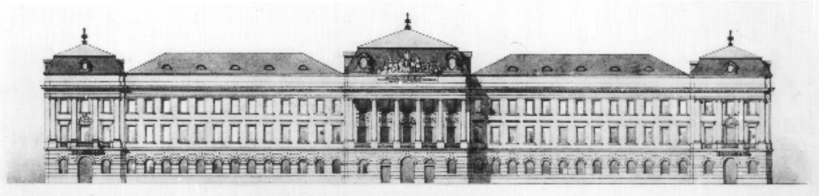
TECHNICAL
REPORT



INSTITUT FÜR LOGIC AND COMPUTATION

ABTEILUNG DATENBANKEN UND ARTIFICIAL INTELLIGENCE

Manipulating Skeptical and Credulous Consequences when Merging Beliefs

DBAI-TR-2019-114

Adrian Haret **Johannes P. Wallner**

Institut für Logic and Computation
Abteilung Datenbanken und
Artificial Intelligence
Technische Universität Wien
Favoritenstr. 9
A-1040 Vienna, Austria
Tel: +43-1-58801-18403
Fax: +43-1-58801-918403
sek@dbai.tuwien.ac.at
www.dbai.tuwien.ac.at

DBAI TECHNICAL REPORT
2019

TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

DBAI TECHNICAL REPORT

DBAI TECHNICAL REPORT DBAI-TR-2019-114, 2019

Manipulating Skeptical and Credulous Consequences when Merging Beliefs

Adrian Haret [1] **Johannes P. Wallner** [2]

Abstract. Automated reasoning techniques for multi-agent scenarios need to address the possibility that procedures for collective decision making may fall prey to manipulation by self-interested agents. In this paper we study manipulation in the context of belief merging, a framework for aggregating agents' positions, or beliefs, with respect to a set of issues represented by propositional atoms. Within this framework agents submit their positions as propositional formulas that are to be aggregated into a single formula. To reach a final decision, we employ well-established acceptance notions and extract the *skeptical* and *credulous* consequences (i.e., atoms true in all and, respectively, at least one model) of the resulting formula. We find that, even in restricted cases, most aggregation procedures are vulnerable to manipulation by an agent acting strategically, i.e., one that is able to submit a formula not representing its true position. Our results apply when the goal of such an agent is either that of (i) affecting an atom's skeptical or credulous acceptance status, or (ii) improving its satisfaction with the result. With respect to latter task, we extend existing work on manipulation with new satisfaction indices, based on skeptical and credulous reasoning. We also study the extent to which an agent can influence the outcome of the aggregation, and show that manipulation can often be achieved by submitting a *complete* formula (i.e., a formula having exactly one model), yet, the complexity of finding such a formula resides, in the general case, on the second level of the polynomial hierarchy.

[1] Institute of Logic and Computation, TU Wien, Austria. E-mail: haret@dbai.tuwien.ac.at

[2] Institute of Logic and Computation, TU Wien, Austria. E-mail: wallner@dbai.tuwien.ac.at

Acknowledgements: This work was supported by the Austrian Science Fund (FWF): P30168-N31.

1 Introduction

Collective decision making often involves the aggregation of multiple, possibly conflicting viewpoints. Apart from the matter of how to represent and aggregate such viewpoints, a looming concern in any deliberation scenario is that the agents involved may have an incentive to misrepresent their positions, and thus manipulate the aggregation result, if doing so can bring an advantage. Hence, an understanding of the potential for manipulation of any aggregation procedure is a prerequisite to its successful deployment in real world contexts.

If agents deliberate with respect to a small number of independent alternatives, as is the case in a typical election, aggregation [Zwicker, 2016] and manipulation [Conitzer and Walsh, 2016, Faliszewski and Procaccia, 2010] are well understood due to extensive research in the field of Social Choice. But if agents have to decide on multiple interconnected issues at the same time (as when electing a committee, or choosing a product specification), then the number of possible alternatives can grow too large to expect agents to have explicit preferences over the whole set. The problem, known as *combinatorial voting* in Social Choice [Lang and Xia, 2016], acquires a knowledge representation dimension as agents need compact ways to express positions over a large domain, and automatizable procedures to perform reasoning with such preferences.

Here we use belief merging as a framework for aggregating complex positions over all possible assignments of values to a set of propositional atoms [Konieczny et al., 2002, Konieczny and Pérez, 2011]. Propositional atoms, in this setting, encode the issues deliberated upon, while truth-value assignments to atoms, also called *interpretations*, encode combinations of issues that could make it into the final result, and over which agents can have preferences. In this, propositional logic suggests itself as a natural choice for representing the ways in which issues are interconnected, and lends itself naturally to the modeling of aggregation problems inspired by Social Choice [Díaz and Pérez, 2017, Diaz and Perez, 2018, Everaere et al., 2007, Everaere et al., 2015, Gabbay et al., 2009].

Within the belief merging framework each agent i submits a propositional formula K_i, which stands for i's reported belief about what are the best interpretations with respect to the issues being deliberated upon. A merging operator then aggregates the individual reported beliefs, in the presence of an integrity constraint that must be satisfied. Its result is a set of "winning" interpretations, representable as a propositional formula, that respect the integrity constraint of the merging process.

In general, the set of winning interpretations is not always expected to be the final step in a reasoning process: without further means, such a set of interpretations does not give a direct answer to which atoms (alternatives) are to be ultimately accepted. One can view the winning set as a "tie" between all the interpretations in the set. If the decision procedure needs to be explicit about every issue under consideration, then a further reasoning mechanism is required, amounting to a method of breaking ties. To this end, we employ well established *acceptance notions* from the field of knowledge representation and reasoning: skeptical and credulous consequences [Strasser and Antonelli, 2018]. An atom is a skeptical consequence of a set of interpretations if the atom is part of all interpretations, and a credulous consequence if the atom is part of at least one interpretation in the set. With regards to propositional formulas, skeptical reasoning is equivalent to (atom-wise) classical logical entailment.

Example 1. *A collective of four agents must decide who to give an award to. There are three possible candidates, represented by propositional atoms a, b and c, and the collective is operating under the constraint $\mu = a \vee b \vee c$, i.e., at least one (and possibly more) of the candidates can receive the award. The decision is arrived at by first aggregating the agents' beliefs under a known procedure, called a* belief merging operator *(details of which are reserved for later; see, e.g., Ex. 2). This produces a collective belief, potentially satisfiable by more than one interpretation: since this does not lead to an unequivocal decision, an additional tie-breaking step is required. This tie-breaking step can be thought of as a general strategy, or attitude, the collective adopts for dealing with uncertainty. In this case, we assume the collective affects a conservative (or, as we will call it,* skeptical*) approach: if there is any uncertainty with respect to a candidate, the candidate is not given the award.*

The beliefs of the agents are represented by propositional formulas, as follows. Agent 1 believes candidate b should get the award, is against candidate a and has no opinion with respect to candidate c: this is represented by the formula $K_1 = \neg a \wedge b$. Agents 2 and 3 are represented by formulas $K_2 = a \wedge (b \leftrightarrow \neg c)$ and $K_3 = b \wedge (a \rightarrow c)$, respectively. We assume that agent 4 is what we will call a strategic agent*, i.e., it is not itself compelled to submit its true belief. Agent 4's true belief happens to be $K_4^T = a \wedge \neg b \wedge \neg c$ and, were it to actually submit K_4^T, the result under the aggregation procedure would be $b \wedge \neg c$: candidate b surely gets the award, c is ruled out and there is no verdict on a. In other words, this particular aggregation procedure offers up two winning configurations (i.e., the models of the propositional formula $b \wedge \neg c$): one possible world in which a gets the award, another in which a does not get it. Thus, under the conservative tie-breaking procedure mentioned above, the final decision is arrived at by ruling out a: the final verdict is that b is the sole recipient of the award.*

Significantly, if agent 4 reports $K_4^F = a \wedge \neg b \wedge c$ instead of K_4^T, the result becomes $a \wedge c$, with the award now going to a and c. Thus, by misreporting its own belief, agent 4 ensures that its most preferred candidate a is among the recipients of the award.

Example 1 features the main ingredients of the framework we are working in: propositional logic as the language in which agents state their beliefs about the best interpretations to be included in the result, and in which the result is expressed; aggregation *via* merging operators; the need for an additional tie-breaking step; and the possibility that one agent acting strategically can influence the result to its advantage. The example also sets up the main aims of the paper: (i) formalizing strategic goals of possibly untruthful agents with respect to skeptical and credulous reasoning, (ii) investigating vulnerabilities of established merging operators to such strategic manipulation, and (iii) ways in which an agent can change (manipulate) the outcome of the aggregation process, to the extent that this is possible. Our main contributions are as follows:

- We propose to approach manipulation of skeptical or credulous consequences in two ways: (a) by considering what we call *constructive* and *destructive* manipulation, where the aim is to usher a desired atom into (or out of) the skeptical or credulous consequences, and (b) by adapting an earlier approach to manipulation [Everaere et al., 2007] that utilizes satisfaction indices to quantify the satisfaction of agents w.r.t. merged outcomes; our contribution here consists in proposing new indices.

- We give the full landscape of (non-)manipulability: concretely, we show that all main aggregation operators are manipulable (even when enforcing restrictions that yielded non-manipulability in earlier works [Everaere et al., 2007]); the sole exception is the case when aggregation is done using only so-called *complete* bases (i.e., such that each formula has exactly one model) without integrity constraint and using aggregation operator $\Delta_{\top}^{d_H,\Sigma}$ (defined below), under our new satisfaction indices.

- On the question of how an agent can manipulate, we look at general approaches to influencing the aggregation procedure by promoting or demoting interpretations. Further, we show that manipulation under skeptical consequences can be carried out by the strategic agent submitting a complete base, suggesting that manipulation does not require sophisticated propositional structures to succeed; however, in the same light, we show that deciding the existence of such a complete base is a complex problem, namely a Σ_2^P-complete problem, for destructive manipulation.

This paper improves on an earlier workshop version [Haret and Wallner, 2018].

2 Belief Merging

Propositional Logic. We assume a finite set $\mathcal{P}$ of propositional atoms, with $\mathcal{L}$ the set of formulas generated from $\mathcal{P}$ using the usual connectives. A *knowledge base* K is a formula from $\mathcal{L}$. The models of a propositional formula μ are the interpretations which satisfy it, and we write $[\mu]$ for the set of models of μ. We typically write interpretations as words where letters are the atoms assigned to true, e.g., $\{\{a, b\}, \{b, c\}\}$ is written as $\{ab, bc\}$. If $\varphi_1, \varphi_2 \in \mathcal{L}$, we say that $\varphi_1 \models \varphi_2$ if $[\varphi_1] \subseteq [\varphi_2]$, and that $\varphi_1 \equiv \varphi_2$ if $[\varphi_1] = [\varphi_2]$. A knowledge base K is *complete* if it has exactly one model. A formula φ is *consistent* (*satisfiable*), if $[\varphi] \neq \emptyset$. If v and w are interpretations, $v \triangle w$ is their symmetric difference, defined as $v \triangle w = (v \setminus w) \cup (w \setminus v)$.

Aggregation. A *profile* $P = (K_1, \dots, K_n)$ is a finite tuple of consistent bases, representing the reported beliefs of n distinct agents. We say that K_i *is agent* i*'s reported belief.* The qualification that the K_i's stand for *reported* beliefs is important, as we want to allow for the possibility of agents participating in the merging process with beliefs other than their truthful ones. We typically write K_i^T for agent i's truthful belief, and K_i^F for an untruthful belief that i reports in the merging scenario.

If P_1 and P_2 are profiles, we write $P_1 + P_2$ for the profile obtained by appending P_2 to P_1. If K is a base and there is no danger of ambiguity, we write $P + K$ instead of $P + (K)$.

A merging operator Δ is a function mapping a profile P of consistent knowledge bases and a propositional formula μ, called *the constraint*, to a propositional formula, written

$\Delta_\mu(P)$. We focus on semantic operators $\Delta^{d,f}$ from the framework of logic-based merging [Konieczny and Pérez, 2011], the main ingredients of which are a distance d and an aggregation function f. To define these operators we start with a distance d between interpretations. Given a distance d between interpretations, an interpretation w and a propositional formula φ, *the distance* $d(w, \varphi)$ *from* w *to* φ is defined as $d(w, \varphi) = \min\{d(w, v) \mid v \in [\varphi]\}$. This makes it possible to order interpretations w.r.t. bases: $w_1 \leq_K^d w_2$ if $d(w_1, K) \leq d(w_2, K)$. For a profile $P = (K_1, \dots, K_n)$ and an aggregation function f, *the distance* d^f *from* w *to* P is $d^f(w, P) = f(d(w, K_1), \dots, d(w, K_n))$. That is, $d_f(w, P)$ is the result of aggregating, *via* f, the distances between w and each $K_i \in P$.

We assume that distances from interpretations to profiles can be compared using an order $\leq$, such that, for any interpretations w_1 and w_2, we have either $d^f(w_1, P) \leq d^f(w_2, P)$ or $d^f(w_2, P) \leq d^f(w_1, P)$. We say that $w_1 \leq_P^{d,f} w_2$ if $d^f(w_1, P) \leq d^f(w_2, P)$. If d is a distance between interpretations and f is an aggregation function, the *propositional merging operator* $\Delta^{d,f}$ is defined, for any profile P and constraint μ, as $[\Delta_\mu^{d,f}(P)] = \min_{\leq_P^{d,f}}[\mu]$. The result of aggregating the bases in P thus consists of the models of μ, also called *the winning interpretations*, at minimum overall distance to the consistent bases in P, with distances specified via d and aggregation function f.

We will focus on a sample of representative merging operators, constructed using a set of common distance/aggregation functions. If w_1 and w_2 are interpretations, the *drastic* and *Hamming* distances d_D and d_H, respectively, are defined as follows:

$$d_D(w_1, w_2) = \begin{cases} 0, & \text{if } w_1 = w_2, \\ 1, & \text{otherwise,} \end{cases} \qquad\qquad d_H(w_1, w_2) = |w_1 \triangle w_2|.$$

If $X = (x_1, \dots, x_n)$ is an n-tuple of non-negative integers, the Σ, $\texttt{max}$ and $\texttt{gmax}$ aggregation functions are defined as follows:

- $\Sigma(X) = \Sigma_{i=1}^n x_i$,
- $\texttt{max}(X) = \max(\{x_i \mid 1 \leq i \leq n\})$, and
- $\texttt{gmax}(X)$ is X in descending order.

For $f \in \{\Sigma, \texttt{max}\}$ the aggregated value $d^f(w, P)$ is an integer and thus interpretations can be ordered w.r.t. their distance to P. For $f = \texttt{gmax}$, $d^{\texttt{gmax}}(v, P)$ is an n-tuple made up of the numbers $d(w, K_1), \dots, d(w, K_n)$ ordered in descending order. To rank interpretations via $\texttt{gmax}$ we order vectors lexicographically: $(x_1, \dots, x_n) <_{lex} (y_1, \dots, y_n)$ if $x_i < y_i$ for the first i where x_i and y_i differ. We recall that if $X = (x_1, \dots, x_n)$ and $Y = (y_1, \dots, y_n)$ are n-tuples of non-negative integers, $z \in \mathbb{N}$, π is a permutation of $\{1, \dots, n\}$ and $f \in \{\Sigma, \texttt{max}, \texttt{gmax}\}$ is an aggregation function, the following properties hold [Konieczny and Pérez, 2011]: $f(x_1, \dots, x_n) = f(x_{\pi(1)}, \dots, x_{\pi(2)})$ (symmetry); and if $x_i \leq x_i'$, then $f(x_1, \dots, x_i, \dots, x_n) \leq f(x_1, \dots, x_i', \dots, x_n)$ (monotony).

Example 2. *The scenario described in Example 1 features two aggregation tasks: one involving the profile $P^T = (K_1, K_2, K_3, K_4^T)$, containing the true position of agent 4; the other involving the profile $P^F = (K_1, K_2, K_3, K_4^F)$, obtained by agent 4 acting strategically. Both tasks*

Table 1: Example of merging. Gray cells are the permitted models when integrity constraint $\mu = (a \vee b \vee c)$. Column 1 contains all interpretations over the alphabet $\mathcal{P} = \{a, b, c\}$, columns 2-6 show the minimal (Hamming) distances between interpretations and bases (see Example 1 for what the bases are). Columns 7-10 show the aggregated distances under Σ and $\mathtt{gmax}$ with respect to the profiles P^T and P^F. Bold numbers indicate models with minimum distance.

	$[K_1]$	$[K_2]$	$[K_3]$	$[K_4^T]$	$[K_4^F]$				
	$\{b, bc\}$	$\{ab, ac\}$	$\{b, bc, abc\}$	$\{a\}$	$\{ac\}$	$d_H^{\Sigma}(\cdot, P^T)$	$d_H^{\Sigma}(\cdot, P^F)$	$d_H^{\mathtt{gmax}}(\cdot, P^T)$	$d_H^{\mathtt{gmax}}(\cdot, P^F)$
$\emptyset$	1	2	1	1	2	5	6	(2,1,1,1)	(2,2,1,1)
a	2	1	2	0	1	5	6	(2,2,1,0)	(2,2,1,1)
b	0	1	0	2	3	**3**	4	(2,1,0,0)	(3,1,0,0)
c	1	1	1	2	1	5	4	(2,1,1,1)	(1,1,1,1)
ab	1	0	1	1	2	**3**	4	**(1,1,1,0)**	(2,1,1,0)
ac	2	0	1	1	0	4	**3**	(2,1,1,0)	(2,1,0,0)
bc	0	2	0	3	2	5	4	(3,2,0,0)	(2,2,0,0)
abc	1	1	0	2	1	4	**3**	(2,1,1,0)	**(1,1,1,0)**

occur under the same constraint $\mu = a \vee b \vee c$. Table 1 illustrates the results of aggregating profiles P^T and P^F under constraint μ with operators $\Delta_{\mu}^{d_H, \Sigma}$ and $\Delta_{\mu}^{d_H, \mathtt{gmax}}$. The aggregation result is computed by choosing, from the models of μ, the ones with minimum aggregated distance. For instance, we have $K_2 = a \wedge (b \leftrightarrow \neg c)$. Further, we have $[K_2] = \{ab, ac\}$ and $d_H(ab, K_2) = \min\{d_H(ab, ab), d_H(ab, ac)\} = \min\{0, 2\} = 0$. The following holds: $d_H^{\Sigma}(ab, P^T) = d_H(ab, K_1) + d_H(ab, K_2) + d_H(ab, K_3) + d_H(ab, K_4^T) = 3$. The orders $\leq_{P^T}^{d_H, f}$ and $\leq_{P^F}^{d_H, f}$, for $f \in \{\Sigma, \mathtt{gmax}\}$, are obtained by ordering interpretations according to their aggregated distances to P^T and P^F, respectively. Finally, we get that $[\Delta_{\mu}^{d_H, \Sigma}(P^T)] = \{b, ab\}$, $[\Delta_{\mu}^{d_H, \Sigma}(P^F)] = \{ac, abc\}$, $[\Delta_{\mu}^{d_H, \mathtt{gmax}}(P^T)] = \{ab\}$ and $[\Delta_{\mu}^{d_H, \mathtt{gmax}}(P^F)] = \{abc\}$.

It is worth mentioning that $\Delta_{\mu}^{d_D, \Sigma}$ and $\Delta_{\mu}^{d_D, \mathtt{gmax}}$ are equivalent, for any profile P and constraint μ (i.e., $[\Delta_{\mu}^{d_D, \Sigma}(P)] = [\Delta_{\mu}^{d_D, \mathtt{gmax}}(P)]$). Further, the operator $\Delta_{\mu}^{d_D, \mathtt{max}}$ delivers $[\bigwedge P \wedge \mu]$, if consistent, and $[\mu]$ otherwise.

3 Acceptance and Satisfaction Notions

Merging operators output a set of interpretations, all of which can be seen as tied for the winning position. In decision terms, this translates as inconclusiveness with respect to the final verdict (see Example 1). To arrive at a definite opinion on every issue we use well-established notions of *acceptance* with respect to a formula. Further, in order to make sense of the way an agent can manipulate, we need to be able to measure an agent's satisfaction with respect to the result of a merging operator. To this end we introduce a set of *satisfaction indices* that build on the acceptance notions.

Acceptance. An *acceptance function* $\mathtt{Acc}\colon \mathcal{L} \to 2^{\mathcal{P}}$ maps propositional formulas to sets of atoms in $\mathcal{P}$. We say that $\mathtt{Acc}(\varphi)$ *are the accepted atoms of* φ. For a formula φ, we define the following acceptance notions:

$$\mathtt{Skept}(\varphi) = \bigcap_{w \in [\varphi]} w, \qquad \mathtt{Cred}(\varphi) = \bigcup_{w \in [\varphi]} w.$$

For a formula φ, an atom is *skeptically accepted* if it is true in all models of φ (i.e., is in $\mathtt{Skept}(\varphi)$); an atom is *credulously accepted* if it is true in at least one model of φ (i.e., is in $\mathtt{Cred}(\varphi)$).[1] Skeptical acceptance is equivalent to atom-wise logical entailment, and credulous acceptance indicates support of an atom in at least one model.

Example 3. *In Example 2 we obtain that* $[\Delta_{\mu}^{d_H,\Sigma}(P^T)] = \{b, ab\}$. *For the acceptance notions introduced, we have* $\mathtt{Skept}(\Delta_{\mu}^{d_H,\Sigma}(P^T)) = b$ *and* $\mathtt{Cred}(\Delta_{\mu}^{d_H,\Sigma}(P^T)) = ab$.

These acceptance notions focus on positive literals. Thus, we say that $p \in \mathtt{Skept}(\varphi)$ if the atom p is in every model of φ, but we do not treat acceptance of negative literals in a similar fashion: for instance, in Example 3 we do not say something like '$\mathtt{Skept}(\Delta_{\mu}^{d_H,\Sigma}(P^T)) = b\neg c$', even though c is in none of (and hence rejected by) all the models of $\Delta_{\mu}^{d_H,\Sigma}(P^T)$. This asymmetry is not unusual in a Social Choice context, where rejection of a candidate is often assimilated to non-acceptance, but would be worth looking at in a more extensive treatment of acceptance notions.

Satisfaction. A *satisfaction index* $i\colon \mathcal{L} \times \mathcal{L} \to \mathbb{N}^+$ is a function that maps a pair of formulas to a non-negative integer [Everaere et al., 2007]. If φ and ψ are two propositional formulas and $\mathtt{Acc}$ is an acceptance notion, *the satisfaction index* $i_{\mathtt{Acc}}$ is defined as $i_{\mathtt{Acc}}(\varphi, \psi) = |\mathtt{Acc}(\varphi) \triangle \mathtt{Acc}(\psi)|$. For the two acceptance notions introduced above, this gives us the satisfaction indices $i_{\mathtt{Skept}}$ and $i_{\mathtt{Cred}}$.

Example 4. *For* K_4^T *from Ex. 2 we have* $[K_4^T] = \{a\}$ *and* $[\Delta_{\mu}^{d_H,\Sigma}(P^T)] = \{b, ab\}$. *With the indices we can measure agent 4's satisfaction regarding the truthful aggregation result: we have* $i_{\mathtt{Skept}}(K_4^T, \Delta_{\mu}^{d_H,\Sigma}(P^T)) = |\mathtt{Skept}(K_4^T) \triangle \mathtt{Skept}(\Delta_{\mu}^{d_H,\Sigma}(P^T))| = |a \triangle b| = 2$. *Analogously,* $i_{\mathtt{Cred}}(K_4^T, \Delta_{\mu}^{d_H,\Sigma}(P^T)) = |a \triangle ab| = 1$.

For arbitrary formulas the numeric results given by the indices $i_{\mathtt{Skept}}$ and $i_{\mathtt{Cred}}$ are generally not directly correlated, in that each may be higher or lower than the other. However, there is a duality relation between the indices and aggregation operators defined via skeptical and credulous acceptance. *The dual* $\overline{\varphi}$ *of a formula* φ is obtained by replacing every literal in φ with its negation. If $P = (K_1, \ldots, K_m)$ is a profile, then *the dual* $\overline{P}$ *of* P is the profile defined as $\overline{P} = (\overline{K_1}, \ldots, \overline{K_m})$. If w is an interpretation, *the dual* $\overline{w}$ *of* w is the complement of w, i.e., the interpretation $\mathcal{P} \setminus w$. If W is a set of interpretations, *the dual* $\overline{W}$ *of* W is the set of interpretations defined as $\overline{W} = \{\overline{w} \mid w \in W\}$. For a propositional formula φ we have $\overline{[\varphi]} = [\overline{\varphi}]$. This transfers to the indices: it holds that

[1] We note that the notions of *skeptical* (cautious) and *credulous* (brave) consequences are not uniformly used throughout the literature. For instance, skeptical consequences may be defined as those consequences that follow (e.g. by classical logic) from all formulas in a set of formulas, and skeptical acceptance may refer to membership of an object in all sets of a given set of sets. We make use of the latter interpretation.

$i_{\mathtt{Skept}}(\varphi, \psi) = i_{\mathtt{Cred}}(\overline{\varphi}, \overline{\psi})$. Intuitively, this is because an atom p being in the symmetric difference of the skeptical consequences is equivalent to there being a model of one of the formulas not containing p, with the dual having p in at least one model. Interestingly, a duality also holds with respect to merging operators.

Proposition 1. *If P is a profile, μ is a constraint, $d \in \{d_H, d_D\}$ is a distance function, and $f \in \{\Sigma, \mathtt{max}, \mathtt{gmax}\}$ is an aggregation function, then* $\mathtt{Skept}(\Delta^{d,f}_{\mu}(P)) \equiv \mathtt{Cred}(\Delta^{d,f}_{\overline{\mu}}(\overline{P}))$.

Proposition 1 builds on an interesting symmetry exhibited by the merging operators we work with: the result of merging a profile P under a constraint μ and the result of merging $\overline{P}$ under constraint $\overline{\mu}$ turn out to be themselves duals of each other. This allows us, once we have found some instance related to the skeptical index, to automatically adapt it to the credulous index.

Example 5. *For the alphabet $\mathcal{P} = \{a, b\}$, take a profile $P = (K_1, K_2)$, with $K_1 = a \rightarrow b$, $K_2 = \neg a$ and $\mu = a$. We get $[\Delta^{d_H,\Sigma}_{\mu}(P)] = \{ab\}$, and* $\mathtt{Skept}(\Delta^{d_H,\Sigma}_{\mu}(P)) = ab$. *Taking the duals, we have $\overline{K_1} = \neg a \rightarrow \neg b$, $\overline{K_2} = a$ and $\overline{\mu} = \neg a$. Notice that $[K_1] = \{\emptyset, b, ab\}$ and $[\overline{K_1}] = \{ab, a, \emptyset\} = \{\overline{\emptyset}, \overline{b}, \overline{ab}\} = \overline{[K_1]}$, i.e., the models of the dual of K_1 are the duals of the models of K_1. We get that $[\Delta^{d_H,\Sigma}_{\overline{\mu}}(\overline{P})] = \{\emptyset\}$, which is the same as $\overline{[\Delta^{d_H,\Sigma}_{\mu}(P)]}$ (this equality also holds more generally). Lastly,* $\mathtt{Skept}(\Delta^{d_H,\Sigma}_{\mu}(P)) = \mathtt{Cred}(\Delta^{d_H,\Sigma}_{\overline{\mu}}(\overline{P}))$.

4 Manipulability and Strategyproofness

Manipulation occurs when an agent, called *the strategic agent*, can influence the merging result in its favor by submitting a base different from its truthful one. Unless otherwise stated, the agent's truthful position is the base K^T, and the base with which it manipulates as K^F. We represent the strategic agent's contribution by appending its submitted base to a pre-existing profile P (e.g., $P + K^T$): intuitively, it is as if the strategic agent joins the aggregation process *after* everyone else has submitted their positions. This is merely a notational choice, meant to improve readability, and no generality is lost in this way: all aggregation functions used here satisfy the symmetry property (see Section 2) and the result never depends on the merging order.

A profile P, constraint μ, distance d, aggregation function f and acceptance notion $\mathtt{Acc}$ are assumed in most definitions, but, in the interest of concision, are explicitly referred to only under pain of ambiguity. Unless otherwise stated, d ranges over $\{d_D, d_H\}$ and f over $\{\Sigma, \mathtt{gmax}, \mathtt{max}\}$.

4.1 Constructive and destructive manipulation with respect to an atom

One of the most basic forms of manipulation is one in which the strategic agent has a specific atom p that it targets for acceptance: the strategic agent may want to see p get accepted (or rejected) in the final result. This sets up the stage for what we call, along the lines of similar concepts from Social Choice [Conitzer and Walsh, 2016], *constructive* and *destructive* manipulation. The strategic agent *constructively* $\mathtt{Acc}$*-manipulates P w.r.t. p using K^F* if $p \notin \mathtt{Acc}(\Delta_{\mu}(P + K^T))$ and $p \in \mathtt{Acc}(\Delta_{\mu}(P + K^F))$, and *destructively* $\mathtt{Acc}$*-manipulates P w.r.t. p*

using K^F if $p \in \texttt{Acc}(\Delta_\mu(P+K^T))$ and $p \notin \texttt{Acc}(\Delta_\mu(P+K^F))$. Intuitively, an agent constructively Acc-manipulates w.r.t. p if it can make p be in the accepted atoms of the aggregation result by submitting K^F instead of K^T; similarly, an agent destructively manipulates w.r.t. p if it can kick p out of the accepted atoms of the result. We say that an operator Δ is Acc-*strategyproof* if there is no profile P, constraint μ, atom p and bases K^T and K^F s.t. the strategic agent, having K^T as its truthful position, Acc-manipulates P, either constructively or destructively, w.r.t. p using K^F.

We first note that, if K^T is the strategic agent's truthful position, any instance of constructive manipulation with respect to p using K^F is also an instance of destructive manipulation with respect to p, obtained by swapping K^T and K^F as the truthful and manipulating bases, respectively. Next, our results regarding duality (see Proposition 1) imply the following duality for manipulation.

Proposition 2. *A strategic agent constructively (destructively)* Skept*-manipulates* P *with respect to* p *iff it destructively (constructively)* Cred*-manipulates* $\overline{P}$ *with respect to* p *using* $\overline{K^F}$*, with* $\overline{K^T}$ *as its truthful position and* $\overline{\mu}$ *as the constraint.*

In other words, an instance of constructive Skept-manipulation has a direct counterpart, *via* the duals, in an instance of destructive Cred-manipulation, and likewise for destructive Skept-manipulation and constructive Cred-manipulation. This simplifies our study as we can focus on only one acceptance notion, with results for the other notion following by Proposition 2.

Example 6. *In Example 2 agent 4 constructively* Skept*-manipulates the profile* $P = (K_1, K_2, K_3)$ *w.r.t the atom* a *(relative to the operator* $\Delta^{d_H,\Sigma}$ *and constraint* $\mu = a \vee b \vee c$*), in that* $a \notin \texttt{Skept}(\Delta_\mu^{d_H,\Sigma}(P^T))$ *but* $a \in \texttt{Skept}(\Delta_\mu^{d_H,\Sigma}(P^F))$*. Consider, now, a merging scenario where every formula is replaced by its dual. Thus, the truthful position of agent 4 is* $\overline{K_4^T}$*: we get that* $[\overline{K_4^T}] = \{bc\}$*, the constraint is* $\overline{\mu} = \neg a \vee \neg b \vee \neg c$*, with* $[\overline{\mu}] = \{\emptyset, a, b, c, ab, ac, bc\}$*, and the profile is* $\overline{P}$*. We get that* $[\Delta_{\overline{\mu}}^{d_H,\Sigma}(\overline{P^T})] = \{c, ac\}$*, and* $a \in \texttt{Cred}(\Delta_{\overline{\mu}}^{d_H,\Sigma}(\overline{P^T}))$*. However, if agent 4 now submits* $\overline{K_4^F}$*, we get that* $[\Delta_{\overline{\mu}}^{d_H,\Sigma}(\overline{K_4^F})] = \{\emptyset, b\}$*, with* $a \notin \texttt{Cred}(\Delta_{\overline{\mu}}^{d_H,\Sigma}(\overline{K_4^F}))$*. Hence, if agent 4's truthful position is* $\overline{K_4^T}$*, then it destructively* Cred*-manipulates* $\overline{P}$ *w.r.t* a *using* $\overline{K_4^F}$*.*

Examples 2 and 6 already show that $\Delta^{d_H,\Sigma}$ is constructively Skept-manipulable (and destructively Cred-manipulable). Indeed, this extends to all operators introduced so far.

Theorem 1. *For any* $n \in \mathbb{N}$ *and* $p \in \mathcal{P}$*, there exists a profile* $P = (K_1, \ldots, K_n)$ *and bases* K^T*,* K^F *such that the strategic agent constructively (and destructively, respectively)* Acc*-manipulates* P *w.r.t* p *using* K^F*, even if* $\mu = \top$ *and all* K_i*, for* $i \in \{1, \ldots, n\}$*, as well as* K^T *and* K^F*, are complete.*

Theorem 1 suggests that the situation with respect to constructive/destructive manipulation is acute, for two reasons. Firstly, restrictions on the size of the profile or on the specificity of the bases (e.g., requiring that all bases are complete), which ensure strategyproofness in other contexts [Everaere et al., 2007], turn out not to have any effect in this case. Second, instances of manipulation exist for *any* size of the profile P: this is best understood by consulting Example 7.

Example 7. *To constructively* Skept*-manipulate a profile of size* $n = 4$ *w.r.t. the atom* a*, relative to the constraint* $\mu = \top$ *and* $f \in \{\Sigma, \texttt{gmax}\}$*, take* K_i*, for* $i \in \{1, 2, 3, 4\}$*,* K^T *and* K^F *as in Table 2.*

Table 2: Constructive Skept-manipulation of a profile of size 4 w.r.t the atom a

	$[K_1]$	$[K_2]$	$[K_3]$	$[K_4]$	$[K^T]$	$[K^F]$				
	$\{\emptyset\}$	$\{\emptyset\}$	$\{a\}$	$\{a\}$	$\{\emptyset\}$	$\{a\}$	$d^{\Sigma}(\cdot, P^T)$	$d^{\mathtt{gmax}}(\cdot, P^T)$	$d^{\Sigma}(\cdot, P^F)$	$d^{\mathtt{gmax}}(\cdot, P^F)$
$\emptyset$	0	0	1	1	0	1	2	$(1,1,0,0,0)$	3	$(1,1,1,0,0)$
a	1	1	0	0	1	0	3	$(1,1,1,0,0)$	2	$(1,1,0,0,0)$
b	1	1	2	2	1	2	7	$(2,2,1,1,1)$	8	$(2,2,2,1,1)$
...	...	...	...	...	...	...	...	...	...	...

It is straightforward to see that $[\Delta_{\mu}^{d,f}(P^T)] = \{\emptyset\}$ and $[\Delta_{\mu}^{d,f}(P^F)] = \{a\}$, for $d \in \{d_D, d_H\}$ and $f \in \{\Sigma, \mathtt{gmax}\}$ (Table 2 shows results for d_H, but the reasoning for d_D is entirely similar). This example easily generalizes to any even n. If n is odd, which we can write as $n = 2p+1$, for $p \in \mathbb{N}$, we can take $[K_1] = \cdots = [K_p] = \{\emptyset\}$, $[K_{p+1}] = \cdots = [K_n] = \{a\}$, and K^T, K^F as above.

If possible for an agent to constructively or destructively manipulate, it is appropriate to ask *how* it can do it: are intricate formulas needed to achieve the goal, or can a 'simple' base work just as well? In Example 7 the strategic agent manipulates using complete bases, suggesting that the answer lies with the second option. Indeed we can show that, if manipulation is possible at all, then it can be done with a complete base.

Theorem 2. *If the strategic agent constructively/destructively* Acc*-manipulates P w.r.t. p using K^F, for* Acc $\in \{$Skept, Cred$\}$*, then there exists a complete base K_*^F such that $K_*^F \models K^F$ and the agent constructively/destructively* Skept*-manipulates P w.r.t. p using K_*^F.*

We give here the intuition driving the proof for Skept-manipulation, adding as well the fact that the base K_*^F is found in the same way for constructive and destructive manipulation: if manipulation is possible with K^F, then pick a model of K^F that is closest to one of the models of μ crucial for the success of manipulation. In the case of destructive Skept-manipulation, this would be an interpretation v_* that ends up being in $[\Delta_{\mu}^{d,f}(P + K^F)]$ and is such that $p \notin v_*$: v_* must exist, under the assumption that K^F successfully achieves destructive Skept-manipulation. We can then replace K^F with K_*^F, where $[K_*^F] = \{v_*\}$ and still achieve destructive Skept-manipulation.

There is one thing that mitigates the acuteness of the manipulation results. Note that we have not assumed so far that the strategic agent needs to have p among its accepted atoms, i.e., we do not require the agent to actually *believe* p in order to constructively/destructively manipulate with respect to it. Seeing the merging process as aggregation of agents' *reported* beliefs, stressed in Section 2, comes into play, as it allows for agents to participate with bases that can reflect a richer cognitive structure (e.g., the effects of bribery, or influence, motivating an agent to alter its reported beliefs). Thus, here we operate under the assumption that p (its acceptance, or otherwise) figures for the agent as a goal, regardless of whether it is actually part of its beliefs (manipulation furthering the truthful beliefs of the strategic agent is treated in Section 4.2).

Can an agent influence the acceptance of an atom it does not believe? We see in Example 7 that the answer is yes: the strategic agent there is able to constructively Skept-manipulate w.r.t. a even though a is not among the skeptical beliefs of the agent itself. And, in fact, we are able to show that, when $\mu = \top$ and all bases are complete, Skept-manipulation is possible only under this assumption.

Proposition 3. *If the strategic agent constructively* Skept*-manipulates P with respect to an atom p, relative to the constraint $\mu = \top$ and operator $\Delta^{d_H,\Sigma}$, when all bases are complete, then $p \notin$* $\texttt{Skept}(K^T)$.

Proposition 3 can be seen as a positive result, one way of reading it being that if the strategic agent already accepts p (i.e., $p \in \texttt{Skept}(K^T)$), then if it cannot impose p by submitting K^T itself, for the given parameters, then there is no other way of doing it. As such, this is the closest we can come to a strategyproofness result for constructive/destructive manipulation.

4.2 Manipulation with respect to a satisfaction index

Constructive and destructive manipulation deals with the question of whether an agent can affect the acceptance of an atom in the aggregated outcome, regardless of the beliefs of the agent. In this section we look at the case when the agent improves the outcome with respect to its beliefs, where improvement is measured using the skeptical and credulous satisfaction indices $i_{\texttt{Acc}}$, for $\texttt{Acc} \in \{\texttt{Skept}, \texttt{Cred}\}$.

The strategic agent *manipulates P with respect to $i_{\texttt{Acc}}$ using K^F* if it holds that $i_{\texttt{Acc}}(K^T, \Delta_\mu(P + K^F)) < i_{\texttt{Acc}}(K^T, \Delta_\mu(P + K^T))$. In other words, an agent can improve its satisfaction index by submitting K^F instead of K^T. We say that an operator Δ is *strategy-proof with respect to a satisfaction index $i_{\texttt{Acc}}$* if there is no profile P, constraint μ and bases K^T and K^F such that the strategic agent, having K^T as its truthful position, manipulates P with respect to $i_{\texttt{Acc}}$ using K^F.

Our definition of manipulability based on satisfaction indices is inspired by previous work on manipulation of propositional merging operators [Everaere et al., 2007] but differs from it in an important respect: we measure the distance between the *accepted* atoms of the manipulating agent and the result, rather than between the sets of models themselves. A more minor (technical) difference is that, in our case, an agent is more satisfied when its index *decreases*.[2] This reflects the fact that the manipulated result gets closer to the agent's beliefs.

Example 8. *In Example 2, we have* $\texttt{Skept}(K_4^T) = a$ *and* $\texttt{Skept}(\Delta_\mu^{d_H,\Sigma}(P^T)) = b$. *Thus,* $i_{\texttt{Skept}}(K_4^T, \Delta_\mu^{d_H,\Sigma}(P^T)) = |a \triangle b| = 2$. *However, by agent* 4 *submitting* K_4^F *instead of* K_4^T, *we get that* $\texttt{Skept}(\Delta_\mu^{d_H,\Sigma}(P^F)) = ac$ *and* $i_{\texttt{Skept}}(K_4^T, \Delta_\mu^{d_H,\Sigma}(P^F)) = 1$. *Thus, by submitting a position different from its truthful one, agent 4 is able to bring the (skeptically accepted atoms of) the merging result closer to its own position.*

Example 8 shows that manipulation is possible in the general case for the merging operator $\Delta^{d_H,\Sigma}$ and the skeptical index. What is, now, the full picture with respect to manipulability? As for constructive and destructive manipulation, we first note that the identity $i_{\texttt{Skept}}(\varphi, \psi) = i_{\texttt{Cred}}(\overline{\varphi}, \overline{\psi})$

[2] As such, our indices can be interpreted as dissatisfaction indices; nevertheless we stick to the term satisfaction index.

(see Sect. 3) allows us to turn a manipulation instance with respect to $i_{\texttt{Skept}}$ into a manipulation instance with respect to $i_{\texttt{Cred}}$ simply by replacing every formula involved with its dual.

For the operators $\Delta^{d,\texttt{gmax}}$ and $\Delta^{d,\texttt{max}}$ index manipulation turns out to be, like atom manipulation, unavoidable. This stays so even under heavy restrictions (i.e., complete bases and $\mu = \top$), and for any size $n \geq 2$ of the profile.

Theorem 3. *For $d \in \{d_D, d_H\}$, $f \in \{\texttt{gmax}, \texttt{max}\}$ and any $n \geq 2$ there exists a profile $P = (K_1, \dots, K_n)$ and bases K^T and K^F such that the strategic agent manipulates P with respect to $i_{\texttt{Acc}}$, even if $\mu = \top$ and all bases K_i, for $i \in \{1, \dots, n\}$, as well as K^T and K^F, are complete.*

The story is different for the operator $\Delta^{d_H,\Sigma}$: as seen in Proposition 3, constructive manipulation for skeptical acceptance, complete profiles, and $\mu = \top$ can only get an atom p into the result if the agent does not believe p. In other words, the result can be affected for p, but it is worth noting that the skeptical index does not increase by doing so. It turns out that this holds in general for the operator $\Delta^{d_H,\Sigma}_{\top}$, i.e., this operator is strategy-proof with respect to a satisfaction index $i_{\texttt{Acc}}$, for $\texttt{Acc} \in \{\texttt{Skept}, \texttt{Cred}\}$.

Theorem 4. *If all bases in the profile, as well as K^T and K^F, are complete and $\mu = \top$, then the operator $\Delta^{d_H,\Sigma}_{\top}$ is strategy-proof with respect to $i_{\texttt{Skept}}$ and $i_{\texttt{Cred}}$.*

Proof. (sketch) For complete profiles, the operator $\Delta^{d_H,\Sigma}_{\top}$ returns models v that reflect majority opinion, i.e., if an atom p is true in a majority of bases, p is in v; if p is false in a majority of bases, then p is not in v; and if there is no majority (half of the bases have p in their model), then the result contains both a v with p and a v' without p in them. A strategic agent cannot increase its index: adding something to its model can make this skeptically accepted, but this is not in the agent's belief (similarly for other cases). □

The restrictions on $\Delta^{d_H,\Sigma}$ in Theorem 4 are essential: weakening any of them results in the operator being manipulable.

Proposition 4. *If it is does not hold that $\mu = \top$ and all bases in P, as well as the truthful position of the strategic agent, are complete, then $\Delta^{d_H,\Sigma}$ is manipulable with respect to $i_{\texttt{Acc}}$.*

5 Influence of one agent over the outcome

Section 4 addresses the question of whether the strategic agent can modify the merging result to its advantage. But it is useful to take a step back and ask whether the strategic agent can modify the result in the first place, i.e., whether it matters if the strategic agent takes part in the merging process at all and, if yes, how exactly it can influence it. Given a profile P, an operator Δ, a constraint μ and a base K, we say that $\Delta_\mu(P)$ is *the intermediary result*, and $\Delta_\mu(P + K)$ is *the final result*.

There are, *a priori*, two ways in which the agent can change the intermediary result: one is by removing interpretations from $[\Delta_\mu(P)]$; i.e., by turning winning interpretations into non-winning interpretations; the other is by adding interpretations to $[\Delta_\mu(P)]$, i.e., by turning non-winning interpretations into winners. If w is an interpretation, we say that the strategic agent *demotes w from $\Delta_\mu(P)$ using K* if $w \in [\Delta_\mu(P)]$ and $w \notin [\Delta_\mu(P+K)]$, and that it *promotes w with respect to $\Delta_\mu(P)$ using K* if $w \notin [\Delta_\mu(P)]$ and $w \in [\Delta_\mu(P+K)]$.

It turns out that for a significant proportion of the operators we are working with the strategic agent can demote any number of interpretations from the intermediary result, using an easy strategy: focus on the wanted interpretations, and submit a base with those interpretations as models; the unwanted interpretations thus receive a penalty that renders them non-winning in the final result.

Proposition 5. *If P is a profile, μ is a constraint, $d \in \{d_H, d_D\}$, $f \in \{\Sigma, \mathtt{gmax}\}$ and $W \subset [\Delta_\mu^{d,f}(P)]$ is a set of interpretations, then a strategic agent can demote all interpretations in $[\Delta_\mu^{d,f}(P)] \setminus W$ from $\Delta_\mu^{d,f}(P)$ by submitting K_W with $[K_W] = W$.*

On the other hand, promoting interpretations is more difficult: the strategic agent's ability to promote an interpretation w depends on the margin by which w loses out to the winning interpretations. We show this here for the operator $\Delta^{d_H,\Sigma}$.

Proposition 6. *If w is an interpretation such that $w \in [\mu]$ and $w \notin [\Delta_\mu^{d_H,\Sigma}(P)]$, then the strategic agent can promote w with respect to $\Delta_\mu^{d_H,\Sigma}(P)$ iff $d_H^\Sigma(w,P) - d_H^\Sigma(w_i,P) \leq d_H(w,w_i)$, for every $w_i \in [\Delta_\mu^{d_H,\Sigma}(P)]$.*

Intuitively, $d_H^\Sigma(w,P) - d_H^\Sigma(w_i,P)$ is the margin by which w loses out to a winning interpretation w_i in $\leq_P^{d,f}$. Proposition 6 then tells us that the strategic agent can reverse the order between w and w_i if and only if this margin is less than the Hamming distance between w and w_i. In general, the amount of support the strategic agent can give to w relative to w_i is at most $d_H(w,w_i)$ and thus, if w is trailing w_i by more than this amount, there is nothing the strategic agent can do for it. Using this result we note that, if possible for an agent to promote an interpretation w, then it can do so using a complete base.

Corollary 1. *If the strategic agent can promote an interpretation w with respect to $\Delta_\mu^{d,f}(P)$, then it can do so with a base K_w such that $[K_w] = \{w\}$.*

This result is similar in spirit to Theorem 2, and suggests something like a best strategy if the goal is to promote w: the strategic agent can always submit a base K_w with w as the sole model, since if w can be promoted to the final result then K_w is guaranteed to do it; otherwise, it does not matter what the agent submits.

Example 9. *Suppose $[\mu] = \{w_1, w_2, w_3, w_4\}$, $[\Delta_\mu^{d,\Sigma}(P)] = \{w_1, w_2, w_3\}$, for $d \in \{d_H, d_D\}$. The strategic agent submits K with $[K] = \{w_1, w_2\}$. We write $d_H(w_1,P) = d_H(w_2,P) = d_H(w_3,P) = \beta$, $d_H(w_4,P) = \beta + \epsilon_4$ and $\delta_{3*} = \min\{\delta_{31}, \delta_{32}\}$, $\delta_{4*} = \min\{\delta_{41}, \delta_{42}\}$ for the distance from w_3 and w_4, respectively, to K (see Table 3). Notice now that $[\Delta_\mu^{d,\Sigma}(P+K)] = \{w_1, w_2\}$, i.e., the strategic agent demotes w_3 from $\Delta_\mu^{d,\Sigma}(P)$. To promote w_4 to the final result, the obvious strategy is for the strategic agent to submit K', with $[K'] = \{w_4\}$. In this case, promoting w_4 is successful only if $\epsilon_4 \leq \delta_{i4}$, where $\delta_{i4} = d_H(w_i, w_4)$, for $i \in \{1,2,3\}$ (again, see Table 3). The same argument applies to the drastic distance d_D, the only difference being that $\delta_{3*} = \delta_{4*} = \delta_{i4} = 1$, for $i \in \{1,2,3\}$.*

With respect to atoms, an analogous question regarding the influence of an agent asks under what conditions a specific atom can be made part of the final result. The idea here turns out to be that no single agent can overturn majorities w.r.t. skeptical acceptances of the bases in the complete profile and $\mu \equiv \top$: if more than half of the agents skeptically accept a, then no strategic agent can alter this. This is the same fact that underwrites strategyproofness of $\Delta^{d_H,\Sigma}$. For non-complete profiles strategyproofness is lost, but a related result can be shown.

For a profile P, define agents' support for acceptances as $\mathtt{Credsupp}_P(a) = |\{K \in P \mid a \in \mathtt{Cred}(K)\}|$ and $\mathtt{Skeptsupp}_P(a) = |\{K \in P \mid a \in \mathtt{Skept}(K)\}|$. By generalizing a result from [Delobelle et al., 2016], we show that neither a majority of skeptical support nor a majority of credulous non-support can be altered, for aggregation under $\Delta_\top^{d_H,\Sigma}$.

Proposition 7. *Let $P = (K_1, \dots, K_{n-1})$ be a profile, $X = \{x \mid \mathtt{Skeptsupp}_P(x) > \frac{n}{2}\}$, and $Y = \{x \mid \mathtt{Credsupp}_P(x) < \frac{n}{2}\}$. For any base K_n and $M = \Delta_\top^{d_H,\Sigma}(P + K_n)$, it holds that $X \subseteq \mathtt{Skept}(M)$, and $Y \subseteq (\mathcal{P} \setminus \mathtt{Cred}(M))$.*

A similar result does not hold for operator $\Delta_\top^{d_H,\mathtt{max}}$, i.e., when using max instead of Σ. Thus, for max majorities may be overturned, as illustrated in the next example.

Example 10. *Take $[K_1] = \{b\}$, $[K_2] = \{c\}$, and $[K_3^T] = \{abc\}$. With $\Delta_\top^{d_H,\mathtt{max}}$ the result is $\{bc\}$. When agent 3 reports $[K_3^F] = \{ab\}$ instead, the result is $\{\emptyset, a, b, bc, abc\}$. Thus, agent 3 can get a to be true in a model of the output, even if less than half of the agents have a in some model of their base (in fact only agent 3 accepts a credulously).*

6 Complexity of Constructive and Destructive Manipulation

By our results, if an agent can constructively or destructively Skept-manipulate the aggregation process, then it can do so by submitting a complete base (see Theorem 2). By Proposition 2, Cred-manipulation can always be achieved by the dual of a complete base (again a complete base), if manipulation is possible. By these results, for both constructive and destructive manipulation, deciding whether a profile is manipulable is in Σ_2^P. To see this, we first recall that computing the result of the merging process, i.e., whether $\Delta_\mu^{d,f}(P) \models \varphi$ holds, is a problem that can be solved via a deterministic polynomial time algorithm with access to an NP oracle, for all operators considered in this paper [Konieczny et al., 2002, Konieczny et al., 2004]. This implies that one can

Table 3: The agent penalizes w_3 by not including it in the models of its submitted base, and can only promote w_4 if the margin ϵ_4 by which it trails the other interpretations is sufficiently small.

	P	$\{w_1, w_2\}$	$\{w_4\}$	$d_H^\Sigma(\cdot, P + K)$	$d_H^\Sigma(\cdot, P + K')$
w_1	β	0	δ_{14}	β	$\beta + \delta_{14}$
w_2	β	0	δ_{24}	β	$\beta + \delta_{24}$
w_3	β	δ_{3*}	δ_{34}	$\beta + \delta_{3*}$	$\beta + \delta_{34}$
w_4	$\beta + \epsilon_4$	δ_{4*}	0	$\beta + \delta_{4*} + \epsilon_4$	$\beta + \epsilon_4$

check whether an unmodified (non-altered) profile already returns the desired atom skeptically (credulously). If not, a non-deterministic construction ("guess") of a complete base with a subsequent new check of the result decides whether the constructed base results in a manipulation. For operator $\Delta_{\mu}^{d_H,\Sigma}$ and destructive `Skept`-manipulation, we also can show hardness for this class.

Theorem 5. *Deciding whether a profile can be destructively* `Skept`*-manipulated w.r.t. an atom and* μ *for operator* $\Delta_{\mu}^{d_H,\Sigma}$ *by submitting a complete base is* Σ_2^P*-complete.*

7 Related work

Existing work on manipulation of belief merging operators [Diaz and Perez, 2018, Everaere et al., 2007] differs from ours in that satisfaction indices in [Everaere et al., 2007] are not based on skeptical or credulous acceptance but on the models that the strategic agent and the result have in common. To highlight this difference, note that under the indices in [Everaere et al., 2007] the strategic agent in Example 2 would be equally unsatisfied with both the truthful result $\Delta_{\mu}^{d_H,\Sigma}(P^T)$ and $\Delta_{\mu}^{d_H,\Sigma}(P^F)$, since K^T shares no model with either. Under our interpretation of the indices, $\Delta_{\mu}^{d_H,\Sigma}(P^F)$ ends up delivering a better result for the strategic agent than $\Delta_{\mu}^{d_H,\Sigma}(P^T)$, as under $\Delta_{\mu}^{d_H,\Sigma}(P^F)$ the atom a is guaranteed to be in the result, and there is a sense in which this is satisfactory for the strategic agent, as a is an atom that it skeptically accepts. Then, different to both [Diaz and Perez, 2018, Everaere et al., 2007], we also show results for acceptance manipulation (not based on indices), i.e., for constructive and destructive manipulation.

Belief merging invites comparison to multi-winner elections [Amanatidis et al., 2015, Barrot et al., 2017, Faliszewski et al., 2017, Meir et al., 2008], combinatorial voting [Lang and Xia, 2016], and Judgment Aggregation [Baumeister et al., 2015, Baumeister et al., 2017, Endriss, 2016]. We mention here that our use of acceptance notions and satisfaction indices, the compact encoding of sets of interpretations (agents' "top candidates") as propositional formulas, and the fact that we do not require the output to be of a specific size suggest that existing results in this area are not directly applicable to our setting. Our work intersects with social choice in the special case when the profile is complete and the number of bases is odd. In this case the aggregation problem corresponds to a Judgment Aggregation problem, with $\Delta_{\top}^{d_H,\Sigma}$ delivering the majority opinion on the atoms (considered as issues): this corresponds to the observation made in the Social Choice literature [Brams et al., 2007] that the majority opinion minimizes the sum of the Hamming distances to voters' approval ballots. Our strategy-proofness result for $\Delta_{\top}^{d_H,\Sigma}$ dovetails neatly with a similar result in Judgment Aggregation [Baumeister et al., 2017, Endriss, 2016], though our treatment is slightly more general, as it accommodates both an even and an odd number of bases.

8 Conclusions

We have looked at the potential for manipulation in a belief merging framework [Konieczny et al., 2002, Konieczny and Pérez, 2011], when results are obtained considering skeptical or credulous consequences. We have shown that manipulation is not only possible for well-known aggregation operators, but also that manipulation can be achieved by semantically simple (i.e., complete) bases, even if the complexity of doing so is in general high.

For future work, our aim is to extend these results to more merging operators, study best responses (strategies) by agents, manipulability in settings with incomplete information, and to consider extended settings of manipulation studied in Social Choice, e.g., bribery [Baumeister et al., 2015], where sets of agents can be "bribed" to form a joint manipulating coalition. We also want to look at properties from Social Choice used to understand strategyproofness at a more abstract level (e.g., monotonicity), and at how to adapt these properties to the merging framework. This topic has received some attention [Diaz and Perez, 2018, Haret et al., 2016], but more work is needed to establish connections to manipulation and strategyproofness.

References

[Amanatidis et al., 2015] Amanatidis, G., Barrot, N., Lang, J., Markakis, E., and Ries, B. (2015). Multiple Referenda and Multiwinner Elections Using Hamming Distances: Complexity and Manipulability. In Weiss, G., Yolum, P., Bordini, R. H., and Elkind, E., editors, *Proc. AAMAS 2015*, pages 715–723. ACM.

[Barrot et al., 2017] Barrot, N., Lang, J., and Yokoo, M. (2017). Manipulation of Hamming-based approval voting for multiple referenda and committee elections. In Larson, K., Winikoff, M., Das, S., and Durfee, E. H., editors, *Proc. AAMAS 2017*, pages 597–605. ACM.

[Baumeister et al., 2015] Baumeister, D., Erdélyi, G., Erdélyi, O. J., and Rothe, J. (2015). Complexity of manipulation and bribery in judgment aggregation for uniform premise-based quota rules. *Mathematical Social Sciences*, 76:19–30.

[Baumeister et al., 2017] Baumeister, D., Rothe, J., and Selker, A.-K. (2017). Strategic behavior in judgment aggregation. In Endriss, U., editor, *Trends in Computational Social Choice*, pages 145–168. AI Access.

[Brams et al., 2007] Brams, S. J., Kilgour, D. M., and Sanver, M. R. (2007). A minimax procedure for electing committees. *Public Choice*, 132(3):401–420.

[Conitzer and Walsh, 2016] Conitzer, V. and Walsh, T. (2016). Barriers to manipulation in voting. In Brandt, F., Conitzer, V., Endriss, U., Lang, J., and Procaccia, A. D., editors, *Handbook of Computational Social Choice*, pages 127–145. Cambridge University Press.

[Delobelle et al., 2016] Delobelle, J., Haret, A., Konieczny, S., Mailly, J., Rossit, J., and Woltran, S. (2016). Merging of abstract argumentation frameworks. In Baral, C., Delgrande, J. P., and Wolter, F., editors, *Proc. KR 2016*, pages 33–42. AAAI Press.

[Díaz and Pérez, 2017] Díaz, A. M. and Pérez, R. P. (2017). Impossibility in belief merging. *Artif. Intell.*, 251:1–34.

[Diaz and Perez, 2018] Diaz, A. M. and Perez, R. P. (2018). Epistemic states, fusion and strategy-proofness. In Ferme, E. and Villata, S., editors, *Proc. NMR*, pages 176–185.

[Endriss, 2016] Endriss, U. (2016). Judgment Aggregation. In Brandt, F., Conitzer, V., Endriss, U., Lang, J., and Procaccia, A. D., editors, *Handbook of Computational Social Choice*, pages 399–426. Cambridge University Press.

[Everaere et al., 2007] Everaere, P., Konieczny, S., and Marquis, P. (2007). The strategy-proofness landscape of merging. *Journal of Artificial Intelligence Research*, 28:49–105.

[Everaere et al., 2015] Everaere, P., Konieczny, S., and Marquis, P. (2015). Belief merging versus judgment aggregation. In Weiss, G., Yolum, P., Bordini, R. H., and Elkind, E., editors, *Proc. AAMAS 2015*, pages 999–1007. ACM.

[Faliszewski and Procaccia, 2010] Faliszewski, P. and Procaccia, A. D. (2010). AI's war on manipulation: Are we winning? *AI Magazine*, 31(4):53–64.

[Faliszewski et al., 2017] Faliszewski, P., Skowron, P., Slinko, A., and Talmon, N. (2017). Multiwinner voting: A new challenge for social choice theory. In Endriss, U., editor, *Trends in Computational Social Choice*, pages 27–47. AI Access.

[Gabbay et al., 2009] Gabbay, D. M., Rodrigues, O., and Pigozzi, G. (2009). Connections between belief revision, belief merging and social choice. *Journal of Logic and Computation*, 19(3):445–446.

[Haret et al., 2016] Haret, A., Pfandler, A., and Woltran, S. (2016). Beyond IC postulates: Classification criteria for merging operators. In Kaminka, G. A., Fox, M., Bouquet, P., Hüllermeier, E., Dignum, V., Dignum, F., and van Harmelen, F., editors, *Proc. ECAI 2016*, pages 372–380.

[Haret and Wallner, 2018] Haret, A. and Wallner, J. P. (2018). Manipulation of semantic aggregation procedures for propositional knowledge bases and argumentation frameworks. In Ferme, E. and Villata, S., editors, *Proc. NMR*, pages 146–155.

[Konieczny et al., 2002] Konieczny, S., Lang, J., and Marquis, P. (2002). Distance based merging: A general framework and some complexity results. In Fensel, D., Giunchiglia, F., McGuinness, D. L., and Williams, M., editors, *Proc. KR 2002*, pages 97–108. Morgan Kaufmann.

[Konieczny et al., 2004] Konieczny, S., Lang, J., and Marquis, P. (2004). DA^2 merging operators. *Artificial Intelligence*, 157(1-2):49–79.

[Konieczny and Pérez, 2011] Konieczny, S. and Pérez, R. P. (2011). Logic based merging. *Journal of Philosophical Logic*, 40(2):239–270.

[Lang and Xia, 2016] Lang, J. and Xia, L. (2016). Voting in combinatorial domains. In Brandt, F., Conitzer, V., Endriss, U., Lang, J., and Procaccia, A. D., editors, *Handbook of Computational Social Choice*, pages 197–222. Cambridge University Press.

[Meir et al., 2008] Meir, R., Procaccia, A. D., Rosenschein, J. S., and Zohar, A. (2008). Complexity of strategic behavior in multi-winner elections. *Journal of Artificial Intelligence Research*, 33:149–178.

[Strasser and Antonelli, 2018] Strasser, C. and Antonelli, G. A. (2018). Non-monotonic logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2018 edition.

[Zwicker, 2016] Zwicker, W. S. (2016). Introduction to the theory of voting. In Brandt, F., Conitzer, V., Endriss, U., Lang, J., and Procaccia, A. D., editors, *Handbook of Computational Social Choice*, pages 23–56. Cambridge University Press.

Full Proofs

These results concern the acceptance notions and satisfaction indices introduced in Section 3, linking them to the dual formulas.

Lemma 1. *If φ is a propositional formula, then $\overline{[\varphi]} = [\overline{\varphi}]$.*

Proof. Induction on the structure of φ. □

Lemma 2. *If φ is a propositional formula, then $\overline{\mathtt{Skept}(\varphi)} = \mathtt{Cred}(\overline{\varphi})$.*

Proof. If p is an atom, we have $p \in \overline{\mathtt{Skept}(\varphi)}$ iff $p \notin \mathtt{Skept}(\varphi)$ iff $p \notin w$, for some $w \in [\varphi]$ iff $p \in \overline{w}$, for some $\overline{w} \in [\overline{\varphi}]$, the last step obtained using Lemma 1. This is further equivalent to $p \in \mathtt{Cred}(\varphi)$. □

Lemma 3. *If φ and ψ are propositional formulas, then it holds that $\mathtt{Skept}(\varphi) \triangle \mathtt{Skept}(\psi) = \mathtt{Cred}(\overline{\varphi}) \triangle \mathtt{Cred}(\overline{\psi})$.*

Proof. The claim follows by noticing that, if p is an atom, then $p \in (\mathtt{Skept}(\varphi) \setminus \mathtt{Skept}(\psi))$ iff $p \in (\mathtt{Cred}(\overline{\psi}) \setminus \mathtt{Cred}(\overline{\varphi}))$. □

Using Lemma 3, Corollary 2 follows immediately.

Corollary 2. *If φ and ψ are propositional formulas, then $i_{\mathtt{Skept}}(\varphi, \psi) = i_{\mathtt{Cred}}(\overline{\varphi}, \overline{\psi})$.*

Corollary 2 is mentioned in Section 3 and, together with Lemma 7, is used in Section 4.2 to turn instances of index manipulation with respect to one acceptance notion into instances of index manipulation with respect to the other acceptance notion.

Turning to aggregation of propositional bases, we now show a series of results linking our notions of acceptance, the indices and the duals.

Lemma 4. *If w_1 and w_2 are two interpretations and $d \in \{d_H, d_D\}$, then $d(w_1, w_2) = d(\overline{w_1}, \overline{w_2})$.*

Proof. For $d = d_D$ the claim follows immediately, since $w_1 = w_2$ if and only if $\overline{w_1} = \overline{w_2}$. For $d = d_H$, notice first that if p is in atom, then $p \in (w_1 \setminus w_2)$ if and only if $p \in (\overline{w_2} \setminus \overline{w_1})$. It follows from here that $w_1 \triangle w_2 = \overline{w_1} \triangle \overline{w_2}$, which implies the conclusion. □

Lemma 5. *If K is a base, w is an interpretation and $d \in \{d_H, d_D\}$, then $d(w, K) = d(\overline{w}, \overline{K})$.*

Proof. By definition, we have $d(w, K) = \min\{d(w, w') \mid w' \in [K]\}$. Applying Lemma 1 and Lemma 4, we get that $\{d(w, w') \mid w' \in [K]\} = \{d(\overline{w}, \overline{w'}) \mid \overline{w'} \in [\overline{K}]\}$, which implies the conclusion. □

Lemma 6. *If $P = (K_1, \ldots, K_m)$ is a profile, w is an interpretation, $d \in \{d_H, d_D\}$ and $f \in \{\Sigma, \mathtt{max}, \mathtt{gmax}\}$, then $d^f(w, P) = d^f(\overline{w}, \overline{P})$.*

Proof. Using Lemma 5. □

Lemma 7. *If P is a profile, μ is a constraint, $d \in \{d_H, d_D\}$ and $f \in \{\Sigma, \mathtt{max}, \mathtt{gmax}\}$, then $\overline{\Delta^{d,f}_{\mu}(P)} \equiv \Delta^{d,f}_{\overline{\mu}}(\overline{P})$.*

Proof. Using Lemma 1 and Lemma 6. □

Proposition 1. For an atom p, it holds that $p \notin \mathtt{Skept}(\Delta^{d,f}_{\mu}(P))$ iff there exists an interpretation $w \in [\Delta^{d,f}_{\mu}(P)]$ such that $p \notin w$. Using Lemma 7, this is equivalent to $p \in \overline{w}$, for some interpretation $\overline{w} \in [\Delta^{d,f}_{\overline{\mu}}(\overline{P})]$, which is in turn equivalent to $p \in \mathtt{Cred}(\Delta^{d,f}_{\overline{\mu}}(\overline{P}))$. □

The following results concern Section 4.1. First we show that instances of Skept-manipulation are interchangeable with instances of Cred-manipulation.

Proposition 2. Assume an instance of constructive Skept-manipulation with respect to p. If $p \notin \mathtt{Skept}(\Delta^{d,f}_{\mu}(P + K^T))$, then $p \in \overline{\mathtt{Skept}(\Delta^{d,f}_{\mu}(P + K^T))}$. Thus, by Proposition 1, it holds that $p \in \mathtt{Cred}(\Delta^{d,f}_{\overline{\mu}}(\overline{P} + \overline{K^T}))$. Similarly, we get that if $p \in \mathtt{Skept}(\Delta^{d,f}_{\mu}(P + K^F))$, then $p \notin \mathtt{Cred}(\Delta^{d,f}_{\overline{\mu}}(\overline{P} + \overline{K^F}))$. We have obtained, in this way, an instance of destructive Cred-manipulation with respect to p.

The proof going from an instance of destructive Skept-manipulation to an instance of constructive Skept-manipulation with respect to p is entirely analogous. □

Lemma 8. *If K is a base, $d \in \{d_H, d_D\}$, w_1 and w_2 are two interpretations and K_* is a complete base whose model v_* is such that $v_* \in [K]$ and $d(w_1, v_*) = \min\{d(w_1, v) \mid v \in [K]\}$, then it holds that:*

(i) if $w_1 <^d_K w_2$, then $w_1 <^d_{K_*} w_2$;

(ii) if $w_1 \approx^d_K w_2$, then $w_1 \leq^d_{K_*} w_2$.

Proof. We write $[K] = \{v_1, \ldots, v_m\}$ and $d(w_k, v_j) = \delta_{kj}$, for $k \in \{1, 2\}$, $j \in \{1, \ldots, m\}$. Additionally, we write $\delta^{\min}_k = \min\{\delta_{k1}, \ldots, \delta_{km}\}$, for $k \in \{1, 2\}$ (see Table 4). By definition, we have $d(w_k, K) = \delta^{\min}_k$, for $k \in \{1, 2\}$.

We start with claim (i): by assumption, it holds that $\delta^{\min}_1 < \delta^{\min}_2$. We take now an interpretation $v_* \in [K]$ that is closest to w_1 among the models of K, [3] i.e., $d(w_1, v_*) = \min\{d(w_1, v) \mid v \in [K]\}$, and a base K_* such that $[K_*] = \{v_*\}$. We now have $d(w_1, K_*) = \min\{d(w_1, v_*)\} = \min\{\delta_{1*}\} = \delta_{1*} = \delta^{\min}_1$, while $d(w_2, K_*) = \delta_{2*} = \delta^{\min}_2 + \epsilon$, for some $\epsilon \geq 0$. The latter claim is just a rewriting of the fact that $\delta_{2*} \geq \delta^{\min}_2$, and it follows from the fact that $\delta^{\min}_2 = \min\{\delta_{21}, \ldots, \delta_{2*}, \ldots, \delta_{2m}\}$. Since, by assumption, $\delta^{\min}_1 < \delta^{\min}_2$, then it also holds that $\delta^{\min}_1 < \delta^{\min}_2 + \epsilon$, and hence $d(w_1, K_*) < d(w_2, K_*)$. For claim (ii), our assumption is equivalent to the fact $\delta^{\min}_1 = \delta^{\min}_2$, from which it follows that $\delta^{\min}_1 \leq \delta^{\min}_2 + \epsilon$ and hence $d(w_1, K_*) \leq d(w_2, K_*)$.

□

[3] There might be more than one interpretation that is equidistant to w_1 and fits this description; we pick one at random.

Table 4: Replacing K with K_*, with $[K_*] = \{v_i\}$ and v_i being the model of K closest to w_1, preserves the order between w_1 and w_2.

	$[K]$				$[K_*]$	
	v_1	$\dots$	v_m	$d(\cdot, K)$	v_*	$d(\cdot, K_*)$
w_1	δ_{11}	$\dots$	δ_{1m}	$\delta_1^{\min}$	$\delta_1^{\min}$	$\delta_1^{\min}$
w_2	δ_{21}	$\dots$	δ_{2m}	$\delta_2^{\min}$	$\delta_2^{\min} + \epsilon$	$\delta_2^{\min} + \epsilon$

Table 5: Replacing K with K_*, where $[K_*] = \{v_i\}$ and v_i is the model of K closest to w_1, preserves the order between w_1 and w_2.

		$[K]$	$[K_*]$		
	P	$\{v_1, \dots, v_m\}$	$\{v_*\}$	$d^\Sigma(\cdot, P + K)$	$d^\Sigma(\cdot, P + K_*)$
w_1	β_1	$\delta_1^{\min}$	$\delta_1^{\min}$	$\beta_1 + \delta_1^{\min}$	$\beta_1 + \delta_1^{\min}$
w_2	β_2	$\delta_2^{\min}$	$\delta_2^{\min} + \epsilon$	$\beta_2 + \delta_2^{\min}$	$\beta_2 + \delta_2^{\min} + \epsilon$

Lemma 9. *If P is a profile, K is a base, $d \in \{d_H, d_D\}$, $f \in \{\Sigma, \texttt{gmax}, \texttt{max}\}$, w_1 and w_2 are two interpretations and K_* is a complete base whose model v_* is such that $v_* \in [K]$ and $d(w_1, v_*) = \min\{d(w_1, v) \mid v \in [K]\}$, then it holds that:*

(i) if $w_1 <^{d,f}_{P+K} w_2$, then $w_1 <^{d,f}_{P+K_*} w_2$;

(ii) if $w_1 \approx^{d,f}_{P+K} w_2$, then $w_1 \leq^{d,f}_{P+K_*} w_2$;

Proof. We first show the claim for $f = \Sigma$, as it provides a nice illustration of the main ideas. For this, we write $d^\Sigma(w_k, P) = \beta_k$, for $k \in \{1, 2\}$. Assuming that $[K] = \{v_1, \dots, v_m\}$, we write $\min\{d(w_k, v) \mid v \in [K]\} = \delta_k^{\min}$, for $k \in \{1, 2\}$ (see Table 5). By definition, $d(w_k, K) = \delta_k^{\min}$, for $k \in \{1, 2\}$. We take now an interpretation $v_* \in [K]$ that is closest to w_1 among the models of K, i.e., $d(w_1, v_*) = \min\{d(w_1, v) \mid v \in [K]\}$, and a base K_* such that $[K_*] = \{v_*\}$. We now have $d(w_1, K_*) = \min\{d(w_1, v_*)\} = \min\{\delta_{1*}\} = \delta_{1*} = \delta_1^{\min}$, while $d(w_2, K_*) = \delta_{2*} = \delta_2^{\min} + \epsilon$, for some $\epsilon \geq 0$ (see Lemma 8 for more details). Then, we get that $d^f(w_1, P + K) = \beta_1 + \delta_1^{\min}$ and $d^f(w_2, P + K) = \beta_2 + \delta_2^{\min}$. Additionally, we have $d^f(w_1, P + K_*) = \beta_1 + \delta_1^{\min}$ and $d^f(w_2, P + K_*) = \beta_2 + \delta_2^{\min} + \epsilon$. If $\beta_1 + \delta_1^{\min} < \beta_2 + \delta_2^{\min}$, as per the assumption of (i), then $\beta_1 + \delta_1^{\min} < \beta_2 + \delta_2^{\min} + \epsilon$ and hence $d^f(w_1, P + K_*) < d^f(w_2, P + K_*)$. If $\beta_1 + \delta_1^{\min} = \beta_2 + \delta_2^{\min}$, as per the assumption of (ii), then $\beta_1 + \delta_1^{\min} \leq \beta_2 + \delta_2^{\min} + \epsilon$ and hence $d^f(w_1, P + K_*) \leq d^f(w_2, P + K_*)$

For $f \in \{\texttt{gmax}, \texttt{max}\}$ the argument has to be adapted to the output for each aggregation function, but is otherwise entirely similar. If $f = \texttt{gmax}$ then the integers β_1 and β_2 (i.e., the distances from w_1 and w_2, respectively, to P) must be replaced with tuples of integers $B_1 = (\beta_1^1, \beta_2^1, \dots)$ and $B_2 = (\beta_1^2, \beta_2^2, \dots)$. For (i) we then have, by assumption, that $\texttt{gmax}(\beta_1^1, \beta_2^1, \dots, \delta_1^{\min}) <_{lex} \texttt{gmax}(\beta_1^2, \beta_2^2, \dots, \delta_2^{\min})$. Since $\delta_{2*} \leq \delta_{2*} + \epsilon$ and gmax satisfies the MONOTONICITY property (see Section 2), we get that $\texttt{gmax}(\beta_1^1, \beta_2^1, \dots, \delta_{1i}) <_{lex} \texttt{gmax}(\beta_1^2, \beta_2^2, \dots, \delta_{2i} + \epsilon)$ and thus

$d^{\mathtt{gmax}}(w_1, P + K_*) \leq d^{\mathtt{gmax}}(w_2, P + K_*)$. The argument for (ii) is entirely similar, and if $f = \mathtt{max}$ the claim follows analogously. □

Theorem 1. Without loss of generality, we can assume the target atom p is a. We only showcase the constructive $\mathtt{Skept}$-manipulation instances, as corresponding $\mathtt{Cred}$-manipulation instances can be obtained using Proposition 2 and a destructive manipulation instance can be obtained from a constructive manipulation instance by swapping K^T and K^F as the truthful and manipulating base, respectively, of the strategic agent. We assume, throughout, that $\mu = \top$.

The following argument applies to operators $\Delta^{d,f}$, for $d \in \{d_D, d_H\}$ and $f \in \{\Sigma, \mathtt{gmax}\}$. To obtain constructive $\mathtt{Skept}$-manipulation, we take $K^T = \bigwedge_{p \in \mathcal{P}} \neg p$. Thus, $[K^T] = \{\emptyset\}$ and $\mathtt{Skept}(K^T) = \emptyset$. We then do a case analysis depending on whether n is odd or even. In both cases, the agent manipulates using $K^F = a \wedge \bigwedge_{p \in \mathcal{P}, p \neq a} \neg p$, with $[K^F] = \{a\}$. Each operator is analyzed in turn.

Case 1 (n is even). We write $n = 2k$, for $k \in \mathbb{N}$. For the operators $\Delta^{d,f}$, for $d \in \{d_D, d_H\}$ and $f \in \{\Sigma, \mathtt{gmax}\}$ we take the profile $P = (K_1, \dots, K_{2k})$ such that $[K_1] = \dots = [K_k] = \{\emptyset\}$ and $[K_{k+1}] = \dots = [K_{2k}] = \{a\}$. Notice that all bases are complete.

($\Delta^{d_H,\Sigma}$) In the truthful profile $P^T = P + K^T$ we have $d_H^\Sigma(\emptyset, P^T) = k$ and $d_H^\Sigma(\emptyset, P^T) = k + 1$, while for any other interpretation w we get that $d_H^\Sigma(w, P^T) = \sum_{i=1}^{2k} \delta_i + \delta^T$, where $\delta_i = d_H^\Sigma(w, K_i)$ and $\delta^T = d_H^\Sigma(w, K^T)$. It is straightforward to see that $\delta_i \geq 1$, for any $i \in \{1, \dots, 2k\}$ and that $\delta^T \geq 1$ as well. Thus, $\emptyset <_{P+K^T}^{d_H,\Sigma} a$ and $\emptyset <_{P+K^T}^{d_H,\Sigma} w$ for any other interpretation w, i.e., $[\Delta_\top^{d_H,\Sigma}(P^T)] = \{\emptyset\}$.

In the manipulated profile $P^F = P + K^F$ we get that $d_H^\Sigma(\emptyset, P + K^F) = k + 1$ and $d_H^\Sigma(\emptyset, P + K^F) = k$, while for any other interpretation w we get that $d_H^\Sigma(w, P^T) = \sum_{i=1}^{2k} \delta_i + \delta^F$, where $\delta^F = d_H^\Sigma(w, K^F)$. It is straightforward to see that $\delta^F \geq 1$ and thus $a <_{P^F}^{d_H,\Sigma} \emptyset$ and $a <_{P^F}^{d_H,\Sigma} w$ for any other interpretation w, i.e., $[\Delta_\top^{d_H,\Sigma}(P^F)] = \{a\}$.

Since $a \notin \mathtt{Skept}(\Delta_\top^{d_H,\Sigma}(P^T))$ but $a \in \mathtt{Skept}(\Delta_\top^{d_H,\Sigma}(P^F))$, this counts as an instance of constructive manipulation.

($\Delta^{d_D,\Sigma}$) The argument for $\Delta^{d_H,\Sigma}$ works here unchanged, since the argument does not rely on the fact that any of the numbers involved are greater than 1.

($\Delta^{d_H,\mathtt{gmax}}$) We reason analogously as for $\Delta^{d_H,\Sigma}$, using the same profile P. Notice that $d_H^{\mathtt{gmax}}(\emptyset, P^T) = (\underbrace{1, \dots, 1}_{k \text{ times}}, \underbrace{0, \dots, 0}_{(k+1) \text{ times}})$ and $d_H^{\mathtt{gmax}}(a, P^T) = (\underbrace{1, \dots, 1}_{(k+1) \text{ times}}, \underbrace{0, \dots, 0}_{k \text{ times}})$, while $d_H^{\mathtt{gmax}}(w, P^T) = \mathtt{gmax}(\delta_1, \dots, \delta_{2k}, \delta^T)$ for any other interpretation w. It follows then that $[\Delta_\top^{d_H,\mathtt{gmax}}(P^T)] = \{\emptyset\}$, and then that $[\Delta_\top^{d_H,\mathtt{gmax}}(P^T)] = \{a\}$.

($\Delta^{d_D,\mathtt{gmax}}$) This operator is equivalent to the operator $\Delta^{d_D,\Sigma}$ (see Section 2).

Case 2 (n is odd). We write $n = 2k + 1$, for $k \in \mathbb{N}$. For the operators $\Delta^{d,f}$, for $d \in \{d_D, d_H\}$ and $f \in \{\Sigma, \mathtt{gmax}\}$ we take the profile $P = (K_1, \dots, K_{2k+1})$ such that $[K_1] = \dots = [K_k] = \{\emptyset\}$ and $[K_{k+1}] = \dots = [K_{2k+1}] = \{a\}$. Notice that all bases are complete. We apply the same reasoning as above.

For the operators $\Delta^{d_H,\mathtt{max}}$ and $\Delta^{d_D,\mathtt{max}}$ we have to pick a different profile. In this case we do not need an even/odd distinction, as the same kind of profile works for both cases. Take, then, a

	$[K_1]$	$[K_2]$	$[K_3]$	$[K_4]$	$[K^T]$	$[K^F]$				
	$\{\emptyset\}$	$\{\emptyset\}$	$\{a\}$	$\{a\}$	$\{\emptyset\}$	$\{a\}$	$d^{\Sigma}(\cdot, P^T)$	$d^{\mathtt{gmax}}(\cdot, P^T)$	$d^{\Sigma}(\cdot, P^F)$	$d^{\mathtt{gmax}}(\cdot, P^F)$
$\emptyset$	0	0	1	1	0	1	2	$(1,1,0,0,0)$	3	$(1,1,1,0,0)$
a	1	1	0	0	1	0	3	$(1,1,1,0,0)$	2	$(1,1,0,0,0)$
w	δ_1	δ_2	δ_3	δ_4	δ^T	δ^F	$\sum_1^4 \delta_i + \delta^T$	$\mathtt{gmax}(\delta_1, \ldots, \delta_4, \delta^T)$	$\sum_1^4 \delta_i + \delta^F$	$\mathtt{gmax}(\delta_1, \ldots, \delta_4, \delta^F)$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$

Table 6: Constructive Skept-manipulation of a profile of size 4

profile $P = (K_1, \ldots, K_n)$ with $[K_i] = \{a\}$, for $i \in \{1, \ldots, n\}$.

($\Delta^{d_H, \mathtt{max}}$) We get that $d_H^{\mathtt{max}}(\emptyset, P^T) = 1$ and $d_H^{\mathtt{max}}(a, P^T) = 1$, so $\emptyset$ and a are guaranteed to be winning interpretations, which ensures that $a \notin \mathtt{Skept}(\Delta_{\top}^{d_H, \mathtt{max}}(P^T))$. But $d_H^{\mathtt{max}}(\emptyset, P^F) = 1$ and $d_H^{\mathtt{max}}(a, P^F) = 0$, from which we get that $[\Delta_{\top}^{d_H, \mathtt{max}}(P^F)] = \{a\}$, so $a \in \mathtt{Skept}(\Delta_{\top}^{d_H, \mathtt{max}}(P^F))$.

($\Delta^{d_D, \mathtt{max}}$) Similar as for $\Delta^{d_H, \mathtt{max}}$. □

The following example illustrates the main idea of the proof of Theorem 1, for a profile of size 4.

Example 11. *We pick $[K_1] = [K_2] = \{\emptyset\}$ and $[K_3] = [K_4] = \{a\}$, and $[K^T] = \{\emptyset\}$, $[K^F] = \{a\}$. The merging result is summarized in Table 6.*

Theorem 2. We prove the claim for constructive Skept-manipulation first. The fact that the strategic agent Skept-manipulates P using K^F implies that there exist interpretations $w_1, \ldots, w_l$ in $[\mu]$ such that $p \in \mathtt{Skept}(\{w_1, \ldots, w_l\})$, and $[\Delta_{\mu}^{d,f}(P + K^F)] = \{w_1 \ldots, w_l\}$. We pick one of the interpretations in $[\Delta_{\mu}^{d,f}(P + K^F)]$, say w_1. Take, now, $v_* \in [K^F]$ such that $d(w_1, v_*) = \min\{d(w_1, v) \mid v \in [K^F]\}$, i.e., a model of K^F that is closest to w_1. The claim now is that we can constructively Skept-manipulate P with K_*^F, where $[K_*^F] = \{v_*\}$. This follows by observing that $w_1 \leq_{P+K^F}^{d,f} w_i$, for all $w_i \in [\mu]$ and thus, by Lemma 9, it follows that $w_1 \leq_{P+K_*^F}^{d,f} w_i$, for all $w_i \in [\mu]$. Thus, w_1 stays part of the aggregation result. Additionally, if $w_1 <_{P+K_*^F}^{d,f} w_i$, for some $w_i \in [\mu]$, then, again by Lemma 9, it follows that $w_1 <_{P+K_*^F}^{d,f} w_i$. In summary, by replacing K^F with K_*^F, w_1 and possibly some other interpretations in $\{w_1, \ldots, w_l\}$ remain winning, and no new winning interpretations are added. Another way of putting this is that $[\Delta_{\mu}^{d,f}(P + K_*^F)] \subseteq [\Delta_{\mu}^{d,f}(P + K^F)]$. Since $p \in \mathtt{Skept}(\Delta_{\mu}^{d,f}(P + K^F))$, we get that $p \in \mathtt{Skept}(\Delta_{\mu}^{d,f}(P + K_*^F))$ as well.

For destructive Skept-manipulation, we get that there exists an interpretation $w_1 \in [\Delta_{\mu}^{d,f}(P + K^F)]$ such that $p \notin w_1$. We pick, as before, a model v_* of K that is closest to w_1 among the models of K, i.e., $d(w_1, v_*) = \min\{d(w_1, v) \mid v \in [K^F]\}$. Using Lemma 9, we again get that $w_1 \in [\Delta_{\mu}^{d,f}(P + K_*^F)]$. This guarantees that $p \notin \mathtt{Skept}(\Delta_{\mu}^{d,f}(P + K_*^F))$.

The case for constructive/destructive Cred-manipulation follows by applying Proposition 2. □

Proof of Proposition 3. The operator $\Delta^{d_H, \Sigma}$, for complete bases and $\mu = \top$ acts as a majority operator. In other words: if an atom p is accepted by a majority of the agents, then p is in the result; if p is not accepted by a majority of the agents, then p is not in the result; and if there is equality with respect to acceptance of p, then the result features a model that contains p and a model that does not. This being said, if an agent can constructively $\texttt{Skept}$-manipulate with respect to atom p, then it means, by definition, that p is not in $\texttt{Skept}(\Delta_{\mu}^{d_H,\Sigma}(P^T))$, but that $p \in \texttt{Skept}(\Delta_{\mu}^{d_H,\Sigma}(P^F))$. This implies that the strategic agent's influence over the result consists in inducing a majority for p: the result (with the base of the strategic agent) goes from being undecided with respect to p (and hence p not being skeptically accepted in the result) when the strategic agent is honest, to being in favor of p when the strategic agent submits a base different from its truthful on. In this, the strategic agent is the decisive agent who tips the balance in favor of p: but this can only happen if p is not in $\texttt{Skept}(K^T)$ to begin with. □

The following results concern Section 4.2, on index manipulation.

Theorem 3. We showcase here instances of manipulation with respect to $i_{\texttt{Skept}}$ for a profile $P = (K_1, \ldots, K_n)$ of size $n \geq 2$. Instances of manipulation with respect to $i_{\texttt{Cred}}$ are obtained by taking the duals of all formulas involved in the instances of manipulation with respect to $i_{\texttt{Skept}}$. We assume that $\mu = \top$.

($\Delta^{d_H,\texttt{gmax}}$) Take $[K_i] = \{a\}$, for $i \in \{1, \ldots, n\}$, $[K^T] = \{\emptyset\}$ and $[K^F] = \{b\}$. We get that $[\Delta_{\mu}^{d_H,\texttt{gmax}}(P + K^T)] = \{a\}$ and $[\Delta_{\mu}^{d_H,\texttt{gmax}}(P + K^F)] = \{\emptyset, ab\}$.

($\Delta^{d_H,\texttt{max}}$) Take $[K_i] = \{a\}$, for $i \in \{1, \ldots, n-1\}$, $[K_n] = \{ab\}$, $[K^T] = \{\emptyset\}$ and $[K^F] = \{c\}$. We get that $[\Delta_{\mu}^{d_H,\texttt{gmax}}(P + K^T)] = \{a\}$ and $[\Delta_{\mu}^{d_H,\texttt{gmax}}(P + K^F)] = \{a, b, ac, abc\}$.

($\Delta^{d_D,\texttt{gmax}}$) We make a distinction according to whether n is odd or even. If $n = 2k$, take $[K_1] = \cdots = [K_k] = \{\emptyset\}$, $[K_{k+1}] = \cdots = [K_{2k}] = \{a\}$, $[K^T] = \{ab\}$ and $[K^F] = \{a\}$. We get that $[\Delta_{\mu}^{d_H,\texttt{gmax}}(P + K^T)] = \{\emptyset, a\}$ and $[\Delta_{\mu}^{d_H,\texttt{gmax}}(P + K^F)] = \{a\}$.

If $n = 2k + 1$, take $[K_1] = \cdots = [K_k] = \{\emptyset\}$, $[K_{k+1}] = \cdots = [K_{2k}] = \{a\}$, $[K_{2k+1}] = \{b\}$, $[K^T] = \{ab\}$ and $[K^F] = \{a\}$. We get that $[\Delta_{\mu}^{d_H,\texttt{gmax}}(P + K^T)] = \{\emptyset, a\}$ and $[\Delta_{\mu}^{d_H,\texttt{gmax}}(P + K^F)] = \{a\}$.

($\Delta^{d_D,\texttt{max}}$) Take $[K_i] = \{a\}$, for $i \in \{1, \ldots, n\}$, $[K^T] = \{ab\}$ and $[K^F] = \{a\}$. We get that $[\Delta_{\mu}^{d_H,\texttt{gmax}}(P + K^T)] = 2^{\mathcal{P}}$ and $[\Delta_{\mu}^{d_H,\texttt{gmax}}(P + K^F)] = \{a\}$. □

Proposition 4. We showcase, again, only instances of manipulation with respect to $i_{\texttt{Skept}}$ for the operator $\Delta^{d_H,\Sigma}$, as instances of manipulation with respect to $i_{\texttt{Cred}}$ are obtained by taking the duals. In the following we exhibit instances of manipulation in three cases, obtained by weakening the conditions of Theorem 4.

Case 1. Suppose $\mu = \top$ and every base except K^T is required to be complete. Then we can find instances of manipulation for every profile of size $n \geq 1$. Take $[K^T] = \{a, b\}$.

For $n = 1$, take a profile $P = (K_1)$, with $[K_1] = \{a\}$. For $n \geq 2$ and $n = 2k$ take a profile $P = (K_1, \ldots, K_{2k})$ where $[K_1] = \cdots = [K_k] = \{\emptyset\}$ and $[K_{k+1}] = \cdots = [K_{2k}] = \{a\}$. For $n \geq 2$ and $n = 2k + 1$ take a profile $P = (K_1, \ldots, K_{2k})$ where $[K_1] = \cdots = [K_k] = \{\emptyset\}$ and $[K_{k+1}] = \cdots = [K_{2k+1}] = \{a\}$. In all cases, the profile P is manipulable with respect to $i_{\texttt{Skept}}$ using $[K^F] = \{\emptyset\}$.

Case 2. Suppose, now, that K^T, K^F and every base in P is required to be complete, except one. Then we can still find instances of manipulation with respect to $i_{\texttt{Skept}}$. For $n = 2$, take $P = (K_1, K_2)$, with $[K_1] = \{a\}$, $[K_1] = \{a, b\}$, $[K^T] = \{\emptyset\}$ and $[K^F] = \{b\}$.

Table 7: The agent penalizes interpretation w_3 by not including it in the models of its submitted base.

	P	$\{w_1, w_2\}$	$d_H^{\Sigma}(\cdot, P+K)$
w_1	β	0	β
w_2	β	0	β
w_3	β	δ_{3*}	$\beta + \delta_{3*}$
w_4	$\beta + \epsilon_4$	δ_{4*}	$\beta + \delta_{4*} + \epsilon_4$

Case 3. If every base in P is complete, as well as K^T and K^F, but we are allowed to choose μ, then examples of manipulation are readily available. If $[\mu] = \{a, bc\}$, then we can take $P = (K_1)$, with $[K_1] = \{\emptyset\}$, $[K^T] = \{\emptyset\}$ and $[K^F] = \{b\}$. For a profile of size $n \geq 1$, taking $[K_1] = \{a\}$ and $[K_2] = \cdots = [K_n] = \{\emptyset\}$, with K^T and K^F as before also results in an instance of manipulation with respect to $i_{\texttt{Skept}}$. □

For the case when all bases in P are complete except one (and, additionally, K^T, K^F are complete and $\mu = \top$), examples of manipulation can be found for n up to 5 and the conjecture is that they exist for any $n \geq 2$, but an example that works for any n is still forthcoming. Similarly, there is more work to be done in finding general examples of manipulation when we are allowed to pick μ.

The following results concern Section 5.

Proposition 5. This follows from classical work on belief merging: operators $\Delta^{d,f}$, for $d \in \{d_H, d_D\}$ and $f \in \{\Sigma, \texttt{gmax}\}$ satisfy the IC postulates IC_{1-8} [Konieczny et al., 2002, Konieczny and Pérez, 2011]. This results follows directly by applying postulates IC_{5-6}. A more intuitive explanation, in semantic terms, is given in Example 12. □

Example 12. *Suppose $[\mu] = \{w_1, w_2, w_3, w_4\}$, $[\Delta_\mu^{d,f}(P)] = \{w_1, w_2, w_3\}$, for $d \in \{d_H, d_D\}$ and $f \in \{\Sigma, \texttt{gmax}\}$. and the agent submits K, with $[K] = \{w_1, w_2\}$. The claim is that $[\Delta_\mu^{d,f}(P+K)] = \{w_1, w_2\}$, for $d \in \{d_H, d_D\}$ and $f \in \{\Sigma, \texttt{gmax}\}$, i.e., the interpretation w_3, which is a winning interpretation in $\Delta_\mu^{d,f}(P)$, becomes non-winning in $\Delta_\mu^{d,f}(P+K)$.*

We write $\delta_{3*} = \min\{\delta_{31}, \delta_{32}\}$ and $\delta_{4*} = \min\{\delta_{41}, \delta_{42}\}$ for the distance from w_3 and w_4, respectively, to K. For $\Delta^{d_H,\Sigma}$, see Table 7. Notice that w_1 and w_2 end up being at a minimal distance to $P+K$, while the formerly winning interpretation w_3 becomes non-winning. Also, the interpretation w_4 stays non-winning. For $\Delta^{d_H,\texttt{gmax}}$, see Table 8. The important thing to notice here is that the final vectors for w_3 and w_4 end up being lexicographically greater than the vector for w_1 and w_2, regardless of where δ_{3*} and δ_{4*} get inserted. The same argument applies to the drastic distance d_D, the only difference being that $\delta_{3*} = \delta_{4*} = 1$.

The main idea is that by not including it in the models of its submitted base K, the agent effectively introduces a penalty for w_3. This penalty becomes the margin by which w_3 loses.

Next, we show a kind of generalized triangle inequality.

Table 8: The agent penalizes interpretation w_3 by not including it in the models of its submitted base.

	P	$\{w_1, w_2\}$	$d_H^{\texttt{gmax}}(\cdot, P + K)$
w_1	$(\beta_1, \ldots, \beta_i, \ldots, \beta_n)$	0	$(\beta_1, \ldots, \beta_i, \ldots, \beta_n, 0)$
w_2	$(\beta_1, \ldots, \beta_i, \ldots, \beta_n)$	0	$(\beta_1, \ldots, \beta_i, \ldots, \beta_n, 0)$
w_3	$(\beta_1, \ldots, \beta_i, \ldots, \beta_n)$	δ_{3*}	$\texttt{gmax}(\beta_1, \ldots, \beta_i, \ldots, \beta_n, \delta_{3*})$
w_4	$(\beta_1, \ldots, \beta_i + \epsilon_4^i, \ldots, \beta_n')$	δ_{4*}	$\texttt{gmax}(\beta_1, \ldots, \beta_i + \epsilon_4^i, \ldots, \beta_n', \delta_{4*})$

Table 9: Reversing the order between w_1 and w_2 by adding K to P is possible only if $\epsilon_1 \leq \delta_{12}$.

	P	K	$d_H^{\Sigma}(\cdot, P + K)$
w_1	$\beta + \epsilon_1$	γ_1	$\beta + \gamma_1 + \epsilon_1$
w_2	β	γ_2	$\beta + \gamma_2$

Lemma 10. *If K is a base and w_1 and w_2 are interpretations, then $d_H(w_1, K) \leq d_H(w_2, K) + d_H(w_1, w_2)$.*

Lemma 11. *If w_1 and w_2 are two interpretations such that $w_2 <_P^{\Sigma} w_1$, then there exists a base K such that $w_1 \leq_{P+K}^{\Sigma} w_2$ iff $d_H^{\Sigma}(w_1, P) - d_H^{\Sigma}(w_2, P) \leq d_H(w_1, w_2)$.*

Proof. ("⇒") We write $d_H^{\Sigma}(w_2, P) = \beta$, $d_H^{\Sigma}(w_1, P) = \beta + \epsilon_1$, with $\epsilon_1 > 0$. $d_H(w_1, K) = \gamma_1$ and $d_H(w_2, K) = \gamma_2$. This fits with the earlier naming convention, as w_2 is a winning interpretation in a direct contest with w_1 (i.e., if $[\mu] = \{w_1, w_2\}$). See Table 9 for a nicer picture of this situation. We have $w_1 \leq_{P+K}^{\Sigma} w_2$ if and only if:

$$\beta + \gamma_1 + \epsilon_1 \leq \beta + \gamma_2. \tag{1}$$

By Lemma 10 we have :

$$\gamma_2 \leq \gamma_1 + \delta_{12}. \tag{2}$$

Chaining inequalities 1 and 2 we get that $\beta + \gamma_1 + \epsilon_1 \leq \beta + \gamma_1 + \delta_{12}$. Simplifying, we get that $\epsilon_1 \leq \delta_{12}$.

("⇐") Take K such that $[K] = \{w_1\}$. Then we get that $d_H(w_1, K) = 0$ and $d_H(w_1, K) = \delta_{12}$. This implies that $d_H^{\Sigma}(w_1, P + K) = \beta + \epsilon_1$ and $d_H^{\Sigma}(w_1, P + K) = \beta + \delta_{12}$. Since $\epsilon_1 \leq \delta_{12}$, it follows that $w_1 \leq_{P+K}^{\Sigma} w_2$.

□

Proposition 6. Follows from Lemma 11, as the agent has to reverse the order between w and every model of $\Delta_{\mu}^{d_H, \Sigma}(P)$. □

Lemma 12. *Let $P = (K_1, \ldots, K_n)$ be a profile of complete bases, and $M = [\Delta_{\top}^{d_H, \Sigma}(P)]$. For any $v \in M$, it holds that $\texttt{Skeptsupp}_P(x) > \frac{n}{2}$ implies $x \in v$ and $\texttt{Skeptsupp}_P(x) < \frac{n}{2}$ implies $x \notin v$.*

Table 10: Reversing the order between w_1 and w_2 by adding K to P, with $[K] = \{w_1\}$, is possible if $\epsilon_1 \leq \delta_{12}$.

	P	$\{w_1\}$	$d_H^{\Sigma}(\cdot, P + K)$
w_1	$\beta + \epsilon_1$	0	$\beta + \epsilon_1$
w_2	β	δ_{12}	$\beta + \delta_{12}$

Proof. Let $X = \{v \mid \texttt{Skeptsupp}_P(x) > \frac{n}{2} \Rightarrow x \in v, \texttt{Skeptsupp}_P(x) < \frac{n}{2} \Rightarrow x \notin v\}$. Suppose $[M] \not\subseteq X$, i.e., there is a $v \in [M]$ s.t. $v \notin X$. This means there is an x s.t.

- $x \notin v$ and $\texttt{Skeptsupp}_P(x) > \frac{n}{2}$, or
- $x \in v$ and $\texttt{Skeptsupp}_P(x) < \frac{n}{2}$.

Suppose the first case, i.e., $x \notin v$ and $\texttt{Skeptsupp}_P(x) > \frac{n}{2}$ (other case symmetric). Then $d(v \cup \{x\}, P) < d(v, P)$, since strictly more than half models of the complete bases assign x to true. Thus, $v \notin [M]$.

Suppose $X \not\subseteq [M]$, i.e., there is a $v \in X$ s.t. $v \notin [M]$. This means there is a $w \in [M]$ s.t. $d(w, P) < d(v, P)$. By the previous reasoning, we know that $[M] \subseteq X$, and that $w \in X$. This means that for all x if strictly more than half models of the complete bases assign x to true (false), then $x \in w$ ($x \notin w$). Thus, $x \in v$ iff $x \in w$ for all x where there is a strict majority, and $x \notin v$ iff $x \notin w$ where there is a strict majority against (both interpretations are equal when there is a majority). Therefore, there is a y s.t. $y \in v$ and $y \notin w$ or $y \notin v$ and $y \in w$ where $\texttt{Skeptsupp}_P(y) = \frac{n}{2}$. We now claim that $d(v, P) = d(w, P)$. For each such y where the interpretations differ it holds that each complete base contributes a distance of 1 or 0 to the sum of Hamming distances (namely 1 to one of the interpretations and 0 to the other). It holds that exactly $\frac{n}{2}$ contribute 1 and $\frac{n}{2}$ contribute 0 for v and the same for w. Since this is the case for each such y, and distances are the same where the interpretations coincide, it holds that they have the same distance. Thus, $v \in [M]$. This concludes that $[M] = X$. □

Theorem 4. This follows from Lemma 12: The skeptical/credulous outcome for each atom is decided dependent only on whether there is a strict majority for or against, or exactly half of the agents in favour or against. In any case, a strategic agent cannot further its index, since it if an atom is in its base's model, but not in the skeptical result, the only option, w.r.t. that atom, is to remove it, which never furthers the skeptical acceptance of that atom. Similar reasoning suffices for the other cases. □

Lemma 13. *Let $P = (K_1, \ldots, K_{n-1})$ be a profile and $a \in \mathcal{P}$. If $\texttt{Credsupp}_P(a) < \frac{n}{2}$, then $a \notin \texttt{Cred}(\Delta_{\top}^{d_H,\Sigma}(P + K_n))$ for any base K_n.*

Proof. Let $P' = (K_1, \ldots, K_{n-1}, K_n)$ for any K_n. Further, let v be an interpretation over $\mathcal{P}$ with $a \in v$ and $\nexists v'$ over $\mathcal{P}$ s.t. $a \in v'$ and $d(v, P') < d(v', P')$ (i.e. v has minimum sum of distances to all agents wrt interpretations that include a). We claim that $v \setminus \{a\} = w$ has strictly lower distance, i.e., we claim that $d(v, P') > d(w, P')$. By definition, we have

$$d(v, P') = d_H(v, K_1) + \cdots + d_H(v, K_{n+1}) + d_H(v, K_n).$$

Without loss of generality, we can assume that we order the bases K_i whether $a \notin \texttt{Cred}(K_i)$ or not (with first those that do not accept a credulously). Let $m = \texttt{Credsupp}(a)$. It holds that

$$\begin{aligned} d(v, P') = & \underbrace{d_H(v, K_1) + \cdots + d_H(v, K_m)}_{\text{strictly more than } \frac{n}{2}} + \\ & \underbrace{d_H(v, K_{m+1}) + \cdots + d_H(v, K_n)}_{\text{strictly less than } \frac{n}{2}}. \end{aligned}$$

By assumption, we get that $d_H(w, K) \leq d_H(v, K) + 1$ if $a \in \texttt{Cred}(K)$. To see this, consider $w' \in [K]$ such that $d_H(v, w') = d_H(v, K)$ (w' is a witness of the distance of v to K, thus distance between v and w' is minimum). We have $d_H(w, w') \leq d_H(v, w') + 1$, since v and w differ only in assignment of a. Thus, w' witnesses that distance (w, K) as at most one higher than (v, K). Similarly, we get that $d_H(w, K) \leq d_H(v, K) - 1$ when $a \notin \texttt{Cred}(K)$ (by a similar line of reasoning, we find that witness w' implies a lower distance). Overall, we get the inequality

$$\begin{aligned} d(w, P') \leq & \, d_H(v, K_1) + \cdots + d_H(v, K_m) - m + \\ & d_H(v, K_{m+1}) + \cdots + d_H(v, K_n) + (n - m). \end{aligned}$$

Thus, $d(w, P') \leq d(v, P') - m + n - m$ and, since $m > \frac{n}{2}$, we have $2 \cdot m > n$ and $d(w, P') < d(v, P')$. This implies that $v \notin [\Delta_{\top}^{d_H, \Sigma}(K_1, \ldots, K_{n-1}, K_n)]$, for any v such that $a \in v$. □

Proposition 7. This follows from Lemma 13 and from [Delobelle et al., 2016]. □

Reduction 1. *Let $\psi = \exists X \forall Y \varphi$ be a closed QBF in prenex form. Define*

- $D = \{d_1, \ldots, d_{3 \cdot |Y| + 1}\}$,
- $X_i = \{x_i \mid x \in X\}$ *for* $1 \leq i \leq n$ *with* $n = 3 \cdot (|D| + |Y|) + 1$, *and*
- *formula* $\chi = \bigwedge_{x \in X} x \leftrightarrow x_1 \leftrightarrow \cdots \leftrightarrow x_n$.

Construct profile $P = (K_1, K_2, K)$ over vocabulary $V = X \cup (\bigcup_{1 \leq i \leq n} X_i) \cup Y \cup D$, with $K_1 = K_2 = \chi \wedge \bigwedge_{z \in Y \cup D} z$ and $K = \top$. Finalize the instance $\texttt{red}(\psi) = (P, \mu)$ with $\mu = \chi \wedge (\varphi \rightarrow \bigwedge_{d \in D} \neg d) \wedge (\neg \varphi \rightarrow \bigwedge_{d \in D} d)$.

Lemma 14. *Let $\psi = \exists X \forall Y \varphi$ be a closed QBF in prenex form. For $\texttt{red}(\psi) = ((K_1, K_2, K), \mu)$ and any complete K' it holds that if $v \in [\Delta_{\mu}^{d_H, \Sigma}(K_1, K_2, K')]$, $v \models \varphi$, and there is a $v' \not\models \varphi$ such that $v' \cap X = v \cap X$ then it holds that for*

$$w = (v \cap (V \setminus (Y \cup D))) \cup (v' \cap Y) \cup D$$

we have $d(w, (K_1, K_2, K')) < d(v, (K_1, K_2, K'))$.

Proof. Assume $v \in [\Delta_{\mu}^{d_H,\Sigma}(K_1, K_2, K')]$ (i.e., v is a model of the merged result), and there exists an interpretation v' not satisfying φ such that $v'|_X = v|_X$, i.e., both interpretations assign the same truth value to variables in X. We show that w is in $[\Delta_{\mu}^{\Sigma,d_H}(K_1, K_2, K')]$, i.e., w is a model of the merged result. First, note that v satisfies μ and that w does not satisfy φ (since v' does not satisfy φ and w and v' have the same truth value assignment on the vocabulary of φ). This means w satisfies μ, since $w \models \chi$ (due to v satisfying χ) and $w \models (\neg\varphi \to \bigwedge_{d \in D} d)$ (the conjunct in the middle of μ is trivially satisfied, since $w \not\models \varphi$). We now show that supposing $d(v, (K_1, K_2, K')) \leq d(w, (K_1, K_2, K'))$ (the contrary to the lemma's claim) leads to a contradiction. By construction, w and v differ in their assignment only on variables in $Y \cup D$. Let us now consider the models of each base with minimum Hamming distance to v and w:

- $x_i \in \{x \in [K_i] \mid d(v, x) \leq d(v, x') \forall x' \in [K_i]\}$,
- $x' \in \{x \in [K'] \mid d(v, x) \leq d(v, x') \forall x'' \in [K']\}$,
- $y_i \in \{y \in [K_i] \mid d(w, y) \leq d(w, y') \forall y' \in [K_i]\}$, and
- $y' \in \{y \in [K'] \mid d(w, y) \leq d(w, y') \forall y'' \in [K']\}$.

By definition, it holds that $d(v, (K_1, K_2, K')) = d(v, x_1) + d(v, x_2) + d(v, x')$ and $d(w, (K_1, K_2, K')) = d(v, y_1) + d(v, y_2) + d(v, y')$. Since the difference is the Hamming distance, we can phrase the distances also as partitions of the whole vocabulary: $d(v, z) = |v \cap X \Delta z \cap X| + |v \cap X_1 \Delta z \cap X_1| \cdots |v \cap X_n \Delta z \cap X_n| + |v \cap Y \Delta z \cap Y| + |v \cap D \Delta z \cap D|$. Let $d^Z(v, z) = |v \cap Z \Delta z \cap Z|$ as an auxiliary notion for considering the Hamming distance on a part of the vocabulary. By construction, we have $d^X(v, z) = d^X(w, z')$ for any $z \in \{x_1, x_2, x'\}$ and $z' \in \{y_1, y_2, y'\}$, since $v \cap (X \bigcup_{1 \leq i \leq n} X_i)$ is equal to $w \cap (X \bigcup_{1 \leq i \leq n} X_i)$ and there are models of K_1 and K_2 for any assignment on X and the values of Y and D are "fixed" in these bases (only one value). Furthermore $x' = y'$ since K' is complete. The same holds for d^{X_i} for all i. This implies that the distance of the v and w to the profile differs only by contributions of variables Y and D to the symmetric differences. Thus,

$$d^Y(v, x_1) + d^Y(v, x_2) + d^Y(v, x') + d^D(v, x_1) + d^D(v, x_2) + d^D(v, x') <$$
$$d^Y(w, y_1) + d^Y(w, y_2) + d^Y(w, y') + d^D(w, y_1) + d^D(w, y_2) + d^D(w, y').$$

Since $v \models \varphi$ and $v \models \mu$ we have v assigns all D to false, and w assigns all D to true, by similar reasoning. Thus, it holds that the first part of the inequality is $d^Y(v, x_1) + d^Y(v, x_2) + d^Y(v, x') + |D| + |D| + d^D(v, x')$ and if all terms different to $|D|$ are 0 (the lowest value possible) then this is equal to $2 \cdot |D|$. On the other hand the right term is $d^Y(w, y_1) + d^Y(w, y_2) + d^Y(w, y') + 0 + 0 + d^D(w, y')$, if all have their highest value we have $3 \cdot |Y| + |D|$. Since $|D| = 3 \cdot |Y| + 1$ we have $2 \cdot |D| = 6 \cdot |Y| + 2 > 6 \cdot |Y| + 1 = 3 \cdot |Y| + |D|$. This contradicts the supposition that $d(v, (K_1, K_2, K')) \leq d(w, (K_1, K_2, K'))$. This implies that $d(v, (K_1, K_2, K')) > d(w, (K_1, K_2, K'))$. □

Theorem 5. For membership, compute whether the merging results in a given atom a being skeptically entailed, which is a problem in $\Delta_P^2 = P^{NP}$ [Konieczny et al., 2002]. If the atom is not entailed, non-deterministically construct a complete base K' over the given vocabulary (which is the same as guessing a truth value assignment), and check whether a is entailed in the merged result, again a check achievable in polynomial time with an NP oracle. Recall that, if there exists a base such that destructive manipulation is possible, then there is a complete base achieving the result, see Proposition 2. For hardness, let $\psi = \exists X \forall Y \varphi$ be a closed QBF in prenex form. Construct $\texttt{red}(\psi) = P = (K_1, K_2, K)$ (see Reduction 1). We claim that ψ is true iff there exists a complete base K' such that $\exists v \in [\Delta_\mu^{d_H,\Sigma}(K_1, K_2, K')]$ with $d_1 \notin v$. Assume ψ is true. Then there is an assignment on the X variables such that for all assignments on the Y variables φ is satisfied. Consider one such assignment v on X, and consider complete base $K' = \bigwedge_{x \in v} x \wedge \bigwedge_{x \in X \setminus v} \neg x \wedge \bigwedge_{z \in Y \cup D} z$. We claim that $\Delta_\mu^{d_H,\Sigma}(K_1, K_2, K') \models \bigwedge_{x \in v} x \wedge \bigwedge_{x \in X \setminus v} \neg x$, i.e,, every model of μ selected in the merged result assigns truth values to the X variables exactly as v. Suppose the contrary, i.e., there is a w with a different assignment on the X variables that is part of the merged result. Due to χ (see Reduction 1), both v and w assign to X_i's the same value as for X (since both are models of μ). Consider the distance of w to the profile: $d(w, P)$ is at least $1 + n$ (since K' assigns all variables of X and X_i's differently than w by assumption). Consider $w' = (w \cap (Y \cup D) \cup (v \cap (V \setminus (Y \cup D))$, i.e., w' assigns all variables in $Y \cup D$ as w, but the X and X_i variables as v (and K'). It holds that $d(w', P) < d(w, P)$, since for both K_1 and K_2 there is a corresponding model with the same distances (i.e., their contribution to the overall distance stays the same), and the distance to K' decreases by $n + 1$. (recall K' is complete). If there is no assignment on the Y variables, given the assignment of v to the X variables, that falsifies φ, it holds that any model of the merged result satisfies φ (since only models of μ are selected with the same assignment on X as v). This implies that there is a model of the result with d_1 false (due to construction of μ).

For the other direction, assume that there exists a complete base K' such that $\exists v \in [\Delta_\mu^{d_H,\Sigma}(K_1, K_2, K')]$ with $d_1 \notin v$. Note that, since $v \models \mu$ (by definition) and due to construction of μ, it holds that $v \models \varphi$ (any model of μ that falsifies an atom in D satisfies φ). We claim that any interpretation v_X with $v_X \supseteq v \cap X$ satisfies φ, i.e., $v_X \in [\varphi]$. Proving this claim shows that ψ is true. Suppose the contrary, i.e., there is a $v' \not\models \varphi$ with $v' \cap X = v_X$ (for the v_X's assignment on X there is an assignment on Y such that φ is falsified). Consider Lemma 14: v and v' satisfy the conditions of that lemma (for the equally named interpretations). This implies that for w, as defined in the lemma, we have $d(w, P) < d(v, P)$. This implies that v cannot be part of the merged result, a contradiction. Thus, there is no v', with $v' \cap X = v_X$, that falsifies φ, implying that ψ is true.

Finally, if φ is not tautological, then with the original profile P (with $K = \top$), it holds that d_1 is a skeptical consequence: a model of μ not satisfying φ has a lower distance w.r.t. all models of μ that satisfy φ. That is, if φ is refutable, then d_1 is not a skeptical consequence and destructive manipulation is possible iff ψ is true. Further, one can modify φ with $\varphi \wedge x'$ for a fresh x' that is existentially quantified such that ψ is true iff $\exists X \cup \{x'\} \forall Y \varphi \wedge x'$ is true. Thus, we can assume w.l.o.g. that φ is not a tautology. □