (Konieczny 2004; Gauwin, Konieczny and Marquis 2005), whose key concept – a conciliation operator – is related to the present concept of a judgment transformation function. The theory of individual belief revision – either in a Bayesian tradition or in the Alchourrón-Gärdenfors-Makinson framework (1985) – addresses a somewhat different question from the present one. It focuses on individual belief change in response to new information rather than the transformation of opinions in group communication. (The latter may be only partly information-driven and sometimes not information-driven at all.)

2 Opinions as judgments on propositions

How can the opinions held by a group of individuals at a given time be modelled? In this section, I explain how to model them as judgments on propositions expressed in logic.[12] I turn to their aggregation and transformation subsequently.

The ingredients of the model are the following. There is a group of individuals.[13] The set of propositions considered by them is called the *agenda*. Propositions are represented by sentences, generally denoted 'p', 'q', 'r', ..., from propositional logic or a more general language.[14] Propositional logic can express atomic propositions, without logical connectives, such as 'a', 'b', 'c', ..., and compound propositions, with the logical connectives 'not', 'and', 'or', 'if-then' and 'if and only if', such as 'a and b' and 'if a or b, then not c'. As is standard in logic, one can distinguish between consistent and inconsistent sets of propositions.[15]

Each individual's opinions at a given time are represented by a *judgment set*: the set of all those propositions in the agenda that the individual accepts.[16] On the standard interpretation, to accept proposition 'p' means to believe 'p'; thus judgments are binary cognitive attitudes. Alternatively, to accept 'p' could mean to desire 'p'; judgments would then be binary emotive attitudes. A judgment set is called *consistent* if it is a consistent set

[12] This follows List and Pettit (2002, 2004) and the generalization in Dietrich (2007).

[13] The group is finite, and individuals are labelled $1, 2, ..., n$.

[14] Formally, the *agenda* is a subset X of the logic, where (i) X is closed under negation (if 'p' is in X, then so is 'not p'), (ii) 'not not p' is identified with 'p', and (iii) X contains no tautological or contradictory propositions. Instead of propositional logic, any logic with some minimal properties can be used, including expressively richer logics such as predicate, modal, deontic and conditional logics (Dietrich 2007).

[15] In propositional logic, a set of propositions is *consistent* if all its members can be simultaneously true, and *inconsistent* otherwise. E.g., {'a', 'a or b'} is a consistent set, whereas {'a', 'not a'} and {'a', 'if a then b', 'not b'} are not. More generally, consistency is definable in terms of a more basic notion of *logical entailment* (Dietrich 2007).

[16] Formally, individual i's *judgment set* is a subset J_i of the agenda X.

of propositions and *complete* if it contains a member of each proposition-negation pair in the agenda. A combination of judgment sets across all the individuals in the group is called a *profile*.[17]

Let me give some examples of agendas of propositions on which groups of individuals may make judgments and to which the theorems presented below apply:

Example 1: Climate change. A panel of experts deliberates about climate change. The agenda on which the experts make judgments contains the following propositions and their negations: 'Global CO_2 emissions are above 30000 million metric tons of carbon per annum' ('a'); 'If global CO_2 emissions are above this threshold, then the global temperature will increase by at least 1.5^{o}C by 2030' ('if a then b'); 'The global temperature will increase by at least 1.5^{o}C by 2030' ('b').[18]

Example 2: A tenure case. A university committee deliberates about whether to grant tenure to a junior academic. The agenda on which the committee members make judgments contains the following propositions and their negations: 'The candidate is excellent at teaching' ('a'); 'The candidate is excellent at research' ('b'); 'Excellence at both teaching and research is necessary and sufficient for tenure' ('c if and only if (a and b)'); 'The candidate should be given tenure' ('c').[19]

Example 3: Ranking candidates or policy options. A political decision-making body (e.g., a legislature, committee or electorate) deliberates about how to rank three or more candidates or policy options in an order of social preference. The agenda on which the individuals make judgments contains all propositions of the form 'x is preferable to y' and their negations, where x and y are distinct candidates or options from some set of

[17] Formally, a *profile* is an n-tuple $(J_1, J_2, ..., J_n)$.

[18] Variants of this example appear across the literature on judgment aggregation.

[19] This example is due to Bovens and Rabinowicz (2006).

available ones and 'is preferable to' is a binary relation, with the rationality constraints on preferences built into the (predicate) logic.[20]

Example 4: Group membership. A club, society or association deliberates about which candidates from a list of three or more available ones should be granted membership, subject to the constraint that some, but not all, candidates should be granted membership. The agenda on which the individuals make judgments contains all propositions of the form 'candidate j should be granted membership' and their negations, where j is any available candidate and the mentioned constraint is built into the logic.[21]

Each of these agendas exhibits certain logical connections between propositions. By contrast, 'trivial' agendas such as those containing only a single proposition-negation pair are not typical in complex decision-making settings. To set them aside, I assume throughout the paper that the agenda is at least *minimally complex* in a sense satisfied in all the examples but whose technical details are not central for the exposition.[22]

[20] For details, see Dietrich and List (2007a), drawing on List and Pettit (2004).

[21] The conjunction of the propositions in quotes is stipulated to be false and their disjunction to be true. The example is due to Kasher and Rubinstein (1997).

[22] Formally, I assume that (i) the agenda has an inconsistent subset of three or more propositions that becomes consistent upon removing any one of its members, and (ii) it is not (nor isomorphic to) a set of propositions whose only logical connectives are 'not' and 'if and only if'. Property (ii) is a variant of *non-affineness* (Dokow and Holzman forthcoming) and *even-number negatability* (Dietrich and List 2007a). Properties (i) and (ii) are met in examples 1 to 4. E.g., the agenda containing 'a', 'if a then b', 'b' and negations (example 1) satisfies (i) because its three-member inconsistent subset {'a','if a then b','not b'} becomes consistent if any one proposition is removed; it obviously satisfies (ii). In examples 2 to 4, a further property is met, which I assume only where explicitly stated: (iii) any proposition in the agenda can be deduced from any other proposition in it via a sequence of pairwise conditional entailments. Property (iii) has been introduced under the name *total blockedness* by Nehring and Puppe (2002).

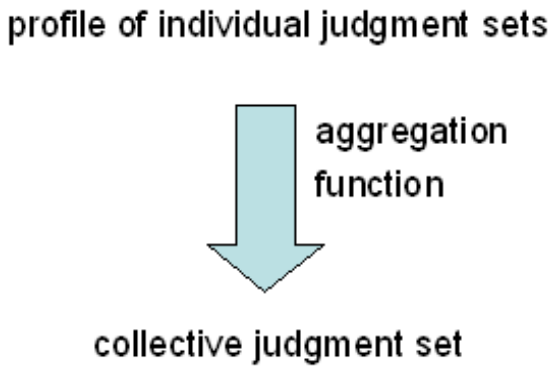

Figure 1: Judgment aggregation

3 The aggregation of judgments

Before I can formally analyze the problem of judgment transformation, it is necessary to recapitulate the problem of judgment aggregation: How can each profile of individual judgment sets on a given agenda be aggregated into a collective judgment set? This problem arises, for example, in referenda involving multiple propositions, in legislatures or committees deciding what factual and normative propositions to accept in legislation, in multi-member courts resolving cases on the basis of several premises, and in expert panels seeking to merge several scientific viewpoints into a collective viewpoint.

As illustrated in Figure 1, an *aggregation function* is a function that maps each profile of individual judgment sets in some domain to a collective judgment set.[23] Examples of aggregation functions are *majority voting*, where each proposition is collectively accepted if and only if it is accepted by a majority of individuals; *supermajority* or *unanimity rules*, where each proposition is collectively accepted if and only if it is accepted by a certain qualified majority of individuals, for example, two thirds, three quarters,

[23] While a judgment aggregation function, as defined in List and Pettit (2002), goes back to Arrovian social choice theory, a related concept is that of a merging operator in belief merging (Konieczny and Pino Pérez 2002). Parallels are discussed in Pigozzi (2006).

or all of them; and *dictatorships*, where the collective judgment set is always the individual judgment set of the same antecedently fixed individual, the *'Arrovian' dictator*, named after Kenneth Arrow's classic impossibility theorem.[24] Many other aggregation functions have been proposed.

Although the possibilities seem abundant, it is surprisingly difficult to find an aggregation function that guarantees consistent collective judgment sets. Notoriously, majority voting can produce inconsistent collective judgment sets even when all individual judgment sets are consistent.[25] Consider the climate change example above (example 1), and suppose there are three experts on the panel, with opinions as shown in Table 1. The first expert judges that 'a', 'if a then b' and 'b'; the second judges that 'a', but 'not (if a then b)' and 'not b'; and the third judges that 'if a then b', but 'not a' and 'not b'. Clearly, each expert holds an individually consistent judgment set. Yet, the majority judgments are inconsistent: Majorities accept 'a', 'if a then b' and 'not b', an inconsistent set of propositions in the standard sense of logic. The same problem can arise in each of the other examples given above.

	'a'	'if a then b'	'b'
Individual 1	True	True	True
Individual 2	True	False	False
Individual 3	False	True	False
Majority	True	True	False

Table 1: A profile of individual judgment sets

Can we find aggregation funtions that are immune to this problem? The recent literature on judgment aggregation has explored this question in great

[24] See Arrow (1951/1963). Note that Arrow's theorem itself concerns preference aggregation, not judgment aggregation. The relationship between preference and judgment aggregation is discussed in List and Pettit (2004) and Dietrich and List (2007a).

[25] This is the *discursive paradox* (Pettit 2001, extending Kornhauser and Sager 1986), which generalizes Condorcet's paradox of majority voting (List and Pettit 2004).

generality. One of its generic findings is that there exist no democratically appealing aggregation functions satisfying the following conditions:

Universal domain. The aggregation function accepts as admissible input any possible profile of consistent and complete individual judgment sets.

Collective rationality. The aggregation function produces as output a consistent and complete collective judgment set.

Consensus preservation. If all individuals hold the same judgment set, this is also the collective judgment set.

Independence/systematicity. The collective judgment on any proposition 'p' on the agenda depends only on individual judgments on 'p' [and the pattern of dependence is the same across propositions]. (Independence omits, and systematicity includes, the neutrality clause in square brackets.)

Theorem 1 *Any aggregation function satisfying universal domain, collective rationality, consensus preservation and independence/systematicity is a dictatorship of one individual.*[26] *(Whether the result requires independence or systematicity depends on how the minimal complexity of the agenda is defined.*[27]*)*

A lot could be said about how to interpret this theorem, which generalizes Arrow's original impossibility theorem.[28] To avoid the dictatorship

[26] This theorem was proved by Dietrich and List (2007a) and Dokow and Holzman (forthcoming), building on earlier results by List and Pettit (2002), Nehring and Puppe (2002), Pauly and van Hees (2006), Dietrich (2006).

[27] If the agenda meets only properties (i) and (ii), systematicity is needed for the result; if it also meets property (iii), independence is enough.

[28] This is because Theorem 1 also applies to the special case of preference aggregation, representable in the judgment aggregation model, as in the case of ranking candidates or options in example 3 above. Explicit derivations of Arrow's theorem (1951/1963) as a corollary of Theorem 1 are given in Dietrich and List (2007a) and Dokow and Holzman (forthcoming). For earlier derivations of Arrow-like results from judgment and abstract aggregation results, see Wilson (1975), List and Pettit (2004) and Nehring (2003).

conclusion, we must relax one of universal domain, collective rationality, consensus preservation or independence/systematicity. Given the present focus on group communication, however, I set these issues aside for the moment and return to analogous issues when I present the new theorem on the transformation of judgments.

4 The transformation of judgments

To model the transformation of judgments, I introduce the new concept of a *judgment transformation function*. As illustrated in Figure 2, this is defined as a function that maps each profile of individual judgment sets in some domain to a profile of individual judgment sets in some co-domain, possibly the same as the domain.[29] The input profile represents the individuals' judgments before communication, the output profile their judgments after communication. The output judgments may or may not differ from the input judgments, and the transformation may or may not lead to consensus.

A simple example of a transformation function is *deference to the majority*, where, after communication, each individual accepts all those propositions that a majority accepts before communication. But just as majority voting as an aggregation function fails to guarantee consistent collective judgments, so deference to the majority as a transformation function fails to guarantee consistent output judgments.[30] If each expert in the climate change example were to defer to the majority judgments in Table 1, for instance, the resulting post-communication judgments would be inconsistent.

An alternative to deference to the majority is *deference to a supermajority* or *unanimity*: Here each individual accepts all those propositions after

[29] Just as a judgment aggregation function is related to a belief merging operator, so a judgment transformation function is related to a belief conciliation operator (Gauwin, Konieczny and Marquis 2006). Pigozzi's (2006) insights on the parellels between judgment aggregation and belief merging apply, *mutatis mutandis*, to revision too.

[30] For a critique of deference to a majority, see Pettit (2006).

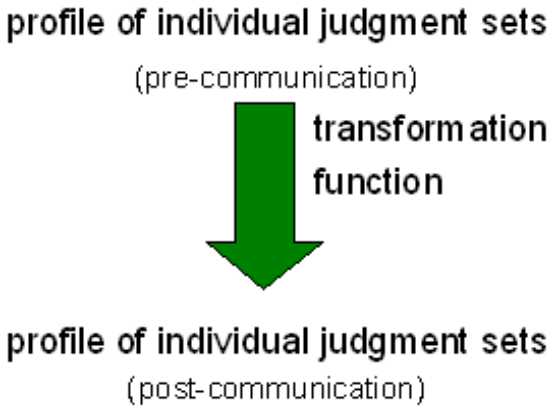

Figure 2: Judgment transformation

communication that a certain qualified majority – perhaps everyone – accepts before communication. If the supermajority threshold is sufficiently large, such a transformation function performs better than deference to a majority at securing consistency. If the propositions are as in the climate change example, for instance, any threshold greater than two thirds guarantees consistent output judgments.[31] But such a transformation function has problems of its own. First, the individuals' post-communication judgments will be incomplete on all those issues on which there is no supermajority consensus; and second, they may violate deductive closure: An individual may come to accept 'a' and 'if a then b', because each receives the required supermajority support, and yet fail to accept 'b', because there is no supermajority consensus on 'b'. Moreover, it is hard to solve these two problems together. Only a unanimity threshold can generally prevent violations of deductive closure,[32] but it also amplifies the incompleteness problem, because it permits the acceptance of only those propositions on which there is total agreement.

[31] To make this distinct from unanimity deference, the group size must be greater than three. For a discussion of deference to a supermajority, see List (2006b).

[32] For a proof, see Dietrich and List (2007b).

Other examples of transformation functions are *opinion leader functions*, where each individual adopts as his or her output judgment set the input judgment set of an antecedently fixed individual, called the individual's *opinion leader*. The opinion leader may differ for different individuals or be the same across individuals. In the latter case, the opinion leader function is the communicative analogue of a dictatorial aggregation function. An opinion leader function may represent not only the presence of one or several particularly persuasive individuals but also the effects of indoctrination, propaganda or, if different individuals cluster around different opinion leaders, group fragmentation. Finally, an entirely degenerate transformation function is the *identity function*, where the output profile is always the same as the input profile: Nobody ever changes his or her judgments.

None of these examples of judgment tranformation functions appear to be particularly 'deliberative'. This is not accidental. Just as many (in fact, most) possible aggregation functions do not qualify as 'democratic' – think of Arrovian dictatorships as the most extreme examples – so many (again, most) transformation functions are far from 'deliberative' in the sense of the normative literature on deliberative democracy. The main purpose of defining judgment transformation functions in such general terms is to have a flexible concept available which allows us to represent a large spectrum of possible communication processes, ranging from degenerate ones without any opinion change and ones involving indoctrination to deliberation. Below I introduce some other, arguably more compelling judgment transformation functions. In particular, I discuss the class of so-called *constrained minimal revision functions*, which may be of some relevance for theoretically explaining the empirically observed phenomenon of deliberation-induced 'meta-agreement'. Generally, a transformation function may depend on the individuals, their context and the agenda of propositions under consideration.

Just as the theory of judgment aggregation seeks to characterize the log-

ical space of possible aggregation functions satisfying various conditions, so I now want to explore the logical space of possible transformation functions satisfying certain conditions. This exercise is illuminating from two perspectives. From a normative perspective, deliberative democrats have proposed a number of desiderata that a group communication process should meet in order to count as properly 'deliberative'. Habermas's conditions on an 'ideal speech situation' are well-known desiderata of this kind. I do not analyze Habermas's own conditions here, but by formalizing such desiderata as conditions on a transformation function – e.g., as conditions on its inputs, outputs or the relationship between inputs and outputs – we may ask whether they can be met together and what a transformation function looks like that meets them all. From a positive perspective, several effects of group communication on individual opinions are empirically known, ranging from 'meta-agreement' to 'group polarization'.[33] By formally describing such effects as properties of the underlying transformation function, we may investigate what transformation functions explain those empirically observed effects. While my model is consistent with either of these interpretations – normative or positive – the results provable in it must obviously be viewed differently depending on whether the conditions on a transformation function are interpreted normatively or positively.

5 An impossibility result

Let me introduce five conditions on a transformation function.[34] Although each condition can be made plausible, I do not suggest that they are all equally compelling; indeed, I relax some of them below. However, they are

[33] For evidence of these two kinds of effects, see List, Luskin, Fishkin and McLean (2000/2006) and Sunstein (2002), respectively.

[34] As will be apparent, four of these conditions have analogues in the context of aggregation; one condition, *miminal relevance*, has no established counterpart in the literature on aggregation.

useful for analyzing the logical space of possible transformation functions.

Universal domain. The transformation function accepts as admissible input any possible profile of consistent and complete individual judgment sets.

Universal domain requires the transformation function to cope with conditions of pluralism on the input side, subject to the constraint of full individual rationality. A more demanding input condition would require it to cope also with less than fully rational individual judgments; but the theorem below shows that even the present requirement is far from undemanding.[35]

Rational co-domain. The transformation function produces as output a profile of consistent and complete individual judgment sets.

Rational co-domain requires the transformation function to generate outputs that also meet the constraint of full individual rationality. Later I consider a weaker variant of this requirement.

Consensus preservation. The transformation function maps any unanimous profile to itself.

Consensus preservation requires that, if all individuals agree on all propositions on the agenda before communication, this all-way consensus be preserved after communication. This is distinct from either of the following, arguably less plausible conditions. The first is *consensus generation*, the requirement that the transformation function map every profile to a unanimous profile, as captured by Elster's quote above. Although endorsed by many deliberative democrats, especially those of a Habermasian orientation,

[35] The present model certainly allows us to study, for example, transformation functions that map profiles of less than fully rational individual judgment sets into profiles of fully rational ones. Indeed, a frequently mentioned goal of deliberation is to correct rationality violations in individual judgments.

this requirement seems unduly demanding and empirically unrealistic. The second condition from which consensus preservation is distinct is *proposition-wise unanimity preservation*, the requirement that if all individuals agree on a particular proposition 'p' before deliberation, without necessarily agreeing on anything else, this unanimity on 'p' be preserved after deliberation. This requires that even an incompletely theorized agreement on 'p' be preserved in communication, even if different individuals agree on 'p' for incompatible reasons. Such a requirement is neither normatively compelling nor empirically realistic.[36] For example, upon noticing that you and I support 'p' for incompatible reasons, we may each decide to give up our belief in 'p'. By contrast, consensus preservation is the much milder requirement that an all-way consensus on everything – in those rare cases in which it occurs – be stable under communication.

To state the next condition, call two profiles *variants for a given individual* if they coincide for all individuals except the given one.

Minimal relevance. For each individual, there exists at least one admissible pair of variant input profiles for which the individual's output judgment sets differ.

Minimal relevance requires that individuals do not always ignore their pre-communication judgments. This is a very mild requirement: It only rules out that an individual's pre-communication judgments *never* make any difference to his or her post-communication judgments. It does not require those pre-communication judgments to make a difference more than once, nor does it say anything about *how* they should make a difference. Consistently with minimal relevance, the individual's post-communication judgments could even respond negatively to his or her pre-communication judgments.

[36] On incompletely theorized agreements, see Sunstein (1994) and, in the context of judgment aggregation, List (2006a). For critiques of propositionwise unanimity preservation, see Bradley (2007), Mongin (2005) and Nehring (2005).

Independence/systematicity. Each individual's output judgment on any proposition on the agenda depends only on the input judgments on that proposition across the group [and the pattern of dependence is the same across propositions]. (Again, independence omits, and systematicity includes, the neutrality clause in square brackets.)

Independence is a requirement of 'local' as opposed to 'holistic' communication: The post-communication judgments on any proposition should be determined by pre-communication judgments on that proposition and should not depend on pre-communication judgments on other propositions. Systematicity adds to this a neutrality requirement across propositions. In the climate change example, independence requires, for instance, that individuals' post-communication judgments on whether emissions above the relevant threshold would lead to the specified temperature increase ('if a then b') depend only on pre-communication judgments on this proposition and not on pre-communication judgments on, say, whether emissions are in fact above the threshold ('a'). In the group membership example, to give another illustration, independence requires that post-communication judgments on whether a particular candidate should be granted membership depend only on pre-communication judgments regarding this candidate, not on pre-communication judgments regarding other candidates.

Whether one considers independence plausible seems to depend, from a normative perspective, on whether a focus on one proposition at a time is deemed desirable in group communication and, from a positive one, on whether real-world communication processes display such a focus. However, the most compelling justification of independence is that it is a necessary condition for *strategy-proofness*. A communication process is *strategy-proof* if truthful expression of judgments is a weakly dominant strategy for every participant. Under sufficiently permissive assumptions about individual incentives, the requirement of strategy-proofness is met if and only if the transformation function satisfies independence and another condition called

monotonicity.[37] If one considers strategy-proofness desirable, as many deliberative democrats do, one may therefore have to endorse independence too.[38] Moreover, independence also prevents various forms of *agenda manipulability*, in which an agenda setter can manipulate the judgments on some propositions by including others in the agenda or excluding them from it.[39]

Although these five requirements on group communication may seem initially plausible, I now show that only a maximally conservative communication process can meet them all. Although the proof turns out to be remarkably simple *ex post*, the result is nonetheless surprising *ex ante*. Recall that the *identity function* is the transformation function that maps every profile to itself.

Theorem 2 *The only transformation function satisfying universal domain, rational co-domain, consensus preservation, minimal relevance and independence/systematicity is the identity function. (As before, whether the result requires independence or systematicity depends on how the minimal complexity of the agenda is defined.*[40]*)*

[37] This follows from related results on aggregation (Dietrich and List 2007c, Nehring and Puppe 2002). Monotonicity requires that any individual's post-communication acceptance of a given proposition should not be reversed if the pre-communication profile changes such that one additional individual supports the proposition in question and all other individuals' judgments remain the same.

[38] Independence (with monotonicity) is also equivalent to *non-manipulability* (Dietrich and List 2007c). A transformation function is *non-manipulable* if there exist no profile, individuals i, j, and proposition 'p' on the agenda such that i can manipulate j on 'p', i.e., (i) if i expresses his/her pre-communication judgment set truthfully, then j's post-communication judgment on 'p' disagrees with i's pre-communication judgment on 'p'; and (ii) if i misrepresents his/her pre-communication judgment set, then j's post-communication judgment on 'p' agrees with i's pre-communication judgment on 'p'. (The case $i = j$ rules out self-manipulation.)

[39] Variants of this point have been established by List (2004) and Dietrich (2006) in the context of judgment aggregation, but carry over to judgment transformation.

[40] Again, if the agenda meets only properties (i) and (ii), systematicity is needed for the

Proof. Consider any transformation function satisfying the conditions of Theorem 2. Notice that this transformation function can be decomposed into n separate functions, where the i-th such function maps each profile of individual judgment sets in the domain of the transformation function to individual i's output judgment set. Formally, each of these n functions – being a mapping from profiles of judgment sets to single judgment sets – is an aggregation function. Its interpretation is obviously different from the standard one: It is not the group that faces an *interpersonal* aggregation problem here, but each individual who faces an *intrapersonal* one, namely the problem of how to reconcile the judgments of the other individuals with his or her own judgments. Since the underlying transformation function satisfies universal domain, rational co-domain, consensus preservation and independence/systematicity – the condition of minimal relevance is not yet used – each induced aggregation function satisfies universal domain, collective rationality (here meaning rationality of the output judgment sets), consensus preservation and independence/systematicity. By Theorem 1 above, it is therefore a dictatorship of one individual.[41] This already shows that the underlying transformation function must be an opinion leader function, where each individual adopts as his or her output judgment set the input judgment set of some antecedently fixed individual, his or her opinion leader (the *dictator* in the terminology of the induced aggregation function). Could any individual's opinion leader be distinct from the individual him- or herself? Now the condition of minimal relevance comes into play. If any individual had another individual as his or her opinion leader, minimal relevance would be violated contrary to the proof's assumption: The individual's output judgment set would be invariant under any changes of his or her input judgment set. Each individual must therefore be his or her own opinion leader. Consequently, the transformation function is the identity

result; if it also meets property (iii), independence is enough.

[41] The qualifications regarding independence and systematicity in Theorem 1 (note 27) apply here too and thus carry over to Theorem 2.

function. This completes the proof.

Theorem 2 is an impossibility result, showing that 'effective' group communication is impossible under the given five conditions. In particular, they imply the 'unchanging minds hypothesis': Under them, there is no opinion change in communication. In consequence, the result casts doubt on these conditions.

From a normative perspective, one does not want to impose conditions on group communication that are so restrictive as to be met only by a degenerate communication process in which nobody ever changes his or her judgments. This would be against the spirit of the normative literature on deliberative democracy. Further, Theorem 2 implies that the five introduced conditions are inconsistent with the further condition of *consensus generation* discussed above, which is implicit in many writings on deliberative democracy, as illustrated by Elster's opening quote. If one did expect communication to produce consensus, one could not also expect it to meet the five introduced conditions.

From an empirical perspective, although 'it is frequently observed', as Mackie notes, 'that public deliberation on a pending item seldom seems to change anyone's mind',[42] group communication does not always exhibit the extreme conservatism implied by the theorem. There is plenty of empirical evidence that opinions do change in deliberative settings.[43] Let me therefore go through the conditions one by one and consider relaxing them.

6 Mapping out the possibilities

6.1 Relaxing universal domain

Universal domain requires the transformation function to cope with any level of pluralism in its input, subject only to the constraint of individual

[42] See Mackie (2006, p. 279), as quoted above.

[43] See, among many contributions, Luskin, Fishkin and Jowell (2002).

rationality. What happens if this is weakened to the requirement that it should cope only with those input profiles that exhibit a certain amount of cohesion among the individuals? Then there exist transformation functions other than the identity function that satisfy all the other conditions. An example is deference to the majority, which guarantees consistent post-communication judgments provided no profiles are deemed admissible in which distinct majorities support mutually inconsistent propositions. Could pre-communication judgments exhibit this amount of cohesion?

Suppose, for example, that even before communication the individuals agree on some cognitive or ideological dimension in terms of which to think about the propositions on the agenda – a 'meta-agreement' – and that, in consequence, the individuals can be aligned from left to right on that dimension such that, for each proposition on the agenda, the individuals accepting the proposition are either all to the left, or all to the right, of those rejecting it.[44] Deference to the majority is then guaranteed to yield consistent and – absent ties – complete post-communication judgments. Consider, for example, the individual judgments over the agenda containing 'a', 'if a then b' and 'b', as shown in Table 2, where the required left-right alignment of the individuals – here from 1 to 5 – holds.

	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5
'a'	True	False	False	False	False
'if a then b'	False	True	True	True	True
'b'	False	False	False	True	True

Table 2: Unidimensionally aligned judgments

Notice that the majority judgments in Table 2 coincide with the judgments of the median individual relative to the left-right alignment, here individual 3. Generally, given any profile of the form described, no proposition can be supported by a majority unless it is also supported by the median in-

[44] For a formal treatment of this kind of 'meta-agreement', see List (2002, 2003).

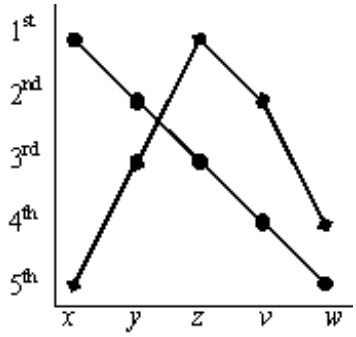

Figure 3: Single-peaked ranking judgments

dividual. So, by deferring to the majority, individuals inherit the consistent judgments of the median individual.[45] In this way, communication moves individual opinions in a centrist direction.

Another type of cohesion sufficient for consistent majority judgments applies to ranking judgments, as in example 3 above. Note that an individual's set of ranking judgments can be viewed as expressing a ranking of the given options (or candidates) from most to least preferable. Let some left-right ordering of these options be given; this could order them from most socialist to most capitalist, from most secular to most religious, from most urban to most rural, or in any other way. An individual's set of ranking judgments is called *single-peaked relative to that left-right ordering* if the individual has a most highly ranked option somewhere on the ordering with a decreasing ranking as options get more distant from it in either direction. This is illustrated by the two rankings in Figure 3 of the options x, y, z, v, w from most (1st) to least (5th) preferable. A profile (across individuals) is called *single-peaked* if there exists a left-right ordering of the options relative to which *all* individuals' ranking judgment sets are single-peaked. A classic result by Duncan Black shows that, for any single-peaked profile of ranking

[45] Assuming full rationality of that individual, in accordance with universal domain.

judgments, the resulting majority judgments are consistent.[46] Consequently, if pre-communication ranking judgments are single-peaked, individuals can form consistent post-communication judgments by deferring to the majority.

Does such pre-communication cohesion provide a plausible escape route from the impossibility result on judgment transformation? Although pre-communication profiles may sometimes exhibit the required amount of cohesion, this cannot generally be assumed to be the case. Often the aim of communication is precisely to deal with pluralism. While consensus, or some other form of cohesion, may ideally be the *output* of communication, requiring it as its *input* appears to miss the point of communication. Nonetheless, one possible interpretation of the impossibility result is that, *if* rational co-domain, consensus preservation, minimal relevance and independence/systematicity are required, *then* non-degenerate judgment transformation is possible only if individuals enter the process with sufficient initial cohesion.

6.2 Relaxing rational co-domain

Rational co-domain requires the individuals' output judgment sets to be both consistent and complete. Suppose this is weakened to the requirement that output judgment sets be merely consistent and deductively closed, where deductive closure means that individuals accept the implications of other accepted propositions, at least when they are also included in the agenda. Deductive closure is much less demanding than completeness,[47] as it is satisfied, for example, even by an empty judgment set. Requiring deductive closure, particularly in a deliberative setting, is plausible[48] because a frequently stated aim of proper deliberation is not just to lead people to form considered judgments on the propositions on the agenda but also to

[46] See Black (1948). Single-peakedness is one particular sufficient condition for consistent majority ranking judgments. A more general condition is *value-restriction* (Sen 1966).

[47] In the presence of consistency.

[48] At least when confined to propositions on the agenda, as assumed here.

make them aware of the implications of their judgments.

What happens if rational co-domain is relaxed in this way? Unfortunately, it does not open up a compelling escape route from the impossibility result. Any transformation function satisfying the weakened co-domain condition together with the other conditions – universal domain, consensus preservation, minimal relevance, independence/systematicity – is of the following form. For each individual, there exists a fixed subset of individuals in which he or she is included – his or her *peer group* (in the limiting case, this could be the singleton set containing only the individual him- or herself) – such that the individual's output judgment set is always the intersection of the input judgment sets among the individual's peers.[49] Arguably, such a transformation function is no better, and possibly worse, than the identity function: It has the property that each individual's output judgment set is always a subset of his or her input judgment set. At best an individual's judgment set remains unchanged after communication, at worst it shrinks. How much it shrinks depends on the size of the individual's peer group and the amount of disagreement among the peers. Such a transformation function perhaps instantiates the combination of a conservative and a sceptical attitude: An individual never comes to accept a proposition he or she did not accept in the first place and never continues to accept a proposition unless everyone in his or her peer group agrees with it.

6.3 Relaxing consensus preservation

Consensus preservation is the requirement that the transformation function map any unanimous profile to itself. Relaxing this requirement is not a very promising route. First, the requirement is already very mild, as argued

[49] This follows from a result on judgment aggregation without full rationality (Dietrich and List 2008, generalizing Gärdenfors 2006; see also Dokow and Holzman 2006). It still holds if the transformation function admits as input any profile of consistent and deductively closed judgment sets (not requiring completeness); a weakened independence/systematicity condition suffices for the result.

above. But, secondly, even if one were prepared to drop it, this would not lead very far: Under slightly stronger assumptions about how the propositions on the agenda are interconnected, Theorem 2 continues to hold even without consensus preservation.[50] I therefore set this route aside here.

6.4 Relaxing minimal relevance

Minimal relevance rules out that an individual's pre-communication judgment set never makes any difference to his or her post-communication judgment set. As in the case of consensus preservation, relaxing minimal relevance does not lead to a strong escape route from the impossibility theorem. Not only is minimal relevance a mild requirement, but, as shown in the proof of Theorem 2, its relaxation makes possible only a very restrictive class of transformation functions, namely that of opinion leader functions. Under an opinion leader function, each individual adopts as his or her output judgment set the input judgment set of an antecedently fixed individual, the individual's opinion leader. As noted, an opinion leader function is analogous to a dictatorial aggregation function except that different individuals may defer to different opinion leaders. Obviously, such a transformation function is plausible at most in special circumstances, for example when individuals have reasons to think that their opinion leaders have a special expertise on the agenda of propositions under consideration.

6.5 Relaxing independence/systematicity

Independence requires the transformation function to determine each output judgment on any proposition solely on the basis of the individuals' input judgments on that proposition, with systematicity requiring in addition that

[50] If systematicity is required, the relevant agenda assumption is the conjunction of (i), (ii) and a property called *asymmetry*; this follows from a result by Dietrich (2007). If only independence is required, the relevant agenda assumption is *atomic closure* or *atomicity*; this follows from results by Pauly and van Hees (2006) and Dietrich (2006).

the pattern of dependence be the same across propositions. As noted above, this can be viewed as a constraint of 'local' deliberation, requiring the consideration of one proposition at a time. Given the limited appeal of the previous escape routes from the impossibility result, it seems natural to relax independence.

If we give it up, one possibility is for each individual to designate some propositions on the agenda as 'premises' and others as 'conclusions' and to generate his or her post-communication judgments by deferring to the pre-communication majority judgment on each premise and then deriving the judgments on other propositions from these majority judgments on the premises. If the premises are chosen as a 'logical basis' for the entire agenda – that is, they are mutually independent and any assignment of truth-values to them settles the truth-values of all other propositions – the resulting transformation function guarantees consistent and complete post-communication judgments and satisfies all the other conditions introduced above. The choice of premises and conclusions need not be the same across individuals.

While the present class of transformation functions is the judgment-transformation analogue of the 'premise-based procedures' in the context of aggregation, a more general class of transformation functions draws on the 'sequential priority procedures' of aggregation.[51] Here each individual determines a particular order of priority among the propositions on the agenda, interpreting earlier propositions in that order as epistemically (or otherwise) prior to later ones. The individual then considers the propositions one-by-one in the chosen order and forms his or her post-communication judgment on each proposition as follows. If the pre-communication majority judgment on the proposition is consistent with the judgments the individual has made on propositions considered earlier, then he or she defers to that majority judgment; but if it is inconsistent with those earlier judgments, then he or she accepts the implications of those earlier judgments. In the case of Table

[51] On these two kinds of procedures of aggregation, see Pettit (2001) and List (2004), respectively.

1, for example, an individual may consider the propositions in the order 'a', 'if a then b', 'b' (with negations interspersed) and then accept 'a' and 'if a then b' by deferring to the pre-communication majority judgments while accepting 'b' by logical inference. The output profile under such a transformation function is sensitive to each individual's chosen order of priority among the propositions. This property of the transformation function can be seen as a virtue or as a vice, depending on one's perspective. On the one hand, it takes into account the fact that different propositions may have a different status.[52] But on the other hand, it makes individuals manipulable by a cunning Rikerian 'heresthetician' who leads them to consider propositions in a strategically adjusted order.[53] In the next section, I discuss a third and arguably particularly interesting class of transformation functions that becomes possible once independence/systematicity is dropped.

What is the cost of violating independence? As already noted, a transformation function violating it may be susceptible to strategic agenda setting as well as provide incentives for strategic misrepresentation of pre-communication judgments. To illustrate the latter, consider the climate change example with individual judgments as shown in Table 1 above, and suppose the experts form their post-communication judgments in the sequential manner just described, considering the propositions in the order 'a', 'if a then b', 'b'. Suppose, further, that the second expert, who does not accept that there will be the specified temperature increase (proposition 'b'), does not want his or her colleagues to accept that proposition either. Under the sequential transformation function, he or she may lead them to accept 'not b' by misrepresenting his or her judgment on 'a', strategically expressing the view that 'not a'. Such possibilities of strategic manipulation arise as soon as the transformation function violates independence.[54] The impossibility theorem presented can therefore be seen as describing a dilemma be-

[52] E.g., Pettit (2001) and Chapman (2002).

[53] See Riker (1986) and List (2004).

[54] See notes 37 and 38.

tween two problems of a group communication process: 'unchanging minds' on the one hand and strategic manipulability on the other.

7 Away from consensus

While proponents of deliberative democracy sometimes expect group communication to bring about consensus, there is very little empirical evidence of this effect.[55] It is also unclear whether achieving consensus is always normatively desirable. Moreover, if we do require transformation functions to satisfy consensus generation, then the problem of judgment transformation formally collapses into that of judgment aggregation, and the only remaining transformation functions satisfying universal domain, rational co-domain, consensus preservation and independence/systematicity – dropping minimal relevance – are those opinion leader functions in which all individuals defer to the same opinion leader, the equivalent of an Arrovian dictator. This is particularly ironic in so far as the possibility of a deliberation-induced consensus is often proposed as a solution to, not a variant of, the notorious problem of aggregation.

Could group communication bring about something less than consensus that is still helpful for democratic decision-making – for example, by facilitating the consistent aggregation of post-communication judgments? A recent literature suggests that the kind of group communication envisaged by deliberative democrats may have this effect, at least under favourable conditions.[56] Recall the earlier discussion of the possibility that individuals agree on some cognitive or ideological dimension in terms of which to think about the relevant propositions or, in the case of ranking judgments,

[55] For evidence from deliberative polls, see List, Luskin, Fishkin and McLean (2000/2006).

[56] See notes 2 and 10. William Riker (1982, p. 128) conceded that '[i]f, by reason of discussion, debate, civic education, and political socialization, voters have a common view of the political dimension ..., then a transitive outcome is guaranteed.'

the candidates or policy options. While such 'meta-agreement' may be too demanding as a pre-condition for communication, it can more plausibly be expected as its outcome. In an empirical study using data from James Fishkin's deliberative polls, such an effect has been identified.[57] Groups of between 150 and 350 randomly sampled participants were interviewed on their ranking judgments over multiple policy options both before and after a period of group deliberation. Deliberation increased the proximity of these ranking judgments to single-peakedness, as defined above – a deliberation-induced movement towards 'meta-agreement'.

Can we find an empirically plausible class of transformation functions to explain this effect? Let me introduce the class of *constrained minimal revision functions*. Under such a function, the transformation of judgments takes place in two stages. An input profile is given. At the first stage, the group identifies a particular set of judgment sets that are deemed admissible as output judgment sets conditional on the given input profile. Formally, this can be modelled as the application of a *focusing function*, which maps the input profile to a set of admissible output judgment sets. The latter set should ideally have the property that any profile constructible from it leads to consistent majority judgments. At the second stage, each individual selects an output judgment set from the identified set of admissible ones. Formally, this can be modelled as the application of a *minimal judgment revision policy*, under which each individual chooses an output judgment set from the set of admissible ones that is as close as possible to his or her input judgment set, relative to some distance metric over judgment sets.[58] This fixes the output profile and thereby completes the definition.

Informally, the first stage involves the identification of the opinions that can 'reasonably' be held after group communication given the opinions before communication; and the second stage involves a change of individual opinions such that each individual ends up holding one of the 'reasonable'

[57] See List, Luskin, Fishkin and McLean (2000/2006).

[58] On the notion of minimal revision, see, e.g., Schulte (2005).

ones. Whether an individual's post-communication opinion coincides with his or her pre-communication opinion depends on whether it was already among the ones identified as 'reasonable' during the first stage of the process.

Crucially, two ingredients of this definition allow a number of different specifications: the group's focusing function at the first stage, and each individual's distance metric over judgment sets at the second. Thus the definition specifies an entire class of transformation functions, one for each possible specification of these two ingredients.

To illustrate how a constrained minimal revision function can bring about a 'meta-agreement', suppose again a group deliberates about how to rank three or more policy options in an order of social preference (example 3), as in the deliberative polls studied empirically. The following constrained minimal revision function generates single-peaked output profiles. For a given pre-communication profile, it is first determined which left-right ordering of the options renders a maximal number of individuals' ranking judgment sets single-peaked, as defined above.[59] Now a ranking judgment set is deemed admissible if and only if it is single-peaked relative to the identified left-right ordering. This specifies the group's focusing function and completes the first stage. Each individual then minimally revises his or her ranking judgment set so as to adopt one of the admissible ones; here an individual's distance metric could be the *Hamming distance*, whereby the distance between any two judgment sets is the number of propositions on the agenda on which these judgment sets disagree.[60] This determines the post-communication profile and completes the second stage. By construction, this transformation function guarantees a single-peaked output profile.

Further empirical research is needed to test whether a suitable con-

[59] In List, Luskin, Fishkin and McLean (2000/2006), such a left-right ordering is called a *largest structuring dimension*.

[60] The Hamming distance has been applied to judgment aggregation by Pigozzi (2006). In the case of ranking judgment sets it captures the so-called *Kemeny distance* between the underlying preference rankings.

strained minimal revision function can explain the precise patterns of deliberation-induced opinion change observed in deliberative polls and other experiments. To the best of my knowledge, however, the present approach is the first attempt to model a deliberation-induced 'meta-agreement' formally.

From a normative perspective, further questions need to be asked on whether a suitable constrained minimal revision function captures the requirements of a good communication process as discussed in the literature on deliberative democracy. A constrained minimal revision function satisfies universal domain, rational co-domain, consensus preservation and minimal relevance, while violating independence/systematicity, but does it also satisfy some other desiderata of properly 'deliberative' communication?

Obviously, it does not generally satisfy consensus generation, unless the focusing function always picks out only one admissible output judgment set for each input profile. What about some other conditions? As already illustrated, a constrained minimal revision function may satisfy *cohesion generation*, where a profile is defined to be *cohesive* if it generates consistent majority judgments. Cohesion generation is a particularly appealing condition when the communication process precedes a majority decision.

Another condition is *stability under repeated rounds of communication*, the requirement that the transformation function map any output profile (that is, any profile in the function's range) to itself, or equivalently, that repeated applications of the function lead to the same output as a single application.[61] A constrained minimal revision function satisfies this condition so long as the group's focusing function and the individuals' distance

[61] Any transformation function satisfying both consensus preservation and consensus generation also satisfies this condition, because its first application leads to a unanimous profile and subsequent applications preserve this unanimity. By contrast, suppose the members of a group sit around a circular table and each individual defers to his or her neighbour on the right in forming his or her post-communication judgments. This transformation function – a special kind of opinion leader function – violates the stability condition, as each round of transformation yields a further permutation of the given profile across individuals.

metrics are sufficiently well-behaved.[62] Whether stability under repeated rounds of communication is a plausible requirement depends on how the judgment transformation function is interpreted. If it is meant to capture opinion change in a single round of group communication, then there is no reason to expect subsequent rounds of communication to leave opinions fixed. But if it is meant to capture a complete communication process up to the point of 'reflective equilibrium',[63] then the stability condition is very plausible, arguably more so than consensus generation.

8 Conclusion

I have formalized the problem of judgment transformation and proved a baseline impossibility result. My approach opens up a new way of analyzing group communication processes axiomatically, which allows us to determine which combinations of conditions on such processes are compatible with effective opinion change and which are not. Among the five basic conditions introduced, the first four, I have argued, can be relaxed only in special cases. The most plausible candidate for relaxation – under both normative and positive interpretations of the model – is the fifth condition: independence/systematicity.

This observation suggests that effective group communication processes as envisaged by deliberative democrats must exhibit a certain kind of holism: The objects of judgment transformation cannot generally be single propositions in isolation, but must be larger 'webs' of interconnected propositions. As noted, this holistic property of group communication is analogous to the holistic property of theory testing in science. In empirical science, too,

[62] The focusing function must have the property that it maps any output profile constructed from any set of judgment sets in its range to a new set of judgment sets that still contains all the judgment sets in the given output profile. The distance metric must have the standard property that the distance of any judgment set from itself is uniquely zero.

[63] E.g., Rawls (1971).

we cannot generally adjudicate single propositions in isolation, but only in conjunction with other logically connected propositions.

At one level, this is perhaps an unsurprising result. However, the cost of this holism in communication, as I have pointed out, is strategic manipulability: Individuals may be able to influence post-communication judgments on some propositions in a preferred manner by misrepresenting their pre-communication judgments on others. The communicative process may also be open to agenda manipulation. Thus we are faced with a trade-off between two problematic features of group communication: 'unchanging minds' on the one hand and strategic manipulability on the other. This trade-off, I think, is quite fundamental: Under a broad range of conditions, group communication processes *either* fail to be properly deliberative *or* violate strategy-proofness. We cannot generally have both: effective communication that changes minds *and* strategy-proofness.

Let me conclude with some remarks about how the present approach is related to game-theoretic approaches to studying group communication. While game-theoretic approaches seek to come up with a behavioural theory of group communication, analyzing individuals' incentives in communicative processes and predicting their behaviour on this basis,[64] the goal of the present approach is to map out the logical space of possible functional relations between pre- and post-communication judgments. The present approach is thus more akin to social choice theory in the tradition of Arrow than to game theory, and the relationship between the two approaches is similar to that between social choice theory and the theory of mechanism design, which has recently received a lot of attention following the award of the 2007 Nobel Prizes in Economics.[65] The former investigates possible functional relations between individual inputs and social outputs and the latter investigates the various mechanisms available (or unavailable) for implementing these functional relations under certain incentive constraints.

[64] See note 3.

[65] See the survey article by the Royal Swedish Academy of Sciences (2007).

Therefore I see the present approach as complementary to game-theoretic approaches, not as competitive.

In addition, there are important bridges between the two approaches. Since the condition of independence can be motivated game-theoretically – as a requirement for the incentive-compatibility of truthfulness – my conclusion that realistic group communication processes are likely to violate it reinforces a central game-theoretic question about group communication: How can we design communication processes that induce participants to reveal their judgments truthfully? Broadly, there are at least two ways to tackle this question. One may either go along the mechanism-design route and ask what communication processes ensure truthfulness by eliminating opportunities to benefit from strategic misrepresentation. Or one may go along a psychological route and ask under what conditions individuals are truthful even in the presence of strategic opportunities. Which of these routes – or which combination of them – is most promising remains a question for future research.

9 References

Alchourrón, C. E., P. Gärdenfors and D. Makinson. 1985. On the Logic of Theory Change: Partial Meet Contraction and Revision Functions. *Journal of Symbolic Logic*, 50 (2, June), 510-30.

Arrow, K. 1951/1963. *Social Choice and Individual Values*. New York: Wiley.

Austen-Smith, D., and T. J. Feddersen. 2006. Deliberation, Preference Uncertainty, and Voting Rules. *American Political Science Review*, 100 (2, May), 209-17.

Black, D. 1948. On the Rationale of Group Decision-making. *Journal of Political Economy*, 56 (1, February), 23-34.

Bohman, J., and W. Rehg (eds). 1997. *Deliberative Democracy: Essays on Reason and Politics*. Cambridge, MA: MIT Press.

Bovens, L., and W. Rabinowicz. 2006. Democratic Answers to Complex Questions – An Epistemic Perspective. *Synthese*, 150 (1, May), 131-53.

Bradley, R. 2007. Reaching a Consensus. *LSE Choice Group Working Papers*, 3 (3).

Calvert, R. 2006. Deliberation as Coordination through Cheap Talk. Working paper, Washington University, St. Louis.

Chapman, B. 2002. Rational Aggregation. *Politics, Philosophy and Economics*, 1 (3, October), 337-54.

Cohen, J. 1989. Deliberation and Democratic Legitimacy. In A. Hamlin and P. Pettit (eds), *The Good Polity: Normative Analysis of the State*. Oxford: Blackwell.

Dietrich, F. 2006. Judgment Aggregation: (Im)Possibility Theorems. *Journal of Economic Theory*, 126 (1, January), 286-98.

Dietrich, F. 2007. A generalised model of judgment aggregation. *Social Choice and Welfare*, 28 (4, June), 529-65.

Dietrich, F., and C. List. 2007a. Arrow's theorem in judgment aggregation. *Social Choice and Welfare*, 29 (1, July), 19-33.

Dietrich, F., and C. List. 2007b. Judgment aggregation by quota rules. *Journal of Theoretical Politics*, 19 (4, October), 391-424.

Dietrich, F., and C. List. 2007c. Strategy-proof judgment aggregation. *Economics and Philosophy*, 23 (3, November), 269-300.

Dietrich, F., and C. List. 2008. Judgment aggregation without full rationality. *Social Choice and Welfare,* 31, 15-39.

Dokow, E., and R. Holzman. Forthcoming. Aggregation of binary evaluations. *Journal of Economic Theory*.

Dokow, E., and R. Holzman. 2006. Aggregation of binary evaluations with abstentions. Working paper, Technion Israel Institute of Technology.

Dryzek, J. 1990. *Discursive Democracy: Politics, Policy and Political Science*. New York: Cambridge University Press.

Dryzek, J. 2000. *Deliberative Democracy and Beyond*. Oxford: Oxford

University Press.

Dryzek, J., and C. List. 2003. Social Choice Theory and Deliberative Democracy: A Reconciliation. *British Journal of Political Science*, 33 (1, January), 1-28.

Elster, J. 1986. The Market and the Forum. Pp. 103-32 in Elster, J., and A. Hylland (eds), *Foundations of Social Choice Theory*. Cambridge: Cambridge University Press.

Fishkin, J. S. 1991. *Democracy and Deliberation*. New Haven: Yale University Press.

Gärdenfors, P. 2006. A Representation Theorem for Voting with Logical Consequences. *Economics and Philosophy*, 22 (2, July), 181-90.

Gauwin, O., S. Konieczny and P. Marquis. 2005. Conciliation and Consensus in Iterated Belief Merging. Pp. 514-26 in *Proceedings of the Eighth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty* (ECSQARU'05).

Gutman, A., and D. Thompson. 1996. *Democracy and Disagreement*. Cambridge, MA: Harvard University Press.

Hafer, C., and D. Landa. Forthcoming. Deliberation as Self-Discovery and Institutions for Political Speech. *Journal of Theoretical Politics*.

van Hees, M. 2007. The limits of epistemic democracy. *Social Choice and Welfare,* 28 (4, June), 649-66.

Kasher, A., and A. Rubinstein. 1997. On the Question 'Who is a J?': A Social Choice Approach. *Logique et Analyse*, 40 (160), 385-95.

Knight, J., and J. Johnson. 1994. Aggregation and Deliberation: On the Possibility of Democratic Legitimacy. *Political Theory*, 22 (2, May), 277-96.

Konieczny, S. 2004. Belief base merging as a game. *Journal of Applied Non-Classical Logics*, 14 (3), 275-94.

Konieczny, S. and R. Pino Pérez. 2002. Merging Information Under Constraints: A Logical Framework. *Journal of Logic and Computation*, 12 (5, October), 773-808.

Kornhauser, L. A., and L. G. Sager. 1986. Unpacking the Court. *Yale Law Journal*, 96 (1, November), 82-117.

Landa, D., and A. Meirowitz. 2006. Information, Game Theory, and Deliberative Democracy. Working paper, New York University.

Lehrer, K. and C.Wagner. 1981. *Rational Consensus in Science and Society*. Dordrecht/Boston: Reidel.

List, C. 2002. Two Concepts of Agreement. *The Good Society*, 11 (1), 72-9.

List, C. 2003. A Possibility Theorem on Aggregation over Multiple Interconnected Propositions. *Mathematical Social Sciences*, 45 (1, February), 1-13.

List, C. 2004. A Model of Path Dependence in Decisions over Multiple Propositions. *American Political Science Review*, 98 (3, August), 495-513.

List, C. 2006a. The Discursive Dilemma and Public Reason. *Ethics,* 116 (2, January), 362-402.

List, C. 2006b. When to defer to supermajority testimony – and when not. Working paper, London School of Economics.

List, C., R. C. Luskin, J. Fishkin and I. McLean. 2006/2006. Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy: Evidence from Deliberative Polls. Working paper, London School of Economics.

List, C., and P. Pettit. 2002. Aggregating Sets of Judgments: An Impossibility Result. *Economics and Philosophy*, 18 (1, April), 89-110.

List, C., and P. Pettit. 2004. Aggregating Sets of Judgments: Two Impossibility Results Compared. *Synthese*, 140 (1-2, May), 207-35.

Luskin, R. C., J. S. Fishkin, and R. Jowell. 2002. Considered Opinions: Deliberative Polling in Britain. *British Journal of Political Science*, 32 (3, July), 455-87.

Mackie, G. 2006. Does democratic deliberation change minds? *Politics, Philosophy and Economics*, 5 (3, October), 279-303.

Mongin, P. 2005. Spurious Unanimity and the Pareto Principle. *LSE Choice Group Working Papers*, 1 (5).

Miller, D. 1992. Deliberative Democracy and Social Choice. *Political Studies*, 40 (Special Issue), 54-67.

Nehring, K. 2003. Arrow's theorem as a corollary. *Economics Letters*, 80 (3, September), 379-82.

Nehring, K. 2005. The Impossibility of a Paretian Rational. Working paper, University of California at Davies.

Nehring, K., and C. Puppe. 2002. Strategyproof Social Choice on Single-Peaked Domains: Possibility, Impossibility and the Space Between. Working paper, University of California at Davies.

Nehring, K. and C. Puppe. 2008. Consistent judgement aggregation: the truth-functional case. *Social Choice and Welfare*, 31, 41-57.

Pauly, M., and M. van Hees. 2006. Logical Constraints on Judgment Aggregation. *Journal of Philosophical Logic*, 35 (6, December), 569-85.

Pettit, P. 2001. Deliberative Democracy and the Discursive Dilemma. *Philosophical Issues*, 11 (October), 268-99.

Pettit, P. 2006. When to defer to majority testimony – and when not. *Analysis*, 66 (3, July), 179-87.

Pigozzi, G. 2006. Belief merging and the discursive dilemma: an argument-based account to paradoxes of judgment aggregation. *Synthese*, 152 (2, September), 285-98.

Quine, W. V. O. 1951. Two Dogmas of Empiricism. *Philosophical Review*, 60 (1, January), 20-43.

Rawls, J. 1971. *A Theory of Justice*. Oxford: Oxford University Press.

Riker, W. H. 1982. *Liberalism against Populism*. San Francisco: W. H. Freeman.

Riker, W. H. 1986. *The Art of Political Manipulation*. New Haven: Yale University Press.

Royal Swedish Academy of Sciences. 2007. Mechanism Design Theory:

Scientific background on the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2007. Online paper, Stockholm. Available at: http://nobelprize.org/nobel_prizes/economics/laureates/2007/ecoadv07.pdf

Rubinstein, A., and P. C. Fishburn. 1986. Algebraic Aggregation Theory. *Journal of Economic Theory*, 38 (1, February), 63-77.

Schulte, O. 2005. Minimal belief change, Pareto-optimality and logical consequence. *Economic Theory*, 19 (1, January), 105-44.

Sen, A. K. 1966. A Possibility Theorem on Majority Decisions. *Econometrica*, 34 (2, April), 491-99.

Sunstein, C. R. 1994. Political Conflict and Legal Agreement. Tanner Lectures on Human Values, Harvard University.

Sunstein, C. R. 2002. The Law of Group Polarization. *Journal of Political Philosophy*, 10 (2, June), 175-95.

Wilson, R. 1975. On the Theory of Aggregation. *Journal of Economic Theory*, 10 (1, February), 89-99.