

VOIDS OR FRAGMENTATION:
MORAL RESPONSIBILITY FOR COLLECTIVE OUTCOMES*
[Forthcoming in the *Economic Journal*]

Matthew Braham and Martin van Hees

Abstract

Institutional rules create difficulties for the allocation of moral responsibility. One problem is the existence of *responsibility voids*, i.e. situations in which an outcome results from individual interactions but for which no one is responsible. Another is that responsibility can be *fragmented* in the sense that responsibility-bearing individuals may be responsible for different features of the outcome. This study examines both problems together. We show that for a large class of situations the two problems are logically dependent. More precisely, non-dictatorial decision procedures can only ensure the absence of voids if they allow for the fragmentation of responsibility.

1. The Question

Responsibility attributions are a very elementary means by which we regulate our social, political, and economic affairs. The root idea of holding a person responsible is to declare that the person is a legitimate target for adverse or beneficial treatment, signalling that she should modify or maintain her conduct. As a matter of justice it is quite obvious that we need to get our ‘responsibility system’ right. Simply put: a distorted allocation of responsibility will not only be considered unfair but can easily be seen as sending the wrong signal about the appropriateness of behaviour.

Getting a responsibility allocation right is, however, anything but straightforward. Broadly speaking, there are two main problems to be solved. The first is the choice of a first-order normative system, which is the set of principles that yield the general conditions for assigning moral responsibility. These conditions define the

*Corresponding author: Martin van Hees, Department of Philosophy, VU Amsterdam, de Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Email: m.van.hees@vu.nl

We thank an anonymous referee, the Editor, Itai Sher, and the participants at the ‘Normative Ethics and Welfare Economics Conference’ held at the Becker Friedman Institute, University of Chicago, 24–25 October 2014, the ‘Responsibility and Causation Conference’ held in 2014 in Strasbourg, the 2013 Mancept Workshop, as well as seminar audiences at the Universities of Amsterdam, Rotterdam, Lund, Saarbrücken, York, and the London School of Economics and Political Science.

grounds of responsibility: who is to be held to account and for what. Determination of these conditions has been the focus of most of the philosophical analysis of moral and legal responsibility. The second problem has received less attention. It concerns the way in which institutional structures affect responsibility assignments. Here the question to be answered is, given an account of the grounds of responsibility, how does the context of interaction – the second-order normative system – determine the resulting responsibility assignment.

It is important to clearly demarcate these two parts of a responsibility assignment. The grounds of responsibility refer to general conditions that are defined independently from particular circumstances. The grounds should be stable across both time and a wide variety of conditions. Yet the verdicts that the grounds return will vary with the structure of interaction. For instance, a committee may deliver different outcomes depending on the decision rule in use, even if the opinions or votes of the individual committee members remain stable. This is because different decision rules aggregate the individual inputs in different ways and thus the resulting responsibility assignments may also differ. The precise question we address in this paper is this: Given a first-order normative system of the canonical ‘folk ethics’ principles of moral responsibility, do second-order systems exist that allocate moral responsibility in an ideal way?

This question needs to be acuminated further. When your conduct becomes interwoven with others, two fundamental problems will arise from the fact that the consequences will now depend not only on what you have done and the information and options you have, but also on actions and information of others. The first problem refers to what we have previously coined as *responsibility voids* (Braham and van Hees 2011). The other is what we denote here as the *fragmentation of responsibility*. A responsibility void is a situation in which a combination of actions by different individuals leads to an outcome for which none of the individuals involved can be said to be responsible. Fragmented responsibility occurs when a combination of actions by different individuals leads to an outcome for which at least some of the responsible individuals are responsible for different features of the outcome. In a rough sense, both these issues can be considered as different facets of Dennis Thompson’s ‘problem of many hands’: ‘[B]ecause 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’ (Thompson 1980: 905).¹

To deliver a first answer to our question we examine in this paper the existence of forms of strategic interaction that are immune to both responsibility voids and fragmented responsibility. We demonstrate that under a set of very general conditions voids and fragmentation are logically dependent. More concretely, we show that institutional safeguards against responsibility voids have important implications for the fragmentation of responsibility (how responsibility is differentially allocated). The result is, therefore, to be understood in terms of mechanism design. We investigate the possibility of constructing modes of strategic interaction that generate

¹See also Bovens (1998) for a detailed institutional study of the problem.

assignments of responsibility that respect certain properties.

There are three methodological features of our analysis to be flagged at the outset. Firstly, by a decision-making mechanism we mean the institutional rules that combine individual actions into a collective decision. We will describe these rules in an abstract game-theoretical framework. An institutional setting will be described by a *game form*. We will also focus on classes of game forms rather than any particular one, permitting us to hone in on the structural aspects of such a mechanism in which responsibility voids or fragmentation occur. Furthermore, we understand these institutional mechanisms in a very general way: they can be centralised or decentralised aggregations of choices; they can be a single committee vote or a chain of individual decisions or committee decisions. Our model places no restrictions on how a multiplicity of individuals interact with each other.

Secondly, our analysis is *individualistic* in two senses. The first concerns the bearers of moral responsibility, which we take to be concrete individuals, and not collections of individuals even though we are dealing with multi-agent interactions (the responsibility of such collections will not be the focus of our attention here).² This is individualistic only in a weak sense because we only assume that the concept of individual moral responsibility does not *derive* from the concept of collective moral responsibility.³ The second sense of individualistic is that when we determine the allocation of individual responsibility, no independent weight will be given to considerations outside the conduct of individuals. That is, a responsibility allocation is to be based solely on the way the mechanisms of social interaction combine individual conduct into outcomes.

Thirdly, our study is an idealised one. The object of our analysis is the structure of a decision-making mechanism and we abstract away from many of the things that characterise actual human beings and institutions, such as their mutual attractions, repulsions, and influence as well as formal roles, offices, and institutionalised patterns of behaviour. We assume, however, that the agents are autonomous in a minimal sense and that they possess the relevant knowledge about the decision situation at hand. Additionally, we restrict ourselves to what is best termed as ‘normal’ decision-making. This means we will bracket out the consequences of ‘unreasonable’ (or ‘ineligible’) actions, the standard of which we leave as a primitive.⁴

We start our analysis in section 2 by laying out the principles of what we call the ‘folk ethics’ conception of moral responsibility for outcomes. In section 3 we define and discuss what we mean by responsibility voids and fragmentation and set out three fundamental properties for responsibility allocations for decision-making mechanisms. Our main result and a discussion of its implications follows in section 4. In sections 5 to 7 we interrogate the framework and the result. Section 8 is a

²See List and Pettit (2011) for a recent contribution on the responsibility of collections of individuals.

³See for instance the classic expositions in Lewis (1948) and Hart (2008). See also the discussions in Sverdlik (1987) and Mellema (1997).

⁴These actions may include maiming oneself or ending one’s life. For a classic discussion on unreasonable actions, see Benn and Weinstein (1971).

conclusion.

Before taking up our analysis we would like to remark that the chief contribution of this paper is based on a formal result. In order to communicate the result to the widest possible audience we have relegated the entire formal framework and the proofs to an appendix. No special technical knowledge is presupposed in the main body of the text.

2. The Grounds of Moral Responsibility

The concept of responsibility is used in many different ways. The way in which we will use it here, and which we believe to be the relevant one for studying institutionalised interactions, is in terms of individual *accountability*. This conception singles out how our conduct as concrete individuals is related to a given outcome.⁵ Deeper factors such as our identity, attitudes, values, and community relations play no explicit role in this notion. Accountability refers to what we have done as a concrete individual and not for what we are or what we represent. Furthermore, we take the scope of accountability to be limited to the consequences (the outcomes) of our actions.

There are various conceptions of this notion of responsibility. The one that we will rely on here is based on a regimentation of the so-called ‘folk ethics’ version of responsibility that we have developed elsewhere.⁶ The nub of this conception says that a person can be ascribed responsibility for a given outcome, or some aspect of it, if three conditions are satisfied. The first concerns *agency*: to be an apt object of a responsibility attribution an individual must be minimally autonomous. There are different views about the nature of the required autonomy, but usually it means that the individual is able to act intentionally, to plan, and to distinguish right and wrong and good and bad. The second condition concerns *causal relevance*. This says that there should be a causal relation between the conduct of the agent and the resultant outcome. The third condition is what we call the *avoidance condition*: it concerns a person’s opportunity to do otherwise. This refers to the assumption that a person should have had the possibility to avoid causally contributing to the outcome and that at least some of those alternative courses of action can be said to be reasonable ones.⁷

If any of these conditions are not satisfied it means that we are not responsible for the consequences of our actions. You are not responsible for the mess you made when you were sleepwalking (lack of agency). You are not responsible for the death of President Kennedy if you were not involved in any way in his assassination (no causal contribution). And as a bank teller you are not responsible for the bank robbery if you could not have stopped it (no alternative) or only by seriously risking

⁵See the discussion in Watson (2004: 260ff). See also Cane (2002).

⁶See Braham and van Hees (2012). This conception is also known as the ‘tripartite conception’, see Feinberg (1970: 196).

⁷The condition is motivated by the views developed by McKenna (1997), Wyma (1997), Otsuka (1998), and Hetherington (2003). A very recent expression is Sartorio (2012). These views are, however, not undisputed. For a critique, see Fischer (1999).

your life (no reasonable alternative).

These conditions need to be fleshed out because there are many ways in which we can say that a person has agency, has made a causal contribution, or has had a reasonable opportunity to avoid a causal contribution. Setting aside agency completely (we assume that our agents satisfy the relevant conditions), our criterion for a causal contribution is a weak form of counterfactual dependence known as the NESS-test.⁸ This test denotes c as a causal condition for e if there is an event that is sufficient for e such that: (i) c is a member of the set; (ii) all elements of the event obtain; (iii) c is necessary for the sufficiency of the event. In short, c is a causal condition for e if c is a *necessary element* of a sufficient set of conditions for e ; or: c is a cause of e if e depends on c under *some* circumstances that were present on the occasion.

The avoidance condition has a number of features that require elaboration.⁹ Firstly, this is a refinement of the principle that ought implies can. If we are responsible for an outcome given that we have performed some action with a causal connection to that outcome, then an alternative action must have been available. Secondly, the condition targets alternative courses of action, not alternative outcomes. It is about doing your best to prevent an outcome by choosing a strategy that reduces the probability that your action is a NESS condition for an outcome; it is not about you having full control over that outcome. Thirdly, we only focus on the part of the condition that stipulates that there should be some alternative. We set aside the requirement that the alternative should be reasonable by simply assuming that all alternatives available to an agent are reasonable.¹⁰

Having set out the requirements for saying that someone bears moral responsibility, we have to explicate the objects for which she is or is not responsible. This is the scope of responsibility. Our target here is *outcomes*. But what does this mean? At first blush, it is a situation that has resulted from some prior events such as voting, shooting, or traveling. So, for example, we may say that outcomes are ‘Ben was offered the job in the economics department’, ‘Jones was killed’, or ‘you are in Paris’. Yet, each of these statements is still imprecise because they only isolate certain features of an outcome. An outcome actually has many other features: Ben was offered the job but Alice and Charlie were not; Jones was killed but Smith got his kidney and his life was thus saved; you are in a cheerful mood whilst climbing the stairs of the Eiffel Tower.

⁸Hart and Honoré (1959) and Wright (1988). The NESS-test is closely related to Mackie’s (1974) notion of an INUS-condition. For an extensive discussion of the use of the INUS/NESS test for very similar contexts as the ones we will discuss here, see Tuck (2008). The reason for choosing the NESS-test, instead of, say, David Lewis’s (1973) account of counterfactual dependence, is that this conception resolves basic quandaries of causal overdetermination. For a discussion, see Braham and van Hees (2009).

⁹We provide an extensive discussion of this criterion in Braham and van Hees (2012).

¹⁰The restriction is mainly for reasons of simplification as our framework can be extended by introducing the distinction between reasonable and unreasonable alternatives. Such an extension would, however, leave our main result untouched. It would only complicate the presentation and proof of it.

We take an outcome to be a possible world (or a social state in the terminology of economics), with its features being described by the sets of possible worlds to which it belongs. We call these sets *states of affairs*. Thus, you being in Paris is a state of affairs described by a set of worlds in which you are in Paris. A possible world can belong to different sets of possible worlds, and it is for this reason that an outcome can be seen as a combination of features. Hence, a particular outcome may not only belong to the set of outcomes of being in Paris, but also to the one in which you are in a cheerful mood, climb the Eiffel Tower, etc. Note that not all of the features of an outcome will be morally relevant. For example, the fact that Ben wears glasses or is bald might not be a morally relevant feature of the outcome of hiring him; the fact that he is a man and that his appointment maintains an historical gender imbalance might be. For our purposes we will simply assume that any set of possible worlds to which an outcome belongs can be relevant (we elaborate on this in Section 5).

3. Three Properties

The focus of our study is the allocation of responsibility for outcomes that are brought about by multi-agent interactions. We call these ‘collective outcomes’ for short (there is no requirement that the collection of agents form a ‘collectivity’ in any sense of joint, common, or organic agency). This focus means that we want to examine decision-making mechanisms in which no single agent has all the decision-making power. We call this a non-dictatorship property.

Denote a *dictator* as an agent who can force any possible outcome in the decision problem, whether it be a decision on who is appointed to a job, that your company will join a cartel, or that the colour of your kitchen walls will be pink. If you are a dictator you can adopt a course of action that will lead to your chosen outcome irrespective of what others do.

PROPERTY 1 (NON-DICTATORSHIP). *There is no dictator.*

Our second property concerns responsibility voids, which are outcomes in which none of the individuals are responsible for a given state of affairs. Pettit (2001, 2007) suggested this possibility in his discussion of the so-called ‘discursive dilemma’. This is the situation in which a committee has to make a judgement about a set of interconnected questions. A yes–no vote is taken on each of the questions. The combination of the majority judgements subsequently yields a decision for which the committee members are unanimously opposed even though there is a majority in favour for each of the questions. If the final decision is made on these majorities then we can have a situation in which despite unanimous opposition to the final verdict in favour it will still go through. And since this opposition is unanimous, Pettit claims that none of the individuals can be held responsible. There is apparently a void. (We return to this case in Section 6.)

For an example of voids outside the setting of committees, consider a classic model of a pure coordination game. Two friends have an appointment to meet each other but they forgot to discuss the meeting place. There are two equally likely

venues, Grand Central Station and Café Central, and they must choose to go to one or the other without knowing what the other will choose. As it happened, they made different choices and failed to meet. Now, while they both clearly are responsible for the particular outcome of having gone to Grand Central Station or Café Central, neither of them is morally responsible for the fact that they did not meet each other: they did not have a strategy with a smaller probability of leading to that state of affairs.

Ideally, a responsibility system should not have voids for the very reason that the purpose of such a system is to aid our self-governance. It is like a ‘loophole in our morality’.¹¹ We formulate a property to this effect. We say that a mechanism is *complete* if it can never lead to a responsibility void. This means that for any state of affairs there is at least one person who is morally responsible for its occurrence.

PROPERTY 2 (COMPLETENESS). *The decision-making mechanism has no voids.*

Our third property is novel. It airs and gives explicit formal substance to an implicit thought that can be found in discussions about the allocation of moral responsibility for collective outcomes. The property concerns what we have already denoted as the *fragmentation* of responsibility. Fragmentation and its converse, what we call *uniformity*, is a subtle and complex property that can be interpreted in various ways. For that reason we will first state the property and then examine it more fully in the sections that follow as part of the discussion of our main result.

What, precisely, is a fragmented allocation of responsibility? On our account, responsibility should only be attributed to you for those aspects of a given outcome in which you were involved. These are features of the outcome for which your conduct made a causal contribution. Your causal involvement in the realisation of some particular feature of the outcome need not, however, put you in the dock or on the pedestal for *all* of its features. To see this, consider the hiring committee of an economics department that appointed Ben rather than one of two other candidates, Alice or Charlie. Assume that Ben and Alice are social choice theorists, whereas Charlie is an econometrician. Suppose the committee used a plurality rule with a tie-breaker and as it turned out there was a tie among all three candidates and that the chairman’s vote, which was for Ben, broke the tie. According to our account, the committee member who voted for Alice is responsible for the hiring of a social choice theorist. This is easily checked. The vote for Alice contributed to the tie and thus also to the election of Ben, who is a social choice theorist. Hence, the requirement of causal contribution in Ben’s appointment is met. Furthermore, the person could have voted for Charlie, a vote that had a lower probability of leading to the appointment of a social choice theorist. Clearly, this committee member is *not* responsible for the fact that a male candidate is hired since the vote for Alice minimised that probability. On the other hand, the committee member who voted for Charlie *is* responsible for a man being appointed, although *not* for the appointment being in social choice theory, given that Charlie is an econometrician.

¹¹For a discussion of loopholes in morality, see Pogge (1992).

Now, despite all being part of a common and institutionalised procedure and contributing to a common outcome – Ben being hired – we see that the committee members are in fact responsible for different features of the outcome. Only the chairperson is responsible for *all* the features. At first glance this seems reasonable as it is an exact match with the intuition behind the folk ethics conception: each one of us is responsible for the consequences that flow from our own conduct and not the consequences that flow from the conduct of others (insofar as we do not have special vicarious relationships to others). Thus, this allocation of responsibility can be defended on the grounds of the submitted votes (voting for different candidates yields different responsibility) or the different roles of the committee members. For instance, with respect to the chairman’s responsibility, it seems plausible that the extent of responsibility should track additional decision-making power. One is reminded of Spiderman’s rendition of an insight supposedly first formulated by Voltaire: ‘With great power comes great responsibility.’

Yet, there are cases and contexts in which we may want to avoid such an allocation (especially when it is clear to the agents involved that the outcome is the result of the interaction of individual decisions). That is, we want to structure the institutional environment and modes of interaction so as to *guarantee* responsibility will not be fragmented. We shall discuss these reasons later, but it is useful to note that there are at least three normative reasons. Firstly, uniformity makes the responsibility allocation more transparent. Secondly, uniformity may capture an ideal of procedural fairness. And thirdly, the fragmentation of responsibility is associated with another and ‘deeper’ and undesirable form of the fragmentation of our value system.

So, given our account of responsibility does there exist a decision-making mechanism such that the following property is satisfied?

PROPERTY 3 (UNIFORMITY). *A decision-making mechanism never displays any fragmentation of responsibility: all individuals that are responsible for some feature of the outcome are responsible for the very same features.*

As a point of clarification, while we will refer to the properties as conditions, we are using them here primarily in a descriptive sense. That is, we do not claim that every mechanism should possess these three features. Rather we consider it useful to know which decision-making mechanisms possess them as it will impart at a very general level information about the allocation of responsibility that we can expect.

4. The Theorem

The three conditions and our account of moral responsibility can be formulated in a game theoretic framework. This permits us to capture multi-agent interactions in a very general way (see Appendix). In this framework we call the decision-making mechanisms we are examining *responsibility games*. The following theorem holds:

THEOREM *If a responsibility game is uniform and complete, then it is dictato-*

rial.¹²

The theorem says that there is a logical interdependence between our three conditions. We arrive, so to speak, at possibilities and impossibilities. It is possible to have uniform and complete decision-making mechanisms, but this comes at the cost of concentrating the decisions in the hand of a single agent – a dictator – who has complete and full responsibility. Alternatively, if we have, or are required to have, a non-degenerate multi-agent decision-making mechanism, then it is impossible to simultaneously satisfy completeness and uniformity. In other words, some fragmentation or some voids are built into the warp and woof of multi-agent decision-making. There appears to be no perfect responsibility system for the folk ethics account of responsibility.

5. Technicalities

Is there a way out? Can we drop any of the conditions? Before examining each in turn we need to dispel possible technical objections to the analytical framework as such. One possibility we have already referred to is that the source of the trouble lies with our methodological individualism. Instead, what is required for determining responsibility in multi-agent interactions is another method altogether. Rather than beginning with the individual and her actions, we should begin with groups or collectives and examine the causal contribution that such ‘collective agents’ make to the outcome.

On this view, responsibility voids may disappear when we allow for the assignment of responsibility to collective agents. In other words, although it may be the case that none of the individuals are responsible for some feature of the outcome, some group of individuals that possess the characteristics of collective agency might said to be so. For instance, in a coordination game it is obviously true that the coalition of all players can realise any outcome. As explained above, we do not pursue this route here since our project is to interrogate an individualistic approach to responsibility.¹³ One interpretation of the theorem could be that it highlights the boundaries of methodological individualism for the assignment of responsibility.

A second objection concerns the assumption that any (non-empty and proper) subset of the set of outcomes is a relevant state of affairs. This is a rather demanding assumption indeed. The number of states of affairs rapidly becomes very large if the number of outcomes increases, and we will not be interested in many of them. We have already mentioned that some of the features of an outcome are irrelevant (e.g. the baldness of a candidate or the fact that the number of items in his tax return is a prime). If some of the states of affairs with respect to which voids arise only concern such irrelevant features, then there is no reason for concern. We do believe this is a legitimate concern. Yet it should be pointed out that the force of the argument is

¹²The theorem is a generalization of Proposition 3 in van Hees (2010).

¹³For a defence of the view that the coordination failure is to be solved at the level of the group, see Bacharach (2006). For a more general discussion of group agency and responsibility see Pettit (2007); List and Pettit (2011).

weakened by the fact that the proof of our result does not require the entire set of features. It is, therefore, unlikely that the voids only arise in unproblematic cases. Establishing exactly which features yield the impossibility result is a natural next step, but is beyond the scope of this paper.

Finally, one could object to the way the scope of the result is defined. We have shown that for *any* game form and *any* probability distribution, there is *some* play that yields either a void or fragmentation of responsibility. Yet we do not make any assumptions about the relation between the probabilities and the play in question. This entails that the impossibility may be hypothetical only. The general result merely states that there is *some* play of this responsibility game in which there is a void or fragmentation. To illustrate with an abstract example, consider the following three-player, non-dictatorial game form with three outcomes x , y and z . According to the rules of the game, a single choice for x is sufficient for x to be the outcome; but if no player chooses x , then a single choice for y is sufficient for y to be the outcome; z will be the outcome only if all individuals choose z . Now, consider the scenario in which all players attach a probability of 1 to all of the others choosing z and everyone does indeed choose z . This play of the game does not exhibit voids or fragmentation. To see this, note that all players are responsible for $\{z\}$ and $\{y, z\}$ because a choice for x had a larger avoidance potential for those sets than choosing z , and all are responsible for $\{x, z\}$ because voting for y had a larger avoidance potential for it than a choice for z . Hence, uniformity and completeness are both satisfied. In contrast, uniformity is violated in the very same game form if, despite the beliefs that each player has that the others will choose z , one player did not vote for z but for x . This generates fragmented responsibility: in this play, the player who chose x is responsible for all features of the outcome (i.e. for $\{x\}$, $\{x, y\}$ and $\{x, z\}$), whereas all the others are only responsible for the fact that y is rejected (i.e. for $\{x, z\}$).

6. Dictators or Voids?

Permitting dictatorial rules will of course give us an unambiguous allocation of responsibility. But this option can be dismissed with relative ease. First and foremost there is an obvious practical issue: it is impossible to broadly implement it in our social life, much of which consists of decentralised multi-agent decisions of the kind that our coordination game mentioned above exemplifies. Secondly, for those circumstances for which we can design our interactions, a single authoritative decision-maker is also unattractive. One cluster of normative problems concerns the adverse consequences of concentrations of power. We lose, for instance, the epistemic and representative advantages that ‘democratic’ participation brings (which implies voter choices can be effective). A second cluster of problems refers to the intrinsic disvalue of such a concentration of power. One significant source of disvalue is that having a single decision-maker runs counter to another form of responsibility that is important in our social life and self-governance, that of ‘taking

responsibility'.¹⁴

Can we drop completeness? Such a strategy certainly has morally disturbing implications. It is tantamount to saying that it is permissible to have institutions that can create outcomes in which no one will be responsible. One need only think of major corporate harms such as the Union Carbide India gas leak in Bhopal, the sinking of the Townsend Thoresen ferry Herald of Free Enterprise, or now the Volkswagen emissions scandal, to see that a responsibility void would let culpable managers off the hook.

One tack to drop completeness is to play down the significance of voids in terms of the likelihood of their occurrence. This is similar to the way in which social choice theorists play down the likelihood of voting paradoxes. We could then conclude that voids are a very special kind of phenomena. This is something that we have in fact previously suggested (Braham and van Hees 2011). We demonstrated that in the specific setting of institutionalised committee decision-making voids are in fact unlikely. First, one significant kind of void discussed in the literature has already been excluded by our analysis. This is the void that can occur in the so-called discursive dilemma that Pettit refers to and that we mentioned above. In our account, for an individual to be held as morally responsible she must have had a reasonable opportunity to have minimised her causal contribution to the outcome. Now, a void emerges in Pettit's analysis, and not in ours, because he presupposes that acting strategically is not a reasonable (or eligible) alternative for the committee members. In our analysis, the void would not have arisen if the members had voted 'non-truthfully' or better put, 'strategically' (not in accordance with what they really believed to be true). In our view, voting in this way is indeed a reasonable alternative. Hence, we call this a *normative void* because it occurs due to a purported moral impermissibility of voting in this way. Adhering to this moral rule is what generates the void. If, on the other hand, strategic behaviour of this kind in the discursive dilemma is morally permissible, and one can argue that it is, such a void will not arise.

We also have discussed a second class of voids, which we have labelled as *epistemic voids*. These, as the label suggests, occur due to the factual beliefs of the individuals involved. In particular, they arise in a circumstance in which each individual has a false but justified belief about the situation at hand. The belief leads them to incorrectly assign zero probability to alternative contingencies. Since their beliefs are justified, however, the relevant avoidance potential is zero and we do not hold the agents responsible. We argued, however, that the exact conditions yielding such epistemic voids are rather peculiar and thus we need not worry about them either.

The problem here is that these arguments about normative and epistemic voids are not very convincing outside the institutional setting of committee decision-

¹⁴This argument could be developed along Kantian lines. Alternatively, it could emphasise the importance of character traits that are necessary for human development and flourishing. In the latter case, the touchstone text is Mill's well known discussion of the problems of dictators in his *Considerations on Representative Government* (2008/1859). See also Dewey (1960).

making (and it was exactly that kind of decision-making that we were addressing). One example of a void was given by the coordination game in the previous section. The example might suggest that such voids should be equated with blameless harms. It is unfortunate that the two friends were unable to meet each other, but coordination failures are a fact of social life and that here, too, we need not be overly concerned about them. But it is not straightforward that most coordination games do belong to this moral category.¹⁵ Bowles (2004: 23–24), for instance, recounts a real-life share-cropping example from India which matches the well-known Stag Hunt game, and argues that it is one of the sources of poverty. A discussion about the frequency with which morally non-trivial coordination problems occur is beyond the scope of this paper. But even if they are uncommon, there are instrumental reasons for wanting to ensure that they cannot arise: if the outcome is a serious harm, a complete (no voids) decision-making mechanism guarantees that the decision-makers will have reasons to act in such a way as to prevent the harm occurring again.

We conclude that we do not want to give up the non-dictatorship condition. Most institutional structures already prevent the emergence of voids and, insofar as they do not, we have reason to ensure that they do. Hence, our result entails that we have to accept the possibility of non-uniform or fragmented allocations of responsibility.

7. The Problem of Fragmentation

Should we be worried about the possibility of fragmentation? We already suggested that there are various social settings in which we are not only prepared to accept fragmentation, but in which it is in fact a desirable property. These are the cases in which collective outcomes are the result of decentralised or even random interactions. Here, indeed, we may want our responsibility allocation to reflect the differences among the contributions of the members. Yet, the fact that we are willing to accept fragmentation in *some* or even *many* institutional settings does not yield the conclusion that we should *always* be willing to do so. One need only think of formal institutional settings (company, school, department) in which different individuals have the same roles and tasks that are all tied together by an internal decision-structure. Here it is uniformity and not fragmentation that is the desideratum.

The problem of fragmented responsibility is that it lacks transparency – it is opaque – for those seeking to hold individuals responsible for some collective outcome. That is, if any two individuals are said to be morally responsible for some outcome and owe compensation or are due punishment, then uniformity means that they will do so in exactly the same measure. Uniform allocations of responsibility circumvents the need for contestable judgements about how much more responsibility one individual bears in comparison to another for an outcome that they jointly authored. This must be understood as more than just a pragmatic consideration.

¹⁵See Braham and van Hees (2014) for a recent example of the normative problems that such games generate.

Consider once more the case of the hiring committee: is it really true that the committee member who voted for Charlie is less responsible for the outcome in which Ben is appointed than the chairperson? To determine this would mean that we have to weigh up the moral importance of different features and each allocation could easily be contested by those wanting to shirk any blame. If a system is uniform, then no such problem would occur. Note that the condition of uniformity is not an attribution rule: it does not pin one person's wrongdoing on another. Rather, to say that a decision-making mechanism is uniform simply describes how it allocates responsibility, given the inputs. Another way of expressing the importance of the property is to say that despite the individual features of your conduct, if two individuals did something of equal moral significance, then each are equally responsible. That is to say, morality sometimes requires that a responsibility allocation should be neutral with respect to the individual features of conduct.

A second reason for desiring a uniform allocation of responsibility relates to equality of influence. This is an ideal that has wide scope, from decision-making in small organisations (firms, a university) to large ones (a national polity). We can express this ideal in terms of a norm of procedural justice. Why should individuals who are tied together via a decision-making mechanism and have an *ex ante* equal opportunity to influence the outcome of a collective undertaking be held differentially responsible? One view to take is that *ex ante* equality of influence should translate into *ex post* equality of responsibility.

How exactly this aspect of 'decisional' or political equality should be implemented is open for debate. We can go from a simple 'one person, one vote' to a more elaborate account of the way in which individual choices contribute to collective outcomes. Alternatively, we can take an *ex post* or retrospective view and focus on the actual impact that people have had on an outcome. Here responsibility enters the equation because it tracks this impact. To say that you are responsible for a feature of the outcome is to say that you were influential in bringing it about. Uniformity aligns this impact with the assignment of responsibility and guarantees the equality of this impact. Stated differently, uniformity is an ideal of equality. It does not express that our institutions guarantee that we have equal opportunity for influence, but that any influence is always exercised equally. Of course, this is a radical ideal. In fact, our result shows that it may indeed be *too* radical because in the absence of voids, such equality cannot be guaranteed.

Finally, we need to take into account that fragmentation may be symptomatic of a problem at the heart of the ethics of voting. The issue at hand is that the fragmentation of responsibility can be associated with another form of fragmentation, that between *what we are responsible for* and *the values we hold*. Consider the infamous US presidential elections in 2000 in which Bush won his first term of office. Recall the following schematic account. Ignoring the fringe candidates and abstentions, we can distinguish three possible outcomes: Bush wins, Gore wins, and Nader wins. Now hone in on Florida and New Hampshire, two states won by Bush but in which Nader and Gore together had the majority of the popular vote. For these states, we arrive at a distribution of responsibility in which the Bush voters were responsible

for all features of the outcome (Bush winning, Gore losing, and Nader losing) and the Gore voters for Nader losing but not for Bush winning or Gore losing.¹⁶

Now turn to the Nader voters. They were causally effective for Bush winning but had a strategy – to vote for Gore – which would have decreased their *ex ante* probability of being effective for Bush winning. They were thus responsible for Bush carrying their state and, since the New Hampshire and Florida results both formed a NESS-condition, for Bush winning the presidency. It is this responsibility of the Nader voters that we see that fragmentation can be associated with a failure of moral responsibility to track a voter's value system. For most of the Nader voters, the ideological distance from Bush was very large. Their responsibility for Bush becoming president therefore shows there to be a gap between their responsibility and their values.¹⁷ The problem is sharpened if one looks at the alternative course of action that a Nader voter in swing states had. If a Nader wanted to avoid the responsibility for a Bush victory, that voter would have had to vote for Gore and in so doing would not have had the opportunity to let his or her vote fully express their political values. Moreover, had the Nader voter voted in this way, he or she would have made themselves morally responsible for Nader losing – precisely what should not happen in view of the values that voter holds (and assuming of course that this voter is a rational and coherent personality).¹⁸

This is what we take to be the third problem of fragmentation, namely, that accountability may no longer be in song with a person's value system. That is, the example of the US presidential elections highlights a potential and irreconcilable trade-off in democratic systems: that between our moral responsibility for collective outcomes and a core component of our moral life in a democracy, the public expression of our political values. Our theorem says that we can only guarantee that these values run together at the expense of completeness (accepting the possibility of responsibility voids).

¹⁶Note that by the NESS test, *all* voters, including those who voted for Gore, made a causal contribution to Bush winning. The reason is that when determining the causal contribution that a voter makes, one looks for the subset of the actual votes in which an actual vote happens to be critical. The logic of the framework implies that all non-Bush voters could have formed a coalition for Nader and thus could have prevented Bush from winning and hence every Gore voter will in fact be in one such subset. This is not as counter-intuitive as it may seem, because it is precisely the same reasoning that determines that the Nader voters were causally effective for Bush winning (which is more obvious). Furthermore, the causal contribution of a voter is not to be conflated with the moral responsibility of the voter. In this case, despite the Gore voters causally contributing to Bush's victory they were not responsible for Bush becoming president because in our framework voting for Gore was the strategy that maximised their avoidance potential for Bush's victory.

¹⁷See Wolf (1990) and Watson (2004) for a discussion of this 'self-disclosure' aspect of moral responsibility. For the importance of expressive voting, see Brennan and Lomasky (1993).

¹⁸A similar phenomenon occurred in the 2002 French presidential elections. Everyone who in the first round did not vote for Jospin (including those who voted for candidates more on the left) were responsible for the eventual outcome of Chirac being re-elected.

8. Conclusion

The institutional design result posed here says that for the individualistic folk-ethics conception of moral responsibility there is no way of structuring social interactions that have the three properties: non-dictatorship, completeness (no responsibility voids), and uniformity (no fragmentation). Either we plump for a dictator, which is undesirable or infeasible, or we have to accept voids or fragmentation. This result is foundational and inescapable.

Guaranteeing the absence of voids entails the inevitability of fragmentation of responsibility. In many cases fragmentation will be unproblematic since it merely reflects the different roles that individuals happen to play in bringing about an outcome. Given that we believe there to be better reasons to organise our collective affairs around rules that do not permit voids, we must accept fragmentation. But, this entails that injustices may arise, that a radical ideal of equality will sometimes be violated, and that an allocation will not always track individual values. At this stage our practical guidance for institutional designers can only be very broad. To sharpen it we would have to examine the exact conditions under which fragmentation can be accepted. Also, can uniformity be modified in such a way that voids are still excluded without paying the price of procedural fairness, inequality of influence or inadequate value-tracking? If so, which institutions correspond with it? Our analysis has isolated what we believe to be the real challenge posed by the problem of moral responsibility for collective outcomes.

Appendix A: Formal Statement and Proof

A.1. Definitions and Notation

X is a finite set of outcomes (alternatives, possible worlds) with at least three elements and $N = \{1, \dots, n\}$ is a finite set of individuals. The set of all non-empty proper subsets of X is denoted by $[X]$ and its elements are called *states of affairs*.

DEFINITION 1 A game form G is an $n + 1$ -tuple $G = (S_1, \dots, S_n, \pi)$ where each S_i denotes the finite set of strategies of $i \in N$. A play $s_n = (s_1, \dots, s_n)$ is a particular element of $S_1 \times \dots \times S_n$, and π is an outcome function, i.e. a mapping from the set of all plays to X .

For all $T \subseteq N$, we call an element s_T of $\prod_{i \in T} S_i$ an *event*. We write (s_T, s_{N-T}) to denote the play which consists of the combination of the (mutually exclusive) events s_T and s_{N-T} . If $T \subseteq U$ and if all $i \in T$ adopt the same strategy in s_T as in s_U , s_T is called a *subevent* of s_U . The set $\pi(s_T)$ denotes the set of outcomes that can result from the event s_T : $\pi(s_T) = \{\pi(s_T, s_{N-T}) \mid s_{N-T} \in \prod_{i \notin T} S_i\}$. If A is a singleton set, say $A = \{x\}$, we omit the set brackets and write $\pi(s_T) = x$ just like we will write $X - x$ (rather than $X - \{x\}$) or $N - i$ (instead of $N - \{i\}$).

DEFINITION 2 For any play s_N and $i \in N$, an individual strategy s of i is said to be a *cause* of A (and *i* causally effective) if, and only if, there is a subevent s_T of s_N such that (i) i adopts s in s_T , (ii) $\pi(s_T) \subseteq A$, (iii) $\pi(s_{T-i}) \not\subseteq A$.

With each game form G and each individual we associate a probability distribution p_i over the set of the strategy combinations of the others. For any i , $A \in [X]$ and $s \in S_i$, $\rho_i(A|s)$ denotes s 's *avoidance potential* for A . Let for each individual i and each $s \in S_i$, $h_i(s, A) = \{s_{-i} \mid i \text{ is not effective for } A \text{ in } (s, s_{-i})\}$. We then define:

$$\rho_i(A, s) = \sum_{s_{-i} \in h_i(s, A)} p_i(s_{-i}).$$

DEFINITION 3 *Let $A \in [X]$. We call a strategy $s \in S_i$ an A -minimal strategy (A -maximal) if, and only if, the strategy's avoidance potential for A is at least as large as (smaller than or equal to) that of any other strategy. A player i is called a dummy player if for all $x \in X$, any $s \in S_i$ is x -maximal and x -minimal. A coalition T is decisive for some $x \in X$ if $\pi(s_T^x) = x$ for any x -maximizing T -event s_T^x .*

We write s^A to denote that a strategy s is A -maximal and s_A to denote that it is A -minimal. We do not make any assumptions about how the various probability functions p_i are to be inferred from the game form or from other information (e.g. preferences). We only make one regularity assumption and that is that any individual strategy s that is A -maximal for some A will be $X - A$ -minimal, that is, for any such s we have $s = s^A = s_{X-A}$.¹⁹

The combination of a game form G with an n -tuple of probability distributions p is denoted by G_p and is called a *responsibility game*.

DEFINITION 4 *Given a responsibility game G_p , and a play $s_N = (s_1, \dots, s_n)$, an individual $i \in N$ is responsible for $A \in [X]$ if and only if,*

1. i is causally effective for A in s_N ;
2. $\rho_i(A|s') > \rho_i(A|s_i)$ for some $s' \in S_i$.

Given some responsibility game G_p and play s_N , an individual i is *responsibility-bearing* if he is responsible for some $A \in [X]$.

DEFINITION 5 *A responsibility game is uniform iff for any play s_N it holds that all responsibility-bearing individuals in that play are responsible for the very same set of features of the outcome: if i is responsible in that play for some $A \in [X]$ and j is responsible for some $B \in [X]$, then j is also responsible for A , and i for B .*

DEFINITION 6 *A responsibility game displays a responsibility void iff there is a play s_N and an $A \in [X]$ such that $\pi(s_N) \in A$ and none of the individuals is responsible for A . A responsibility game is complete if it does not display a responsibility void.*

DEFINITION 7 *A responsibility game is dictatorial iff there is some $i \in N$ such that for all $x \in X$ there is some $s \in S_i$ with $\pi(s) = x$.*

¹⁹As an anonymous referee has pointed out the exact conditions under which the assumption holds may be difficult to establish. We do not examine this issue further, leaving it to future work.

A.2. Proof of Main Result

THEOREM *If a responsibility game is uniform and complete, then it is dictatorial.*

We start with three useful lemmata.

LEMMA 1. *Let G_p be a responsibility game.*

1. *For all s_N and all i , if i adopts an A -minimal strategy in s_N , then i is not responsible for A in s_N .*
2. *If some strategy of an individual is both A -maximal and A -minimal for some A , then any strategy of the individual is so.*
3. *If there exists strategies $s_i, s'_i \in S_i$, and an event s_{N-i} , and outcomes $x, y \in X$ with $x \neq y$, $\pi(s_{N-i}, s_i) = x$, and $\pi(s_{N-i}, s'_i) = y$, then i is effective for $X - y$ in (s_{N-i}, s_i) and for $X - x$ in (s_{N-i}, s'_i) .*
4. *No two different agents can be decisive for two different outcomes.*

Proof. The first two clauses follow directly from the definitions.

3. As $\pi(s_{N-i}, s_i) = x$, and $\pi(s_{N-i}, s'_i) = y$, it follows that $x, y \in \pi(s_{N-i})$ implying $\pi(s_{N-i}) \not\subseteq X - x$ and $\pi(s_{N-i}) \not\subseteq X - y$. Since $\pi(s_{N-i}, s'_i) = y \in X - x$ and since $\pi(s_{N-i}) \not\subseteq X - x$, i is effective for $X - x$ in (s_{N-i}, s'_i) . Similarly, $\pi(s_{N-i}, s_i) = x \in X - y$ and $\pi(s_{N-i}) \not\subseteq X - y$ mean that i is effective for $X - y$ in (s_{N-i}, s_i) .
4. Suppose to the contrary, there exist i, j ($i \neq j$) such that i is decisive for x , and j for y , with $x \neq y$. Decisiveness of i for x implies $\pi(s_i^x, s_{N-i}) = x$ for any s_{N-i} . Similarly, decisiveness of j for y implies $\pi(s_j^y, s_{N-j}) = y$ for any s_{N-j} . But then $\pi(s_i^x, s_j^y, s_{N-\{i,j\}}) = x = y$, contradicting $x \neq y$.

LEMMA 2. *Let G_p be a responsibility game that is complete and uniform.*

1. *For all s_N , if some individual is responsible in s_N for some $A \in [X]$, then he is fully responsible, that is, he is responsible for all $A \in [X]$ containing $\pi(s_N)$.*
2. *N is decisive for each $x \in X$.*
3. *For all i , (a) if there are $x, y \in X$ ($x \neq y$) and there is an $s \in S_i$ that is both x -maximal and y -maximal, or (b) if there is an $x \in X$ and an $s \in S_i$ that is x -maximal as well as x -minimal, then i is a dummy player.*

Proof.

1. Let i be responsible for some A in some s_N . Take any other state of affairs B containing the outcome $\pi(s_N)$. Since there are no voids, some individual is responsible for B . By uniformity, i is then so as well.
2. Let s_N^x be a play in which all individuals adopt one of their x -maximizing strategies and let v denote the outcome of the play. By completeness, some $i \in N$ is responsible for v . By (1), such i is fully responsible, i.e. responsible for any $A \in [X]$ containing v . He is not responsible for $X - x$ since an x -maximal strategy by assumption always is $X - x$ -minimal. The set $X - x$ therefore cannot contain v , which entails $x = v$.
3. (a) Let $s = s^x = s^y$ for some $i, x, y \in X$ ($x \neq y$) and $s \in S_i$. If a strategy is both maximal and minimal for some outcome, then all strategies are so for that outcome. It therefore suffices to show that s is both z -maximal and z -minimal

for all $z \in X$. Suppose this is not so. For some $z \in X$, s then is either (a1) not z -minimal or (a2) not $X - z$ -minimal.

(a1) Let s_N be a play in which i adopts s , which has z as its outcome, and in which i is effective for z . Since s is not z -minimal, such a play must exist. Player i is then responsible for z at s_N : he is effective for z and he adopts a strategy that is not z -minimal. We have $z \in X - x$ or $z \in X - y$. Since x -maximality and y -maximality of s by assumption entails $X - x$ - and $X - y$ -minimality, i is not responsible for $X - x$ and $X - y$, contradicting (1) above.

(a2) Now let s_N be a play in which i adopts s , but which has an element of $X - z$ as its outcome, and in which i is effective for $X - z$. Because s is not $X - z$ -minimal, such a play exists. Player i is responsible for $X - z$ at s_N . By full responsibility he should be so for all subsets of X containing $\pi(s_N)$. However, since x -maximality and y -maximality of s by assumption entails $X - x$ - and $X - y$ -minimality, and since $\pi(s_N) \in X - x$ or $\pi(s_N) \in X - y$, i is not responsible for at least one set containing $\pi(s_N)$.

(b) Let s be a strategy that is x -maximal and x -minimal for some $x \in X$. Take $y \in X - x$ and let s' be a y -maximal strategy. Since s is x -maximal and x -minimal, s' also is x -maximal. From (a) it now directly follows that i is a dummy player.

LEMMA 3. *Let G_p be a responsibility game that is complete and uniform. If $\pi(s_S^x, s_{N-S}^*) = x$ for some proper non-empty subset S of N and some s_{N-S}^* in which each $i \in N - S$ adopts a strategy s^{y_i} that is maximal for some $y_i \neq x$, then S is decisive for $\{x\}$.*

Proof. We first prove, by contradiction, that the result holds if $N - S$ does not contain any dummy players. Let $\pi(s_S^x, s_{N-S}^*) = x$ where s_{N-S}^* is the event as described. Define $N - S = \{1, \dots, t\}$ and assume that for some play (s_S^x, s_1, \dots, s_t) we have $\pi(s_S^x, s_1, \dots, s_t) \neq x$. Consider the following sequence of plays:

$$\begin{aligned} s_N^0 &= (s_S^x, s_1, \dots, s_t) && \text{and } \pi(s_N^0) \neq x \\ s_N^1 &= (s_S^x, s^{y_1}, s_2, \dots, s_t) \\ s_N^2 &= (s_S^x, s^{y_1}, s^{y_2}, s_3, \dots, s_t) \\ &\vdots \\ s_N^t &= (s_S^x, s^{y_1}, \dots, s^{y_t}) = (s_S^x, s_{N-S}^*) && \text{and } \pi(s_N^t) = x \end{aligned}$$

Let m be the largest m such that $\pi(s_N^{m-1}) = v \neq x$ for some $v \in X$. That is,

$$\begin{aligned} s_N^{m-1} &= (s_S^x, s^{y_1}, \dots, s^{y_{m-1}}, s_m, s_{m+1}, \dots, s_t) && \text{and } \pi(s_N^{m-1}) = v \neq x \\ s_N^m &= (s_S^x, s^{y_1}, \dots, s^{y_{m-1}}, s^{y_m}, s_{m+1}, \dots, s_t) && \text{and } \pi(s_N^m) = x \end{aligned}$$

Since $\pi(s_N^t) = x \neq \pi(s_N^0)$, such m exists. We show that for any $s^* \in S_m - \{s^{y_m}\}$:

$$\pi(s_S^x, s^{y_1}, \dots, s^{y_{m-1}}, s^*, s_{m+1}, \dots, s_t) \in \{x, y_m\} \quad (\text{A.1})$$

and thus that $v = y_m$. Suppose to the contrary that there is some $z \neq x, y_m$ and some $s^* \in S_m$ such that

$$\pi(s_S^x, s^{y_1}, \dots, s^{y_{m-1}}, s^*, s_{m+1}, \dots, s_t) = z.$$

By Lemma 1.3, individual m is effective for $X - z$ in s_N^m . Since his strategy in s_N^m is $X - y_m$ -minimal, $x \in X - y_m$ implies by uniformity and m 's effectivity for $X - z$ that the strategy s^{y_m} is $X - z$ -minimal, which means that $s^{y_m} = s^z$. Lemma 2.3 then implies that m is a dummy player. This is a contradiction, which proves A.1.

Next assume for some $z \neq x, y_m$ we have

$$\pi(s_S^x, s^{y_1}, \dots, s^{y_{m-1}}, s_m^z, s_{m+1}, \dots, s_t) = x.$$

Individual m is not responsible for outcome x because $x \in X - z$ and because s_m^z is $X - z$ -minimal. By Lemma 2.1 he is then also not responsible for any other subset containing x . In particular he is not for $X - y_m$. Since he is effective for it, s_m^z must be $X - y_m$ -minimal. This means it is y_m -maximal, contradicting Lemma 2.3, case (a), and absence of dummy players in $N - S$. Hence $\pi(s_S^x, s^{y_1}, \dots, s^{y_{m-1}}, s_m^z, s_{m+1}, \dots, s_t) \neq x$. By A.1 it follows that its outcome must be y_m , i.e. for all $z \neq x, y_m$ we have

$$\pi(s_S^x, s^{y_1}, \dots, s^{y_{m-1}}, s_m^z, s_{m+1}, \dots, s_t) = y_m. \quad (\text{A.2})$$

To conclude the first part of the proof, we note that from $X - z$ -minimality of s_m^z , it follows that m is not responsible for $X - z$. By Lemma 2.1 and $y_m \in X - z$, he is not responsible for any other subset containing y_m either. In particular he is not so for $X - x$. Since $x = \pi(s_N^m)$, he is effective for $X - x$. Therefore s_m^z must be $X - x$ minimal. That means it is not only z -maximal but also x -maximal, which contradicts Lemma 2.3 and absence of dummy players.

Now consider cases in which $N - S$ does contain dummy players. Let U be the set of all dummy players in $N - S$. If $U = N - S$, the result follows from the proof of Lemma 2.2 and the fact that any strategy of a dummy player is maximal for any element. Assume U is a proper subset of $N - S$. Since any strategy of $i \in U$ is x -maximal, applying the first part of the proof to $S \cup U$ rather than to S shows that $S \cup U$ is decisive for x . From decisiveness of $S \cup U$ for x , and from the fact that, by definition of a dummy player, any play of which s_S^x is a subevent also has $s_{S \cup U}^x$ as a subevent, it follows that S is decisive for x . \square

Proof of theorem. We show that for any $x \in X$ there is an individual i who is decisive for x . Since no two individuals can be decisive for different alternatives (Lemma 1.4), this suffices to yield the theorem: i then is decisive for all $x \in X$.

Take $x \in X$. Let U be the set of all dummy players. If $N = U$, no one would be responsible for the outcome of any play, violating completeness. Decisiveness

of N for x , Lemma 3 and $N \neq U$, together entail that $N - U$ is also decisive for x . Hence, there is a coalition of non-dummy players that is decisive for x . Let S be the smallest such set of non-dummy players, i.e. S does not have a proper subset that is also decisive for x .

If S is a singleton set, then some individual is decisive for x , which we want to show. Suppose therefore S is not a singleton set. We show that this leads to a contradiction. Take $i \in S$ and consider the play $(s_{S-i}^x, s_i^y, s_{N-S}^z)$, where x, y , and z are distinct elements of X . The outcome of the play must be x, y or z : each individual would otherwise fail to be responsible for at least one $A \in [X]$ containing the outcome (the members of $S - i$ for $X - x$, i for $X - y$ and the members of $N - S$ for $X - z$), which given Lemma 2.1 would entail that no one bears any responsibility for features of the outcome, contradicting completeness. If the outcome is z , $N - S$ is decisive for z by Lemma 3, contradicting decisiveness of S for x . Hence the outcome is x or y . If the outcome is x , then Lemma 3 implies that $S - i$ is decisive for x , contradicting the way S was defined. Hence, the outcome must be y and, by Lemma 3, i is decisive for y .

Now consider the play (s_{N-i}^z, s_i^x) . The outcome must be x or z . If were some outcome v other than z , $X - z$ -minimality of the strategies of all $j \neq i$ and absence of voids, would mean that i would be the only person responsible for $X - z$, implying either a violation of uniformity or decisiveness of i by Lemma 3. If the outcome is z , then $N - i$ is decisive for z by Lemma 3, contradicting decisiveness of i for y . The outcome thus is x and Lemma 3 shows that i is decisive for x , contradicting that S is the smallest coalition that is decisive for x . \square

University of Hamburg
VU Amsterdam

References

- Bacharach, M. (2006). *Beyond Individual Choice*, Princeton: Princeton University Press.
- Benn, S.I. and Weinstein, W.L. (1971). 'Being free to act, and being a free man', *Mind*, vol. 80, pp. 194–211.
- Bovens, M. (1998). *The Quest for Responsibility*, Cambridge: Cambridge University Press.
- Bowles, S. (2004). *Microeconomics: Behavior, Institutions and Evolution*, Princeton: Princeton University Press.
- Braham, M. and van Hees, M. (2009). 'Degrees of causation', *Erkenntnis*, vol. 73, pp. 323–344.
- Braham, M. and van Hees, M. (2011). 'Responsibility voids', *Philosophical Quarterly*, vol. 61, pp. 6–15.
- Braham, M. and van Hees, M. (2012). 'An anatomy of moral responsibility', *Mind*, vol. 121(483), pp. 601–634.
- Braham, M. and van Hees, M. (2014). 'The impossibility of pure libertarianism', *Journal of Philosophy*, vol. 111, pp. 420–436.
- Brennan, G. and Lomasky, L. (1993). *Democracy and Decision: The Pure Theory of Electoral Preference*, Cambridge: Cambridge University Press.
- Cane, P. (2002). *Responsibility in Law and Morality*, Oxford: Oxford University Press.
- Dewey, J. (1960). *Theory of the Moral Life*, New York: Holt.
- Feinberg, J. (1970). *Doing and Deserving*, Princeton: Princeton University Press.
- Fischer, J.M. (1999). 'Recent work on moral responsibility', *Ethics*, vol. 10(1), pp. 93–139.

- Hart, H.L.A. (2008). *Punishment and Responsibility*, Oxford: Oxford University Press, 2nd edn.
- Hart, H.L.A. and Honoré, A.M. (1959). *Causation in the Law*, Oxford: Oxford University Press.
- Hetherington, S. (2003). 'Alternate possibilities and avoidable moral responsibility', *American Philosophical Quarterly*, vol. 40(4), pp. 229–239.
- Lewis, D. (1973). 'Causation', *Journal of Philosophy*, vol. 70, pp. 556–569.
- Lewis, H.D. (1948). 'Collective responsibility', *Philosophy*, vol. 23(84), pp. 3–18.
- List, C. and Pettit, P. (2011). *Group Agency*, Oxford: Oxford University Press.
- Mackie, J.L. (1974). *The Cement of the Universe*, Oxford: Oxford University Press.
- McKenna, M. (1997). 'Alternative possibilities and the failure of the counterexample strategy', *Journal of Social Philosophy*, vol. 28(3), pp. 71–85.
- Mellema, G.F. (1997). *Collective Responsibility*, Amsterdam: Rodopi.
- Mill, J.S. (2008/1859). *On Liberty*, Oxford: Oxford University Press.
- Otsuka, M. (1998). 'Incompatibilism and the avoidability of blame', *Ethics*, vol. 108, pp. 685–701.
- Pettit, P. (2001). *A Theory of Freedom: From the Psychology to the Politics of Agency*, Cambridge: Polity.
- Pettit, P. (2007). 'Responsibility incorporated', *Ethics*, vol. 117, pp. 171–201.
- Pogge, T. (1992). 'Loopholes in moralities', *Journal of Philosophy*, vol. 89, pp. 79–98.
- Sartorio, C. (2012). 'Causation and freedom', *Journal of Philosophy*, vol. 109, pp. 629–651.
- Sverdlik, S. (1987). 'Collective responsibility', *Philosophical Studies*, vol. 51, pp. 61–76.
- Thompson, D.F. (1980). 'Moral responsibility of public officials: The problem of many hands', *American Political Science Review*, vol. 74(4), pp. 905–916.
- Tuck, R. (2008). *Free Riding*, Cambridge MA: Harvard University Press.
- van Hees, M. (2010). 'Some general results on responsibility for outcomes', in (A. van Deemen and A. Rusinowska, eds.), *Collective Decision Making*, pp. 99–109, Heidelberg: Springer.
- Watson, G. (2004). *Agency and Answerability*, Oxford: Oxford University Press.
- Wolf, S. (1990). *Freedom Within Reason*, Oxford: Oxford University Press.
- Wright, R. (1988). 'Causation, responsibility, risk, probability, naked statistics, and proof: Pruning the bramble bush by clarifying the concepts', *Iowa Law Review*, vol. 73, pp. 1001–1077.
- Wyma, K. (1997). 'Moral responsibility and leeway for action', *American Philosophical Quarterly*, vol. 34(1), pp. 57–70.