From degrees of belief to beliefs: Lessons from judgment-aggregation theory

Dietrich, Franz and List, Christian

Paris School of Economics, London School of Economics

August 2014

From Degrees of Belief to Beliefs: Lessons from Judgment-Aggregation Theory*

Franz Dietrich
CNRS & University of East Anglia

Christian List
London School of Economics

August 2014

Abstract

What is the relationship between degrees of belief and (all-or-nothing) beliefs? Can the latter be expressed as a function of the former, without running into paradoxes? We reassess this "belief-binarization" problem from the perspective of judgment-aggregation theory. Although some similarities between belief binarization and judgment aggregation have been noted before, the literature contains no general study of the implications of aggregation-theoretic impossibility and possibility results for belief binarization. We seek to fill this gap. At the centre of this paper is an impossibility theorem showing that, except in simple cases, there exists no belief-binarization rule satisfying four baseline desiderata ("universal domain", "belief consistency and completeness", "propositionwise independence", "certainty preservation"). We show that this result is a corollary of the judgment-aggregation variant of Arrow's impossibility theorem and explore several escape routes from it.

1 Introduction

We routinely make belief ascriptions of two kinds. We speak of an agent's *degrees of belief* in some propositions and also of the agent's *beliefs* simpliciter. On the standard picture, *degrees of belief* (or *credences*) take the form of subjective probabilities the agent assigns to the propositions in question, such as when he or she assigns a subjective probability of $\frac{1}{2}$ to the proposition that a coin, which has been tossed but not yet observed, has landed

*This paper was presented at a seminar at the Australian National University, Canberra, in July 2014. We are very grateful to the participants for their comments. We also thank Geoffrey Brennan, John Broome, David Chalmers, Alan Hájek, James Joyce, Hanti Lin, Aidan Lyon, and Daniel Stoljar for helpful conversations. Christian List further wishes to thank the Australian National University and the Leverhulme Trust for supporting this work.

"heads". *Beliefs* take the form of the agent's overall acceptance of some propositions and non-acceptance of others, such as when the agent accepts that the sun will rise tomorrow or that $2 + 2 = 4$, while he or she does not accept that there are presently humans on mars. The agent's *belief set* consists of all the propositions that he or she accepts. What is the relationship between degrees of belief and beliefs? Can the latter be expressed as a function of the former and, if so, what does this function look like, formally speaking? Call this the *belief-binarization problem.*

A widely studied class of belief-binarization rules is the class of *threshold rules.* According to a threshold rule, an agent believes a proposition (in the all-or-nothing sense) if and only if he or she has a high-enough degree of belief in it. Threshold rules, however, run into well-known problems such as *lottery paradoxes* (e.g., Kyburg 1961). Suppose, for example, that an agent believes of each lottery ticket among a million tickets that this ticket will not win, since his or her degree of belief in this proposition is 0.999999 (which, for the sake of argument, counts as "high enough"). The believed propositions then jointly imply that none of the tickets will win. However, the agent knows that this proposition is false, and indeed has a degree of belief of 1 in its negation: some ticket will win. This illustrates that, under a threshold rule, the agent's belief set may be neither implication-closed (the agent fails to believe some implications of propositions he or she believes) nor logically consistent (the agent believes an inconsistent set of propositions).

The belief-binarization problem has recently received renewed attention. For example, Leitgeb (2014) defends the view that rational belief is formally equivalent to the assignment of a stably high rational degree of belief (where this is understood as a joint constraint on degrees of belief and beliefs, not as a reduction of one side to the other), and he develops a relevant notion of stability. Lin and Kelly (2012a, 2012b) use a combination of geometric and logical insights to defend a class of belief-acceptance rules that avoid lottery paradoxes, and investigate whether and when reasoning with beliefs can track reasoning with degrees of belief. In earlier work, Hawthorne and Bovens (1999) investigate how to make threshold rules consistent, and Douven and Williamson (2006) prove some triviality results, showing that belief-binarization rules based on what they call a "structural" criterion for the acceptance of any proposition must require a threshold of 1 for belief *or* fail to ensure consistency.

Our aim is to reassess the problem of belief binarization from a different perspective: that of judgment-aggregation theory. This is the branch of social choice theory that investigates how we can aggregate several individuals' judgments on logically connected propositions, such as p, q, and $p \wedge q$, into corresponding collective judgments.[1] A multi-

[1] For the purposes of this paper, salient contributions include List and Pettit (2002, 2004), Pauly and

member court, for example, may wish to aggregate its members' judgments on whether a defendant did some action (p), whether that action was prohibited by a contract (q), and whether the defendant is liable for breach of contract (for which $p \wedge q$ is necessary and sufficient). Finding methods of aggregation that secure consistent collective judgments and satisfy some other desiderata is surprisingly difficult. In our example, there might be a majority for p, a majority for q, and yet a majority against $p \wedge q$, which illustrates that majority rule is not generally satisfactory. Majority rule's failure to secure consistent and implication-closed collective judgments is reminiscent of a threshold rule's failure to secure consistent and implication-closed beliefs in the case of belief binarization. We will show that this reminiscence is not accidental: several key results in judgment-aggregation theory have important implications for belief binarization, which are straightforwardly derivable once the formal apparatus of judgment-aggregation theory is suitably adapted.

Of course, some similarities between lottery paradoxes and the paradoxes of judgment aggregation have been discussed before (especially by Levi 2004, Douven and Romeijn 2007, and Kelly and Lin 2011). But so far the focus has been on identifying lessons for judgment aggregation that can be learnt from existing work on the lottery paradox, not the other way round. The literature contains no general study of the implications of aggregation-theoretic impossibility and possibility results for belief binarization. We seek to fill this gap.[2]

At the centre of this paper is a "baseline" impossibility theorem, which we use to map out the space of possible solutions to the belief-binarization problem. The theorem says that, except in limiting cases (which we characterize precisely), there exists no belief-binarization rule satisfying four desiderata:

(i) *universal domain*: the rule should always work;

(ii) *consistency and completeness of beliefs*: beliefs should be logically consistent

van Hees (2006), Dietrich (2006, 2007), Dietrich and List (2007a, b, 2013), Nehring and Puppe (2010), and Dokow and Holzman (2010a, b). For a survey, see List (2012). Judgment-aggregation theory was inspired by earlier work on the "doctrinal paradox" in law and economics (Kornhauser and Sager 1986). Social choice theory in its traditional form, which goes back to Nicolas de Condorcet in the 18th century and Kenneth Arrow in the 20th, does not address the aggregation of judgments on logically connected propositions, but focuses on the aggregation of preferences over two or more options (e.g., rankings of electoral candidates or policy proposals). It has recently been shown, however, that several classic social-choice-theoretic results can be generalized from preference aggregation to judgment aggregation.

[2]Douven and Romeijn (2007, p. 318) conclude their paper with an invitation to conduct the kind of study we embark on here: "given the liveliness of the debate on judgement aggregation, and the many new results that keep coming out of that, it is not unrealistic to expect that at least some theorems originally derived, or still to be derived, within that context can be applied fruitfully to the context of the lottery paradox, and will teach us something new, and hopefully also important, about this paradox."

and complete, as explained in more detail later;

(iii) *propositionwise independence*: the belief in any proposition should depend only on the degree of belief in it; and

(iv) *certainty preservation*: in the special case in which the degrees of belief are already "binary", taking only the values 0 or 1, these beliefs should be preserved as the all-or-nothing beliefs.

To find a possible belief-binarization rule, we must relax at least one of these desiderata. Notably, this result is a corollary of the judgment-aggregation variant of Arrow's classic impossibility theorem (in the form proved in Dietrich and List 2007a, Dokow and Holzman 2010a, and Nehring and Puppe 2010). Arrow's theorem, which was originally proved in preference-aggregation theory (1951/1963), demonstrates the non-existence of any non-dictatorial methods of aggregation satisfying some plausible desiderata. Our analysis shows that the Arrovian impossibility carries over to belief binarization and, in consequence, that the lottery paradox and the paradoxes of social choice can be traced back to a common source.

The present paper does not compete with, but rather complements, existing work on the relationship between degree of belief and belief. Our aim is not to defend a particular solution to the belief-binarization problem, but to map out the space of possible solutions in a novel way. Although some of our findings no doubt replicate – or provide alternative derivations of – earlier insights, our approach offers a new perspective on the relevant logical space, and, to the best of our knowledge, our central impossibility theorem, in its full generality, is new. Furthermore, we believe that there is value in linking two different areas of study: judgment aggregation and belief binarization.

One important qualification is due. We here explore the belief-binarization problem as a *formal* problem. When we ask whether an agent's all-or-nothing beliefs can be expressed as a *function* of his or her degrees of belief, we use the term "function" in the mathematical sense: a function is simply a special kind of relation and should not be interpreted in any metaphysically loaded way. Nonetheless, our investigation is relevant to a number of metaphysical, psychological, and epistemological questions. We may be interested, for example, in whether an agent really has *both* degrees of belief *and* all-or-nothing beliefs, or whether one of the two modes of ascription – say, that of all-or-nothing beliefs – is just a simplified shorthand for the other – say, a summary of the agent's degrees of belief. Furthermore, even if an agent has both kinds of belief, we may be interested in whether one kind – say all-or-nothing belief – is reducible to the other – say degree-of-belief – or whether no such reduction is possible. And even if neither kind of belief can be reduced to the other, we may still be interested in whether there is some systematic connection between the two (say, a connection of supervenience), or whether they are, in principle, quite independent from one another. Finally, we may be interested in how *rational* beliefs relate to *rational* degrees of belief, even if, in the absence of rationality, the two could come apart. A formal analysis of the belief-binarization problem is relevant to all of these questions. It can tell us what formal constraints the relationship between degrees of belief and beliefs could, or could not, satisfy, thereby constraining the substantive philosophical views one can consistently hold on this matter.

2 The parallels between belief binarization and judgment aggregation

To give a first flavour of the parallels between belief binarization and judgment aggregation, we begin with a simple example of a judgment-aggregation problem, which echoes our earlier example of the multi-member court. Suppose a committee of three experts has to make collective judgments on the propositions p, q, r, $p \wedge q \wedge r$, and their negations on the basis of the committee members' individual judgments. The individual judgments are as shown in Table 1.

Table 1: A judgment-aggregation problem

	p	q	r	$p \wedge q \wedge r$	$\neg(p \wedge q \wedge r)$
Individual 1	True	True	False	False	True
Individual 2	True	False	True	False	True
Individual 3	False	True	True	False	True
Proportion of support	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	0	1

The difficulty lies in the fact that there are majorities (in fact, two-thirds majorities) in support of each of p, q, and r, but the conjunction of these propositions, $p \wedge q \wedge r$, is unanimously rejected and its negation, $\neg(p \wedge q \wedge r)$, unanimously accepted. Majority voting, or any supermajority rule under which a quota of $\frac{2}{3}$ is sufficient for the collective acceptance of any proposition, yields a set of collective judgments that is neither implication-closed (it fails to include $p \wedge q \wedge r$ despite the inclusion of p, q, and r) nor consistent (it includes all of p, q, r, and $\neg(p \wedge q \wedge r)$). Pettit (2001) has called problems of this sort *discursive dilemmas*, though they are perhaps best described simply as *majority inconsistencies*. A central goal of the theory of judgment aggregation is to find aggregation rules that generate consistent and/or implication-closed collective judgments while also satisfying some other desiderata (List and Pettit 2002).

A belief-binarization problem can take a similar form. Suppose an agent seeks to convert his or her degrees of belief in the propositions p, q, r, $p \wedge q \wedge r$, and their negations into all-or-nothing beliefs. Suppose, specifically, the agent assigns an equal subjective probability of $\frac{1}{3}$ to each of three distinct possible worlds, in which p, q, and r have different truth-values, as shown in Table 2. Each world renders two of p, q, and r true and the other false. The bottom row of the table shows the agent's overall degrees of belief in the propositions.

Table 2: A belief-binarization problem

	p	q	r	$p \wedge q \wedge r$	$\neg(p \wedge q \wedge r)$
World 1 (subj. prob. $\frac{1}{3}$)	True	True	False	False	True
World 2 (subj. prob. $\frac{1}{3}$)	True	False	True	False	True
World 3 (subj. prob. $\frac{1}{3}$)	False	True	True	False	True
Degree of belief	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	0	1

Here the difficulty lies in the fact that while the agent has a relatively high degree of belief (namely $\frac{2}{3}$) in each of p, q, and r, his or her degree of belief in the conjunction of these propositions is 0, and the agent's degree of belief in its negation is 1. Any threshold rule under which a degree of belief of $\frac{2}{3}$ suffices for full belief in any proposition (and, *a fortiori*, any rule with a lower threshold) yields a belief set that is neither implication-closed nor consistent. On the other hand, if we demand a higher threshold for including a proposition in the agent's belief set, that belief set will include only $\neg(p \wedge q \wedge r)$ and will therefore be incomplete with respect to many proposition-negation pairs (accepting neither p, nor $\neg p$, for example). Other examples can be constructed in which more demanding threshold rules also lead to inconsistencies.

If we identify individual voters in Table 1 with possible worlds in Table 2, the parallels between our two problems should be evident. In this simple analogy, possible worlds in a belief-binarization problem play the role of individuals in a judgment-aggregation problem, and the agent's degree of belief in any proposition plays the role of the proportion of individuals accepting that proposition. In fact, the function that assigns to each proposition in a judgment-aggregation problem the proportion of individuals supporting it behaves formally like a probability function over these propositions: it satisfies the constraints of probabilistic coherence (assuming individual judgment sets are consistent and complete), though it is interpretationally very different. This already suggests that belief-binarization and judgment-aggregation problems are structurally similar (for earlier discussions of this similarity, see Levi 2004 and Douven and Romeijn 2007).

Yet, there is an important difference in format. In a judgment-aggregation problem, we are usually given the entire *profile of individual judgments,* i.e., the full list of the individuals' judgment sets, as in the first three rows of Table 1. In a belief-binarization-problem, by contrast, we are only given an agent's *degrees of belief in the relevant propositions*, i.e., the last row of Table 2, summarizing his or her overall subjective probabilities. The possible worlds underpinning these probabilities are hidden from view. Thus the input to a belief-binarization problem corresponds, not to a fully specified *profile of individual judgment sets*, but to a *propositionwise anonymous profile*, i.e., a specification of the *proportions* of individuals supporting the various propositions under consideration. This gives us, not a full table such as Table 1, but only its last row.[3] Indeed, in our subsequent formal analysis, possible worlds drop out of the picture.

In sum, a belief-binarization problem corresponds to a *propositionwise anonymous* judgment-aggregation problem, the problem of how to aggregate the final row of a table such as Table 1 into a single judgment set. We can view this as an aggregation problem *with a special restriction*: namely that when we determine the collective judgments, we must pay attention only to the *proportions* of individuals supporting each proposition and must disregard, for example, who holds which judgment set. A belief-binarization problem will then have a solution of a certain kind if and only if the corresponding propositionwise anonymous judgment-aggregation problem has a matching solution.

Of course, the theory of judgment aggregation has primarily focused, not on the aggregation of propositionwise anonymous profiles (final rows of the relevant tables), but on the aggregation of fully specified profiles (lists of judgment sets across all individuals, without the special restriction we have mentioned). We will see, however, that despite the more restrictive informational basis of belief binarization several results from judgment-

[3]There is another notion of an anonymous profile, with which the notion of a *propositionwise* anonymous profile should not be confused. While a *propositionwise anonymous profile* specifies the proportion of individuals supporting each proposition, an *anonymous profile* simpliciter specifies the proportion of individuals supporting each *combination of judgments* across the relevant propositions. The fully specified profile in Table 1, for instance, gives rise to an anonymous profile in which the judgment sets $\{p,q,\neg r,\neg(p \wedge q \wedge r)\}$, $\{p,\neg q,r,\neg(p \wedge q \wedge r)\}$, and $\{\neg p,q,r,\neg(p \wedge q \wedge r)\}$ each receive the support of $\frac{1}{3}$ of the individuals, which, in turn, gives rise to a propositionwise anonymous profile in which p, q, and r each receive the support of $\frac{2}{3}$ of the individuals, $p \wedge q \wedge r$ receives the support of none of them, and $\neg(p \wedge q \wedge r)$ receives the support of all of them. Propositionwise anonymous profiles correspond to equivalence classes of anonymous profiles, which, in turn, correspond to equivalence classes of fully specified profiles. Degree-of-belief functions are structurally equivalent to propositionwise anonymous profiles.

aggregation theory carry over.[4] We will now make this precise.

3 Belief binarization formalized

We begin with a formalization of the belief-binarization problem. Let X be the set of propositions on which beliefs are held (where propositions are subsets of some underlying set of worlds).[5] We call X the *proposition set*. For the moment, our only assumption about the proposition set is that it is non-empty and closed under negation (i.e., for any proposition p in X, its negation $\neg p$ is also in X). In principle, the proposition set could be an entire *algebra* of propositions, i.e., a set of propositions that is closed under negation and conjunction (and thereby also under disjunction).

A *degree-of-belief function* is a function Cr that assigns to each proposition p in X a number $Cr(p)$ in the interval from 0 to 1, where this assignment is probabilistically coherent.[6] A *belief set* is a subset $B \subseteq X$. It is called *consistent* if B is a consistent set; it is called *complete* (relative to X) if it contains a member of each proposition-negation pair $p, \neg p$ in X; and it is called *implication-closed* (relative to X) if it contains every proposition p in X that is entailed by B. Consistency and completeness jointly imply implication-closure.

A *belief-binarization rule* for X is a function f that maps each degree-of-belief function Cr on X (within some domain of admissible degree-of-belief functions) to a belief set $B = f(Cr)$. An important class of binarization rules is the class of *threshold rules.*

[4] Douven and Romeijn (2007) and Kelly and Lin (2011) proceed the other way round and derive some impossibility results for anonymous judgment aggregation from analogous results on belief binarization. These results differ from the canonical "Arrovian" impossibility result on judgment aggregation, on which we focus here (Dietrich and List 2007a, Dokow and Holzman 2010a, and Nehring and Puppe 2010). The latter cannot be derived from any results on belief binarization because of the richer informational basis of judgment aggregation. As noted, belief-binarization problems correspond to *propositionwise anonymous* judgment-aggregation problems, not to judgment-aggregation problems simpliciter.

[5] The following background definitions apply. Let Ω be some non-empty set of possible worlds. A *proposition* is a subset $p \subseteq \Omega$. For any proposition p, we write $\neg p$ to denote the complement (*negation*) of p, i.e., $\Omega \backslash p$. For any two propositions p and q, we write $p \wedge q$ to denote the intersection (*conjunction*) of p and q, i.e., $p \cap q$; and we write $p \vee q$ to denote their union (*disjunction*), i.e., $p \cup q$. A set S of propositions is *(logically) consistent* if its intersection is non-empty, i.e., $\bigcap_{p \in S} p \neq \emptyset$; S *(logically) entails* another proposition q if the intersection of all propositions in S is a subset of q, i.e., $\bigcap_{p \in S} p \subseteq q$. A proposition p is *tautological* if $p = \Omega$ and *contradictory* if $p = \emptyset$.

[6] Formally, Cr is a function from X into $[0, 1]$ which is extendable to a probability function (with standard properties) on the algebra generated by X. The *algebra generated by* X is the smallest algebra including X.

Here there exists some threshold t in $[0, 1]$, which can be either *strict* or *weak*, such that, for every admissible degree-of-belief function Cr, the belief set B is the following:

$$B = \{p \in X : Cr(p) \text{ exceeds } t\},$$

where "$Cr(p)$ exceeds t" means

$$\begin{aligned} & Cr(p) > t \text{ in the case of a strict threshold} \\ \text{and} \quad & Cr(p) \geq t \text{ in the case of a weak threshold.} \end{aligned}$$

More generally, we can relativize thresholds (and their designations as *strict* or *weak*) to the propositions in question. To do this, we must replace t in the expressions above with t_p, the *proposition-specific* threshold for proposition p, where each proposition-specific threshold can again be deemed to be either *strict* or *weak*. If the acceptance threshold (or its designation as *strict* or *weak*) differs across different propositions, we speak of a *non-uniform* threshold rule, to distinguish it from the *uniform* rules with an identical specification of the threshold for all propositions. Threshold rules are by no means the only possible belief-binarization rules; later, we consider other examples.

We now introduce four desiderata that we might, at least initially, expect a belief-binarization rule to meet. The first desideratum says that the belief-binarization rule should always work, no matter which degree-of-belief function is fed into it as input.

Universal domain. The domain of f is the set of all degree-of-belief functions on X.

The second desideratum says that the belief set generated by the belief-binarization rule should always be consistent and complete (relative to X). While consistency is a plausible requirement on the belief set B (though we later consider its relaxation too), one may object that completeness is too demanding, since it rules out suspending belief on some proposition-negation pairs. Indeed, it would be implausible to defend completeness as a general requirement of rationality. However, for the purpose of characterizing the logical space of possible belief-binarization rules, it is a useful starting point. Remember that completeness is required only relative to X, the proposition set under consideration. We relax the completeness requirement in our subsequent discussion.

Belief consistency and completeness. For every Cr in the domain of f, the belief set $B = f(Cr)$ is consistent and complete.

The third desideratum is another useful baseline requirement. It says that the belief in any proposition p should depend only on the degree of belief in p, not on the degree of belief in other propositions.

Propositionwise independence. For any Cr and Cr' in the domain of f and any p in X, if $Cr(p) = Cr'(p)$ then $p \in B \Leftrightarrow p \in B'$, where $B = f(Cr)$ and $B' = f(Cr')$.

The final desideratum is quite minimal. It says that, in the special case in which the degree-of-belief function is already binary (i.e., it only ever assigns degrees of belief 0 or 1 to the propositions in X), the resulting all-or-nothing beliefs should be exactly as specified by that degree-of-belief function.

Certainty preservation ("no change if beliefs are already binary"). For any Cr in the domain of f, if Cr already assigns extremal degrees of belief (0 or 1) to all propositions in X, then, for every proposition p in X, B contains p if $Cr(p) = 1$ and B does not contain p if $Cr(p) = 0$, where $B = f(Cr)$.

It is easy to see that, at least in simple cases, these desiderata can be met by a suitable threshold rule. For example, if the proposition set X contains only a single proposition p and its negation $\neg p$ (or many logically independent proposition-negation pairs), the four desiderata are met by any threshold rule that uses a (strict) threshold t for p and a (weak) threshold $1 - t$ for $\neg p$, where $0 \leq t < 1$. As we will see below, things become more difficult once the proposition set X is more complex.

4 Judgment aggregation formalized

We now move on to the formal definition of a judgment-aggregation problem (following the logic-based model developed in List and Pettit 2002 and Dietrich 2007). The proposition set X remains as defined in the last section and is now interpreted as the set of propositions on which judgments are to be made. In judgment-aggregation theory, this set is also called the *agenda*. Let there be a finite set $N = \{1, 2, ..., n\}$ of individuals (with $n \geq 2$). Each individual i holds a *judgment set*, labelled J_i, which is defined just like a belief set in the previous section (the name "judgment set" is purely conventional). So J_i is a subset of X, which is called *consistent*, *complete*, and *implication-closed* if it has the respective properties, as defined above. As before, consistency and completeness jointly imply implication-closure. A combination of judgment sets across the n individuals, $(J_1, ..., J_n)$, is called a *profile*. An example of a profile is given by the first three rows of Table 1 above. Here, the underlying set X consists of p, q, r, $p \wedge q \wedge r$, and their negations.

A *judgment-aggregation rule* for X is a function F that maps each profile of individual judgment sets (within some domain of admissible profiles) to a collective judgment set

J. Like the individual judgment sets, the collective judgment set J is a subset of X. The best-known example of a judgment-aggregation rule is *majority rule*: here, for each profile $(J_1, ..., J_n)$, the collective judgment set is the set of majority-accepted propositions in X, formally

$$J = \{p \in X : |\{i \in N : p \in J_i\}| > \frac{n}{2}\}.$$

As we have seen, an important deficiency of majority rule is that, when the propositions in X are logically connected in certain ways, the majority judgment set may be inconsistent, as already illustrated in Table 1.

We now state some desiderata that are often imposed on a judgment-aggregation rule. They are generalizations of the desiderata that Arrow (1951/1963) imposed on a preference-aggregation rule – more on this later. The first desideratum says that the judgment-aggregation rule should accept as input any profile of consistent and complete individual judgment sets.

Universal domain. The domain of F is the set of all profiles of consistent and complete individual judgment sets on X.

The second desideratum says that the collective judgment set produced by the aggregation rule should always be consistent and complete, where completeness is again required only relative to X. In many (though not all) judgment-aggregation problems, the completeness requirement is reasonable insofar as propositions are included in the proposition set X (the "agenda") precisely because they are supposed to be adjudicated. We also consider relaxations of this requirement below.

Collective consistency and completeness. For every profile $(J_1, ..., J_n)$ in the domain of F, the collective judgment set $J = F(J_1, ..., J_n)$ is consistent and complete.

The third desideratum says that the collective judgment on any proposition p should depend only on the individual judgments on p, not on the individual judgments on other propositions.

Propositionwise independence. For any profiles $(J_1, ..., J_n)$ and $(J'_1, ..., J'_n)$ in the domain of F and any p in X, if $p \in J_i \Leftrightarrow p \in J'_i$ for every individual i in N, then $p \in J \Leftrightarrow p \in J'$, where $J = F(J_1, ..., J_n)$ and $J' = F(J'_1, ..., J'_n)$.

The final desideratum says that if all individuals hold the same individual judgment set, this judgment set should become the collective judgment set.

Consensus preservation. For any unanimous profile $(J, ..., J)$ in the domain of F, $F(J, ..., J) = J$.

As in our discussion of the baseline desiderata on a belief-binarization rule, it is important to note that, at least in simple cases, the present desiderata can be met by a familiar judgment-aggregation rule such as majority rule. For example, if the proposition set X contains only a single proposition p and its negation $\neg p$ (or many logically independent proposition-negation pairs), then majority rule satisfies all four desiderata. The same is true for a suitable super- or sub-majority rule.

5 The correspondence between belief binarization and judgment aggregation

We are now in a position to describe the relationship between belief binarization and judgment aggregation more precisely. Let f be a belief-binarization rule for the proposition set X. For any group size n, we can use f to construct a corresponding judgment-aggregation rule F for X. The construction is in two steps.

In the first step, we convert any given profile of consistent and complete individual judgment sets into the corresponding propositionwise anonymous profile, i.e., the specification of the proportion of individual support for each proposition in X. Formally, for each profile $(J_1, ..., J_n)$, let $Cr_{(J_1,...,J_n)}$ be the function (from X into $[0, 1]$) that assigns to each proposition p in X the proportion of individuals accepting it:

$$Cr_{(J_1,...,J_n)}(p) = \frac{|\{i \in N : p \in J_i\}|}{n}.$$

Although the function $Cr_{(J_1,...,J_n)}$ is a "proportion-of-support" function on X, it behaves formally like a degree-of-belief function and can thus be mathematically treated as such a function. In particular, it is probabilistically coherent, since each individual judgment set in $(J_1, ..., J_n)$ is consistent and complete.

In the second step, we apply the given belief-binarization rule f to the constructed proportion function $Cr_{(J_1,...,J_n)}$ so as to yield an all-or-nothing belief set, which can then be reinterpreted as a collective judgment set. As long as $Cr_{(J_1,...,J_n)}$ is in the domain of f, the judgment set $J = f(Cr_{(J_1,...,J_n)})$ is well-defined, so that $(J_1, ..., J_n)$ is in the domain of the judgment-aggregation rule that we are constructing.

These two steps yield the judgment-aggregation rule F which assigns to each admissible profile $(J_1, ..., J_n)$ the collective judgment set

$$F(J_1, ..., J_n) = f(Cr_{(J_1,...,J_n)}).$$

Call this *the judgment-aggregation rule induced by the given belief-binarization rule.* In essence, it aggregates judgments by binarizing the proportion function induced by any given profile of individual judgment sets.

Proposition 1. The judgment-aggregation rule F induced by a belief-binarization rule f is *anonymous*, where anonymity is defined as follows.

Anonymity. F is invariant under permutations (relabellings) of the individuals. Formally, for any profiles $(J_1, ..., J_n)$ and $(J'_1, ..., J'_n)$ in the domain of F which are permutations of one another, $F(J_1, ..., J_n) = F(J'_1, ..., J'_n)$.

Proposition 1 is a consequence of the fact that the proportion of individuals accepting each proposition is not affected by permutations of those individuals. Formally, we have $Cr_{(J_1,...,J_n)} = Cr_{(J'_1,...,J'_n)}$ whenever the profiles $(J_1, ..., J_n)$ and $(J'_1, ..., J'_n)$ are permutations of one another. Furthermore, the following result holds:

Proposition 2. If the belief-binarization rule f satisfies universal domain, belief consistency and completeness, propositionwise independence, and certainty preservation, then, for any group size n, the induced judgment-aggregation rule F satisfies universal domain, collective consistency and completeness, propositionwise independence, and unanimity preservation.

To show this, suppose the binarization rule f satisfies all four desiderata, and let F be the induced aggregation rule for a given group size n. Then:

- F satisfies universal domain because, for every profile $(J_1, ..., J_n)$ of consistent and complete individual judgment sets, the function $Cr_{(J_1,...,J_n)}$ is in the domain of f, and so $F(J_1, ..., J_n) = f(Cr_{(J_1,...,J_n)})$ is well-defined.

- F satisfies collective consistency and completeness because, for every profile $(J_1, ..., J_n)$ in its domain, $f(Cr_{(J_1,...,J_n)})$ is consistent and complete.

- F satisfies propositionwise independence because, for any profiles $(J_1, ..., J_n)$ and $(J'_1, ..., J'_n)$ in its domain, if $p \in J_i \Leftrightarrow p \in J'_i$ for every individual i in N, then $Cr_{(J_1,...,J_n)}(p) = Cr_{(J'_1,...,J'_n)}(p)$, and so $p \in J \Leftrightarrow p \in J'$, where $J = f(Cr_{(J_1,...,J_n)})$ and $J' = f(Cr_{(J'_1,...,J'_n)})$ (by propositionwise independence of f).

- F satisfies consensus preservation because, for any unanimous profile $(J, ..., J)$ in its domain, $Cr_{(J,...,J)}$ assigns extremal degrees of belief (0 or 1) to all propositions in X (namely 1 if $p \in J$ and 0 if $p \notin J$), and so we must have $f(Cr_{(J,...,J)}) = J$ (by certainty preservation of f).

In sum, the existence of a belief-binarization rule satisfying our four baseline desiderata guarantees (for every group size n) the existence of an anonymous jugment-aggregation rule satisfying the four corresponding aggregation-theoretic desiderata. In the next section, we discuss the consequences of this fact.

6 An impossibility theorem

As noted above, when the proposition set X is sufficiently "simple", such as $X = \{p, \neg p\}$, we can indeed find belief-binarization rules for X that satisfy our four desiderata. Similarly, for such a set X, we can find judgment-aggregation rules satisfying the corresponding aggregation-theoretic desiderata. We now show that this situation changes dramatically when X is more complex. In this section, we state and prove the simplest version of our impossiblity result. We present some more general versions in subsequent sections. To state our first result, let us call a proposition p *contingent* if it is neither tautological, nor contradictory.

Theorem 1. For any proposition set X that is closed under conjunction or disjunction and contains more than one contingent proposition-negation pair (i.e., any non-trivial algebra), there exists no belief-binarization rule satisfying universal domain, belief consistency and completeness, propositionwise independence, and certainty preservation.

To prove this result, let X be a proposition set with the specified properties. Suppose, contrary to Theorem 1, there exists a belief-binarization rule for X satisfying all four desiderata. Call it f. Consider the judgment-aggregation rule F induced by f via the construction described in the last section, for some group size $n \geq 2$. By Proposition 1, F satisfies anonymity. Further, by Proposition 2, since f satisfies the four baseline desiderata on belief binarization, F satisfies the corresponding four aggregation-theoretic desiderata. However, the following result is well known to hold, as explained and referenced in detail in the next section:

Background Result 1. For any proposition set X that is closed under conjunction or disjunction and contains more than one contingent proposition-negation pair, any judgment-aggregation rule satisfying universal domain, collective consistency and completeness, propositionwise independence, and consensus preservation is *dictatorial*, i.e., there exists some antecedently fixed individual i in N such that, for each profile $(J_1, ..., J_n)$ in the domain, $F(J_1, ..., J_n) = J_i$.

So, there could not possibly exist an *anonymous* (and thereby non-dictatorial) aggregation rule satisfying all four conditions, and hence the belief-binarization rule f on which the aggregation rule F was based could not satisfy our four desiderata on belief binarization, contrary to our supposition. This completes the proof of Theorem 1.

Since subjective probability functions are normally defined on sets of propositions that are algebras, Theorem 1 immediately implies that our four baseline desiderata are mutually inconsistent when we seek to binarize a degree-of-belief function that constitutes a full-blown subjective probability function. However, we are able to prove something more general, even for cases in which the proposition set X is not an algebra of propositions.

7 The judgment-aggregation variant of Arrow's theorem and its corollary for belief binarization

Background Result 1, from which we have derived Theorem 1, is a special case of a more general theorem. It establishes the non-existence of non-dictatorial judgment-aggregation rules satisfying the four desiderata not only for proposition sets that are algebras, but more generally. By implication, as we will see, our impossibility result on belief binarization holds more generally too.

The significance of the more general background result on judgment aggregation lies in the fact that it has Arrow's classic impossibility theorem as an immediate corollary, as explained in the next section: it can thus be viewed as the judgment-aggregation variant of Arrow's theorem. For the general theorem to apply, the proposition set X must have two combinatorial properties, which are jointly weaker than the previous requirement that X be a non-trivial algebra:[7]

Path-connectedness. For any two contingent propositions p, q in X, there exists a path of conditional entailments from p to q (as explicated in a footnote).[8]

Pair-negatability. There exists a subset Y of X which is *minimally inconsistent* (i.e,

[7] The first property was first introduced by Nehring and Puppe (2010) in a different formalism under the name "total blockedness". The second was introduced, in different variants, by Dietrich (2007), Dietrich and List (2007a), and Dokow and Holzman (2010a) (the latter introduced an algebraic variant under the name "non-affineness").

[8] Formally, a proposition p *conditionally entails* a proposition q if there exists some subset Y of X, consistent with each of p and $\neg q$, such that $\{p\} \cup Y$ entails q. A *path of conditional entailments* from p to q is a sequence of propositions $p_1, p_2, ..., p_k$ in X with $p_1 = p$ and $p_k = q$ such that p_1 conditionally entails p_2, p_2 conditionally entails p_3, ..., and p_{k-1} conditionally entails.

Y is inconsistent but all its proper subsets are consistent) and contains two distinct propositions p and q such that replacing each of p and q with their negations renders Y consistent.[9]

A simple example of a proposition set X that is both path-connected and pair-negatable is the set consisting of p, q, $p \wedge q$, $p \vee q$, and their negations.[10] Another example is a set X consisting of *binary ranking propositions* of the form "x is preferable to y", "y is preferable to z", "x is preferable to z", and so on, where x, y, z, ... are three or more options (e.g., electoral alternatives, candidates, states of affairs), as explained in more detail in the next section. A third example is a set X which constitutes an algebra with more than one contingent proposition-negation pair.

The judgment-aggregation variant of Arrow's theorem can now be stated as follows:

Background Result 2. For any proposition set X that is path-connected and pair-negatable, any judgment-aggregation rule satisfying universal domain, collective consistency and completeness, propositionwise independence, and consensus preservation is dictatorial (Dietrich and List 2007a, Dokow and Holzman 2010a, building on Nehring and Puppe 2010).[11]

Put differently, when X is path-connected and pair-negatable, there exists no non-dictatorial aggregation rule satisfying the four desiderata. Given our discussion up to this point, it should be clear that this result has a direct corollary for belief binarization, which can be derived in exact analogy to Theorem 1 in the previous section. The non-existence of any non-dictatorial judgment-aggregation rule for X satisfying the four desiderata immediately implies the non-existence of any *anonymous* such rule, and since any belief-binarization rule for X satisfying our four binarization desiderata would induce such an aggregation rule, there cannot exist a belief-binarization rule of this kind. In sum, the following theorem holds:

[9] Formally, $Y \backslash \{p, q\} \cup \{\neg p, \neg q\}$ is consistent.

[10] To establish path-connectedness, note that, from any one of these propositions, we can find a path of conditional entailments to any other (e.g., we can find a path from p to q via $p \vee q$: p entails $p \vee q$, conditional on the empty set; and $p \vee q$ entails q, conditional on $\neg p$). To establish pair-negatability, note that the minimally inconsistent set $Y = \{p, q, \neg(p \wedge q)\}$ becomes consistent if we replace p and q with their negations $\neg p$ and $\neg q$.

[11] This theorem was proved independently, and in different formal frameworks, by Dietrich and List (2007a) and Dokow and Holzman (2010a). The latter proved, in addition, that (if X is finite) the two combinatorial properties of X are not merely sufficient, but also necessary for the theorem's conclusion to hold. Both of the cited papers build on earlier work by Nehring and Puppe, reported in Nehring and Puppe (2010). Nehring and Puppe's version of the theorem imposes an additional monotonicity desideratum on the aggregation rule, but does not require the pair-negatability property of X.

Theorem 2. For any proposition set X that is path-connected and pair-negatable, there exists no belief-binarization rule satisfying universal domain, belief consistency and completeness, propositionwise independence, and certainty preservation.

In the Appendix, we present a further refinement of this result, in particular a logically tight characterization of the minimal conditions on the proposition set X under which the theorem's negative conclusion – the non-existence of any belief-binarization rule satisfying the four desiderata – holds. (Theorem 2 itself gives only a sufficient condition on X for that conclusion to hold, not a necessary condition.)

8 The relationship with Arrow's original theorem

To illustrate that Background Result 2 is indeed a generalization of Arrow's original impossibility theorem – and hence that our impossibility theorem on belief-binarization and Arrow's theorem can be traced back to a common source – it is useful to revisit Arrow's original result (1951/1963).[12]

As we have already noted, Arrow considered the aggregation of preferences, rather than judgments. Suppose there is a finite set $N = \{1, 2, ..., n\}$ of individuals (with $n \geq 2$), each of whom holds a preference ordering, P_i, over some set $K = \{x, y, ...\}$ of options. A combination of preference orderings across the n individuals, $(P_1, ..., P_n)$, is called a *profile* of preference orderings. We are looking for a *preference-aggregation rule*, $\mathcal{F}$, which is a function that maps each profile of individual preference orderings (within some domain of admissible profiles) to a collective preference ordering P. Arrow imposed four conditions on a preference-aggregation rule.

Universal domain. The domain of $\mathcal{F}$ is the set of all profiles of rational individual preference orderings. (We here call a preference ordering *rational* if it is a transitive, irreflexive, and complete binary relation on K; for expositional simplicity, we thus restrict our attention to indifference-free preferences.)

Collective rationality. For every profile $(P_1, ..., P_n)$ in the domain of $\mathcal{F}$, the collective preference ordering $R = \mathcal{F}(P_1, ..., P_n)$ is rational.

Pairwise independence. For any profiles $(P_1, ..., P_n)$ and $(P'_1, ..., P'_n)$ in the domain of $\mathcal{F}$ and any pair of options x and y in K, if P_i and P'_i rank x and y in the same way for every individual i in N, then P and P' also rank x and y in the same way, where $R = \mathcal{F}(P_1, ..., P_n)$ and $R' = \mathcal{F}(P'_1, ..., P'_n)$.

[12] We here follow the analysis in Dietrich and List (2007a). For a precursor, see List and Pettit (2004).

The Pareto principle. For any profile $(P_1, ..., P_n)$ in the domain of $\mathcal{F}$ and any pair of options x and y in K, if P_i ranks x above y for every individual i in N, then P also ranks x above y, where $R = \mathcal{F}(P_1, ..., P_n)$.

Arrow's original theorem (1951/1963) can be stated as follows:

Arrow's theorem. For any set K of three or more options, any preference-aggregation rule satisfying universal domain, collective rationality, pairwise independence, and the Pareto principle is *dictatorial*, i.e., there exists some antecedently fixed individual i in N such that, for each profile $(P_1, ..., P_n)$ in the domain, $\mathcal{F}(P_1, ..., P_n) = P_i$.

To confirm that Background Result 2 from the previous section is indeed a generalization of Arrow's theorem, we note that the latter can be derived from the former. The key observation is that, setting aside interpretational differences, we can represent any preference-aggregation problem formally as a special kind of judgment-aggregation problem. To construct this representation, we define the set X of propositions on which judgments are made as the set of pairwise ranking propositions of the form xPy ("x is preferable to y"), where x and y are options in K and P represents pairwise preference. Formally,

$$X = \{xPy : x, y \in K \text{ with } x \neq y\}.$$

Call this proposition set the *preference agenda* for K. Under the simplifying assumption of irreflexive preferences, we can interpret yPx as the negation of xPy, and so the constructed proposition set is negation-closed. We call any subset Y of X *consistent* if Y is a consistent set of binary ranking propositions relative to the rationality constraints on preferences introduced above (transitivity etc.).[13] For example, the set $Y = \{xPy, yPz, xPz\}$ is consistent in this sense, while the set $Y = \{xPy, yPz, zPx\}$ is not, because it involves a breach of transitivity.

Since any preference ordering P over K is just a binary relation, it can be uniquely represented by a subset of X, namely the subset consisting of all pairwise ranking propositions validated by P.[14] In this way, rational preference orderings over K stand in a one-to-one correspondence with consistent and complete judgment sets for the preference agenda X. Furthermore, preference-aggregation rules (for preferences over K) stand in a one-to-one correspondence with judgment-aggregation rules (for judgments on the associated preference agenda X).

[13] Technically, the set Y is consistent if and only if there exists at least one rational (here: transitive, irreflexive, and complete) preference ordering over K that validates all the binary ranking propositions contained in Y.

[14] As defined here, this representation assumes strict preferences.

Now, applied to X, the judgment-aggregation desiderata of universal domain, collective consistency and completeness, and propositionwise independence collapse into Arrow's original desiderata of universal domain, collective rationality, and pairwise independence. Consensus preservation collapses into a weaker counterpart of Arrow's Pareto principle, which says that whenever all individuals hold the same preference ordering over all options, this preference ordering should become the collective one.[15]

It can be verified that when K contains more than two options, the corresponding preference agenda X is path-connected and pair-negatable, and so Background Result 2 applies, yielding Arrow's original theorem as a corollary.

Corollary of Background Result 2. For any preference agenda X defined for a set K of three or more options, any judgment-aggregation rule satisfying universal domain, collective consistency and completeness, propositionwise independence, and consensus preservation is dictatorial.

Figure 1 displays the logical relationships between (i) the jugdment-aggregation variant of Arrow's theorem, (ii) Arrow's original theorem, and (iii) our baseline impossibility theorem on belief binarization. In short, Arrow's theorem and our result on belief binarization, which are at first sight very different from one another, can both be derived from the same common impossibility theorem on judgment aggregation.

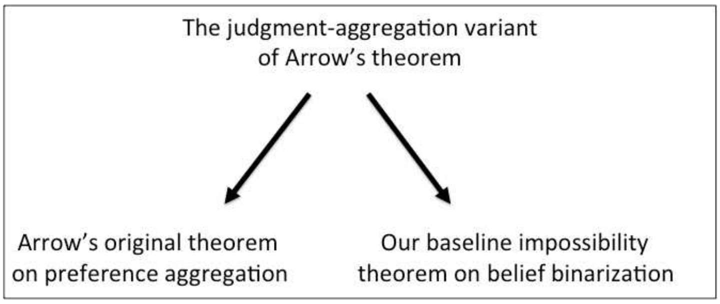

Figure 1: The common source of two distinct impossibility results

9 Escape routes from the impossibility

If we wish to avoid the impossibility of belief binarization, we must relax at least one of the four baseline desiderata we have introduced. In what follows, we discuss a number

[15]This weaker desideratum is implied by Arrow's Pareto principle (in the presence of collective rationality), but does not generally imply it. The converse implication holds in the presence of universal domain and pairwise independence.

of possible escape routes from our impossibility result, roughly presented in an order of how obvious they seem (though that order is largely a matter of taste and nothing hinges on it).[16]

9.1 Relaxing completeness of beliefs, but retaining implication-closure

The first and perhaps most obvious response to our impossibility result is to argue that the requirement of completeness of all-or-nothing beliefs is too strong. This suggests dropping that requirement, while retaining the familiar requirement that all-or-nothing beliefs should be consistent as well as closed under logical implication (within the proposition set X).

Belief consistency and implication-closure. For every Cr in the domain of f, the belief set $B = f(Cr)$ is consistent and implication-closed (relative to X).

This permits suspending belief on some proposition-negation pairs in X. Indeed, even an empty belief set satisfies consistency and implication-closure (assuming X contains no tautology). Surprisingly, however, the new desideratum does not get us very far if we insist on the other three desiderata. The new set of desiderata permits only a single, extremely conservative binarization rule, namely a uniform threshold rule with threshold 1 for all propositions. This can be viewed as a triviality result, along the lines of – though in some respects more general than – other triviality results in the literature (e.g., Douven and Williamson 2006).[17]

Theorem 3. For any proposition set X that is path-connected and pair-negatable, any belief-binarization rule satisfying universal domain, belief consistency and implication-closure, propositionwise independence, and certainty preservation is a threshold rule with a uniform threshold of 1 for the acceptance of any proposition, i.e., for any degree-of-belief function Cr in the domain, $f(Cr) = \{p \in X : Cr(p) = 1\}$.

This result, too, is a consequence of a result on judgment aggregation, though the proof is a bit more complicated than our earlier proofs. Let the proposition set X be path-connected and pair-negatable, and suppose f is a belief-binarization rule satisfying the desiderata listed in Theorem 3. As before, for any group size n, f induces an

[16] For simplicity, we assume that the proposition set X is finite in Sections 9.2 to 9.5.

[17] Unlike Douven and Williamson's result, our result does not require the set of propositions under consideration to be an algebra, and it does not presuppose that the sufficient condition for belief acceptance is what Douven and Williamson call "structural"; rather, our desideratum of propositionwise independence allows, in principle, the use of different acceptance criteria for different propositions.

anonymous judgment-aggregation rule F. By the analogue of Proposition 2, F satisfies universal domain, collective consistency and implication-closure, propositionwise independence, and consensus preservation. But the following result holds:

Background Result 3. For any proposition set X that is path-connected and pair-negatable, any judgment-aggregation rule satisfying universal domain, collective consistency and implication-closure, propositionwise independence, and consensus preservation is *oligarchic*: i.e., there exists some antecedently fixed non-empty set M of individuals in N such that, for each profile $(J_1, ..., J_n)$ in the domain, $F(J_1, ..., J_n) = \bigcap_{i \in M} J_i$ (Dietrich and List 2008, Dokow and Holzman 2010b).

Consistently with this result, the set M of "oligarchs" could be any non-empty subset of N, ranging from a singleton set, where $M = \{i\}$ for some individual i, to the set of *all* individuals, where $M = N$. In the first case, the aggregation rule is dictatorial; in the last, it is the *unanimity rule*. Since any aggregation rule induced by a belief-binarization rule is anonymous, and an anonymous aggregation rule can be oligarchic *only if* it is the unanimity rule, Background Result 3 immediately implies that the induced rule F is the unanimity rule. So, no proposition is collectively accepted under F with less than 100% support.

Could the belief-binarization rule f on which F is based differ from a threshold rule with threshold 1? In what follows, we show that if f were distinct from such a rule, this would contradict what we have just learnt from Background Result 3. Less technically inclined readers may skip the rest of this proof and move on to Section 9.2.

To complete the proof, suppose that f is *not* a threshold rule with threshold 1. There must then exist a proposition q in X and a degree-of-belief function Cr with $Cr(q) < 1$ such that $q \in B$, where $B = f(Cr)$.[18] In the Appendix, we show that, under the present conditions, f must be *monotonic*: if $q \in f(Cr)$, then $q \in f(Cr')$ for any other credence function Cr' with $Cr'(q) > Cr(q)$.

Now consider what this implies for any induced aggregation rule F. Pick two consistent and complete judgment sets $J, J' \subseteq X$ such that $q \in J$ and $q \notin J'$, and construct a profile $(J_1, ..., J_n)$ for some sufficiently large group size n such that a proportion of more than $Cr(q)$ of the individuals in N, but fewer than all, have the judgment set J and the rest have the judgment set J'. By the construction of F, we have $F(J_1, ..., J_n) = f(Cr_{(J_1,...,J_n)})$, where for each proposition p in X,

[18]The proposition q must be contingent. If q were tautological, we could not have $Cr(q) < 1$, and if it were contradictory, we could not have $q \in B$ (given consistency of B).

$$Cr_{(J_1,...,J_n)}(p) = \frac{|\{i \in N : p \in J_i\}|}{n}.$$

Since $Cr_{(J_1,...,J_n)}(q) > Cr(q)$ and our binarization rule f is monotonic, we must have $q \in f(Cr_{(J_1,...,J_n)})$ and hence $q \in F(J_1, ..., J_n)$, despite the lack of unanimous support for q. This contradicts the fact that F is the unanimity rule. We conclude that f is a threshold rule with a uniform threshold of 1 for the acceptance of any proposition.

9.2 Relaxing implication-closure of beliefs

A second escape route from our impossibility result is to relax the requirements on all-or-nothing beliefs further and to require only their consistency, while no longer requiring implication-closure (compare also Kyburg 1961).

Belief consistency. For every Cr in the domain of f, the belief set $B = f(Cr)$ is consistent.

This opens up a number of non-trivial possibilities, even in the presence of the other desiderata. In particular, the following result holds (a version of which has also been proved by Easwaran and Fitelson forthcoming):

Theorem 4. Let k be the size of the largest minimally inconsistent subset of the proposition set X. Any threshold rule with a strict threshold of $\frac{k-1}{k}$ (or higher[19]) for each proposition satisfies universal domain, belief consistency, propositionwise independence, and certainty preservation.

It is worth explaining the significance of k, the size of the largest minimally inconsistent subset of X. This parameter can be interpreted as a simple measure of the interconnectedness between propositions in X. If X contains only one or several unconnected proposition-negation pairs, then the largest minimally inconsistent subsets of X are of the form $\{p, \neg p\}$, so k is 2.[20] If X contains p, q, $p \wedge q$, and their negations, then the largest minimally inconsistent subset is $\{p, q, \neg(p \wedge q)\}$, so k is 3. If X contains p, q, r, $p \wedge q \wedge r$, and their negations, as in our example in Section 2, then the largest minimally inconsistent subset is $\{p, q, r, \neg(p \wedge q \wedge r)\}$, so k is 4. In consequence, the binarization

[19] The highest admissible threshold in this theorem is a weak threshold of 1 (under which the acceptance criterion for any proposition is a degree of belief of 1).

[20] This assumes that some proposition p in X is contingent. If X contains no contingent propositions, then the largest minimally inconsistent subset of X is the singleton set consisting of the contradiction.

threshold $\frac{k-1}{k}$ required in Theorem 5 increases with the complexity of these cases, from $\frac{1}{2}$ to $\frac{2}{3}$ to $\frac{3}{4}$.[21]

To prove Theorem 4, let f be a threshold rule with a strict threshold of $\frac{k-1}{k}$ (or higher) for each proposition. It is easy to see that f satisfies universal domain, propositionwise independence, and certainty preservation.[22] Suppose, for a contradiction, that $B = f(Cr)$ is inconsistent for some degree-of-belief function Cr in the domain. Then B must have at least one minimally inconsistent subset Y, whose size, in turn, is at most k. For every proposition p in Y to be accepted by f, we must have $Cr(p) > \frac{k-1}{k}$. Since Cr is probabilistically coherent, it is extendable to a well-defined probability function Pr on the algebra generated by X. This algebra contains the conjunction of all propositions in Y. Since Y is an inconsistent set, this conjunction is a contradiction and must be assigned probability 0 by Pr. But we now show that this is contradicted by the fact that $Pr(p) > \frac{k-1}{k}$ for every p in Y, which has to hold because Pr is an extension of Cr and so $Pr(p) = Cr(p)$ for every p in X. First note that the probability of the conjunction of any two propositions – no matter how negatively correlated – must exceed 0 when each proposition has a probability greater than $\frac{1}{2}$. Similarly, the probability of the conjunction of any three propositions must exceed 0 when each has a probability greater than $\frac{2}{3}$. Generally, the probability of the conjunction of any k propositions that each have a probability greater than $\frac{k-1}{k}$ must exceed 0, a contradiction. This completes the proof.

Again, this result has a counterpart in judgment-aggregation theory. A *qualified majority rule* is a judgment-aggregation rule which assigns, to each profile $(J_1, ..., J_n)$, the collective judgment set

$$J = \{p \in X : |\{i \in N : p \in J_i\}| \text{ exceeds } qn\},$$

where "exceeds" can be read either strictly (as "$>$") or weakly (as "$\geq$"), and q is some acceptance threshold between $\frac{1}{2}$ and 1. It is a *supermajority rule* when the acceptance threshold requires more than a simple majority of the individuals ("50%+1"). The following result holds:

Background Result 4. Let k be the size of the largest minimally inconsistent subset of the proposition set X. Any qualified majority rule with a strict threshold of

[21] Formulas similar to $\frac{k-1}{k}$ have been used by Bovens and Hawthorne (1999), though without explicitly invoking the notion of minimally inconsistent sets of propositions.

[22] The latter excludes a strict threshold of 1 for any non-contradictory proposition, but permits a weak threshold of 1.

$\frac{k-1}{k}$ (or higher[23]) for each proposition satisfies universal domain, collective consistency, propositionwise independence, and unanimity preservation (Dietrich and List 2007b).

It should be evident that the qualified majority rule in Background Result 4 is simply the judgment-aggregation rule induced by the binarization rule in Theorem 4.

9.3 Relaxing propositionwise independence

Another escape route from our impossibility result involves giving up the requirement that the all-or-nothing belief in any proposition p in X depend only on the degree of belief in p and "going holistic", by allowing the all-or-nothing belief in p to depend also on the degrees of belief in other propositions in X. If we go along this route, several binarization rules become possible. Again, they all have counterparts in judgment-aggregation theory. We here give only a few examples.

Premise-based rules. Under a premise-based rule, we first designate some subset Y of the proposition set X as a set of "premises" (where Y is closed under negation). We then (i) form the all-or-nothing beliefs on these premises by means of a suitable propositionwise independent binarization rule (such as a threshold rule), restricted to the premises, and (ii) derive the all-or-nothing beliefs on all other propositions by logical inference. Formally, for every degree-of-belief function Cr in the domain, we have

$$f(Cr) = \{p \in X : g(Cr) \cap Y \text{ entails } p\},$$

where Y is the set of premises and g is the binarization rule applied to them (e.g., a threshold rule). As long as the set Y and the rule g are chosen so as to guarantee a consistent set $g(Cr) \cap Y$ (e.g., by making sure that the premises in Y are logically independent from one another), the premise-based rule will always yield consistent and implication-closed belief sets. Premise-based rules have been extensively discussed in judgment-aggregation theory, building on an earlier work in law and economics (e.g., Kornhauser and Sager 1986 and Kornhauser 1992, Pettit 2001, List and Pettit 2002, Chapman 2002, Bovens and Rabinowicz 2006, List 2006, Dietrich 2006; for more recent generalizations, see Dietrich and Mongin 2010). Here, a group makes its collective judgments by taking majority votes only on some logically independent premises (e.g., "Did the defendant do a particular action?", "Was he or she contractually obliged not to do that action?") and deriving its judgments on all relevant conclusions by logical

[23] The highest admissible threshold in this theorem is a weak threshold of 1 (under which the acceptance criterion for any proposition is unanimous support).

interference (e.g., "Is the defendant liable for breach of contract?"). The downside of a premise-based rule in judgment aggregation is that a proposition can end up being collectively accepted by logical inference even if none of the individuals accepts that proposition. Similarly, in belief binarization, a proposition could end up being included in the all-or-nothing belief set despite the assignment of a very low, or even zero, degree of belief to it. Furthermore, the output of a premise-based rule depends crucially on what the specified set of premises is. Premise-based belief binarization, like premise-based judgment aggregation, is plausible only to the extent that we have a non-arbitrary way of identifying the premises and are prepared to generate all our overall beliefs (or judgments) on the basis of considering these premises alone.

Sequential priority rules. Sequential priority rules are generalizations of premise-based rules. To define a sequential priority rule, we specify some *order of priority* among the propositions in X (formally a linear order over the elements of X). For each degree-of-belief function Cr, we then determine the belief set $B = f(Cr)$ as follows. The propositions in X are considered in the given order of priority, and the belief set B is built up sequentially. For each proposition under consideration, say p, we begin by asking whether p is entailed by propositions that we have included in B in earlier steps. If the answer is yes, we embrace this entailment and include p in B. If the answer is no, we determine whether or not to include p in B by applying some propositionwise binarization criterion to $Cr(p)$, for example some threshold. By construction, the resulting belief set is always consistent. Whether it is also implication-closed depends on the propositionwise binarization criterion and the order of priority. The downside of a sequential priority rule, like that of a premise-based rule, is that it sometimes mandates the inclusion of a proposition in the belief set B even when the agent's degree of belief in it is very low or perhaps zero (namely when that proposition is entailed by other accepted propositions). Sequential priority rules for belief binarization are analogous to sequential priority rules for judgment aggregation (List 2004, Dietrich and List 2007b). The only difference lies in the use of a propositionwise *binarization* criterion (such as a propositionwise threshold) instead of a propositionwise *aggregation* criterion (such as propositionwise majority voting). In both belief binarization and judgment aggregation, sequential priority rules may be *path-dependent*: the output they generate is not generally invariant under changes of the order of priority among the propositions. This means that the defensibility of such a rule depends, in part, on our ability to specify that order non-arbitrarily.

Distance-based rules. We have already encountered two classes of judgment-aggregation rules violating propositionwise independence that have direct analogues for belief binarization: premise-based rules and sequential priority rules. A third class is the class of distance-based rules. In judgment-aggregation theory, such a rule is defined as follows. We first introduce some *distance metric* over judgment sets, which specifies how "distant" any two judgment sets are from one another (formally, a *distance metric* assigns to each pair of judgment sets a non-negative number, interpreted as the distance between them). For each profile of individual judgment sets, we then select a collective judgment set that minimizes the sum of the distances from the individual judgment sets, according to that distance metric. Some distance-based aggregation rules require more information than what is contained in a propositionwise anonymous profile, and hence have no counterpart in the case of belief binarization, but others naturally carry over to belief binarization. The best-known distance-based aggregation rule is the *Hamming rule* (e.g., Konieczny and Pino Pérez 2002, Pigozzi 2006). Here the distance between any two judgment sets is given by the number of propositions in X on which the two judgment sets disagree (which means the proposition in question is contained in one set but not in the other). For each profile of individual judgment sets $(J_1, ..., J_n)$, we select a consistent and complete (or perhaps consistent and implication-closed) collective judgment set J which minimizes

$$\sum_{i \in N} |\{p \in X : p \in J \not\Leftrightarrow p \in J_i\}|.$$

Such a judgment set need not be unique; so we must either define the Hamming rule as a multi-function (under which more than one collective judgment set can be assigned to any given profile of individual judgment sets), or introduce a tie-breaking rule. The details need not concern us here. Note that minimizing the total Hamming distance is equivalent to minimizing

$$\sum_{p \in X} |\{i \in N : p \in J \not\Leftrightarrow p \in J_i\}|.$$

This, in turn, is equivalent to minimizing

$$\sum_{p \in X} |J(p) - Cr_{(J_1,...,J_n)}(p)|,$$

where, for each p in X,

$$J(p) = \begin{cases} 1 & \text{if } p \in J \\ 0 & \text{if } p \notin J \end{cases}$$

and $Cr_{(J_1,...,J_n)}$ is the function that assigns to each proposition p in X the proportion of individuals accepting p within the given profile $(J_1, ..., J_n)$, as defined earlier. This suggests the following definition of a *Hamming rule for belief binarization*: for each degree-of-belief function Cr, let $f(Cr)$ be a consistent and complete (or alternatively, consistent and implication-closed) belief set B which minimizes

$$\sum_{p \in X} |B(p) - Cr(p)|,$$

where $B(p)$ is defined in exact analogy to $J(p)$. Informally speaking, the Hamming rule binarizes any given degree-of-belief function by selecting an all-or-nothing belief set that is "minimally distant" from it, subject to the constraints of consistency and completeness (or alternatively, consistency and implication-closure). Like its judgment-aggregation counterpart, the Hamming rule for belief binarization must be defined as a multi-function, since there may be more than one distance-minimizing belief set, or we require some tie-breaking rule.

Other non-independent binarization rules. It is worth noting that the relationship between anonymous judgment-aggregation rules and belief-binarization rules can be used not only to derive binarization rules from aggregation rules but also to derive aggregation rules from binarization rules that have been proposed in the literature. In this way, some insights from belief-binarization theory carry over to judgment-aggregation theory (recall Levi 2004). Given space constraints, we here discuss only one class of rules for which this reverse translation is possible: Leitgeb's *P-stability-based rules*. Leitgeb (2014) offers a method of constructing, for each degree-of-belief function Cr (defined on some algebra of propositions), a specific acceptance threshold such that the set of all propositions for which the agent's degree of belief exceeds the threshold is consistent and implication-closed. Crucially, the threshold may differ for different degree-of-belief functions. The key idea is to identify a so-called *P-stable* proposition; this is a proposition p for which $Cr(p|q)$ exceeds $\frac{1}{2}$ for any proposition q consistent with p. The agent then accepts all those propositions in which he or she has a degree of belief greater than or equal to $t = Cr(p)$, where p is the identified P-stable proposition. If we assign to each degree-of-belief function the belief set generated through this process, we obtain a well-defined belief-binarization rule. Since the acceptance threshold may differ for different degree-of-belief functions, the present binarization rule does not satisfy propositionwise independence. Leitgeb recognizes this when he acknowledges that the cost of his proposal is "a strong form of sensitivity of belief to context" (p. 168), and he gives some arguments in defence of this context sensitivity. Using the construction discussed earlier in this paper, we can use this belief-binarization rule to define a corresponding anonymous judgment-aggregation rule. To the best of our knowledge, this aggregation rule has not previously been investigated in the literature on judgment aggregation. It inherits its potential interest-value from the arguments that Leitgeb has offered in support of the underlying belief-binarization rule. A similar translation is possible for Lin and Kelly's *camera-shutter rules* for belief binarization (2012a, 2012b). These, too, are non-independent rules that could be used to generate corresponding judgment-aggregation rules (see also Kelly and Lin 2011).

9.4 Relaxing universal domain

A fourth escape route from our impossibility result opens up once we relax the desideratum of universal domain. Recall that universal domain requires the belief-binarization rule to work for every well-defined degree-of-belief function. If, instead, we suitably restrict the domain of admissible degree-of-belief functions, it becomes easier to find a belief-binarization rule satisfying the other desiderata.

Suppose, for example, that a degree-of-belief function Cr is deemed admissible only if it has the property that, for every minimally inconsistent subset Y of X, there is at least one proposition in p in Y with $Cr(p) \leq \frac{1}{2}$. It then follows that even a permissive binarization rule such as "more-likely-than-not binarization" (a threshold rule with a strict threshold of $\frac{1}{2}$ for all propositions) will never generate an inconsistent belief set B. If B were inconsistent for some Cr in the restricted domain, then B would have to have at least one minimally inconsistent subset Y, which, in turn, would have to contain at least one proposition p for which $Cr(p) \leq \frac{1}{2}$ (because Cr is in the restricted domain). But then p could not be accepted under a threshold rule with a strict threshold of $\frac{1}{2}$. More generally, if every admissible credence function Cr has the property that, for every minimally inconsistent subset Y of X, there is at least one proposition p in Y with $Cr(p) \leq t$, then any threshold rule with a strict threshold of t or higher will never run into any inconsistencies.

When translated into restrictions on admissible profiles of judgment sets in judgment-aggregation theory, the domain restrictions just mentioned match exactly the domain restrictions required for the consistency of majority rule and supermajority rule with threshold t, respectively. In judgment-aggregation theory, domain restrictions are often associated with situations in which the groups of individuals whose judgments are aggregated are reasonably "cohesive": disagreements betweeen them are limited.

In belief-binarization theory, it is harder to justify the required domain restrictions in a non-ad-hoc way. Nonetheless, it remains a possibility to argue that successful belief binarization is feasible if the agent's degree-of-belief function falls into a sufficiently "well-behaved" domain.

9.5 Relaxing consistency of beliefs

A more revisionist response to our impossiblity result is to argue that all-or-nothing beliefs need not be consistent. Of course, if outright inconsistency of beliefs is permitted, the problems we have identified go away immediately. But since inconsistent beliefs go against standard requirements of rationality, the present response may not seem very promising. However, there is a notion of less-than-fully-consistent belief which captures the idea that some inconsistencies are less "blatant" than others, so that we might opt for a belief-binarization rule that avoids "blatant" inconsistencies, while not securing full consistency (indeed, typical human beings are unlikely to hold fully consistent beliefs).

To introduce the relevant notion of less-than-full consistency (drawing on List 2014), we begin with a few intuitive observations. If someone believes a proposition that is self-contradictory, such as $p \wedge \neg p$, he or she seems rather blatantly inconsistent. If someone believes two propositions, neither of which is self-contradictory, but which are jointly inconsistent, such as p and $\neg p$, he or she is still fairly inconsistent, but less so than in the previous case. If someone believes three jointly inconsistent propositions, any two of which are mutually consistent, such as p, $p \rightarrow q$, and $\neg q$, his or her belief set is still relatively inconsistent, but not as much as in the two previous cases. If someone's belief set contains ten jointly inconsistent propositions, any nine of which are mutually consistent, this is nowhere near as bad as the previous inconsistencies. Now the key idea is to interpret the size of the smallest inconsistent set of believed propositions as a measure of the agent's inconsistency.

Formally, let us say that a belief set B is k-*inconsistent* if it has an inconsistent subset of size less than or equal to k. In our examples, a belief set that includes the proposition $p \wedge \neg p$ is 1-inconsistent; a belief set that includes the propositions p and $\neg p$ is 2-inconsistent, and so on. Similarly, we say that a belief set B is k-*consistent* if it is free from any inconsistent subsets of size up to k. As the value of k increases, k-consistency becomes more demanding, and any residual inconsistencies become less "blatant". Full consistency is the limiting case of k-consistency as k goes to infinity. Suppose we replace the requirement of belief consistency with the following:

Belief k-consistency (for some fixed value of k). For every Cr in the domain of f, the belief set $B = f(Cr)$ is k-consistent.

We then obtain a possibility result:

Theorem 5. Any threshold rule with a strict threshold of $\frac{k-1}{k}$ (or higher[24]) for each proposition satisfies universal domain, belief k-consistency, propositionwise independence, and certainty preservation.

The proof of this theorem, which we omit for brevity, is very similar to that of Theorem 4 above. The key point is that a probabilistically coherent function Cr could never assign a degree of belief greater than $\frac{k-1}{k}$ to each of k or fewer mutually inconsistent propositions; the implied subjective probability of their conjunction would then have to be greater than 0, which would violate probabilistic coherence. Like our other results, Theorem 5 has an analogue in judgment-aggregation theory.

Background Result 5. Any qualified majority rule with a strict threshold of $\frac{k-1}{k}$ (or higher[25]) for each proposition satisfies universal domain, collective k-consistency, propositionwise independence, and unanimity preservation (List 2014).

In sum, agents who are prepared to settle for less-than-full consistency in their beliefs can safely use threshold rules with a sufficiently high threshold.

9.6 Relaxing certainty preservation

A final logically possible escape route from our impossibility result is to relax certainty preservation. However, this escape route is of little interest. First, certainty preservation is a very plausible requirement and thus hard to relax. It only says that, in the special case in which the agent's degrees of belief are already "binary" (meaning the agent assigns credence 0 or 1 to all propositions in X), these binary beliefs should be preserved as the all-or-nothing beliefs.

Second, even if we were prepared to give up certainty preservation, this would not get us very far. For a large class of proposition sets X, we would still be faced with an impossibility result. To state this result, call a proposition p an *atom* of X if p is non-contradictory and, for every proposition q in X, p entails q or p entails $\neg q$. Further, call the proposition set X *atomic* if it contains an exhaustive set of atoms.[26] It is easy to see that any proposition set X that forms an algebra (if finite) is atomic. The following result holds:

[24] As before, the highest admissible threshold in this theorem is a weak threshold of 1 (under which the acceptance criterion for any proposition is a degree of belief of 1).

[25] As before, the highest admissible threshold in this theorem is a weak threshold of 1 (under which the acceptance criterion for any proposition is unanimous support).

[26] Formally, X is atomic if the set $\{\neg p \in X : p \text{ is an atom of } X\}$ is inconsistent.

Theorem 6. For any proposition set X that is atomic and contains more than one contingent proposition-negation pair, any belief-binarization rule satisfying universal domain, belief consistency and completeness, and propositionwise independence is *constant*, i.e., it delivers as its output the same antecedently fixed belief set B, no matter which degree-of-belief function Cr is fed into it as input.

Of course, such a binarization rule is totally useless. According to it, the agent's all-or-nothing beliefs are completely unresponsive to his or her degrees of belief. Like our earlier results, Theorem 6 is a corollary of an analogous theorem on judgment aggregation.

Background Result 6. For any proposition set X that is atomic and contains more than one contingent proposition-negation pair, any judgment-aggregation rule satisfying universal domain, collective consistency and completeness, propositionwise independence, and consensus preservation is either dictatorial or constant (Dietrich 2006; for a related result, see Pauly and van Hees 2006).

It is fair to conclude, then, that the relaxation of certainty preservation offers no compelling escape route from our impossibility result.

10 Concluding remarks

We have investigated the relationship between degrees of belief and all-or-nothing beliefs from a formal perspective, drawing on insights from the theory of judgment aggregation. We have proved a baseline impossibility theorem, which turns out to be a cousin of Arrow's classic impossibility theorem on preference aggregation. The two results are each corollaries of a single, mathematically more general impossibility theorem on judgment aggregation.

The central message of our analysis is that the possibilities of expressing all-or-nothing beliefs as a function of degrees of belief are rather limited. Any such possibility requires the relaxation of at least one of four baseline desiderata, and typically this relaxation comes at a cost. If we relax universal domain, we must use a binarization rule that does not work for all possible degree-of-belief functions. If we relax belief consistency and completeness, then, depending on whether or not we retain the requirement of implication-closure, we must either use an extremely conservative acceptance criterion for every proposition – namely a degree-of-belief threshold of 1 – or live with violations of implication-closure or consistency. If we relax propositionwise independence, we must accept what Leitgeb (2014, p. 168) has described as "a strong form of sensitivity of belief to context", where an agent's belief in one proposition may be affected by changes in his or her degrees of belief in others. If we relax certainty preservation, we face another impossibility result: for a large class of proposition sets, the only binarization rules satisfying the other three desiderata are constant rules, under which the agent's beliefs are not responsive at all to his or her degrees of belief.

If one is reluctant to embrace any of these possibilities, one may be led to conclude that there is no simple formal relationship to be found between degrees of belief and all-or-nothing beliefs. The most radical version of this conclusion would be the denial that agents genuinely have both kinds of belief. Extreme Bayesians, for instance, might hold that agents have no all-or-nothing beliefs. On that view, beliefs always come in degrees. The opposite view would be that degrees of belief are theoretical constructs of probability theory and that, in reality, agents have *only* all-or-nothing beliefs. On this picture, degrees enter at most in the *content* of a belief. An agent might have a full belief in the proposition that the epistemic probability of another proposition is x. Crucially, his or her attitude towards the "outer" proposition would then be an all-or-nothing attitude, which does not come in degrees. It would just so happen that that proposition asserts a probability assignment to another proposition (the "inner" one).

A less radical conclusion would be that agents have degrees-of-belief as well as all-or-nothing beliefs, but that the two kinds of belief may come apart: they may be two distinct aspects of an agent's credal state, none of which is determined by the other. To defend that picture, one would have to say more about what such a multi-faceted credal state would look like – a topic well beyond the scope of this paper (for a recent relevant discussion, see Easwaran and Fitelson forthcoming). We need not take a stand here on what the correct conclusion is. Our aim has simply been to lay out some salient options, and to offer an analysis of the logical space in which they are located, in the hope that this exercise will inspire further exploration.

References

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

Bovens, L., and W. Rabinowicz (2006) "Democratic answers to complex questions—an epistemic perspective." *Synthese* 150(1): 131–153.

Chapman, B. (2002) "Rational Aggregation." *Politics, Philosophy and Economics* 1: 337–354.

Dietrich, F. (2006) "Judgment Aggregation: (Im)Possibility Theorems." *Journal of Economic Theory* 126: 286–298.

Dietrich, F. (2007) "A generalised model of judgment aggregation." *Social Choice and Welfare* 28: 529–565.

Dietrich, F., and C. List (2007a) "Judgment aggregation by quota rules: majority voting generalized." *Journal of Theoretical Politics* 19: 391–424.

Dietrich, F., and C. List (2007b) "Arrow's theorem in judgment aggregation." *Social Choice and Welfare* 29: 19–33.

Dietrich, F., and C. List (2013) "Propositionwise judgment aggregation: the general case." *Social Choice and Welfare* 40(4): 1067–1095.

Dietrich, F. and P. Mongin (2010) "The premise-based approach to judgment aggregation." *Journal of Economic Theory* 145: 562–582.

Dokow, E., and R. Holzman (2010a) "Aggregation of binary evaluations." *Journal of Economic Theory* 145: 495–511.

Dokow, E., and R. Holzman (2010b) "Aggregation of binary evaluations with abstentions." *Journal of Economic Theory* 145: 544–561.

Douven, I., and J.-W. Romeijn (2007) "The Discursive Dilemma as a Lottery Paradox." *Economics and Philosophy* 23(3): 301–319.

Douven, I., and T. Williamson (2006) "Generalizing the Lottery Paradox." *British Journal for the Philosophy of Science* 57(4): 755–779.

Easwaran, K., and B. Fitelson (forthcoming) "Accuracy, Coherence and Evidence." In T. Szabó Gendler and J. Hawthorne (eds.), *Oxford Studies in Epistemology* 5.

Hawthorne, J., and L. Bovens (1999) "The preface, the lottery, and the logic of belief." *Mind* 108(430): 241–264.

Kelly, K. T., and H. Lin (2011) "Judgment aggregation: A geometrical impossibility proof." Presentation at the 2011 Episteme Conference, Carnegie Mellon University, June 2011.

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

Kornhauser, L. A. (1992) "Modeling collegial courts. II. Legal doctrine." *Journal of Law, Economics and Organization* 8: 441–470.

Kornhauser, L. A., and L. G. Sager (1986) "Unpacking the Court." *Yale Law Journal* 96: 82–117.

Kyburg, H. E. (1961) *Probability and the Logic of Rational Belief.* Middletown: Wesleyan University Press (1961)

Leitgeb, H. (2014) "The Stability Theory of Belief." *Philosophical Review Volume* 123(2): 131–171.

Levi, I. (2004) "List and Pettit." *Synthese* 140(1/2): 237–242.

Lin, H., and K. T. Kelly (2012a) "Propositional Reasoning that Tracks Probabilistic Reasoning." *Journal of Philosophical Logic* 41(6): 957–981.

Lin, H., and K. T. Kelly (2012b) "A Geo-logical Solution to the Lottery Paradox." *Synthese* 186: 531–575.

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

List, C. (2006) "The Discursive Dilemma and Public Reason." *Ethics* 116: 362–402.

List, C. (2012) "The Theory of Judgment Aggregation: An Introductory Review." *Synthese* 187: 179–207.

List, C. (2014) "When to defer to supermajority testimony – and when not." In J. Lackey (ed.), *Essays in Collective Epistemology.* Oxford: Oxford University Press (in press).

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

List, C., and P. Pettit (2004) "Aggregating Sets of Judgments: Two Impossibility Results Compared." *Synthese* 140(1/2): 207–235.

Nehring, K., and C. Puppe (2010) "Abstract Arrovian Aggregation." *Journal of Economic Theory* 145: 467–494.

Pauly, M., and M. van Hees (2006) "Logical Constraints on Judgment Aggregation." *Journal of Philosophical Logic* 35: 569–585.

Pettit, P. (2001) "Deliberative Democracy and the Discursive Dilemma." *Philosophical Issues* 11: 268–299.

Pigozzi, G. (2006) "Belief merging and the discursive dilemma: an argument-based account to paradoxes of judgment aggregation." *Synthese* 152: 285–298.

Appendix

Minimal conditions for the impossibility result on belief binarization

In the case of the judgment-aggregation variant of Arrow's theorem, it is known that the two combinatorial properties on the proposition set X, path-connectedness and pair-negatability, are not only sufficient, but also necessary for the theorem's negative conclusion, as long as X is finite (Dokow and Holzman 2010a, building on Nehring and Puppe 2010). In other words, as soon as the set X is either not path-connected or not pair-negatable, there exist non-dictatorial judgment-aggregation rules satisfying the four desiderata; we no longer face an impossibility.

Interestingly, the same combinatorial properties, while sufficient for our impossibility result on belief binarization, are not necessary for it. We can derive the impossibility even under weaker assumptions about the proposition set X. This is because, as we have seen, belief binarization corresponds to a particularly restrictive case of judgment aggregation: the case of propositionwise anonymous aggregation. In the presence of this special restriction, judgment aggregation also runs into an impossibility more easily.

Consider the following combinatorial property, which is weaker than path-connectedness (Nehring and Puppe 2010).

Blockedness. There is at least one proposition p in X such that there exists a path of conditional entailments from p to $\neg p$ and also a path of conditional entailments from $\neg p$ to p. (Paths of conditional entailments are defined exactly as in the definition of path-connectedness.)

For example, the set consisting of p, q, $p \leftrightarrow q$, and their negations is blocked (with $\leftrightarrow$ understood as a material biconditional), while the set consisting of p, q, $p \wedge q$, and their negations is not.[27] The following background result holds:

[27]To show that the first set is blocked, it suffices to observe that there exist paths of conditional entailments from p to $\neg p$ and back. To see the former, note that $\{p\} \cup \{p \leftrightarrow q\}$ entails q, and $\{q\} \cup \{\neg(p \leftrightarrow q)\}$ entails $\neg p$. To see the latter, note that $\{\neg p\} \cup \{p \leftrightarrow q\}$ entails $\neg q$, and $\{\neg q\} \cup \{\neg(p \leftrightarrow q)\}$ entails p. To show that the second set is not blocked, it suffices to observe that there is no path of conditional entailments from $\neg p$ to p, from $\neg q$ to q, or from $\neg(p \wedge q)$ to $p \wedge q$.

Background Result 7. There exist judgment-aggregation rules satisfying universal domain, collective consistency and completeness, propositionwise independence, consensus preservation, and anonymity for all group sizes n if and only if the proposition set X is not blocked (Dietrich and List 2013, building on Nehring and Puppe 2010).[28]

We now use this result to the derive the following:

Theorem 7. For any proposition set X that is blocked, there exists no belief-binarization rule satisfying universal domain, belief consistency and completeness, propositionwise independence, and certainty preservation. Conversely, for any finite proposition set X that is *not* blocked, there exists such a belief-binarization rule.

Before we prove this result, it is worth commenting on its significance. The present theorem establishes exact minimal conditions on X for the impossibility of belief binarization to hold. To illustrate, even for a proposition set as simple as the one consisting of p, q, $p \leftrightarrow q$, and their negations, there exists no belief-binarization rule satisfying our four baseline desiderata (because the proposition set is blocked). By contrast, for the proposition set consisting of p, q, $p \wedge q$, and their negations, there exists such a rule (because the set is not blocked), although, as we explain below, the rule is fairly "degenerate". It would accept any proposition among p, q, and $p \wedge q$ if and only if the agent assigns degree of belief 1 to it, and would accept its negation otherwise.

To prove Theorem 7, let us first assume that the proposition set X is blocked, and suppose, contrary to Theorem 7, there exists a belief-binarization rule for X satisfying all four desiderata. Call it f. We have seen that, for any group size n, f induces an anonymous judgment-aggregation rule F, via the construction described in Section 5, and this aggregation rule satisfies all four aggregation-theoretic desiderata. However, the existence of such a judgment-aggregation rule contradicts Background Result 7. So, there cannot exist a belief-binarization rule of the specified kind. This completes the negative part of the proof.[29]

Conversely, let us assume that the proposition set X is finite and not blocked. The following result holds:

Background Lemma 1. If the proposition set X is finite and not blocked, there exists at least one consistent and complete subset $B^* \subseteq X$ which has at most one proposition in common with every minimally inconsistent subset $Y \subseteq X$ (Nehring and Puppe 2010).

[28] If "if" claim assumes that the proposition set X is finite.

[29] Alternatively, some of the additional formal results in the second part of the Appendix could be used to give a direct proof of this negative result.

To illustrate, recall that the set X consisting of p, q, $p \wedge q$, and their negations is not blocked. And indeed, we can find a subset $B^* \subseteq X$ which has at most one proposition in common with every minimally inconsistent subset $Y \subseteq X$. Simply take $B^* = \{\neg p, \neg q, \neg(p \wedge q)\}$. The minimally inconsistent subsets of X are firstly all the proposition-negation pairs, with which B^* obviously has only one proposition in common, secondly the sets $\{\neg p, p \wedge q\}$ and $\{\neg q, p \wedge q\}$, with which B^* again has only one proposition in common, and finally $Y = \{p, q, \neg(p \wedge q)\}$, with which B^* also has only proposition in common.

To establish the existence of a belief-binarization rule satisfying all four desiderata, we construct one such rule. Define f as follows. For every degree-of-belief function Cr on X, let $f(Cr) = B$, where

$$B = \{p \in X : Cr(p) = 1 \text{ or } [p \in B^* \text{ and } Cr(p) \neq 0]\},$$

where B^* is as specified in Background Lemma 1. To give an intuitive flavour of this binarization rule, let us interpret B^* as a "default belief set". The present binarization rule then "accepts" a proposition p if and only if the agent assigns degree of belief 1 to that proposition or the proposition belongs to the default set and the agent does not assign degree of belief 0 to it. (This is in fact a special kind of non-uniform threshold rule, with a strict threshold of 0 for all propositions in B^* and a weak threshold of 1 for all other propositions.) Why does this binarization rule satisfy the four desiderata? Let us begin with the desiderata that are easy to check:

- f satisfies universal domain because it is well-defined for every degree-of-belief function on X.

- f satisfies propositionwise independence because, under its definition, the all-or-nothing belief in any proposition p (i.e., whether or not p is in B) depends only on the degree of belief in p (i.e., $Cr(p)$), not on the degree of belief in other propositions.

- f satisfies certainty preservation because, for any degree-of-belief function Cr that assigns extremal degrees of belief (0 or 1) to all propositions in X, the definition of f ensures that B contains all propositions p in X for which $Cr(p) = 1$ and does not contain any for which $Cr(p) = 0$.

- f satisfies belief completeness because, for every degree-of-belief function Cr and every proposition-negation pair p, $\neg p$ in X, one of the two propositions always satisfies the criterion for membership in $B = f(Cr)$.

Finally, to see that f satisfies belief consistency, suppose, for a contradiction, that $B = f(Cr)$ is inconsistent for some degree-of-belief function Cr on X. Then B, being a finite inconsistent set of propositions, has at least one minimally inconsistent subset Y. By Background Lemma 1, at most one proposition in Y occurs in B^*. Consider first the case in which there is no such proposition. For all propositions q in Y, it then follows immediately that $Cr(q) = 1$; otherwise those propositions could not have met the membership criterion for B (of which Y is a subset). But since Cr is probabilistically coherent, the fact that $Cr(q) = 1$ for all q in Y contradicts the inconsistency of Y. So let us turn to the alternative case in which exactly one proposition in Y occurs in B^*. Call it p. For all propositions q in Y distinct from p, it follows again that $Cr(q) = 1$; otherwise those propositions could not have met the membership criterion for B (of which Y is a subset). But since Y is inconsistent and Cr is probabilistically coherent, the fact that $Cr(q) = 1$ for all q in Y distinct from p implies that $Cr(p) = 0$, and so p cannot meet the membership criterion for B, a contradiction. This completes our proof that f does indeed satisfy the four desiderata on belief binarization.

To emphasize, we do not claim that f is a substantively interesting binarization rule. The point of its construction is merely to show that, as soon as the set X violates the combinatorial property of blockedness, the impossibility result of Theorem 7 no longer goes through.

Some additional formal results

In this final section, we prove some additional formal results to give some further insights into the consequences of our desiderata on the belief-belief-binarization rule f. We have already used one of these insights (about the *monotonicity* of f) as a lemma in one of our earlier proofs (in Section 9.1). Moreover, the present results allow us to establish some facts about belief binarization "directly", i.e., not as corollaries of background results on judgment aggregation.

Let f be a belief-binarization rule satisfying propositionwise independence. This implies that, for every proposition p in X, the question of whether or not p is included in the belief set B depends only on the degree of belief in p. For each proposition p, let C_p be the set of those credence values in $[0, 1]$ for which p is accepted into B, i.e.,

$$C_p = \{x \in [0, 1] : p \in f(Cr) \text{ for some admissible } Cr \text{ with } Cr(p) = x\}.$$

Call the elements of C_p the *acceptance credences* for p. We can then represent f in terms of the family $(C_p)_{p \in X}$ of sets of acceptance credences:

for any admissible Cr, $f(Cr) = \{p \in X : Cr(p) \in C_p\}$.

Claim 1. If f satisfies, in addition, universal domain and certainty preservation, then, for every proposition p in X (where $p \neq \varnothing$), we must have $1 \in C_p$.

To see this, simply consider any {0,1}-valued degree-of-belief function Cr such that $Cr(p) = 1$. Since $p \neq \varnothing$, a well-defined (i.e., probabilistically coherent) such degree-of-belief function exists, and since f satisfies universal domain, Cr is in the domain of f. Certainty preservation then implies that p must be contained in $B = f(Cr)$, and hence $1 \in C_p$.

Claim 2. If f satisfies, in addition, implication-closure, then, for any two propositions p, q in X, if p conditionally entails q, then every acceptance credence for p is also an acceptance credence for q, and thus $C_p \subseteq C_q$.

To show this, suppose p conditionally entails q, and x is an acceptance credence for p. Because of this conditional entailment, there exists a subset $Y \subseteq X$, consistent with each of p and $\neg q$, such that $\{p\} \cup Y$ entails q. It is easy to see that $\{p, q\} \cup Y$ and $\{\neg p, \neg q\} \cup Y$ are each consistent sets. For this reason, there exist well-defined degree-of-belief functions Cr' and Cr'' such that Cr' assigns credence 1 to all the propositions in the first set, and Cr'' assigns credence 1 to all the propositions in the second set. Since any linear average of well-defined degree-of-belief functions is probabilistically coherent, the degree-of-belief function $Cr = xCr' + (1 - x)Cr''$ is well-defined. Let $B = f(Cr)$. Note the following. First, we have $p \in B$, because $Cr(p) = x$, and x is an acceptance credence for p. Second, we have $Y \subseteq B$, because $Cr(r) = 1$ for every $r \in Y$, and 1 is an acceptance credence for every proposition (by Claim 1). Finally, since $\{p\} \cup Y \subseteq B$ and $\{p\} \cup Y$ entails q, we must have $q \in B$, because B is implication-closed. But $Cr(q) = x$; so x is an acceptance credence for q too.

Claim 3. If f is as in Claim 2, then for any two propositions p, q in X that are connectable, in both directions, by a sequence of pairwise conditional entailments, we have $C_p = C_q$.

This follows immediately from Claim 2. Note further that, if all propositions connectable, in both directions, by a sequence of pairwise conditional entailments, then the binarization rule f is representable by a single set $C \subseteq [0, 1]$ of acceptance credences, which are applied to every proposition in X.

Claim 4. If f is representable by a single C and X is pair-negatable, then f is monotonic, meaning that, for any x, y in [0,1] with $y > x$, if x is in C, then y is also in C. This implies that f is a threshold rule with $t = \inf(C)$. The threshold is weak if $\inf(C) \in C$ and strict otherwise.

To prove this, consider some x in C, and consider any $y > x$. Suppose X is pair-negatable. Then X has a minimally inconsistent subset Y in which we can find two distinct propositions p, q such that $Y \backslash \{p, q\} \cup \{\neg p, \neg q\}$ is consistent. Since the sets $Y \backslash \{q\} \cup \{\neg q\}$, $Y \backslash \{p, q\} \cup \{\neg p, \neg q\}$, and $Y \backslash \{p\} \cup \{\neg p\}$ are each consistent (the first and last because of Y's minimal inconsistency), there exist well-defined degree-of-belief functions Cr', Cr'', and Cr''' that assign credence 1 to all propositions in the first set, to all propositions in the second, and to all propositions in the third set, respectively. Now consider the function $Cr = xCr' + (y - x)Cr'' + (1 - y)Cr'''$. Because Cr is a linear average of three well-defined degree-of-belief functions, Cr is itself a well-defined degree-of-belief function. Let $B = f(Cr)$. Since $x \in C$, we must have $p \in B$. Since all elements of $Y \backslash \{p, q\}$ are assigned credence 1 by Cr and $1 \in C$ (by Claim 1), we must have $Y \backslash \{p, q\} \subseteq B$. Since Y in its entirety is inconsistent, $Y \backslash \{p\}$ entails $\neg q$, and hence B entails $\neg q$. By implication-closure of B, $\neg q$ must be in B. But $Cr(\neg q) = y$, and hence y is in C, as required.