



The problem of no hands: responsibility voids in collective decisions

Hein Duijf^{1,2} · Frederik Van De Putte^{3,4}

Received: 29 September 2020 / Accepted: 28 August 2021 / Published online: 9 November 2021
© The Author(s) 2021

Abstract

The problem of no hands concerns the existence of so-called responsibility voids: cases where a group makes a certain decision, yet no individual member of the group can be held responsible for this decision. Criteria-based collective decision procedures play a central role in philosophical debates on responsibility voids. In particular, the well-known discursive dilemma has been used to argue for the existence of these voids. But there is no consensus: others argue that no such voids exist in the discursive dilemma under the assumption that casting an untruthful opinion is eligible. We argue that, under this assumption, the procedure used in the discursive dilemma is indeed immune to responsibility voids, yet such voids can still arise for other criteria-based procedures. We provide two general characterizations of the conditions under which criteria-based collective decision procedures are immune to these voids. Our general characterizations are used to prove that responsibility voids are ruled out by criteria-based procedures involving an atomistic or monotonic decision function. In addition, we show that our results imply various other insights concerning the logic of responsibility voids.

Hein Duijf and Frederik Van De Putte contributed equally to the paper.

✉ Frederik Van De Putte
vandeputte@esphil.eur.nl

Hein Duijf
hein.duijf@lrz.uni-muenchen.de

- ¹ Munich Center for Mathematical Philosophy, Faculty of Philosophy, Philosophy of Science and Study of Religion, Ludwig-Maximilians-Universität München, Munich, Germany
- ² Department of Philosophy, Faculty of Humanities, Free University of Amsterdam, Amsterdam, The Netherlands
- ³ Erasmus Institute for Philosophy and Economics, Erasmus School of Philosophy, Erasmus University of Rotterdam, Rotterdam, The Netherlands
- ⁴ Department of Philosophy and Moral Science, Faculty of Arts and Humanities, Centre for Logic and Philosophy of Science, Ghent University, Ghent, Belgium

1 Introduction

The *problem of no hands* concerns the existence of so-called responsibility voids in collective decision-making: cases where a group makes a certain decision, yet no individual member of the group can be held responsible for this decision.¹ In arguments for and against the possibility of such voids, one particular example of collective decision-making plays a central role, viz. the *discursive dilemma*.² On the one hand, Pettit and others have argued that the decision procedure used in the discursive dilemma is vulnerable to responsibility voids (Copp 2006; Pettit 2007; List and Pettit 2011). On the other hand, Braham and van Hees (2011) argue that, if voting untruthfully is eligible, it is immune to responsibility voids.³ In the present paper, we investigate whether such voids can arise *even if* untruthful voting is eligible. More precisely, we consider a wide class of collective decision procedures that includes the discursive dilemma as a special case and study the conditions under which responsibility voids can arise.

Our investigation is important for several reasons. First, criteria-based collective decision procedures play a central role in democratic and legal institutions, and responsibility voids may indicate a flaw in their regulation (Thompson 1980; Bovens 1998).⁴ Second, if it turns out to be difficult or impossible to hold any participant accountable for certain outcomes, then we may need to resort to an irreducible notion of collective responsibility (List and Pettit 2011).⁵ Third, a clear allocation of responsibility for outcomes may incentivize participants to act responsibly and may thereby improve collective decision-making (Weaver 1986).⁶

¹ This problem is distinct from the familiar “problem of many hands”; Thompson (1980, p. 905) writes: “Because many different officials contribute in many ways to decisions and policies of government, it is difficult even in principle to identify who is morally responsible for political outcomes. This is what I call the problem of many hands”.

² The discursive dilemma was introduced by Pettit (2001). It is related to the *doctrinal paradox*, which originates from legal theory (see Kornhauser 1992; Chapman 1998, the former introduced the term ‘doctrinal paradox’). It is moreover related to the so-called *Ostrogorski paradox* (Rae and Daudt 1976). In each of these paradoxes, there is a mismatch between the majority opinion on a range of criteria, legal questions, or policy issues on the one hand, and the majority opinion on the final decision, verdict, or preferred candidate on the other hand.

³ Braham and van Hees (2011, p. 6) write: “the conditions for these voids are so restrictive as to reduce the philosophical or institutional significance they might be thought to possess”.

⁴ Bovens (1998, p. 49) writes: “A meaningful calling to account is rendered extremely difficult and sometimes even impossible by the above variations on the problem of many hands. This frustrates the need for compensation and retribution on the part of the victims (‘the guilty still run around freely’). More important still, such outcomes signify that what began as a conscious and rational human act can be transformed by the structure and the dynamics of complex organisations into a sort of ‘act of God’ with its own dynamic which seems to be independent of any specific individual human action (‘there are no guilty ones’).”

⁵ On collective responsibility and responsibility voids, Pettit (2007, pp. 196–197) writes: “Let group agents be freed from the burden of being held responsible, and the door will open to abuses: there will be cases where no one is held responsible for actions that are manifestly matters of agential responsibility.”

⁶ Weaver (1986, p. 371) opens his paper with: “Politicians are motivated primarily by the desire to avoid blame for unpopular actions rather than by seeking to claim credit for popular ones.” Whether blame-avoiding strategies actually yield positive or negative effects is up for debate (see Hood 2011, especially

To explain what is at stake in this paper, let us introduce a particular version of the discursive dilemma. Suppose a committee consisting of three members—A, B, and C—has to decide on whether to invite a candidate for a job interview (for an academic position). To form their collective decision, the committee members do not vote directly on the issue of whether to invite the candidate; rather, they give their opinion as to whether the candidate satisfies three basic criteria: (R) she is a good researcher; (T) she is a good teacher; and (D) she will fulfil her administrative duties. The collective decision procedure is as follows. In the first stage, the committee members independently form their own opinion on each of the criteria and, then, share their opinions. In the second stage, the committee considers the majority opinion on each of the separate criteria and the collective decision is based on the majority opinion: the candidate is invited if and only if for each criterion there is a majority who expressed the opinion that she satisfies that criterion.

Consider the scenario in which all but A submit that the candidate is a good researcher, all but B submit that the candidate is a good teacher, and all but C submit that the candidate will fulfil her administrative duties. Hence, the candidate satisfies each criterion according to the majority opinion and, as a result, the committee decides to invite the candidate. This scenario can be represented by the matrix in Table 1. What happens if it turns out that the candidate should not have been invited? In particular, can any of the committee members be held responsible for the decision to invite the candidate? Pettit (2007, p. 197f) defends a negative answer in a case similar to our example and concludes that responsibility voids are possible.

However, Braham and van Hees (2011, 2012) argue that responsibility voids do not arise in the discursive dilemma if voting untruthfully is an eligible option. Simply stated, on their conception of responsibility, an individual can be held (co-) responsible for the group's decision to invite the candidate if there is some other eligible action available to that individual that would have been better for avoiding that decision (Braham and van Hees 2011, 2012, condition AOC). In the particular case of the discursive dilemma, for each committee member, the best way to avoid the decision to invite the candidate is by voting “no” on each of the criteria. Hence, since no committee member does this, *if* voting untruthfully is eligible, *then all* committee members can be held responsible for the decision to invite the candidate. Braham and van Hees conclude that “the relevance of these voids pales in the face of the fact that voting in many cases takes place in what is essentially a strategic environment. In such a context, it is eligible to express an opinion other than what one holds” (Braham and van Hees 2011, p. 14).

These considerations by Braham and Van Hees trigger three questions about responsibility voids that call for a general and systematic analysis. First, what exactly makes it the case that the discursive dilemma is immune to responsibility voids, under this assumption? Second, are there perhaps other decision procedures that are vulnerable to responsibility voids, even if untruthful voting is allowed? Third,

Footnote 6 (continued)

Chapter 8). In any case, these studies show that blame and credit play a crucial role in politician's decision-making and should therefore be incorporated in a formal account of collective decision-making.

under which conditions is a collective decision procedure vulnerable to responsibility voids?

To answer these questions, we propose and specify a conception of individual responsibility that applies to a wide class of scenarios in which a group of agents is to form its decision on the basis of what we call a *criteria-based decision procedure*. That is, we consider cases where the decision is functionally determined by the group opinion on the criteria, in the sense that as soon as one fixes that group opinion, the group's decision is also fixed.⁷ In turn, the group opinion on each criterion is fully determined by the individual opinions on that criterion. The discursive dilemma is an example where the criteria-based decision procedure can be described as follows: the group opinion on each criterion is obtained by applying the simple majority rule, and the candidate is invited if and only if according to the group opinion, she satisfies all criteria.

Within this setting, we establish three key observations. First, our analysis vindicates the logical possibility of responsibility voids (even if untruthful voting is eligible). Second, it supports the conclusion that the discursive dilemma is immune to such voids (at least, under the assumption that untruthful voting is eligible): there cannot be a responsibility void, irrespective of the votes of the committee members. Third, and more generally, we show that if the underlying decision function of a criteria-based decision procedure is *atomistic* or, equivalently, *monotonic*, then that procedure is immune to responsibility voids (Theorem 5).⁸

Each of these observations follows from our central results that concern the characterization of immunity to responsibility voids. That is, we establish necessary and sufficient conditions for the immunity to responsibility voids, for the class of all criteria-based decision procedures. We characterize such immunity in two conceptually distinct ways.⁹ First, we characterize it in terms of whether and how each criterion contributes to the decision in question (Theorem 3). Second, we characterize immunity in terms of the shape of the set of opinions that do not support the decision, when this set is conceived as a subspace of the entire opinion space (Theorem 4). Both characterizations are insightful in their own way. Moreover, both can be used to establish, in a relatively straightforward way, that atomistic (or monotonic) decision functions guarantee immunity to responsibility voids.

Our analysis can be used as a springboard for investigating a number of further topics from the general logic of responsibility voids. To illustrate this point, we will briefly cover two such topics.¹⁰ First, one may wonder whether the fact that a given collective decision procedure is vulnerable to responsibility voids with respect to a *particular* decision entails that that procedure is vulnerable to responsibility voids with respect to *any possible* decision. We show that this entailment *does not hold*. Second, one can view the problem of no hands as a problem concerning the

⁷ The underlying decision function may be part of the institutional setting in which the group operates, or they may be the result of an earlier collective decision.

⁸ More details are given in Sect. 7.

⁹ More details can be found in Sect. 6.

¹⁰ See Sect. 8 for a more elaborate discussion.

Table 1 The discursive dilemma

	Good researcher?	Teaching skills?	Admin duties?	Invite?
A	No	Yes	Yes	No
B	Yes	No	Yes	No
C	Yes	Yes	No	No
Committee	Yes	Yes	Yes	Yes

distribution of responsibility. Responsibility voids concern the case where no individual can be held responsible for a decision, but one may also consider *majoritarian responsibility voids*: the case where no majority of the group can be held responsible for a given group decision. What is the logical relation between majoritarian responsibility voids and responsibility voids simpliciter? We show that, for criteria-based decision procedures in which the group opinion is obtained by simple majority over the criteria, immunity to majoritarian responsibility voids is equivalent to immunity to responsibility voids *simpliciter* (Theorem 6).

The paper is organised as follows. We start by introducing the class of criteria-based decision procedures (Sect. 2) and specify our conception of responsibility voids in the context of such procedures (Sect. 3). We then show that the discursive dilemma is immune to responsibility voids yet there are other criteria-based decision procedures that are vulnerable to such voids (Sect. 4). In Sect. 5, we provide a characterisation of responsibility and avoidance in virtue of whether and how the criteria may contribute to a given decision. Our central results regarding the characterisation of responsibility voids are presented in Sect. 6. In Sect. 7 we show that if the underlying decision function is atomistic or, equivalently, monotonic, then that procedure is immune to responsibility voids. We discuss two further topics in the general logic of responsibility voids in Sect. 8. In the concluding section, we discuss prospects for future work. All results are proved in the appendix.

2 Criteria-based decision procedures

The purpose of this section is to introduce a broad class of collective decision procedures that includes the procedure of the discursive dilemma. Generally, a criteria-based decision procedure enables a finite group of agents to make a collective decision, based on the opinions cast by each individual member on a finite set of criteria. (We will often omit the adjective “collective” and simply write “decision”.) The criteria are taken to be binary, with their two possible values denoted by 0 and 1. We moreover presuppose that the actual collective decision is induced by the individual opinions on the criteria, in two stages (cf. the Introduction). First, starting from the individual opinions, we fix the *group opinion* on every criterion, using criteria-specific thresholds. More precisely, for each criterion, its threshold represents how many members of the group should vote “1” on that criterion in order for the group

opinion to also be “1” for that criterion.¹¹ Different criteria can hence have different thresholds. In the second stage, the decision is determined by the group opinion, using a function that determines, for every opinion, the decision that it supports.

For technical reasons, we impose one additional restriction: if all except one member of the group agree on the value of a given criterion, then the exceptional individual cannot change the group opinion on that criterion. In other words, we do not allow for procedures in which unanimity is required for the group opinion to be 1, resp. 0 on a given criterion. Technically, we assume that the threshold for each criterion is higher than 1 and smaller than the total number of individuals in the group. As a result of this restriction, the group needs to contain at least three agents. We return to this restriction in Sect. 9, explaining how our results fare when it is dropped. Nevertheless, our analysis covers a broad class of collective decision procedures, as given by the following definition:

Definition 1 A *criteria-based decision procedure* (abbreviation: CDP) is a tuple $\Phi = \langle N, C, D, T, f \rangle$, where N is a finite set of at least three *agents*, C is a finite set of *criteria*, D is a finite set of *decisions* with $|D| \geq 2$, $T : C \rightarrow \{2, \dots, |N| - 1\}$ is a *threshold assignment*, and $f : \{0, 1\}^C \rightarrow D$ is a surjective *decision function*.¹²

We will use c, c', \dots to refer to members of C and d, d', \dots to refer to members of D . Given a CDP, an *opinion* $O : C \rightarrow \{0, 1\}$ specifies for each criterion whether it is satisfied. That is, $O(c) = 1$ means that according to O , c is satisfied; $O(c) = 0$ means that according to O , c is not satisfied. We will use \mathbb{O} to denote the *opinion space*, i.e. the set of all opinions. Note that formally, $\mathbb{O} = \{0, 1\}^C$. Hence, the opinion space can be viewed as a coordinate system with $|C|$ dimensions and two values for each dimension. An *opinion profile* $\mathcal{P} \in \mathbb{O}^N$ is a tuple of opinions, one for each agent in N . (Notice that every procedure we consider is symmetric: every individual agent has the same set of opinions available to her.) Given an opinion profile $\mathcal{P} \in \mathbb{O}^N$, we let $O_{\mathcal{P}}^i$ denote *the vote cast by individual i in \mathcal{P}* .

Definition 2 Where $\Phi = \langle N, C, D, T, f \rangle$ is a CDP and \mathcal{P} is an opinion profile:

1. the *group opinion* $O_{\mathcal{P}}^{\Phi}$ induced by \mathcal{P} is defined as follows: for all $c \in C$,

$$O_{\mathcal{P}}^{\Phi}(c) := \begin{cases} 1, & \text{if } |\{i \in N \mid O_{\mathcal{P}}^i(c) = 1\}| \geq T(c) \\ 0, & \text{otherwise.} \end{cases}$$

2. the *collective decision induced by \mathcal{P}* is $\Phi(\mathcal{P}) := f(O_{\mathcal{P}}^{\Phi})$.

¹¹ Note that here, “group opinion” is a purely technical term that refers to an intermediate step in a decision procedure; it does not presuppose that groups have beliefs or anything alike.

¹² The requirement that f is surjective simply means that D contains no redundancies: in principle, every decision $d \in D$ could be reached, depending on the group opinion. Together with the requirement that $|D| \geq 2$, it also means that the procedure is non-vacuous, in that the decision is not settled beforehand.

The decision procedure of the example from the Introduction can be formalized by the CDP $\Phi_0 = \langle N, C, D, T, f \rangle$ where N consists of three agents, C consists of three criteria, and D consists of two possible decisions, i.e. to invite or not to invite the candidate. Let us denote these two decisions by “1” and “0” respectively. Moreover, each criterion has threshold 2, and the decision function f is given by: $f(O) = 1$ if and only if for all $c \in C$ it holds that $O(c) = 1$. We will often denote opinions as tuples of 0s and 1s; for instance, our example involves the respective opinions $(0, 1, 1)$, $(1, 0, 1)$, and $(1, 1, 0)$. So, $(0, 1, 1)$ is the opinion where only the second and third criterion are satisfied. The opinion profile discussed in the Introduction is $\mathcal{P}_0 = \langle (0, 1, 1), (1, 0, 1), (1, 1, 0) \rangle$. Note that we have $O_{\mathcal{P}_0}^{\Phi_0} = (1, 1, 1)$, and $\Phi_0(\mathcal{P}_0) = f(O_{\mathcal{P}_0}^{\Phi_0}) = 1$. In other words, the decision is that the candidate be invited.

The CDP Φ_0 is one instance of a *majoritarian* criteria-based decision procedure, i.e. a CDP in which the threshold for each criterion is given by simple majority. Note that the use of threshold functions allows us to also include decision procedures where $|N|$ is even and the threshold represents a tie-breaking rule in favour of either 0 or 1. The CDP Φ_0 is moreover *binary*, in the sense that it only leaves room for two decisions.¹³ In what follows, we will illustrate our results using examples of binary, majoritarian CDPs, but unless stated otherwise, they apply to CDPs in general.

Groups often need to make a decision between more than two alternatives. To illustrate non-binary CDPs, consider a variant of our example from the Introduction where candidates for an academic position are to be divided into three categories: “definitely invite”, “discuss at the meeting”, and “definitely do not invite”. Consider the criteria-based decision procedure which postulates that a given candidate is definitely invited if she satisfies all three criteria; definitely not invited if she satisfies no criterion; and discussed at the meeting otherwise. This would then be a CDP involving three criteria and three possible decisions.¹⁴

This example gives rise to a broad class of collective decision procedures that our formalism can accommodate: consider examples where a job candidate is directly hired if they satisfy a sufficiently high number of criteria k , will be directly rejected if they do not satisfy a sufficiently high number of criteria l , and will be reconsidered if they satisfy a number of criteria between k and l (where $|C| \geq k > l \geq 0$).¹⁵ Remark that whenever $k \neq l + 1$, then there are three possible group decisions: hire the candidate, reject the candidate, or reconsider the candidate. We propose to represent these group decisions by 1, 0, and $\frac{1}{2}$, respectively. More precisely, the decision

¹³ From the perspective of the literature on judgement aggregation, the criteria of a binary CDP can be viewed as premises of an agenda and the decision of such a CDP as a (unique, designated) conclusion in the agenda, where the conclusion is a truth-function of the premises (Dokow and Holzman 2009; Nehring and Puppe 2008). Our formalism can then be viewed as implying the restriction that the premises are logically independent. The truth-functional relation between the premises and the conclusion is represented by the decision function f .

¹⁴ Similar procedures are commonly applied in the context of scientific conferences, where chairs make an initial categorization of submitted papers into “definitely accept”, “definitely reject”, and “discuss among the chairs”, based on the scores given by each of the individual reviewers.

¹⁵ This class of examples is inspired by—but different from—(Nehring and Puppe 2010, Example 1 on page 7).

function $f_{(k,l)}$ is given by: $f_{(k,l)}(O) = 1$ if and only if $|\{c \mid O(c) = 1\}| \geq k$; $f_{(k,l)}(O) = 0$ if and only if $|\{c \mid O(c) = 1\}| \leq l$; and $f_{(k,l)}(O) = \frac{1}{2}$ otherwise. This gives us a general class of CDPs of the form $\Phi_{(k,l)} = \langle N, C, D, T, f_{(k,l)} \rangle$ where N, C, T, k , and l can be chosen arbitrarily such that $|C| \geq k > l \geq 0$.¹⁶ Note that the previous example falls within this class of examples, where $k = 3$ and $l = 0$. The discursive dilemma falls within this class of examples, where $k = 3$ and $l = 2$.

For any CDP $\Phi = \langle N, C, D, T, f \rangle$ and any $d \in D$, let us use \mathbb{F}_d^Φ to denote the set of opinions that support decision d in Φ , and let us use $\overline{\mathbb{F}_d^\Phi}$ to denote the set of opinions that do not support decision d in Φ . That is, $\mathbb{F}_d^\Phi := \{O \in \mathbb{O} \mid f(O) = d\}$ and $\overline{\mathbb{F}_d^\Phi}$ is the complement of \mathbb{F}_d^Φ relative to \mathbb{O} . We will suppress the superscript Φ and simply write \mathbb{F}_d ($\overline{\mathbb{F}_d}$) in case it does not give rise to ambiguity. To illustrate, for any CDP $\Phi_{(k,l)}$ with $k \neq l + 1$ it holds that $\mathbb{F}_1^{\Phi_{(k,l)}} = \mathbb{F}_0^{\Phi_{(k,l)}} \cup \mathbb{F}_{\frac{1}{2}}^{\Phi_{(k,l)}}$.

Note that for the case where $D = \{0, 1\}$, i.e. for binary CDPs, every decision function f corresponds to a unique set of opinions \mathbb{F}_1 ; and, conversely, every set $\mathbb{F} \subseteq \mathbb{O}$ can be associated with a unique binary decision function f given by $f(O) = 1$ iff $O \in \mathbb{F}$. Consequently, for binary CDPs, the decision function f can be represented by the subset \mathbb{F}_1 of the entire opinion space. To see how this works, let us reconsider the example of the Introduction. Since $|C| = 3$, the available opinions can be represented by a three-dimensional cube and the decision function can be represented as a subset of the vertices of that cube. We adopt the convention to represent decision functions by underlining the opinions that support the decision 1. It should hence be obvious that Fig. 1 represents the set of possible opinions and decision function of the example from the Introduction.

To prevent misunderstandings, it is important to emphasize one feature of the generality of the notion of CDPs. Foreshadowing terminology defined in Sects. 5 and 7, CDPs can involve a non-monotonic or (equivalently) a holistic decision function. That is, there exist CDPs where the “valence” of some criteria may depend on which other criteria are satisfied.¹⁷ At this stage, this is best illustrated using an example.

Suppose the national government assembles a committee of three experts to decide on whether to implement a given policy for agriculture. For simplicity’s sake, let us assume that there are only three relevant questions: Will the policy create any significant inequalities among farmers (I)? Will the policy generate a surplus in

¹⁶ Note that, given k and l , $\Phi_{(k,l)}$ is strictly speaking not a specific CDP but a class of CDPs, since N, C , and T can still be chosen arbitrarily. Nonetheless, since our analysis does not depend on the exact nature of N, C , and T , we loosely speak of $\Phi_{(k,l)}$ as depicting a CDP.

¹⁷ There are some similarities between our notion of a binary criteria-based decision procedure and the “multi-dimensional decision-making rules” from Courtin and Laruelle (2020). The latter rules also work with a finite set of criteria (called “dimensions”), where the opinion profile (called “configuration”) is used to determine a decision on a binary issue. Multi-dimensional decision-making rules are more general in one respect, since they do not presuppose that opinions are first aggregated per criterion, before the decision is made. In another respect, binary CDPs are more general, since they do not presuppose a monotonic relation between the criteria and the decision.

terms of food production (S)? And third, is there a societal need for generating such a surplus (N)? It is plausible to assume that the policy should only be adopted if the answer to the first question is negative and if the answers to the second and third question are either both positive or both negative. After all, in these circumstances, the policy meets the societal need regarding generating such a surplus and does not generate inequality. In contrast, if exactly one of the answers to (S) and (N) is positive, then the policy should not be endorsed since it would not meet the societal demands.

This case can be represented by the CDP $\Phi_1 = \langle N, C, D, T, f \rangle$, where $|N| = 3$, $|C| = 3$, $D = \{0, 1\}$, $T(c) = 2$ for all $c \in C$, and f is such that $f(O) = 1$ if and only if $O = (0, 1, 1)$ or $O = (0, 0, 0) = 1$. Note that this is a binary, majoritarian CDP. The set of opinions and the decision function can be pictured as in Fig. 2. Notice that the impact of the second criterion c_2 of generating a surplus (corresponding to the y -axis) is not monotonic: if the other criteria are both not satisfied, then the satisfaction of c_2 leads to the decision 0; but if only the third criterion of societal need (corresponding to the z -axis) is satisfied, then the satisfaction of c_2 leads to the decision 1.

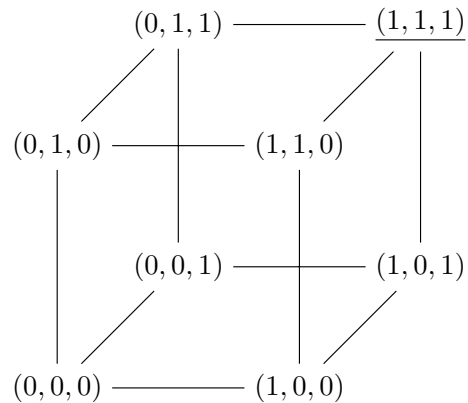
3 Responsibility voids and avoidance

To study responsibility voids, it is important to elaborate on the notion of individual responsibility. In particular, the aim of this section is to specify when we can hold an individual responsible (or at least co-responsible) for some collective decision that is obtained by means of a CDP. Our analysis imposes three assumptions. First, as mentioned in the Introduction, *every* opinion is eligible. Second, it is a situation of complete information: the individuals know the details of the collective decision procedure. Third, the individuals are fully uncertain about the opinions cast by the other members of the group: every opinion of the others is considered possible and their probabilities are unknown. While it would be interesting to investigate whether these assumption can be relaxed, this is left for future work.

Although the concept of moral responsibility is complex and heavily debated within philosophy, there is a commonly recognized intuition that an individual cannot be held responsible for a given outcome if that outcome was unavoidable.¹⁸ To account for this intuition, and similar to the analysis of Braham and van Hees (2012), we propose that an individual cannot be held responsible for a given outcome if she maximally avoided that outcome. The main distinction with the work of Braham and van Hees is that we spell out the notion of maximal

¹⁸ The idea that the ability to do otherwise plays a key role in responsibility attributions goes back to Aristotle's *Nicomachean Ethics* (see Aristotle 2019, Book VI) and it is closely related to the principle of alternative possibilities (see however (Frankfurt 1969; Van Inwagen 1999) for a critical discussion of that principle). More precisely, our formalization is related to recent work on moral responsibility and the avoidance of blame (McKenna 1997; Wyma 1997; Otsuka 1998; Hetherington 2003). Formalisms that capture this idea in a somewhat different way than ours are presented by Vallentyne (2008) and Braham and van Hees (2012).

Fig. 1 A three-dimensional representation of the possible opinions and decision function in Φ_0



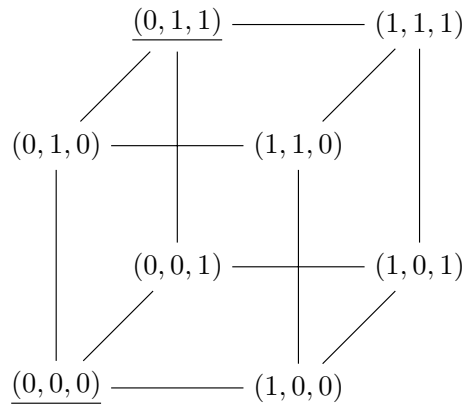
avoidance in non-probabilistic terms, in line with our assumption of full uncertainty mentioned before. Our non-probabilistic setting is very similar to Horty's (2001) analysis of obligations. One can view our conception of responsibility in CDPs as an adaptation of Horty's theory.

To understand our notion of maximal avoidance, it is important to consider the counterfactual scenarios where an individual unilaterally deviated from her actual vote. That is, we may ask whether there is an alternative vote by which that individual would have changed the decision *while fixing the votes of the other committee members*. Reconsider the example of the Introduction where the three members of the hiring committee voted in such a way that the group has made the decision to invite the candidate (Table 1). We see that any committee member would have avoided the decision to invite the candidate if she had unilaterally deviated from her actual vote by voting against all three criteria.

To talk about such counterfactual scenarios, it is useful to introduce some extra notation. Given an opinion O , an individual i and an opinion profile \mathcal{P} , we let $\mathcal{P}_{i \rightarrow O}$ be the opinion profile that would result from replacing the opinion of individual i in \mathcal{P} with O . Accordingly, the collective decision that would be made, had agent i cast vote O , is given by $\Phi(\mathcal{P}_{i \rightarrow O})$. For example, where $\mathcal{P} = \langle (0, 1, 1), (1, 0, 1), (1, 1, 0) \rangle$, the fact that the first committee member could have avoided the decision to invite the candidate, if she had unilaterally deviated from her actual vote by voting against all three criteria, is expressed by $\Phi_0(\mathcal{P}_{1 \rightarrow (0,0,0)}) = f(1, 0, 0) = 0$.

In line with our assumption of full uncertainty above, we moreover propose to consider not only the *actual* votes of the other committee members, but all their *possible* votes. That is, to maximally avoid a given decision, a given committee member needs to consider all possible votes she could have cast, and the decisions these would induce relative to all possible votes of the other members.

Fig. 2 A three-dimensional representation of the possible opinions and decision function in Φ_1



To make these ideas precise, we follow Harty’s account (Harty 2001), which is itself based on the familiar decision-theoretic notions of dominance and admissibility.¹⁹ In line with our assumption of full uncertainty, dominance and admissibility make most sense when individuals take all opinions of the others into consideration. To compare two opinions O and O' with regard to their potential to avoid a given decision, we say that O *simply dominates* O' for avoiding the decision if and only if O would avoid that decision whenever O' would, regardless of the votes of the others. Furthermore, O *weakly dominates* O' if and only if O simply dominates O' but not conversely. Finally, we submit that an opinion O *maximally avoids* a given decision if and only if there is no opinion O' such that O' weakly dominates O for avoiding that decision. More formally:

Definition 3 (Dominance and Avoidance) Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP, let $d \in D$, and let $O, O' \in \mathbb{O}$. Then,

1. O *simply dominates* O' for avoiding d in Φ (notation: $O \leq_d^\Phi O'$) if and only if for every individual $i \in N$ and for every opinion profile \mathcal{P} it holds that if $\Phi(\mathcal{P}_{i \rightarrow O'}) \neq d$, then $\Phi(\mathcal{P}_{i \rightarrow O}) \neq d$;
2. O *weakly dominates* O' for avoiding d in Φ (notation: $O <_d^\Phi O'$) if and only if $O \leq_d^\Phi O'$ and $O' \not\leq_d^\Phi O$;

Finally, O *maximally avoids* d in Φ if and only if there is no $O' \in \mathbb{O}$ such that $O' <_d^\Phi O$. We let \mathbb{A}_d^Φ denote the set of opinions that maximally avoid d in Φ .

We will suppress the superscript Φ and simply write \leq_d , $<_d$ and \mathbb{A}_d when it is clear from the context.

¹⁹ Weak dominance relates to Savage’s (1972) “sure-thing principle”. Savage writes: “I know of no other extralogical principle governing decisions that finds such ready acceptance” (21). Harty in fact uses “weakly dominates” for what we call “simply dominates”, and “strongly dominates” for what we call “weakly dominates”. Our choice of terminology follows common practice in decision and game theory.

It is important to note that our definition of dominance can be simplified: we can remove the quantifier ranging over individuals in N and replace the occurrences of individual i with the first agent 1. After all, every individual has the same set of opinions available to her and every criteria-based collective decision procedure treats every individual alike. More formally:

Observation 1 Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP and let $d \in D$. Let agent 1 be the first agent in N . Then, for any two opinions O, O' it holds that $O \leq_d^\Phi O'$ if and only if for every opinion profile \mathcal{P} it holds that if $\Phi(\mathcal{P}_{1 \rightarrow O'}) \neq d$, then $\Phi(\mathcal{P}_{1 \rightarrow O}) \neq d$.

In what follows, we will freely rely on this observation.

In line with the above considerations and given our specification of avoidance, our conception of responsibility is as follows:

Definition 4 (Responsibility) Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP, let $i \in N$, and let $d \in D$. Let \mathcal{P} be an opinion profile. Then, given \mathcal{P} , an individual i can be held responsible for d if and only if $\Phi(\mathcal{P}) = d$ and $O_p^i \notin \mathbb{A}_d^\Phi$.

Let us briefly consider how this theory of responsibility and maximal avoidance addresses familiar cases of overdetermination. These cases can be illustrated by considering the decision procedure of the example from the Introduction, but where the committee members each cast a slightly different vote. Consider the scenario where each committee member casts a vote in favour of all three criteria and the candidate is invited as a result. This case illustrates a crucial feature of cases of overdetermination: no individual member can change the decision by unilaterally changing her vote. Let us examine whether any of them is responsible for inviting the candidate. In other words, did the committee members maximally avoid the decision to invite the candidate? Consider the alternative to cast a vote against all three criteria. Then, it is easy to verify that this alternative vote weakly dominates their actual vote with respect to avoiding the decision to invite the candidate. More precisely, we get $(0, 0, 0) <_1 (1, 1, 1)$. As a consequence, *each* committee member can be held responsible for the decision. Hence, our theory of responsibility adequately addresses cases of overdetermination.

It may be helpful to note that for any two opinions O and O' and any decision d there are only three possibilities: (1) they are *equivalent* for avoiding d , i.e. $O \leq_d O'$ and $O' \leq_d O$; (2) they are *incomparable* for avoiding d , i.e. $O \not\leq_d O'$ and $O' \not\leq_d O$; and (3) one *weakly dominates* the other for avoiding d , i.e. $O <_d O'$ or $O' <_d O$. The second option where the two opinions are incomparable is often overlooked. We leave it to the reader to verify that this possibility is demonstrated by opinions $(1, 1, 0)$ and $(0, 1, 1)$ in the example from the Introduction. Moreover, these dominance orderings induce a preorder on the set of opinions:

Observation 2 Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP and let $d \in D$. Then, \leq_d^Φ is a preorder on \mathbb{O} . That is, \leq_d^Φ is reflexive and transitive.

It can be cumbersome to check whether a given opinion maximally avoids a given decision. For example, in cases where $|C| = 3$ and $|N| = 3$, each individual has $2^3 = 8$ available opinions and this gives rise to $8^3 = 512$ possible opinion profiles. To determine whether a given opinion O simply dominates another opinion O' with respect to avoiding a given decision d in Φ , we may need to survey all 64 combinations of opinions for the other individuals. Moreover, to check whether O maximally avoids d in Φ , we would have to check for all 7 other available opinions whether they weakly dominate O with respect to avoiding d in Φ . As a result, in the worst-case scenario, we would have to cover $7 \times 64 = 448$ cases. It is therefore important to announce that we will provide a very neat characterization of both \leq_d and \mathbb{A}_d in Sect. 5.

Our conception of responsibility has an important implication: relative to any CDP, every agent is able to avoid being responsible for a given decision. That is, there always is an opinion that maximally avoids a given decision:²⁰

Observation 3 Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP and let $d \in D$. Then there exists an opinion O in \mathbb{A}_d^Φ .

Notice that there are cases where an agent can maximally avoid a given decision in multiple distinct ways. Nevertheless, to avoid being responsible for a given decision, agents may need to vote untruthfully. This should not come as a surprise, since our conception of responsibility gives no special weight to an agent's personal convictions.²¹

Let us conclude by returning to the main aim of this paper: the study of responsibility voids in collective decision-making. In a given criteria-based decision procedure, a profile of opinions *creates a responsibility void* when it induces a decision while none of the individual members of the group can be held responsible for that decision. From the perspective of our theory of responsibility, this means that a responsibility void occurs when a particular decision is made while every individual member of the group casts a vote that maximally avoids that decision. More formally:

Definition 5 (*Responsibility Voids*) Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP and let $d \in D$. Then,

1. Where \mathcal{P} is an opinion profile, \mathcal{P} *creates a responsibility void with respect to* d if and only if $\Phi(\mathcal{P}) = d$ while for every i in N it holds that $O_{\mathcal{P}}^i \in \mathbb{A}_d^\Phi$.

²⁰ This follows from the fact that the dominance relation \prec_d is irreflexive and transitive and the assumption that an agent has only finitely many opinions available to her.

²¹ Although our analysis resembles that of Braham and van Hees (2012), it is important to note that they allow for the possibility that some alternatives are not eligible and, therefore, their framework includes models that give special weight to an agent's personal convictions—in particular, by postulating that casting the opinion that coincides with the agent's personal conviction is the only eligible alternative.

2. Φ is *immune to responsibility voids with respect to d* if and only if there are no opinion profiles \mathcal{P} that create a responsibility void with respect to d .

Φ is *immune to responsibility voids* if and only if Φ is immune to responsibility voids with respect to every $d \in D$.

We say that a CDP Φ is *vulnerable to responsibility voids with respect to d* if it is not immune to responsibility voids with respect to d ; and that it is *vulnerable to responsibility voids* if and only if it is vulnerable to responsibility voids with respect to some $d \in D$.

4 The possibility of responsibility voids

The discursive dilemma from the Introduction plays a central role in the philosophical debate on irreducible collective responsibility and responsibility voids. Let us investigate whether that example is vulnerable to responsibility voids. It can be verified that the only way for the committee members to maximally avoid the decision to invite the candidate is by voting against all three criteria.²² That is, for every $O \neq (0, 0, 0)$ it holds that $(0, 0, 0) <_1 O$. Moreover, whenever the committee collectively decides to invite the candidate, then at least two committee members must have voted in favour of some criteria. Hence, this criteria-based decision procedure is immune to responsibility voids with respect to the decision to invite the candidate.

Since this key example is immune to responsibility voids with respect to the decision to invite the candidate, it is natural to ask whether responsibility voids are ruled out *tout court*. Our formalism can be used to vindicate the logical possibility of responsibility voids.

To do so, it is helpful to mention a necessary condition for immunity to responsibility voids. Consider a CDP that is immune to responsibility voids with respect to a given decision d , and let O be an opinion that maximally avoids d . This means that, when every individual casts opinion O , then every individual casts an opinion that maximally avoids d . Consequently, since the CDP is immune to responsibility voids, the decision supported by that specific opinion profile must differ from d . But note that the group opinion induced by that profile is O , and hence the decision supported by the profile is $f(O)$. More generally, if a CDP is immune to responsibility voids, then every opinion that maximally avoids d must not support the decision d .²³

Observation 4 Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP and let $d \in D$. If Φ is immune to responsibility voids with respect to d , then $\mathbb{A}_d^\Phi \subseteq \mathbb{F}_d^\Phi$.

To illustrate the logical possibility of responsibility voids, let us reconsider the example where the national government assembles a committee of three experts to

²² As explained in Sect. 3, verifying this can be rather cumbersome. However, the claim readily follows from Theorems 1 and 2 in Sect. 5.

²³ Recall that $\overline{\mathbb{F}_d^\Phi} = \mathbb{O} \setminus \mathbb{F}_d^\Phi$.

decide on whether to implement a given policy for agriculture, represented by Φ_1 and depicted in Fig. 2 on page 11. Given Observation 4, we see that if $\mathbb{F}_0 \cap \mathbb{A}_0 \neq \emptyset$, then Φ_1 is vulnerable to responsibility voids with respect to 0. So, in particular, Φ_1 is immune to responsibility voids with respect to 0 only if $(0, 1, 0) (\in \mathbb{F}_0)$ is weakly dominated by either $(0, 0, 0)$ or $(0, 1, 1)$.²⁴ Consider the opinion profile $\mathcal{P} = \langle O_1, O_2, O_3 \rangle$ where $O_1 = (0, 1, 0)$, $O_2 = (0, 1, 1)$, and $O_3 = (0, 0, 1)$. Then, we see that $\Phi_1(\mathcal{P}) = f((0, 1, 1)) = 1$ while $\Phi_1(\mathcal{P}_{1 \rightarrow (0,0,0)}) = f((0, 0, 1)) = 0$. Hence, $(0, 0, 0) \not\leq_0 (0, 1, 0) = O_1$. By similar reasoning, the opinion profile $\mathcal{P}' = \langle O_1, O'_2, O_3 \rangle$ where $O'_2 = (0, 0, 0)$ demonstrates that $(0, 1, 1) \not\leq_0 (0, 1, 0) = O_1$. In fact, in virtue of Theorems 1 and 2 in Sect. 5, one can easily verify that $(0, 1, 0) \in \mathbb{A}_0$. Hence, Φ_1 is vulnerable to responsibility voids with respect to 0.

Let us make this more explicit. Consider the particular scenario where all three experts submit that only the second criterion is satisfied, i.e. submit the opinion $(0, 1, 0)$. The resulting group opinion is that only the second criterion is satisfied and, hence, the group decides not to implement the policy. Can any expert be held responsible for the decision 0? Since each expert could have voted differently, we need to ask whether her actual vote is weakly dominated for avoiding the decision 0. One might, for instance, think that her actual vote $(0, 1, 0)$ is weakly dominated for avoiding 0 by casting the opinion that none of the criteria are satisfied, i.e. $(0, 0, 0)$. However, the opinion profile depicted by \mathcal{P} , specified in the previous paragraph, demonstrates that there exists a counterfactual scenario where casting $(0, 0, 0)$ would have lead to the decision 0 while her actual vote $(0, 1, 0)$ would not have done so. As a consequence, her actual vote is not weakly dominated for avoiding 0 by the opinion $(0, 0, 0)$. More generally, each expert maximally avoided the decision not to implement the policy and, therefore, none of them can be held responsible for that decision.

We conclude that there exist criteria-based decision procedures that are vulnerable to responsibility voids. In such cases the group could make a certain decision even if all agents cast a vote that maximally avoids that decision. This answers the second main question from the Introduction positively: there are indeed examples of collective decision-making that are vulnerable to responsibility voids. In particular, not every CDP is immune to responsibility voids. In Sect. 6 we will characterize the class of criteria-based decision procedures that are immune to responsibility voids.

5 The contribution of criteria

To arrive at our central characterization result, we first introduce a classification of whether and how criteria may contribute to a given decision in a given criteria-based decision procedure (Sect. 5.1). This classification allows us to characterize the

²⁴ After all, given that $\mathbb{F}_1^{\Phi_1} = \{(0, 0, 0), (0, 1, 1)\}$, if $(0, 1, 0)$ is weakly dominated for avoiding 0 by neither $(0, 0, 0)$ nor $(0, 1, 1)$, then either $(0, 1, 0)$ maximally avoids 0 or some other opinion in \mathbb{F}_0 maximally avoids 0. In either case, we have $\mathbb{F}_0 \cap \mathbb{A}_0 \neq \emptyset$.

dominance relation that plays a key role in our conception of individual responsibility (Sect. 5.2).

5.1 Whether and how a criterion may contribute

As we have seen in Sect. 3, determining whether one opinion simply dominates another opinion for avoiding a given decision is routine, but may be cumbersome. To simplify this task, we propose to focus on each criterion separately and determine whether it plays a role in the outcome of the decision function. In the example of the Introduction, the majority opinion is that the candidate satisfies all three criteria. We may wonder whether unilaterally changing the value of any of these criteria changes the outcome of the decision function. In this particular case and for this particular majority opinion, it is easy to see that every criterion plays a crucial role: if we were to change its value, then the decision would change from inviting the candidate to not inviting the candidate. To determine whether and how a given criterion affects the decision, we investigate for each opinion whether only changing the value of that criterion would affect the decision. By focusing on each criterion separately, we greatly simplify the task of determining whether one opinion simply dominates another for avoiding a given decision.

To make these ideas precise, we introduce some new terminology. Given opinions O and O' , we let $\Delta(O, O') = \{c \in C \mid O(c) \neq O'(c)\}$ be the *difference between O and O'* . We say that O *agrees with O'* on criterion c if and only if $O(c) = O'(c)$ (and hence if and only if $c \notin \Delta(O, O')$).

In a given CDP, a given criterion *does not contribute* to that decision if and only if, for any two opinions that only differ on that criterion, it holds that the one supports the decision if and only if the other does. Conversely, a given criterion *does contribute* to that decision if and only if there are opinions that only differ on that criterion, where one supports the decision in question, and the other does not.

Definition 6 (Contribution) Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP, let $c \in C$, and let $d \in D$. Then, c *does not contribute* to d in Φ if and only if for all opinions O and O' with $\Delta(O, O') = \{c\}$ it holds that $f(O) = d$ if and only if $f(O') = d$.

Given a CDP Φ , we use $C_{\Phi, d}^c$ to denote the set of criteria that do not contribute to d in Φ . We will simply write C_d^c when Φ is clear from the context.

Let us reconsider the example of the Introduction. In line with our previous observations, each criterion contributes to the decision to invite the candidate.

Whenever a given criterion contributes to a given decision, we propose to distinguish between whether that criterion contributes positively or negatively, and between whether it contributes strictly or not. For instance, a criterion contributes positively to a given decision if and only if there exist opinions that (i) only differ

on that criterion, (ii) support different decisions, and (iii) the opinion according to which the criterion is satisfied supports that decision. More precisely:²⁵

Definition 7 (*Contribution: positive, negative, strict*) Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP, let $c \in C$, and let $d \in D$. Then,

1. c contributes positively to d in Φ if and only if there are opinions O and O' with $\Delta(O, O') = \{c\}$, $f(O) \neq f(O')$, $O'(c) = 1$ and $f(O') = d$.
2. c contributes negatively to d in Φ if and only if there are opinions O and O' with $\Delta(O, O') = \{c\}$, $f(O) \neq f(O')$, $O'(c) = 0$ and $f(O') = d$. Moreover,
3. c contributes strictly positively to d in Φ if and only if c contributes positively to d in Φ and does not contribute negatively to d in Φ .
4. c contributes strictly negatively to d in Φ if and only if c contributes negatively to d in Φ and does not contribute positively to d in Φ .
5. c contributes strictly to d in Φ if and only if c contributes strictly positively or strictly negatively to d in Φ .

Let us reconsider the example of the Introduction. In line with our previous observations, it is easy to see that each criterion contributes strictly positively to the decision to invite the candidate. That is, for each criterion there are opinions where changing the value of that criterion from ‘not satisfied’ to ‘satisfied’ will change the decision from ‘don’t invite the candidate’ to ‘invite the candidate’; and, moreover, unilaterally changing the value of any criterion from ‘satisfied’ to ‘not satisfied’ will never change the decision in this way.

To demonstrate the possibility that a criterion contributes both positively and negatively, let us reconsider CDP Φ_1 from Sect. 2 (see also Fig. 2). In Φ_1 , the second and third criterion contribute both positively and negatively to the decision 1. For example, since $f(0, 1, 0) = 0$ and $f(0, 0, 0) = 1$, it follows that c_2 contributes negatively to 1. Since $f(0, 0, 1) = 0$ and $f(0, 1, 1) = 1$, it follows that c_2 also contributes positively to 1.

Finally, the terminology above yields a classification of whether and how a given criterion could contribute to a given decision. To clarify this, we introduce a few notational conventions. Consider any CDP Φ and any decision d . We use $C_{\Phi,d}^+$ to denote the set of criteria that contribute strictly positively to d in Φ ; likewise, $C_{\Phi,d}^-$ denotes the set of criteria that contribute strictly negatively to d in Φ . We let $C_{\Phi,d}^\sigma$ denote the set of all criteria that contribute strictly to d in Φ , i.e. $C_{\Phi,d}^\sigma = C_{\Phi,d}^+ \cup C_{\Phi,d}^-$. The set of criteria that contribute both positively and negatively to d in Φ is denoted by $C_{\Phi,d}^\pm$. We suppress the subscript Φ and simply write $C_d^e, C_d^+, C_d^-, C_d^\sigma, C_d^\pm$ when it is clear from the context. Note that $\{C_d^e, C_d^+, C_d^-, C_d^\pm\}$ is a partition of C . Table 2 gives an overview of our classification of criteria in terms of whether and how they contribute to d in Φ .

²⁵ Nehring and Puppe (2008) use the terms “positively relevant”, resp. “negatively relevant” for what we call “contributing positively”, resp. “contributing negatively”. Their setting in the cited paper is more general, since they only assume a relational (not a functional) connection between the premises (criteria) and the conclusion (decision).

5.2 Contribution, right values, and avoidance

Once we know whether and how each criterion contributes to a given decision, the task of determining whether a given opinion dominates another opinion for avoiding that decision is greatly simplified. To make a start, consider criteria that do not contribute to a given decision. For any such criterion, any two opinions that only differ on that criterion will simply dominate each other for avoiding that decision. After all, under these assumptions, for any opinion profile, casting the first opinion will support the same decision as casting the second opinion would. More formally:

Observation 5 Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP, let O and O' be opinions, and let $d \in D$. If $\Delta(O, O') \subseteq C_{\Phi, d}^e$, then $O \leq_d^\Phi O'$ and $O' \leq_d^\Phi O$.

What about opinions that only differ on criteria that contribute strictly? Given an opinion O that assigns value 1 to a given criterion that contributes strictly positively to a given decision, the opinion O' that differs from O only on that criterion will weakly dominate O . More precisely:

Observation 6 Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP, let $c \in C$, and let $d \in D$. Then,

- If c contributes strictly positively to d in Φ , then for any opinions O, O' with $\Delta(O, O') = \{c\}$ and $O'(c) = 0$ it holds that $O' <_d^\Phi O$.
- If c contributes strictly negatively to d in Φ , then for any opinions O, O' with $\Delta(O, O') = \{c\}$ and $O'(c) = 1$ it holds that $O' <_d^\Phi O$.

This suggests that the best way to avoid a given decision is by assigning the value 0 to any criterion that contributes strictly positively to that decision. By similar reasoning, this observation suggests that the best way to avoid a given decision is by assigning the value 1 to any criterion that contributes strictly negatively to that decision. In accordance with these suggestions, we introduce a function ρ_d^Φ that assigns the *right value* for avoiding d in Φ to any criterion that contributes strictly to d .

Definition 8 (Right Values) Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP and let $d \in D$. Then, we define $\rho_d^\Phi : C_d^\sigma \rightarrow \{0, 1\}$ by:

$$\rho_d^\Phi(c) := \begin{cases} 0, & \text{if } c \in C_d^+ \\ 1, & \text{if } c \in C_d^- \end{cases}$$

Moreover, let $O \in \mathbb{O}$ and let $c \in C_d^\sigma$ be a criterion that strictly contributes to d in Φ . Then,

1. O assigns the right value for avoiding d in Φ to c if and only if $O(c) = \rho_d^\Phi(c)$;
2. O is right for avoiding d in Φ if and only if O assigns the right value for avoiding d to every criterion that contributes strictly to avoiding d .

Table 2 Classification of criteria in terms of whether and how they contribute to a given decision d (suppressing the subscript Φ)

		Does c contribute positively to d ?	
		Yes	No
Does c contribute negatively to d ?	Yes	$c \in C_d^\pm$	$c \in C_d^-$
	No	$c \in C_d^+$	$c \in C_d^\epsilon$

Lastly, we let $R_d^\Phi(O)$ denote the set of criteria that contribute strictly and for which the opinion O assigns the right value for avoiding d in Φ ; and we let \mathbb{R}_d^Φ denote the set of all opinions that are right for avoiding d in Φ .

It is easy to verify that $R_d^\Phi(O) = \{c \in C_d^\sigma \mid O(c) = \rho_d^\Phi(c)\}$ and that an opinion O is right for avoiding d in Φ if and only if $R_d^\Phi(O) = C_d^\sigma$. In line with previous conventions, we suppress the superscript Φ and simply write ρ_d , $R_d(O)$, and \mathbb{R}_d when it is clear from the context.

With the help of these auxiliary definitions, we can characterize the simple dominance relation \leq_d^Φ in terms of whether and how criteria contribute to d in Φ and the values assigned to the criteria by the opinions. An opinion O simply dominates opinion O' for avoiding decision d in CDP Φ if and only if (a) O and O' agree on all criteria that contribute non-strictly, and (b) for all the criteria that contribute strictly, O assigns the right value whenever O' does.

Theorem 1 *Let Φ be a CDP, let $O, O' \in \mathbb{O}$, and let $d \in D$. Then,*

$$O \leq_d^\Phi O' \text{ iff } (a) \Delta(O, O') \cap C_{\Phi, d}^\pm = \emptyset \text{ and } (b) R_d^\Phi(O') \subseteq R_d^\Phi(O).$$

To better grasp this characterization of the dominance relation, it may be helpful to state three corollaries of this theorem. Consider any CDP Φ and any decision d . Recall from Sect. 3 that for any two opinions O and O' there are three possibilities:

- (1) they are *equivalent* for avoiding d , i.e. each simply dominates the other for avoiding d ;
- (2) they are *incomparable* for avoiding d , i.e. neither simply dominates the other for avoiding d ; or
- (3) one *weakly dominates* the other for avoiding d .

Each of these possibilities can now be characterized as follows. First, two opinions are equivalent for avoiding a given decision if and only if they differ only on criteria that do not contribute to that decision.²⁶ Second, two opinions are incomparable for

²⁶ Notice that this entails Observation 5.

avoiding a given decision if and only if either (a) they differ on criteria that contribute non-strictly, or (b) there exist two criteria that contribute strictly on which they differ such that the former opinion assigns the right value to the first criterion and the latter opinion assigns the right value to the second criterion. Third, one opinion weakly dominates another opinion for avoiding a given decision if and only if (a) they agree on criteria that contribute non-strictly, (b) whenever the latter opinion assigns the right value to a criterion that contributes strictly, then the former opinion does so too, and (c) they differ on a criterion that contributes strictly.²⁷

Corollary 1 *Let Φ be a CDP, let $O, O' \in \mathbb{O}$, and let $d \in D$. Then,*

- (1) $O \leq_d^\Phi O'$ and $O' \leq_d^\Phi O$ iff $\Delta(O, O') \subseteq C_{\Phi, d}^e$.
- (2) $O \not\leq_d^\Phi O'$ and $O' \not\leq_d^\Phi O$ iff (a) $\Delta(O, O') \cap C_{\Phi, d}^\pm \neq \emptyset$ or
(b) $R_d^\Phi(O') \not\subseteq R_d^\Phi(O)$ and $R_d^\Phi(O) \not\subseteq R_d^\Phi(O')$.
- (3) $O <_d^\Phi O'$ iff (a) $\Delta(O, O') \cap C_{\Phi, d}^\pm = \emptyset$ and
(b) $R_d^\Phi(O') \subset R_d^\Phi(O)$.

To illustrate these characterizations, let us return to the broad class of collective decision procedures of the form $\Phi_{(k,l)}$ presented on page 7 and consider the decision 1. It is easy to see that every criterion contributes to decision 1 and, moreover, every criterion contributes strictly positively to decision 1. In other words, $C_1^e = \emptyset$ and, moreover, $C_1^+ = C$. Consider examples where $|C| = 4$. Then, for instance, following Corollary 1, it is easy to verify that opinions $(0, 0, 0, 1)$ and $(1, 0, 0, 0)$ are incomparable for avoiding 1 and that $(1, 0, 1, 0)$ weakly dominates $(1, 0, 0, 0)$ for avoiding 1.

It is important to remark that Theorem 1 yields a characterization of the opinions that maximally avoid a given decision. Recall that \mathbb{R}_d^Φ is the set of all opinions that assign the right value to all criteria that contribute strictly to d in Φ . The following theorem demonstrates that this set is identical to the set of opinions that maximally avoid decision d in Φ :

Theorem 2 *Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP and let $d \in D$. Then,*

$$\mathbb{R}_d^\Phi = \mathbb{A}_d^\Phi.$$

We can thus determine the opinions that maximally avoid a given decision once we know which criteria contribute strictly. After all, a given opinion maximally avoids a given decision if and only if it assigns the right value to each criterion that contributes strictly.

To illustrate that Theorem 2 greatly simplifies the task of determining the opinions that maximally avoid a given decision, let us reconsider the broad class of collective decision procedures of the form $\Phi_{(k,l)}$ and reconsider decision 1. As mentioned before, $C_1^+ = C$ and, therefore, $C_1^\sigma = C$. As a result, $\rho_1 : C \rightarrow \{0, 1\}$ is

²⁷ Notice that this entails Observation 6.

given by $\rho_1(c) = 0$ for all $c \in C$ and \mathbb{R}_1 is the singleton set given by the opinion that assigns value 0 to all criteria. In light of Theorem 2, it follows that this opinion is the only one that maximally avoids 1.

Let us end with considering a case where $k > l + 1$ and, hence, $\frac{1}{2} \in D$. It is easy to verify that every criterion contributes to decision $\frac{1}{2}$ and, moreover, every criterion contributes both positively and negatively to decision $\frac{1}{2}$. In other words, $C_{\frac{1}{2}}^e = \emptyset$ and $C_{\frac{1}{2}}^\pm = C$. In particular, $C_{\frac{1}{2}}^\sigma = \emptyset$ and thus $\mathbb{R}_{\frac{1}{2}} = \emptyset$. Hence, Theorem 2 implies that every opinion maximally avoids $\frac{1}{2}$. In what follows we will see that this difference between decisions 1 and $\frac{1}{2}$ has immediate repercussions for whether the CDP is vulnerable to responsibility voids with respect to 1 and $\frac{1}{2}$, respectively.

6 Characterization of responsibility voids

We are now in the position to characterize the class of all criteria-based decision procedures that are immune to responsibility voids with respect to a given decision. In Sect. 6.1, we provide a characterization in terms of the set of opinions that assign all the right values for avoiding that decision (Theorem 3). In Sect. 6.2, we give another characterization in terms of the shape of the set of opinions that do not support the decision, where this set can be conceived as a subspace of the entire opinion space \mathbb{O} (Theorem 4).

6.1 Right values and responsibility voids

A given criteria-based decision procedure is immune to responsibility voids with respect to a given decision *if and only if* all the opinions that are right for avoiding that decision do not support that decision. Or, equivalently, (by Theorem 2) if and only if the opinions that maximally avoid that decision do not support the decision.²⁸

Theorem 3 *Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP and let $d \in D$. Then, the following are equivalent:*

1. Φ is immune to responsibility voids with respect to d ;
2. any opinion that maximally avoids d does not support d , i.e. $\mathbb{A}_d^\Phi \subseteq \overline{\mathbb{F}_d^\Phi}$;
3. any opinion that is right for avoiding d does not support d , i.e. $\mathbb{R}_d^\Phi \subseteq \overline{\mathbb{F}_d^\Phi}$.

As a consequence, a given criteria-based decision procedure is vulnerable to responsibility voids with respect to a given decision if and only if there exists an opinion that maximally avoids that decision while supporting that decision. Or, equivalently, if and only if there exists an opinion that assigns the right values for

²⁸ Notice that this entails Observation 4.

avoiding that decision to all criteria that contribute strictly while supporting that decision.

Let us illustrate the usefulness of this theorem by considering the broad class of collective decision procedures of the form $\Phi_{(k,l)}$. Consider decision 1. We have seen that \mathbb{R}_1 and \mathbb{A}_1 are given by the singleton set containing the opinion that assigns value 0 to all criteria. Since the decision function assigns decision 0 to that opinion, Theorem 3 demonstrates that collective decision procedures of the form $\Phi_{(k,l)}$ are immune to responsibility voids with respect to 1. Similar reasoning shows that $\Phi_{(k,l)}$ is also immune to responsibility voids with respect to 0. Consider decision $\frac{1}{2}$. We have seen that $\mathbb{R}_{\frac{1}{2}} = \mathbb{A}_{\frac{1}{2}} = \mathbb{O}$. Hence, Theorem 3 shows that collective decision procedures of the form $\Phi_{(k,l)}$ are vulnerable to responsibility voids with respect to $\frac{1}{2}$.

In the remainder of this subsection we try to provide some insights into why the theorem obtains. Note first that, by Theorem 2, items (2.) and (3.) of Theorem 3 are equivalent. Also, item (1.) implies item (2.) by Observation 4.

We now explain why (2.) implies (1.). To do so, we adapt the notion of an affine subset from linear algebra. An *affine subset* of \mathbb{O} can be thought of as a set of opinions that is generated by fixing the values of some criteria and by freely choosing the values for the other criteria. More precisely:

Definition 9 (Affine Subset) Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP. A subset $\mathbb{K} \subseteq \mathbb{O}$ is called an *affine subset of \mathbb{O}* if and only if there is a $C_{\mathbb{K}} \subseteq C$ and $g_{\mathbb{K}} : C_{\mathbb{K}} \rightarrow \{0, 1\}$ such that $\mathbb{K} = \{O \in \mathbb{O} \mid O(c) = g_{\mathbb{K}}(c) \text{ for all } c \in C_{\mathbb{K}}\}$.

To illustrate, it may be helpful to consider the case where $|C| = 3$. We can distinguish four classes of affine subsets \mathbb{K} of \mathbb{O} by varying the size of $C_{\mathbb{K}}$:

- (i) $|C_{\mathbb{K}}| = 3$ yields singleton affine subsets of the form $\{O\}$,
- (ii) $|C_{\mathbb{K}}| = 2$ yields affine subsets of the form $\{O, O'\}$ with $\Delta(O, O')$ a singleton, i.e., sets corresponding to one edge of the cube,
- (iii) $|C_{\mathbb{K}}| = 1$ yields affine subsets of the form $\{O_1, O_2, O_3, O_4\}$ which correspond to one face of the cube,
- (iv) $|C_{\mathbb{K}}| = 0$ yields affine subsets that span the entire set \mathbb{O} , hence the entire cube.

Let us briefly reconsider the CDPs Φ_0 and Φ_1 that we have discussed previously (see Figs. 1, 2). Notice that in Φ_0 it holds that \mathbb{F}_1 is an affine subset of \mathbb{O} while \mathbb{F}_0 is not. In comparison, in Φ_1 it holds that neither \mathbb{F}_1 nor \mathbb{F}_0 is an affine subset of \mathbb{O} .

Affine subsets are important because they satisfy a kind of closure property: for every opinion profile that consists of opinions from an affine subset it holds that the group opinion based on that opinion profile will also be in that affine subset.

Observation 7 Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP. Let \mathbb{K} be an affine subset of \mathbb{O} and let $\mathcal{P} \in \mathbb{K}^N$ be an opinion profile. Then, $O_{\mathcal{P}}^{\Phi} \in \mathbb{K}$.

The final crucial observation is that the opinions that are right for avoiding a given decision constitute an affine subset of the set of opinions. After all, for any CDP Φ and decision d , it holds that $C_d^\sigma \subseteq C$ and ρ_d is a function from C_d^σ to $\{0, 1\}$. Since Theorem 2 proves that the set of opinions that are right for avoiding a given decision is identical to the set of opinions that maximally avoid that decision, it follows that the latter constitutes an affine subset.

Observation 8 Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP and let $d \in D$. Then, \mathbb{A}_d^Φ is an affine subset of \mathbb{O} .

These observations jointly entail that whenever all individuals cast a vote that maximally avoids a given decision d , then the group opinion maximally avoids d as well. Under the assumption that all opinions that maximally avoid d do not support d , it follows that whenever all individuals cast a vote that maximally avoids d , then the collective decision does not support d . Hence, under this assumption, the CDP is immune to responsibility voids with respect to the decision in question.

6.2 Opinion space and responsibility voids

Whereas we have characterized responsibility voids with respect to a given decision in terms of the right values in the previous section, we here aim for such a characterization in terms of the opinions that do not support the decision. More specifically, given a CDP Φ and a decision d , we investigate whether there exists a property on \mathbb{F}_d that is both necessary and sufficient for immunity to responsibility voids with respect to decision d .

To do so, we generalize the notion of affine subsets. We first introduce some terminology. For any two opinions, we let the *interval* between them be the set of opinions that only differ on criteria where the initial two opinions differ. More precisely, for any opinions O and O' , we use $[O, O']$ to denote the interval between O and O' , which is given by $[O, O'] := \{O'' \in \mathbb{O} \mid \Delta(O, O'') \subseteq \Delta(O, O')\}$.

It may be helpful to note that one can imagine going from opinion O to O' by stepwise changing one of the values of the criteria where they disagree. Note that, depending on the way the criteria are ordered, this gives different intermediate steps; the interval between O and O' includes *all* possible intermediate steps.

An *affine core* \mathbb{K} of a subset of opinions \mathbb{O}' can be thought of as an affine subset of the set of opinions where for every opinion in \mathbb{O}' there exists an opinion in the affine core \mathbb{K} such that the interval between these opinions is included in \mathbb{O}' . More precisely,

Definition 10 (Affine Core) Let $\Phi = \langle N, C, D, T, F \rangle$ be a CDP and let $\mathbb{O}' \subseteq \mathbb{O}$ be a set of opinions. Then,

1. A set $\mathbb{K} \subseteq \mathbb{O}'$ is called an *affine core of \mathbb{O}'* if and only if (a) \mathbb{K} is an affine subset of \mathbb{O}' and (b) for every $O' \in \mathbb{O}'$ there exists an $O \in \mathbb{K}$ such that $[O, O'] \subseteq \mathbb{O}'$.
2. We say that \mathbb{O}' *has an affine core* if and only if there exists an affine core of \mathbb{O}' .

Every affine subset of the entire set of opinions trivially has an affine core (namely, the affine subset itself). There exist, however, subsets of opinions that have an affine core while they are not themselves an affine subset of the entire set of opinions. To demonstrate this possibility, let us briefly reconsider the CDP Φ_0 (see Fig. 1). Recall that $\mathbb{F}_1^{\Phi_0}$ is an affine subset of \mathbb{O} while $\mathbb{F}_0^{\Phi_0}$ is not. Although $\mathbb{F}_0^{\Phi_0}$ is not an affine subset of \mathbb{O} , it can easily be verified that $\mathbb{K}_0 := \{(0, 0, 0)\}$ is an affine core of $\mathbb{F}_0^{\Phi_0}$. It is also important to recognize that there exist subsets of opinions that do not have an affine core. To demonstrate this possibility, let us briefly reconsider the CDP Φ_1 (see Fig. 2). It is easy to verify that $\mathbb{F}_1^{\Phi_1} = \{(0, 0, 0), (0, 1, 1)\}$ does not have an affine core. After all, any affine core of $\mathbb{F}_1^{\Phi_1}$ needs to be a non-empty subset of it, and we have $[(0, 0, 0), (0, 1, 1)] = \{(0, 0, 0), (0, 1, 0), (0, 0, 1), (0, 1, 1)\} \not\subseteq \mathbb{F}_1^{\Phi_1}$.

Our main characterization result proves that a given CDP is immune to responsibility voids with respect to a given decision *if and only if* the subset of opinions that do not support the decision has an affine core. The only-if implication follows from the more specific result that, for any given CDP that is immune to responsibility voids with respect to a given decision it holds that the opinions that are right for avoiding a given decision constitute an affine core of the set of opinions that do not support the decision.²⁹

Theorem 4 *Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP and let $d \in D$. Then, the following are equivalent:*

1. Φ is immune to responsibility voids with respect to d ;
2. the set of opinions that do not support d has an affine core, i.e. $\overline{\mathbb{F}_d^\Phi}$ has an affine core.

The important distinction with Theorem 3 is that Theorem 4 gives us a means for verifying whether a given CDP Φ is vulnerable to responsibility voids with respect to a given decision d by directly checking a property on $\overline{\mathbb{F}_d^\Phi}$ *without* having to first determine the opinions that maximally avoid that decision \mathbb{A}_d^Φ . Let us highlight the fruitfulness of this theorem by considering some examples. Consider the CDP Φ_0 (see Fig. 1). Above, we illustrated that $\overline{\mathbb{F}_0^{\Phi_0}} (= \mathbb{F}_1^{\Phi_0})$ is an affine subset of \mathbb{O} . This implies that $\overline{\mathbb{F}_0^{\Phi_0}}$ has an affine core and, by Theorem 4, it follows that CDP Φ_0 is immune to responsibility voids with respect to 0. Consider the CDP Φ_1 (see Fig. 2). We have shown that $\overline{\mathbb{F}_0^{\Phi_1}} (= \mathbb{F}_1^{\Phi_1})$ does not have an affine core. Hence, Theorem 4 implies that Φ_1 is vulnerable to responsibility voids with respect to decision 0.

²⁹ One may wonder whether, given a CDP Φ and a decision d , it holds that \mathbb{R}_d is the largest or smallest core of $\overline{\mathbb{F}_d^\Phi}$ (in terms of set inclusion). It can, however, be shown that this is not the case. More generally, it can be shown that the set of affine cores of a given subset need not be closed under intersection or union.

7 Atomism, monotonicity, and immunity

As noted in the Introduction, Theorems 3 and 4 are the central results of this paper. They provide conditions that are both necessary and sufficient for immunity to responsibility voids. We proceed with investigating two conditions that are much more well-known: *atomism* and *monotonicity*. We prove that each is sufficient for immunity to such voids. We will demonstrate that these conditions are equivalent for our class of CDPs and that they naturally link to the two distinct characterizations of immunity to voids.

First, in analogy with the account of reason-based preferences of Dietrich and List (2017) and the debate on moral particularism (Ridge and McKeever 2016), we may distinguish between *atomistic* and *holistic* decision functions.³⁰ Atomism requires that the “valence” of any given criterion does not depend on which other criteria are satisfied. In contrast, holism requires that such interdependencies can arise, at least in some cases: for a given criterion c , there may exist configurations of values for the other criteria such that satisfying c is “better” for realizing a given decision, whereas for other configurations of values for the other criteria, satisfying c is in fact “worse” for realizing that decision.³¹

From the perspective of our formalism, and given the terminology from Sect. 5, these philosophical positions can be made more precise. Atomism boils down to the requirement that, whenever a criterion contributes to a given decision at all, then it contributes strictly. Formally, a decision function f is *atomistic with respect to decision d* if and only if $C_d^\pm = \emptyset$. Or, equivalently, if and only if $C = C_d^\sigma \cup C_d^\epsilon$. Otherwise, f is *holistic with respect to d* .

Any CDP that involves an atomistic decision function is immune to responsibility voids:

Theorem 5 *Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP and let $d \in D$. If f is atomistic with respect to d , then Φ is immune to responsibility voids with respect to d .*

Theorem 5 demonstrates that if one endorses atomism for the relation between criteria and a given decision, then responsibility voids with respect to that decision are ruled out. Let us indicate how this theorem follows from our first characterization result, i.e. Theorem 3. Recall that Theorem 3 states that a given CDP Φ is immune to responsibility voids with respect to a given decision d if and only if $\mathbb{R}_d \subseteq \mathbb{F}_d$. Consider any CDP Φ involving a decision function that is atomistic with respect to d . Take an arbitrary opinion O that is right for avoiding d , i.e. $O \in \mathbb{R}_d$. Assume that $O \notin \mathbb{F}_d$, i.e. $f(O) = d$. Since f is surjective and $|D| \geq 2$, there must be

³⁰ This terminology was introduced in the theory of reasons by Dancy (1993). Importantly, the distinction we discuss here should not be confused with the distinction between particular and general immunity to responsibility voids that we introduce in Sect. 8.1. As Dietrich and List (2017) point out, the property of atomism is also known as “separability” in decision and game theory.

³¹ One may further distinguish between radical holists who argue that this holds for any criterion (or reason), and more moderate holists who would argue this holds at least for some criteria (or reason). In what follows we only consider the moderate version and call that “holism”.

some $O' \neq O$ such that $f(O') \neq d$. But note that, since f is atomistic with respect to d , O' can only differ from O on criteria that either do not contribute to avoiding d at all, or that contribute strictly. Consider $O'' \in \mathbb{R}_d$ such that $O''(c) = O'(c)$ for all $c \notin C_d^\sigma$. Then, O and O'' only differ on criteria that do not contribute to avoiding d . The assumption that $f(O) = d$ entails that $f(O'') = d$ (cf. Observation 5). Since $f(O') \neq d = f(O'')$ and $R_d(O') \subseteq R_d(O'')$, there must be a criterion $c' \in C_d^\sigma$ and opinions O_1 and O_2 such that $\Delta(O_1, O_2) = \{c'\}$, $f(O_1) \neq d = f(O_2)$, and $O_2(c') = \rho_d(c')$. This contradicts the definition of ρ_d and the fact that c' contributes strictly to d . Hence, $O \in \overline{\mathbb{F}_d}$ and, therefore, $\mathbb{R}_d \subseteq \overline{\mathbb{F}_d}$ and Theorem 3 entails that Φ is immune to responsibility voids with respect to d .

Second, the notion of *monotonicity* is commonly considered in the literature on judgement aggregation. When we think of the opinion space \mathbb{O} as a coordinate system, then a decision function is (O, d) -monotonic if and only if, given any opinion O' , whenever one moves from O' towards O , one always preserves decision d . More precisely, given a decision function f , an opinion O and a decision d , we say that f is (O, d) -monotonic if and only if for all O', O'' such that $O'' \in [O', O]$, if $O' \in \mathbb{F}_d$, then $O'' \in \mathbb{F}_d$. More generally, we say that the decision function f is *monotonic with respect to decision d* if and only if there is some opinion O such that f is (O, d) -monotonic.

For the class of collective decision-making procedures that we consider, it can be shown that any decision function is atomistic with respect to a given decision *if and only if* it is monotonic with respect to that decision:

Observation 9 Let f be a decision function and let $d \in D$. Then, f is atomistic with respect to d if and only if f is monotonic with respect to d .

In light of this equivalence, Theorem 5 means that any CDP that involves a monotonic decision function is immune to responsibility voids. This statement can be verified alternatively using the notion of monotonicity and our second characterization result, i.e. Theorem 4. Recall that Theorem 4 states that a given CDP Φ is immune to responsibility voids with respect to a given decision d if and only if $\overline{\mathbb{F}_d}$ has an affine core. Consider any CDP Φ involving a decision function that is monotonic with respect to d . Then, there exists an opinion O such that f is (O, d) -monotonic. Define O^* as the opinion that disagrees with O on every criterion: $O^*(c) = \overline{1}$ if and only if $O(c) = 0$, for all $c \in C$. We show that $\{O^*\}$ is an affine core of $\overline{\mathbb{F}_d}$. First, since f is surjective and $|D| \geq 2$, there exists an opinion O' with $f(O') \neq d$. Notice that $O' \in [O^*, O]$ and, since f is (O, d) -monotonic, it holds that $f(O^*) \neq d$. Second, take any $O'' \in \overline{\mathbb{F}_d}$ and consider any $O_1 \in [O'', O^*]$. Then, $O'' \in [O_1, O]$ and, since f is (O_1, d) -monotonic, it holds that $f(O_1) \neq d$. It follows that $\{O^*\}$ is an affine core of $\overline{\mathbb{F}_d}$ and Theorem 4 entails that Φ is immune to responsibility voids with respect to d .

Given Theorem 5, it is natural to ask whether every CDP that is immune to responsibility voids necessarily involves an atomistic or monotonic decision function. Our formalism can be used to refute this claim. Consider the binary, majoritarian CDP Φ_2 involving three agents, three criteria, and the decision function

depicted in Fig. 3. It is straightforward to verify that c_3 contributes non-strictly to 1 (c_3 corresponds to the z -axis). Therefore, $C_d^\pm \neq \emptyset$ and f is neither atomistic nor monotonic. However, $\mathbb{F}_1^{\Phi_2}$ has an affine core, viz. the set $\{(1, 1, 1), (1, 1, 0)\}$. So, by Theorem 4, while its decision function is not monotonic, this CDP is immune to responsibility voids with respect to decision 0. Hence, immunity to responsibility voids with respect to a given decision does not entail monotonicity of the underlying decision function with respect to that decision. Conversely, that the underlying decision function is holistic or non-monotonic need not imply vulnerability to responsibility voids.

8 Further observations

Our formal framework allows us to study the problem of no hands more rigorously and more generally. To illustrate this prospect, we now briefly consider two further questions that are triggered by our analysis. We show how answers to them can easily be found, relying on our characterization theorems. Since these questions have not been discussed in the literature on responsibility voids, they illustrate that our analysis can offer a novel and general perspective on philosophical issues.

8.1 Particular and general immunity

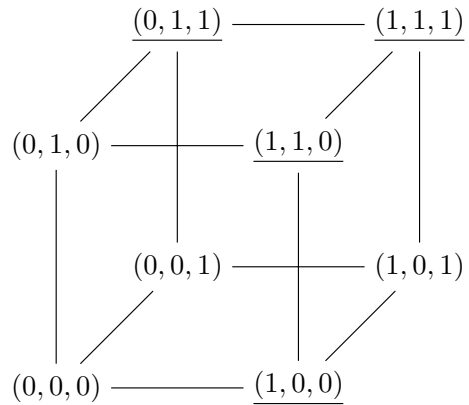
We have investigated the conditions under which collective decision-making can give rise to responsibility voids with respect to a particular decision. Whenever a criteria-based decision procedure is immune to responsibility voids with respect to a given decision, we could say that it satisfies *particular immunity*. Analogously, we could say that a CDP satisfies *general immunity* if it is immune to responsibility voids with respect to any decision.

This distinction is relevant in the following sense. There may be cases where responsibility voids with respect to one decision are considered worse than responsibility voids with respect to another decision. For instance, consider again the example from the Introduction. It may well be that it is more important to invite all good candidates than it is to reject all poor candidates. After all, poor candidates will likely not persuade the committee in the interview. Under these circumstances, while we would like to avoid responsibility voids altogether, a responsibility void with respect to “the candidate is not invited” is considered more problematic than a responsibility void with respect to “the candidate is invited”. As a consequence, we would like to adopt a decision procedure that, among other desiderata, avoids responsibility voids with respect to the decision to reject applicants.

Let us now apply the above distinction to the example from the Introduction. The decision procedure of that example can be represented by CDP Φ_0 (see Fig. 1). Recall from Sect. 6.2 that both $\mathbb{F}_1^{\Phi_0}$ and $\mathbb{F}_0^{\Phi_0}$ have an affine core. By Theorem 4, it follows that the example satisfies both particular and general immunity.

Given this observation, it is natural to ask whether particular immunity *logically entails* general immunity. Our formal analysis can be used to show that it does not.

Fig. 3 A three-dimensional representation of the possible opinions and decision function in Φ_2



To see why, let us reconsider the example Φ_1 , represented in Fig. 2. As noted before (see Sect. 6), $\mathbb{F}_1^{\Phi_1}$ does not have an affine core. In contrast, $\mathbb{F}_0^{\Phi_1}$ does have an affine core, viz. $\{(1, 0, 0), (1, 1, 0), (1, 0, 1), (1, 1, 1)\}$. Consequently, Φ_1 is not vulnerable to responsibility voids with respect to 1. In other words, if the committee of experts decides to adopt the policy for agriculture, then at least one of the group members can be held responsible for that decision. Hence, particular immunity does not entail general immunity:

Observation 10 There exist (binary, majoritarian) CDPs that satisfy particular immunity while not satisfying general immunity.

In light of the previous section on atomism and monotonicity, the above discussion illustrates that CDP Φ_1 shows that *particular* immunity does not entail monotonicity of the underlying decision function. We may ask whether *general* immunity would entail monotonicity. However, CDP Φ_2 can be used to demonstrate the invalidity of this entailment. After all, this CDP satisfies general immunity: notice that $\mathbb{F}_0^{\Phi_2}$ has an affine core, viz. the set $\{(0, 0, 0), (0, 0, 1)\}$, and, therefore, Φ_2 is also immune to responsibility voids with respect to 1. Hence:

Observation 11 There exist (binary, majoritarian) CDPs that involve a non-monotonic and holistic decision function while satisfying general immunity.

8.2 Majoritarian responsibility voids

One can view the problem of responsibility voids as an instance of a more general problem regarding responsibility distributions. One may worry that a given collective decision procedure is open to the possibility that *too little* responsibility can be allocated to the involved individuals. For instance, our notion of responsibility voids concerns cases where the responsibility for a given decision cannot be allocated to *any* individual. More generally, we may consider the problem where the responsibility for a given decision cannot be allocated to *enough* individuals. In this section,

we focus on a particular instance of this more general concern, namely *majoritarian* responsibility voids, and on the question of whether majoritarian CDPs are immune to those majoritarian voids.

To be precise, we say that a group of individuals $G \subseteq N$ is a *majority* if and only if $|G| \geq \frac{|N|+1}{2}$. In a given CDP, a profile of opinions *creates a majoritarian responsibility void* when it brings about a certain decision while the group of individuals that can be held responsible do not constitute a majority group. More formally:

Definition 11 (*Majoritarian responsibility voids*) Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP, and let $d \in D$.

- Where \mathcal{P} is an opinion profile, \mathcal{P} *creates a majoritarian responsibility void with respect to* d if and only if $\Phi(\mathcal{P}) = d$ while the group of all individuals that can be held responsible for d is not a majority.
- Φ is *immune to majoritarian responsibility voids with respect to* d if and only if there are no opinion profiles \mathcal{P} that create a majoritarian responsibility void with respect to d .

We say that a given CDP is *vulnerable to majoritarian responsibility voids* with respect to a given decision if and only if there exists an opinion profile that creates a majoritarian responsibility void with respect to that decision.

Obviously, vulnerability to responsibility voids entails vulnerability to majoritarian responsibility voids. Or, equivalently, immunity to majoritarian responsibility voids entails immunity to responsibility voids. It is, therefore, natural to investigate the converse entailment and ask whether immunity to responsibility voids entails immunity to majoritarian responsibility voids. For the class of majoritarian CDPs, it turns out that this entailment is *valid*.³²

Theorem 6 *Let $\Phi = \langle N, C, D, T, f \rangle$ be a majoritarian CDP and let $d \in D$. Then, Φ is immune to majoritarian responsibility voids with respect to d if and only if Φ is immune to responsibility voids with respect to d .*

9 Conclusion and outlook

Examples of collective decisions such as the familiar discursive dilemma have been central in arguments for the threat of responsibility voids. We have studied the general class of criteria-based decision procedures (which includes the decision procedure of the discursive dilemma) and specified a conception of individual responsibility for collective decisions that are obtained by means of such a procedure. The resulting framework has been used to illustrate the threat of the problem of no hands

³² Theorem 6 can be generalized to arbitrary CDPs, by introducing a notion of m -responsibility voids for any $m \in \{1, \dots, |N|\}$. See Theorem 7 in Appendix C.

for such procedures: there exist criteria-based decision procedures that are vulnerable to responsibility voids. Nevertheless, the procedure of the discursive dilemma has been shown to be immune to responsibility voids (at least, under the assumption that voting untruthfully is eligible).

Our central results provide two conceptually distinct characterizations of immunity to responsibility voids for the class of criteria-based decision procedures. For all these characterizations, the shape of the decision function f is essential; note that the threshold function plays no role in our characterizing conditions (cf. items 2 and 3 of Theorem 3 and item 2 of Theorem 4). We have moreover indicated how our framework and results can be used to investigate a number of topics in the general logic of responsibility voids. In particular, we have proved that any CDP involving a monotonic or atomistic decision function is immune to responsibility voids.

While our formalism is quite general, one could consider further dropping some restrictions in future work. We add four remarks. First, one may ask whether and how our results generalize to a setting in which unanimity thresholds are allowed (i.e. where $T(c) = 1$ or $T(c) = |N|$ for some $c \in C$). Here it should be noted that some of our theorems are no longer valid in this more general setting: the conditions from Theorem 3.3 and Theorem 4.2 are no longer necessary for immunity to responsibility voids.³³ Furthermore, it remains an open question whether those conditions are sufficient for immunity to responsibility voids, and it is unclear whether the condition from Theorem 3.2 is necessary and sufficient for immunity to responsibility voids in this general setting. Lifting the non-unanimity assumption thus seems to invite intricacies that require characterizing conditions other than the ones we spelled out in Theorems 3 and 4.

For a second type of generalization, one could consider dropping our assumption regarding full uncertainty. After all, there will be cases where individuals have at least some clue about the opinions of the others. Our specification of responsibility involves dominance and admissibility. This decision-theoretic principle makes most sense in cases where individuals take all actions of the others into consideration. In other cases, one could consider adopting other decision-theoretic principles to specify a conception of responsibility. Of course, doing so may give rise to other conditions for immunity to responsibility voids.

Third, one could consider dropping the assumption of full eligibility. One could argue that certain opinions are not eligible in some cases. Relaxing this assumption is a delicate affair and may lead to different conditions for responsibility voids.

³³ To see why this is so, consider a procedure $\Phi = \langle N, C, D, T, f \rangle$ with N arbitrary, $C = \{c_1, c_2\}$, $D = \{0, 1\}$, $T(c_1) = T(c_2) = 1$, and $f(O) = 1$ iff $O = (1, 1)$ or $O = (0, 0)$. In this case, both criteria contribute non-strictly and so $\mathbb{R}_0^f = \emptyset$, but given the way the thresholds and the decision function interact, only $(1, 1)$ is maximal for avoiding 0. Indeed, if a voter casts opinion $(1, 1)$, then the group is guaranteed to have that opinion as well and hence to avoid decision 0; the same cannot be said about any of the other opinions. As a result, this procedure is immune to responsibility voids with respect to 0. However, the decision function clearly does not satisfy the characterizing conditions from Theorem 3.3 or Theorem 4.2.

Fourth, another interesting generalization would be to drop the assumption that the criteria have just two possible values. That is, one may consider ranges of values other than $\{0, 1\}$. For example, one could introduce a value for indecision to accommodate cases where individuals can be undecided on some of the criteria. Alternatively, for certain collective decisions one may adopt the Likert scale (typically a five point scale) to express how much one agrees or disagrees with a particular statement.³⁴ In other cases, criteria may even need to be able to take any value from a continuous scale—e.g. if the criterion is specified in terms of the utility a certain agent will receive. Each of these alternatives would require a fundamental change to our present framework, in that there are different ways one might determine the group opinion—depending on the specific interpretation of the criteria and their possible values.

Setting these assumptions and generalizations aside, we would like to emphasize that our framework and results can be the basis for studying the likelihood of responsibility voids. One could investigate the probability that a given collective decision procedure leads to such a void. Of course, this probability will be zero for CDPs that are immune to responsibility voids. As a first step, one could try to approximate this likelihood in a given CDP by assuming that all opinion profiles are equally likely. More generally, one could estimate the likelihood of responsibility voids for the class of CDPs involving a fixed number of individuals (N), criteria (C), or collective decisions (D). In doing so, one could follow up with the question of whether the likelihood of such voids grows or diminishes when the number of agents, criteria, or decisions is increased.

Appendix

In this appendix, we prove all the observations and theorems that are mentioned in the main text. To enhance readability, we hold fixed a given CDP $\Phi = \langle N, C, D, T, f \rangle$ and suppress it in notation whenever possible. We first establish a number of useful lemmas (Section A) and we prove the main results of the paper (Section B). In Section C, we prove the theorems from Sects. 7 and 8.

Appendix A: Some useful lemmas

In what follows, we say that an agent i is *pivotal for criterion c given opinion profile \mathcal{P}* if and only if for every $O \in \mathbb{O}$ it holds that $O_{\mathcal{P}_{i \rightarrow 0}}^\Phi(c) = O(c)$. Intuitively, this means that given \mathcal{P} , i 's opinion regarding c determines whether the group opinion (given threshold assignment T) will settle on value 0 or rather on value 1, for criterion c .

³⁴ A telling example here are the criteria that are used to review papers in the online system EasyChair, with a scale ranging from “−2” to “+2”. On the basis of these criteria one may for instance decide when the paper is accepted or considered for peer review.

Lemma 1 *Each of the following hold:*

1. *If c contributes positively to d , $O(c) = 0$, and $O'(c) = 1$, then $O' \not\leq_d O$.*
2. *If c contributes negatively to d , $O(c) = 1$, and $O'(c) = 0$, then $O' \not\leq_d O$.*

Proof *Ad 1.* Suppose the antecedent holds. By Definition 7.1, there are O_*, O'_* such that $\Delta(O_*, O'_*) = \{c\}$, $O_*(c) = 0$, $O'_*(c) = 1$, $f(O_*) \neq d$, and $f(O'_*) = d$.

Fix an opinion profile \mathcal{P} in which the opinion of the first agent is arbitrary, $T(c) - 1$ other agents choose O_* , and the remaining agents choose O'_* .³⁵ Note that for all $c' \in C \setminus \{c\}$, all agents except from (possibly) the first agent agree. Since $1 < T(c') < |N|$ for all $c' \in C$, this implies that the first agent is only pivotal for c given \mathcal{P} : for all $c' \in C \setminus \{c\}$, $O_{\mathcal{P}_{1 \rightarrow o}}^\Phi(c') = O_{\mathcal{P}_{1 \rightarrow o'}}^\Phi(c') = O_*(c') = O'_*(c')$.

Note also that there are $T(c) - 1$ agents i with $O_{\mathcal{P}_{1 \rightarrow o}}^i(c) = 1$, and there are $T(c)$ agents i with $O_{\mathcal{P}_{1 \rightarrow o'}}^i(c) = 1$. Hence, $O_{\mathcal{P}_{1 \rightarrow o}}^\Phi(c) = 0$ and $O_{\mathcal{P}_{1 \rightarrow o'}}^\Phi(c) = 1$.

From this it follows that $O_{\mathcal{P}_{1 \rightarrow o}}^\Phi = O_*$ and $O_{\mathcal{P}_{1 \rightarrow o'}}^\Phi = O'_*$. Consequently, $\Phi(\mathcal{P}_{1 \rightarrow o}) \neq d$ and $\Phi(\mathcal{P}_{1 \rightarrow o'}) = d$. By Definition 3.1, $O' \not\leq_d O$.

Ad 2. The proof is entirely analogous to the proof for item (1.), relying on Definition 7.2 and swapping the values 0 and 1 throughout. □

Lemma 2 *Each of the following hold:*

1. *If $c \in C_d^+$, $\Delta(O, O') = \{c\}$, and $O'(c) = 1$, then $O <_d O'$.*
2. *If $c \in C_d^-$, $\Delta(O, O') = \{c\}$, and $O'(c) = 0$, then $O <_d O'$.*

Proof *Ad 1.* Suppose the antecedent holds. We first show that $O \leq_d O'$. So let \mathcal{P} be an arbitrary opinion profile. Note that the only difference (if any) between $O_{\mathcal{P}_{1 \rightarrow o}}^\Phi$ and $O_{\mathcal{P}_{1 \rightarrow o'}}^\Phi$ consists in the value they assign to c . If, given \mathcal{P} , the first agent is not pivotal for c , then $O_{\mathcal{P}_{1 \rightarrow o}}^\Phi = O_{\mathcal{P}_{1 \rightarrow o'}}^\Phi$. If the first agent is pivotal for c , then $O_{\mathcal{P}_{1 \rightarrow o}}^\Phi(c) = 0$ while $O_{\mathcal{P}_{1 \rightarrow o'}}^\Phi(c) = 1$. But in that case, since c does not contribute negatively to d and by Definition 7.2, it follows that $\Phi(\mathcal{P}_{1 \rightarrow o'}) \neq d$ implies $\Phi(\mathcal{P}_{1 \rightarrow o}) \neq d$. Since \mathcal{P} was arbitrary, it follows that $O \leq_d O'$.

By the supposition, c contributes positively to d . By Lemma 1.1, $O' \not\leq_d O$. Hence, $O <_d O'$.

Ad 2. The proof is entirely analogous to that for item (2.), switching values 0 and 1 throughout and relying on Definition 7.1 and Lemma 1.2. □

Lemma 3 *If $\Delta(O, O') \subseteq C_d^\sigma$ and $R_d(O) \subset R_d(O')$, then $O <_d O'$.*

³⁵ Note that in this proof, we crucially rely on the property that, for all $c \in C$, $1 < T(c) < |N|$ and hence that $T(c) - 1 \in \{1, \dots, |N| - 2\}$. See also our remark in Sect. 9 concerning unanimity thresholds and the fact that we exclude these.

Proof Suppose the antecedent holds. Fix a repetition-free enumeration $\{c_1, \dots, c_m\}$ of $\Delta(O, O')$. Consider the sequence $\langle O = O_1, O_2, O_3, \dots, O_m = O' \rangle$, where for all $n \in \{1, \dots, m - 1\}$, O_{n+1} is obtained from O_n by switching the value of c_{n+1} . By Lemma 2, $O_1 \prec_d O_2 \prec_d \dots \prec_d O_{m-1} \prec_d O_m$. By the transitivity of \prec_d (cf. Observation 2), $O_1 \prec_d O_m$ and hence $O \prec_d O'$. \square

Lemma 4 *If $\Delta(O, O') \subseteq C_d^\sigma$ and $R_d(O) \subseteq R_d(O')$, then $O \preceq_d O'$.*

Proof Suppose the antecedent holds. If $R_d(O) = R_d(O')$, then $O = O'$ and hence the consequent holds trivially. If $R_d(O) \subset R_d(O')$, then by Lemma 3, $O \preceq_d O'$. \square

Lemma 5 *If $\Delta(O, O') = \{c\}$ for some $c \in C_d^\epsilon$, then $O' \preceq_d O$ and $O \preceq_d O'$.*

Proof Suppose the antecedent holds. Let \mathcal{P} be an arbitrary opinion profile. Note that the only difference between $O_{\mathcal{P}_{1 \rightarrow O}}^\Phi$ and $O_{\mathcal{P}_{1 \rightarrow O'}}^\Phi$ can consist in the value they assign to c . But then, since $c \in C_d^\epsilon$, it follows that $\Phi(\mathcal{P}_{1 \rightarrow O}) = d$ iff $\Phi(\mathcal{P}_{1 \rightarrow O'}) = d$. \square

Lemma 6 *If $\Delta(O, O') \subseteq C_d^\epsilon$, then $O \preceq_d O'$ and $O' \preceq_d O$.*

Proof Analogous to the proof of Lemma 3, but relying on Lemma 5 instead of Lemma 2, and relying on the transitivity of \preceq_d (cf. Observation 2). \square

Lemma 7 *If $\Delta(O, O') \subseteq C_d^\sigma$, $R_d(O) \subseteq R_d(O')$, and $O \in \overline{F}_d$, then $O' \in \overline{F}_d$.*

Proof Analogous to the proof of Lemma 3: construct a sequence of opinions that connect O and O' , each time changing the value for one criterion in $\Delta(O, O')$, and then, relying on the supposition that $R_d(O) \subseteq R_d(O')$, show that if $O \in \overline{F}_d$, then all opinions in this sequence are in \overline{F}_d . \square

Appendix B: Main characterization results

Theorem 1. Let Φ be a CDP, let $O, O' \in \mathbb{O}$, and let $d \in D$. Then,

$$O \preceq_d^\Phi O' \text{ iff (a) } \Delta(O, O') \cap C_{\Phi, d}^\pm = \emptyset \text{ and} \\ \text{(b) } R_d^\Phi(O') \subseteq R_d^\Phi(O).$$

Proof (\Rightarrow) Suppose that (a) is false. Hence, there is some criterion c that contributes both positively and negatively to d , and $O(c) \neq O'(c)$. By Lemma 1, $O \not\preceq_d O'$. Suppose that (b) is false. This can mean only two things. Either there is a $c \in C_d^+$ such that $O'(c) = 0$ and $O(c) = 1$, or there is a $c \in C_d^-$ such that $O'(c) = 1$ and $O(c) = 0$. In either case, by Lemma 1, $O \not\preceq_d O'$.

(\Leftarrow) Suppose that both (a) and (b) hold. Let O_* be such that it agrees with O for all $c \in C \setminus C_d^c$ and with O' for all $c \in C_d^c$. Hence, $\Delta(O, O_*) \subseteq C_d^c$. By Lemma 6, $O \leq_d O_*$ (\dagger). Note that $R_d(O_*) = R_d(O)$ and hence $R_d(O') \subseteq R_d(O_*)$. Moreover, $\Delta(O_*, O') \subseteq C_d^c$. By Lemma 4, $O_* \leq_d O'$. By (\dagger) and the transitivity of \leq_d , $O \leq_d O'$. \square

Theorem 2. Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP and let $d \in D$. Then,

$$\mathbb{R}_d^\Phi = \mathbb{A}_d^\Phi.$$

Proof (\subseteq) Suppose that $O \notin \mathbb{A}_d$. Hence, there is some $O' \in \mathbb{O}$ with $O' <_d O$. By Theorem 1, $\Delta(O, O') \cap C_d^\pm = \emptyset$. Since $O \not\leq_d O'$, it also follows by Theorem 1 that $R_d(O) \subset R_d(O')$. But then $R_d(O) \neq C_d^\sigma$, and hence $O \notin \mathbb{R}_d$.

(\supseteq) Suppose that $O \notin \mathbb{R}_d$. It follows that $R_d(O) \neq C_d^\sigma$. Let $c \in C_d^\sigma \setminus R_d(O)$ and let O' differ from O only in the value of c . It follows that $R_d(O) \subset R_d(O')$ and $\Delta(O, O') \cap C_d^\pm = \emptyset$. By Theorem 1, $O' <_d O$. Hence, $O \notin \mathbb{A}_d$. \square

Theorem 3. Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP and let $d \in D$. Then, the following are equivalent:

1. Φ is immune to responsibility voids with respect to d ;
2. any opinion that maximally avoids d does not support d , i.e. $\mathbb{A}_d^\Phi \subseteq \overline{\mathbb{F}_d^\Phi}$;
3. any opinion that is right for avoiding d does not support d , i.e. $\mathbb{R}_d^\Phi \subseteq \overline{\mathbb{F}_d^\Phi}$.

Proof Note that items 2 and 3 are equivalent in view of Theorem 2. We freely rely on this equivalence in what follows.

(1 \Rightarrow 2) Suppose that $\mathbb{A}_d \not\subseteq \overline{\mathbb{F}_d}$. Fix an $O \in \mathbb{F}_d \cap \mathbb{A}_d$. Let \mathcal{P} be such that, for all $i \in N$, $O_{\mathcal{P}}^i = O$. Then $O_{\mathcal{P}}^\Phi = O$ and hence $\Phi(\mathcal{P}) = d$. However, none of the agents can be held responsible for d , given \mathcal{P} . It follows that Φ is not immune to responsibility voids with respect to d .

(3 \Rightarrow 1) Suppose that Φ is not immune to responsibility voids with respect to d . Hence, there is some \mathcal{P} for which it holds that $\Phi(\mathcal{P}) = d$ and yet, no agent in N can be held responsible for d given \mathcal{P} . This means that, for all agents $i \in N$, $O_{\mathcal{P}}^i \in \mathbb{A}_d$ and hence $O_{\mathcal{P}}^i \in \mathbb{R}_d$. Since R_d is an affine subset of \mathbb{O} , Observation 7 entails that $O_{\mathcal{P}}^\Phi \in \mathbb{R}_d$. Moreover, $f(O_{\mathcal{P}}^\Phi) = \Phi(\mathcal{P}) = d$. It follows that, for some $O \in \mathbb{R}_d$, $f(O) = d$ and, hence, $\mathbb{R}_d \not\subseteq \overline{\mathbb{F}_d}$. \square

Theorem 4. Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP and let $d \in D$. Then, the following are equivalent:

1. Φ is immune to responsibility voids with respect to d ;
2. the set of opinions that do not support d has an affine core, i.e. $\overline{\mathbb{F}_d^\Phi}$ has an affine core.

Proof In what follows, we freely rely on the equivalence between items (1.) and (3.) of Theorem 3.

(\Rightarrow) Suppose that $\mathbb{R}_d \subseteq \overline{\mathbb{F}_d}$. We show that \mathbb{R}_d is an affine core of $\overline{\mathbb{F}_d}$. Notice that \mathbb{R}_d is affine. It thus suffices to prove that \mathbb{R}_d satisfies item (b) in Definition 10.1.

Take an arbitrary $O \in \overline{\mathbb{F}_d}$. Let O' be given by

$$O'(c) = \begin{cases} O'(c) = \rho_d(c) & \text{if } c \in C_d^\sigma \\ O(c) & \text{otherwise.} \end{cases}$$

Notice that $O' \in \mathbb{R}_d$, that $R_d(O) \subseteq R_d(O')$ and that $\Delta(O, O') \subseteq C_d^\sigma$. We need to prove that $[O, O'] \subseteq \overline{\mathbb{F}_d}$.

Let $O'' \in [O, O']$. It follows that $\Delta(O, O'') \subseteq \Delta(O, O') \subseteq C_d^\sigma$ and $R_d(O) \subseteq R_d(O'')$. By Lemma 7, $O'' \in \overline{\mathbb{F}_d}$, as desired.

(\Leftarrow) Assume that \mathbb{K} is an affine core of $\overline{\mathbb{F}_d}$. We need to show that $\mathbb{R}_d \subseteq \overline{\mathbb{F}_d}$. Let $C_K \subseteq C$ and $g_K : C_K \rightarrow \{0, 1\}$ be such that $\mathbb{K} = \{O \mid \text{for every } c \in C_K \text{ it holds that } O(c) = g_K(c)\}$. Let us define, for any $O \in \mathbb{O}$, the set $K(O) \subseteq C_K$ as the set of criteria in C_K for which O assigns the same value as g_K . Formally, $K(O) = \{c \in C_K \mid O(c) = g_K(c)\}$.

First, we show that that (\dagger) for all $O, O' \in \mathbb{O}$: if $\Delta(O, O') \subseteq K(O)$ and $O' \in \overline{\mathbb{F}_d}$, then $O \in \overline{\mathbb{F}_d}$. Let $O, O' \in \mathbb{O}$ be such that $\Delta(O, O') \subseteq K(O)$ and $O' \in \overline{\mathbb{F}_d}$. Since \mathbb{K} is an affine core of $\overline{\mathbb{F}_d}$, there exists an $O^* \in \mathbb{K}$ such that $[O', O^*] \subseteq \overline{\mathbb{F}_d}$. We prove that $O \in [O', O^*]$. Let $c \in \Delta(O', O)$. Then, since $\Delta(O', O) \subseteq K(O)$ and $O^* \in \mathbb{K}$, it follows that $c \in K(O) \subseteq C_K = K(O^*)$ and $c \notin K(O')$. Hence, $c \in K(O^*) \setminus K(O') \subseteq \Delta(O', O^*)$. In sum, $O \in [O', O^*]$ and thus $O \in \overline{\mathbb{F}_d}$, as desired.

Second, we show that (\ddagger) $C_K \cap C_d^\pm = \emptyset$, or, equivalently, that $C_K \subseteq C_d^\sigma \cup C_d^\epsilon$. Let $c \in C_K$. Case 1: $g_K(c) = 1$. We show that c does not contribute positively to d . Take any O and O' with $\Delta(O, O') = \{c\}$, $O'(c) = 0$, $O(c) = 1$, and $f(O') \neq d$. Hence, $O' \in \overline{\mathbb{F}_d}$. We need to show that also $O \in \overline{\mathbb{F}_d}$. Notice that $O(c) = g_K(c)$ and, therefore, $\Delta(O, O') = \{c\} \subseteq K(O)$. So by (\dagger), it follows that $O \in \overline{\mathbb{F}_d}$. Case 2: $g_K(c) = 0$. By an analogous argument, we can show that c does not contribute negatively to d . In sum, $C_K \cap C_d^\pm = \emptyset$, as desired.

Third, we show that (\star) for any $c \in C_K \cap C_d^\sigma$ it holds that g_K assigns the right value for avoiding d to c , i.e. $g_K(c) = \rho_d(c)$. Let $c \in C_K \cap C_d^\sigma$. Case 1: $g_K(c) = 1$. In view of Case 1 of the preceding paragraph, this entails that c does not contribute positively to d . Hence, c contributes strictly negatively to d and hence $g_K(c) = \rho_d(c)$. Case 2, i.e. $g_K(c) = 0$ is again proven analogously, relying on Case 2 of the preceding paragraph.

Finally, we prove that $\mathbb{R}_d \subseteq \overline{\mathbb{F}_d}$. Take an arbitrary $O \in \mathbb{R}_d$. We show that $O \in \overline{\mathbb{F}_d}$. Consider O^* given by:

$$O^*(c) := \begin{cases} g_K(c), & \text{if } c \in C_K \\ O(c), & \text{otherwise.} \end{cases}$$

Then, $O^* \in \mathbb{K} \subseteq \overline{\mathbb{F}_d}$ and by (\ddagger), $\Delta(O, O^*) \subseteq C_K \subseteq C_d^\sigma \cup C_d^\epsilon$. By (\star), it follows that $\Delta(O, O^*) \subseteq C_d^\epsilon$. Hence, $f(O) = d$ iff $f(O^*) = d$. Since $f(O^*) \neq d$, it follows that $f(O) \neq d$. We conclude that $O \in \overline{\mathbb{F}_d}$, as desired. \square

Appendix C: Further results

Observation 9. Let f be a decision function and let $d \in D$. Then, f is atomistic with respect to d if and only if f is monotonic with respect to d .

Proof (\Rightarrow) Suppose f is atomistic with respect to d . Let O be some arbitrary member of \mathbb{R}_d . Then for any $O', O'' \in \mathbb{O}$ such that $O'' \in [O', O]$, it holds that $R_d(O') \subseteq R_d(O'') \subseteq R_d(O)$. By Lemma 7, if $f(O') = d$, then also $f(O'') = d$. Hence, f is (O, d) -monotonic and hence f is monotonic with respect to d .

(\Leftarrow) Assume for contradiction that f is monotonic with respect to d but not atomistic with respect to d . Let O be such that f is (O, d) -monotonic. By our assumption, $C_d^\pm \neq \emptyset$. Let $c \in C_d^\pm$. It follows that there must be O_1, O'_1, O_2, O'_2 such that $f(O_1) = f(O'_2) = d$, $f(O'_1) = f(O_2) \neq d$, $\Delta(O_1, O'_1) = \Delta(O_2, O'_2) = \{c\}$, and $O_1(c) = O_2(c)$, $O'_1(c) = O'_2(c)$. Case 1: $O(c) = O_1(c) = O_2(c)$. Then $O_2 \in [O'_2, O]$, and hence $f(O_2) = d$ —a contradiction. Case 2: $O(c) = O'_1(c) = O'_2(c)$. Then $O'_1 \in [O_1, O]$, and hence $f(O'_1) = d$ —again a contradiction. \square

Theorem 5. Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP and let $d \in D$. If f is atomistic with respect to d , then Φ is immune to responsibility voids with respect to d .

Proof Suppose the antecedent holds. Let $O \in \mathbb{O}$ be such that $f(O) \neq d$.³⁶ Let O' be such that it agrees with O on all $c \in C_d^e$, but $R_d(O') = C_d^\sigma$. Note that $O' \in \mathbb{R}_d$ and $R_d(O) \subseteq R_d(O')$. By Lemma 7, $f(O') \neq d$. Moreover, since f is atomistic with respect to d , for all $O'' \in \mathbb{R}_d$, $\Delta(O', O'') \subseteq C_d^e$ and hence, $f(O'') = d$. It follows that $\mathbb{R}_d \subseteq \overline{\mathbb{F}_d}$. By Theorem 3, Φ is immune to responsibility voids with respect to d . \square

In order to generalize Theorem 6 to the setting with thresholds, we first need a more general notion of responsibility voids.

Definition 12 Where $m \in \{1, \dots, |N|\}$: \mathcal{P} creates an m -responsibility void w.r.t. d iff $\Phi(\mathcal{P}) = d$ and there is a $G \subseteq N$ with $|G| = m$ such that every $i \in G$ can be held responsible for d .

Theorem 7 Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP. Let $d \in D$ and let $m \in \{2, \dots, n - 1\}$ be such that for all $c \in C$, $T(c) \leq m$ and $|N| - T(c) \leq m$. Then Φ is immune to m -responsibility voids w.r.t. d if and only if $\mathbb{R}_d \subseteq \overline{\mathbb{F}_d}$.

Proof (\Rightarrow) Suppose $\mathbb{R}_d \subseteq \mathbb{O}_d$. Consider an arbitrary \mathcal{P} such that, for some $G \subseteq N$ with $|G| = m$, for all $i \in G$, O_p^i maximally avoids d . By Theorem 3, for all $i \in G$ and all $c \in C_d^\sigma$, $O_p^i(c) = \rho_d(c)$. Since $m \geq T(c)$ and $m \geq |N| - T(c)$, this implies that for all $c \in C_d^\sigma$, $O_p^\Phi(c) = \rho_d(c)$. Hence, $O_p^\Phi \in \mathbb{R}_d$ and hence by the supposition,

³⁶ Note that there is such an O , in view of our supposition that $f : \{0, 1\}^C \rightarrow D$ is surjective.

$\Phi(\mathcal{P}) = f(O_{\mathcal{P}}^{\Phi}) \neq d$. Thus, \mathcal{P} cannot create an m -responsibility void. Since \mathcal{P} was arbitrary, this implies that Φ is immune to m -responsibility voids.

(\Leftarrow) Suppose that Φ is immune to m -responsibility voids w.r.t. d . Then a fortiori, Φ is immune to responsibility voids (simpliciter) w.r.t. d . Hence by Theorem 3, $\mathbb{R}_d \subseteq \overline{\mathbb{F}_d}$. \square

Corollary 2 *Let $\Phi = \langle N, C, D, T, f \rangle$ be a CDP. Let $d \in D$ and let $m \in \{2, \dots, n-1\}$ be such that for all $c \in C$, $T(c) \leq m$ and $|N| - T(c) \leq m$. Then Φ is immune to m -responsibility voids w.r.t. d if and only if Φ is immune to responsibility voids (simpliciter).*

To see how Theorem 6 follows from Corollary 2, it suffices to note that when N is odd and $T(c) = \frac{|N|+1}{2}$, then $m = \frac{|N|+1}{2}$ satisfies the conditions in the antecedent of Corollary 2, and m -responsibility voids are majoritarian responsibility voids.

Acknowledgements We are indebted to Stefan Napel, Franz Dietrich, Olivier Roy, Dominik Klein, and two anonymous reviewers for very useful comments on earlier drafts. We also thank the audiences who attended our talks at the University of Amsterdam, the University of Luxemburg and the University of Bayreuth. Special thanks go to Allard Tamminga for extensive comments on the penultimate draft.

Funding Hein Duijf gratefully acknowledges funding from the ERC-2017-CoG Project SEA, No. 771074 and the ERC-2013-CoG project REINS, No. 616512. Frederik Van De Putte's research was funded by a Marie Skłodowska-Curie Fellowship, No. 795329, by a grant from the Research Foundation—Flanders (FWO-Vlaanderen), No. 12Q1918N, and by a grant from the Dutch Research Council, No. VI.Vidi.191.105.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aristotle (2019) *Nicomachean ethics*, 3rd edn. Hackett Publishing Company, Indianapolis
- Bovens M (1998) *The quest for responsibility: accountability and citizenship in complex organisations*. Theories of Institutional Design. Cambridge University Press, Cambridge
- Braham M, van Hees M (2011) Responsibility voids. *Philos Q* 61(242):6–15
- Braham M, van Hees M (2012) An anatomy of moral responsibility. *Mind* 121(483):601–634
- Chapman B (1998) More easily done than said: rules, reasons and rational social choice. *Oxf J Leg Stud* 18(2):293–329
- Copp D (2006) On the agency of certain collective entities: an argument from normative autonomy. *Midwest Stud Philos* 30(1):194–221
- Courtin S, Laruelle A (2020) Multi-dimensional rules. *Math Soc Sci* 103:1–7
- Dancy J (1993) *Moral reasons*. Blackwell, Oxford
- Dietrich F, List C (2017) What matters and how it matters: a choice-theoretic representation of moral theories. *Philos Rev* 126(4):421–479

- Dokow E, Holzman R (2009) Aggregation of binary evaluations for truth-functional agendas. *Soc Choice Welf* 32:221–241
- Frankfurt HG (1969) Alternate possibilities and moral responsibility. *J Philos* 66(23):829–839
- Hetherington S (2003) Alternate possibilities and avoidable moral responsibility. *Am Philos Q* 40(3):229–239
- Hood C (2011) *The blame game: spin, bureaucracy, and self-preservation in government*. Princeton University Press, Princeton
- Horty JF (2001) *Agency and deontic logic*. Oxford University Press, New York
- Kornhauser LA (1992) Modeling collegial courts. II. Legal doctrine. *J Law Econ Organ* 8(3):441–470
- List C, Pettit P (2011) *Group agency: the possibility, design, and status of corporate agents*. Oxford University Press, Oxford
- McKenna MS (1997) Alternative possibilities and the failure of the counterexample strategy. *J Soc Philos* 28(3):71–85
- Nehring K, Puppe C (2008) Consistent judgement aggregation: the truth-functional case. *Soc Choice Welf* 31:41–57
- Nehring K, Puppe C (2010) Justifiable group choice. *J Econ Theory* 145(2):583–602
- Otsuka M (1998) Incompatibilism and the avoidability of blame. *Ethics* 108(4):685–701
- Pettit P (2001) Deliberative democracy and the discursive dilemma. *Philos Issues* 11:268–299
- Pettit P (2007) Responsibility incorporated. *Ethics* 117(2):171–201
- Rae DW, Daut H (1976) The Ostrogorski paradox: a peculiarity of compound majority decision. *Eur J Polit Res* 4(4):391–398
- Ridge M, McKeever S (2016) Moral particularism and moral generalism. In: Zalta EN (ed) *The Stanford encyclopedia of philosophy*, Winter 2016 edn
- Savage LJ (1972) *The foundations of statistics*, 2nd edn. Dover Publications, New York
- Thompson DF (1980) Moral responsibility of public officials: the problem of many hands. *Am Polit Sci Rev* 74(4):905–916
- Vallentyne P (2008) Brute luck and responsibility. *Polit Philos Econ* 7(1):57–80
- Van Inwagen P (1999) Moral responsibility, determinism, and the ability to do otherwise. *J Ethics* 3(4):341–350
- Weaver RK (1986) The politics of blame avoidance. *J Publ Policy* 6(4):371–398
- Wyma KD (1997) Moral responsibility and leeway for action. *Am Philos Q* 34(1):57–70

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.