Synthese
DOI 10.1007/s11229-011-9973-x

An impossibility theorem for amalgamating evidence

Jacob Stegenga

Received: 19 August 2010 / Accepted: 24 June 2011

Abstract Amalgamating evidence of different kinds for the same hypothesis into an overall confirmation is analogous, I argue, to amalgamating individuals' preferences into a group preference. The latter faces well-known impossibility theorems, most famously "Arrow's Theorem". Once the analogy between amalgamating evidence and amalgamating preferences is tight, it is obvious that amalgamating evidence might face a theorem similar to Arrow's. I prove that this is so, and end by discussing the plausibility of the axioms required for the theorem.

Keywords Evidence · Arrow's theorem · Amalgamating evidence · Confirmation

1 Introduction

Often our hypotheses are confirmed or disconfirmed by evidence from multiple methods, or multiple modes of evidence. For example, there are competing hypotheses about how the influenza virus is spread from person to person: the Contact hypothesis is that influenza is spread by direct contact between people, the Droplet hypothesis is that influenza is spread on large droplets expelled by coughs and sneezes, and the Airborne hypothesis is that influenza is spread on tiny airborne particles over large distances. To determine which of these hypotheses is best supported we have multiple modes of evidence available: the epidemiological patterns of influenza spread, evidence from controlled animal experiments using various ingenious designs and different kinds of animals, results from mathematical models, and clinical experience. Some modes of

J. Stegenga (✉)
Department of Philosophy, University of California San Diego, 9500 Gilman Dr, La Jolla, CA 92093, USA
e-mail: jstegenga@ucsd.edu

Published online: 02 August 2011

evidence support the Contact hypothesis while other modes of evidence support one of its competitor hypotheses.

Once evidence is thought of in this way it suggests an analogy between amalgamating preferences from multiple people—a burgeoning topic in social choice theory—and amalgamating evidence from multiple modes. Amalgamating individuals' preferences into a group decision faces several well-known impossibility theorems, including Condorcet's voting paradox and Arrow's impossibility theorem. I describe Arrow's theorem in Sect. 2, and in Sect. 3 I attempt to draw the analogy between amalgamating preferences and amalgamating multimodal evidence as tightly as possible. The analogy suggests that amalgamating multimodal evidence might face an impossibility theorem similar to Arrow's theorem (Sect. 4). The primary contribution of this paper is to demonstrate that this is so—amalgamating multimodal evidence faces an Arrow-like impossibility theorem (proven in Appendix). This paper makes small steps toward delimiting the logical space of possibilities for how multimodal evidence can be amalgamated. More promising, perhaps, is the demonstration that the analogy between preference amalgamation and evidence amalgamation allows for a substantial import of results from the rich literature on amalgamating preferences.[1]

The theorem presented here is limited to amalgamation functions which accept as input only the confirmation ordering of hypotheses by evidence; if an amalgamation function accepted as input the absolute degree of confirmation of a hypothesis, then a key axiom would be violated and so the impossibility result would not apply. This will be explained in due course, but I mention it now since some might think this focus on confirmation ordering to be unorthodox, and will suspect that the impossibility result will have a narrow range of application given the ubiquity of absolute measures of confirmation. That would be hasty. Of the class of evidence–hypothesis relations, some have a trivially determinate likelihood, and so some absolute measures of confirmation are possible (and if concrete prior probabilities are available, then many more absolute measures of confirmation are available). However, many evidence–hypothesis relations make the determination of precise values for likelihoods impossible. In such cases no precise absolute measures of confirmation are possible.[2] Elsewhere I argue that in actual practice, most real cases of confirmation are like this—yielding at best rather vague, imprecise values for likelihoods. If determinations of absolute confirmations are not possible, at least comparative confirmations may be; comparative confirmation can be understood as statements of the form "evidence i supports hypothesis H_1 more than/equally/less than hypothesis H_2." Confirmation orderings like this are the most that can be justified in much of science. Such orderings, explicated in Sect. 4, satisfy the axioms of the impossibility theorem for confirmation presented below.

The primary result presented here is perhaps best thought of as a no-go theorem which directs attention to the general plausibility of its axioms. I discuss the plausibility of each of the axioms in Sects. 5 and 6, and argue that two of the axioms are necessary requirements of evidence amalgamation functions, and the remaining two

[1] Seminal results in social choice theory include Arrow (1951), Black (1958), and Sen (1970). The present paper is not, of course, the first to apply Arrow's Theorem beyond its original context; see also Leitgeb and Segerberg (2007), Okasha (2011), and Dietrich and List (2007).

[2] Vague likelihoods have received little attention, but see Hawthorne (forthcoming).

axioms, while not exceptionless requirements of rationality, are generally desirable features of evidence amalgamation functions.

2 Amalgamating preferences

We often wish to amalgamate the preferences of a set of individuals into an overall group preference. Contemporary social choice theorists call a "social welfare function" any aggregation device which takes as its input the preferences of individuals and generates as its output a group preference. The amalgamation of a set of preferences, given certain minimal conditions, can lead to paradoxes. Here I briefly introduce an infamous example.

2.1 Arrow's theorem

In 1950 the economist Kenneth Arrow published part of his doctoral dissertation as a groundbreaking paper in social choice theory. In one of its strongest forms, Arrow's impossibility theorem shows that if a society has at least two decision makers and three options to choose from, then no social welfare function (SWF) can jointly meet the following desiderata, stated informally:

Non-Dictatorship: The SWF cannot have as its output the preference orderings of a single decision maker, for all possible preference orderings of that decision maker.

Independence of Irrelevant Alternatives: The ordering of choices $\langle A, B, C \ldots N\rangle$ by the SWF should only depend on the individuals' orderings of choices $\langle A, B, C \ldots N\rangle$.

Unrestricted Domain: The SWF must be able to accept as input all preference orderings from all decision makers.

Unanimity: If all individuals prefer A to B, then the SWF must rank A over B.

This is a troubling result, since it is reasonable to want a SWF that meets all of the above desiderata (I further discuss Arrow's axioms in Sect. 5). Here is another way of putting the theorem: any SWF which satisfies Independence of Irrelevant Alternatives, Unrestricted Domain, and Unanimity must be a Dictatorship (that is, must have as its output only the preference orderings of a single individual). It is at first glance a surprising conclusion, since the Independence of Irrelevant Alternatives, Unrestricted Domain, and Unanimity axioms seem to have little to do with one another, but are jointly sufficient to require a SWF to be a Dictatorship. Arrow's theorem has generated an explosion of literature interpreting the theorem and demonstrating other theorems by relaxing, removing, strengthening, or adding axioms.

3 Analogy

3.1 Individual's preference: group's preference:: evidence: confirmation

An individual's preferences are to a group's preference as evidence from a single mode is to a confirmation of a hypothesis by multimodal evidence. Individuals prefer one choice over another choice just as evidence supports one hypothesis over another hypothesis. Multimodal evidence (the set of evidence from all relevant modes) supports one hypothesis over another hypothesis, just as a group's aggregated preferences support one choice over another choice.[3] Finally, just as a set of individual preferences must be combined by an amalgamation function to determine a group's ordering of preferences, so too must multimodal evidence be combined by an amalgamation function to determine an overall ordering of confirmations of hypotheses.

3.2 Social welfare function: preference–choice relation:: multimodal evidence amalgamation function: evidence–confirmation relation

A social welfare function takes as input preference orderings of multiple individuals and delivers as output a group ordering; similarly a multimodal evidence amalgamation function takes as input confirmation orderings from multiple modes and delivers as output an overall confirmation ordering. A social welfare function is the medium between the preference–choice relation just as a multimodal evidence amalgamation function is the medium between the evidence–confirmation relation.

The following table illustrates the analogy between amalgamating preferences in social choice theory and amalgamating multimodal evidence in confirmation.

Social Choice	Confirmation
Individual voter	Single mode of evidence
Set of voters	Set of modes
Preference	Evidence
Preference orderings (input)	Confirmation orderings (input)
Social Welfare Function (operation)	Amalgamation Function (operation)
Preference ordering (output)	Confirmation ordering (output)

4 Impossibility theorem for confirmation

4.1 Notation and definitions

I rely on the following notation.

Multimodal Evidence
A *mode* of evidence i generates evidence e_i. The evidence from all available modes relevant to a set of competing hypotheses **H** $\{H_1 \ldots H_m\}$ I will call *multimodal evidence* $\{e_1, e_2, \ldots e_n\}$.

This account of multimodal evidence is not committed to any particular notion of evidence, or of individuation of modes.[4] A confirmation order is a confirmation relation,

[3] In Sect. 4 I give a more precise exposition of confirmation ordering.

[4] Conceptualizing how modes of evidence should be individuated is surprisingly difficult. One promising approach to individuating modes is to appeal to the background assumptions required for a test of some

denoted by $\succcurlyeq_i$ where i is a mode (the confirmation ordering relation is indexed to the mode of evidence). Thus $H_1 \succcurlyeq_i H_2$ means "evidence from mode i confirmation orders H_1 equally to or above H_2." A confirmation order is reflexive, transitive and connected (but not necessarily anti-symmetric), denoted as follows.

Transitivity
If $H_1 \succcurlyeq_i H_2$ and $H_2 \succcurlyeq_i H_3$ then $H_1 \succcurlyeq_i H_3$, for all H_1, H_2, H_3 in **H** and for all i.[5]

Reflexivity
$H_1 \succcurlyeq_i H_1$ for all H_1 in **H** and for all i.

Connected
$H_1 \succcurlyeq_i H_2$ or $H_2 \succcurlyeq_i H_1$, for all distinct H_1, H_2 in **H** and for all i.[6]

I have not yet said what confirmation ordering means in terms familiar to philosophers of science. Confirmation ordering of multiple hypotheses by evidence from a particular mode can be understood by any account of the evidence–hypothesis relation. A likelihoodist, for example, could understand confirmation ordering as follows:

$$H_1 \succcurlyeq_i H_2 \text{ if and only if } p(e_i|H_1) \geq p(e_i|H_2)$$

Footnote 4 continued
particular hypothesis. Nevertheless, the present argument does not depend on a particular answer to how modes ought to be individuated. Even in what appears to be the clearest domain in which we might be able to individuate modes—sensation—there is little agreement regarding how to individuate sensory modalities (see Keeley 2002).

5 Confirmation of multiple hypotheses by multimodal evidence can be cyclical, or non-transitive, as in Condorcet's voting paradox, which further illustrates the analogy suggested in Sect. 3 and sets the stage for the impossibility theorem below. Suppose we have evidence from three modes, i, j, k, for three hypotheses, H_1, H_2, and H_3. The following confirmation orderings are possible:

$$H_1 \succ_i H_2 \succ_i H_3$$

$$H_2 \succ_j H_3 \succ_j H_1$$

$$H_3 \succ_k H_1 \succ_k H_2$$

No AF which accepts as input only these orderings can determine a best-confirmed hypothesis. If the AF determined H_3 as the best-confirmed hypothesis, then it would seem that H_2 should instead be determined as the best-confirmed hypothesis, since two modes of evidence, i and j (a majority), would confirmation order H_2 over H_3. But if the AF determined H_2 as the best-confirmed hypothesis, then, by the same argument, H_1 should instead be determined as the best-confirmed hypothesis, since two modes of evidence, i and k (a majority), would confirmation order H_1 over H_2. Finally, if the AF determined H_1 as the best-confirmed hypothesis, then, again by the same argument, H_3 should instead be determined as the best-confirmed hypothesis, since two modes of evidence, j and k (a majority), would confirmation order H_3 over H_1. This is an analogue to Condorcet's voting paradox. Similarly, Baumann (2005) shows that theory choice can violate transitivity when multiple 'theoretical virtues', rather than multiple modes of evidence, order competing hypotheses as above. See also Okasha (2011).

6 Every confirmation relation $\succcurlyeq_i$ induces corresponding relations $\succ_i$ ('is more confirmed than') and $\sim_i$ ('is equally confirmed as'). They are defined as usual:

$$H_1 \succ_i H_2 \text{ if and only if } H_1 \succcurlyeq_i H_2 \text{ and not } H_2 \succcurlyeq_i H_1$$

$$H_1 \sim_i H_2 \text{ if and only if } H_1 \succcurlyeq_i H_2 \text{ and } H_2 \succcurlyeq_i H_1$$

A Bayesian could understand confirmation ordering with whatever her preferred confirmation measure happened to be; here is confirmation ordering put in terms of the difference measure of confirmation, for example:

$$H_1 \succcurlyeq_i H_2 \text{ if and only if } p(H_1|e_i) - p(H_1) \geq p(H_2|e_i) - p(H_2)$$

An error-statistical approach could understand confirmation ordering as follows:

> $H_1 \succcurlyeq_i H_2$ if and only if when given e_i the probability that H_1 is false despite concluding that H_1 is true is lower than the probability that H_2 is false despite concluding that H_2 is true.

On these accounts of the evidence–hypothesis relation, see (respectively) Sober (2008), Fitelson (1999), and Mayo (1996).[7]

Confirmation orderings are inherently less informative than absolute measures of evidential support (such as likelihood ratios, likelihood differences, posterior ratios, and posterior differences). Absolute measures of evidential support have been the primary focus of confirmation theory, but comparative measures of evidential support of hypotheses have received at least some attention in modern confirmation theory.[8] Sober (2008) argues that when comparing two hypotheses, if we have little information about the prior probabilities of the hypotheses, given evidence from a single mode we should use the "law of likelihood" to compare their relative support, rather than attempt to compare their posterior probabilities. Comparative measures of support can be derived from absolute measures of support, but not vice versa. In Sect. 6 I sketch an argument which shows that often in scientific practice absolute measures of confirmation are not possible, and in such cases comparative measures of confirmation should be preferred.

I introduce two further notions not broadly used in philosophy of science. A *profile of confirmation orders*, or simply *profile*, is a vector $(\succcurlyeq_1, \ldots, \succcurlyeq_n)$ of confirmation orders. An amalgamation function (AF) is a rule or function which aggregates multimodal evidence. To know how well a hypothesis is supported by multimodal evidence, evidence from particular modes must be combined by an AF, just as multiple individuals' preferences must be amalgamated by a social welfare function in order to determine a group's preference. It is likely that different sciences should have different amalgamation functions. There are functions that combine quantitative evidence from different modes and have quantitative outputs, including Dempster-Shafer Theory, Bayesian conditionalization, and statistical meta-analysis, and there are functions that combine evidence from different modes but have qualitative outputs, such as the evidence hierarchy schemes in evidence-based medicine or consensus conferences in medicine and social policy. Many disciplines currently have AFs, but since the notion has not been studied in depth we have no principles to systematically assess and compare the various AFs currently in use. The notion is perfectly general: an AF is simply any way of considering the support that diverse evidence provides to a hypothesis.

[7] Confirmation orderings can also be understood by a non-inductive theory of scientific testing, such as a view based on Popperian corroboration functions (see, e.g., Rowbottom 2010).

[8] See, e.g., Hacking (1965), Royall (1997) and Sober (2008).

However, the theorem considered here will only consider AFs which map each possible profile of confirmation orders to an output confirmation order.[9] This limitation will be defended in Sects. 5 and 6.

Given a profile $(\succcurlyeq_1, \ldots, \succcurlyeq_n)$, the corresponding aggregate relation is denoted by $\succcurlyeq$ $(= \mathrm{AF}(\succcurlyeq_1, \ldots, \succcurlyeq_n))$. Similarly, given a profile $(\succcurlyeq'_1, \ldots, \succcurlyeq'_n)$, the corresponding aggregate relation is denoted by $\succcurlyeq'$. In short, AF output orders are denoted by dropping the indexes of the input orders.

The goal of much of the literature on preference amalgamation is to specify the logical boundaries on what social welfare functions can do. Just as impossibility theorems serve to delimit the logical space of possibility for preference amalgamation functions, similar impossibility theorems for multimodal evidence amalgamation functions can be constructed, with the same goal. To demonstrate that this is so, I now provide an analogue to Arrow's impossibility theorem.

4.2 Impossibility theorem: Arrow analogue

There are several desiderata one might want a multimodal evidence AF to meet; these are analogues to Arrow's desiderata for social welfare functions. I here informally state the desiderata; I state the desiderata in formal terms in the Appendix.

Unanimity (U): If all modes confirmation order one hypothesis over another, then the AF must do the same.

Non-Dictatorship (D): No mode is dictatorial (i.e. no mode of evidence always trumps all other modes).

Independence of Irrelevant Alternatives (I): The way two hypotheses are confirmation ordered relative to each other by an AF depends only on how the individual modes confirmation order these two hypotheses relative to each other, and not on how the modes confirmation order them relative to other hypotheses.

Ordered Output (O): An AF generates a confirmation order for every profile.

Now it should be clear that the amalgamation of multimodal evidence can be structured in a way perfectly analogous to the amalgamation of preferences. Thus, one should expect an *impossibility theorem for confirmation* analogous to Arrow's impossibility theorem for preference amalgamation. The final condition that must be met for the theorem to hold is that there must be at least two modes of evidence available, and there must be at least three hypotheses in **H**.

Theorem *No AF can jointly satisfy U, I, O and D.*

Proof See Appendix. □

Here is another way to put the theorem: Any amalgamation function which satisfies Unanimity, Independence of Irrelevant Alternatives, and Ordered Output, must be a Dictatorship.

[9] This builds a 'universal domain' criterion into the definition of an aggregation rule; i.e. all profiles of orders are permissible inputs of the AF.

5 Arrow's axioms

The impossibility theorem presented here is a surprising result, just as Arrow's theorem was. At first glance, Ordered Output, Independence, and Unanimity have little to do with each other, but this theorem shows that the only AF which can jointly satisfy these three desiderata is a Dictatorship. In short, the theorem is both valid and surprising, and so it is reasonable to critically evaluate the axioms. Before considering the axioms in the case of amalgamating multimodal evidence it will help to consider the axioms in the case of amalgamating preferences.

Much research after the publication of Arrow's theorem was directed at the axioms that Arrow used, and arguments were proposed to relax some of the axioms, and other theorems were demonstrated by strengthening, relaxing, removing, or adding assumptions.[10] In short, though, Non-Dictatorship, Unanimity, and Ordered Output are intuitively plausible principles for a SWF, especially if one is committed to minimally democratic norms. The axiom most often thought reasonable to relax is Independence of Irrelevant Alternatives. Indeed, some have considered this axiom to be too strong a requirement for a SWF.

Recall what the Independence of Irrelevant Alternatives axiom in Arrow's Theorem requires: The ordering of choices $\langle A, B, C \ldots N \rangle$ by the SWF should only depend on the individuals' orderings of choices $\langle A, B, C \ldots N \rangle$. The axiom limits the information regarding individuals' preferences available to a SWF in at least two ways. First, it disallows information regarding individuals' preferences about options outside the choice set to influence the output ordering of the SWF. Second, it only allows ordinal information regarding individuals' preferences to be amalgamated by the SWF. I will call the first aspect of the axiom simply 'Independence', and the second aspect 'Ordinality'. I address each in turn, since both aspects of the axiom are important to the confirmation analogue.

5.1 Independence

Independence is an intuitively desirable feature of an SWF. For example, if I prefer apples (A) over bananas (B), and bananas over cherries (C), my preference ordering of these fruit is

$$A \succ B \succ C \tag{1}$$

If I then include strawberries (S) in my preference ordering of fruits, then (S) might be more or less preferable to (A) and/or (B) and/or (C), so that one possible ordering might be

$$A \succ S \succ B \succ C \tag{2}$$

[10] See, for example, Black (1958), Sen (1970), and Arrow's 1963 edition of *Social Choice and Individual Values*. For a philosophical exposition and defense of Non-Dictatorship, Unanimity, and Unrestricted Domain, see MacKay (1980).

but the orderings of (A) and (B) and (C) relative to each other should not change upon consideration of (S), so that the following ordering, for example, is prohibited:

$$\mathrm{B} \succ \mathrm{S} \succ \mathrm{C} \succ \mathrm{A} \tag{3}$$

The mere inclusion of (S) in my appreciation of fruit should not change my relative orderings of (A) and (B) and (C). In (3), the mere inclusion of (S) in my rank-ordering of fruit has shifted (A) from my most-preferred fruit to my least-preferred fruit, which seems irrational. In (2) Independence is satisfied whereas in (3) Independence is violated. A natural question to ask is: what does 'irrelevance' mean? Suppose I have an odd condition which makes strawberries react with apples, thereby causing digestive discomfort. Then, supposing my rank-ordering of fruit was for the purpose of making a fruit salad, in which the three highest ranked fruit would be consumed together, then strawberries *would* be relevant to my rank-ordering of apples with respect to other fruit, and so (3) might then be a reasonable ordering of my fruit preferences.

Despite the intuitive plausibility of Independence, voting systems exist which fail to satisfy it. For example, the Borda count method is an election method in which voters rank candidates in order of their preferences: if there are n candidates, then a candidate receives n points for a voter's first preference, $n-1$ points for a second preference, and so on; the points are then summed and the candidate with the most points wins. That Borda count does not satisfy Independence is a commonly recognized fact amongst social choice theorists. My purpose in raising the matter is to urge that the axioms (or at least the Independence axiom) used in Arrow's theorem should be thought of as desiderata rather than as necessary criteria for a SWF. Despite the fact that Borda count fails at least one of the axioms, it is occasionally used in real voting systems: the Pacific Island nation of Kiribati, for example, uses Borda count to elect its politicians. The violation of a desideratum in actual cases of preference amalgamation diminishes the desirability of the amalgamation function, but some desiderata are worse to violate than others, and many have thought that Independence is the most acceptable desideratum to violate. This will be important when considering the evidence analogues of the desiderata.

5.2 Ordinality

The ordinality aspect of Arrow's Independence of Irrelevant Alternatives axiom might at first glance seem unrealistically constraining, since it prohibits information about the intensity of individuals' preferences for the available choices. If we include preference intensity information in a SWF, and we assume we can make meaningful interpersonal comparisons of such information, then we can avoid Arrow's theorem (Sen 1970). One might think that we can elicit preference intensities from individuals. The intensity of individual's preferences could be measured on an interval scale (or an absolute scale), in which the meaning between two equally-sized intervals on the scale is the same across the scale and for all individuals. Such preference intensity measures, if measured in a non-arbitrary and objective way, might allow inter-personal comparisons of preference intensity. This information would be richer than ordinal information: ordinal rankings can be inferred from measures on an interval scale, but interval measures cannot be inferred from ordinal rankings. The Independence axiom in Arrow's theorem explicitly stipulates the exclusion of measures on an interval or absolute scale.

Is this limitation to ordinal information of preferences for a SWF justified? There are at least two reasons to think so. First, it is unlikely that preferences *can* be meaningfully measured on an interval scale. Preferences are just words used to summarize poorly understood mental phenomena. There is no standard, intersubjective scale with which to measure preferences (at least one which is not arbitrary in important respects). A plausible guiding principle is to limit ourselves to that which is both meaningful and possible. Since ordinal rankings of preferences are both meaningful and possible, and since interpersonal interval or absolute measures of preferences are not both meaningful and possible, a SWF should be limited to ordinal measures of preferences. Second, even supposing it were possible to elicit meaningful preference intensity measures, it is not obvious that we would want to include such information when amalgamating individuals' preferences, since including preference intensity information in a SWF would benefit fanatics at the expense of moderates: the preferences of those individuals with higher preference intensities would count more toward the group choice than would the preferences of individuals with less-intense preferences.[11] Thus, it is both possible and desirable to limit the input of an SWF to ordinal rankings of preferences, as opposed to interval or cardinal measures of preference intensity.

6 Axioms of present theorem

How plausible are the axioms for the impossibility theorem for confirmation? For the theorem to be broadly applicable, the axioms would have to be seen as broadly desirable features of an AF. In what follows I give reasons for thinking that, although these desiderata are not exceptionless, inviolable principles of rationality, they are generally desirable features of an AF. Just as with Arrow's theorem, by far the most complicated and controversial axiom is Independence of Irrelevant Alternatives, so I leave it for last and devote the majority of space to it.

6.1 Unrestricted domain

Although not explicitly stated as an axiom in the theorem, the definition of an amalgamation function assumed what social choice theory calls an 'Unrestricted Domain'. The Unrestricted Domain requirement in the case of amalgamating confirmation orderings is nearly a dictate of reason. If an AF could accept as its input only a limited range of possible confirmation orderings, then the AF could be faced with some confirmation ordering by some modes of evidence and would either not be able to include the confirmation orderings or would not be able to return as its output the true confirmation ordering. Such amalgamation functions would be more constrained than they

[11] The considerations in this paragraph skirt over decades of controversy. See MacKay (1980) for a philosophical discussion and defense of the ordinality limitation.

otherwise need be. As with Arrow's Theorem, restrictions to the scope of orderings available to the aggregation function can be devised which allow the impossibility result to be avoided (for example, as Black (1958) showed, limiting the scope of preferences which are made available to a SWF to 'single peaked preferences' thereby avoids Arrow's theorem). However, it is hard to imagine a reason to restrict the domain of a multimodal evidence AF which is independent of the wish to avoid the impossibility result presented here.

6.2 Non-Dictatorship

The Non-Dictatorship axiom is, like Unrestricted Domain, a requirement of rationality. Non-Dictatorship demands that no single mode of evidence determine the confirmation ordering of an AF, for any of the confirmation orderings of the mode. This desideratum is weaker than, but follows logically from, a requirement that an AF consider all evidence from all relevant modes. Carnap's "Principle of Total Evidence" is the similar requirement that a person must consider all available evidence when estimating a probability (1947). If one does not consider all evidence from all available modes, then one is liable to unnecessary inductive risk. Indeed, if one ignores 'defeating' evidence then one is liable to consider a hypothesis true that one would otherwise consider false had one attended to the defeating evidence. In the case of preference amalgamation, the Non-Dictatorship desideratum is a corollary of basic democratic commitments. In the case of evidence amalgamation, the Non-Dictatorship desideratum is a corollary of basic scientific commitments. One of the purposes of a SWF is to take into account the preferences of all decision-makers—if we did not have the goal of accommodating the preferences of all (or at least of most), then there would be no need for a SWF in the first place. Similarly, one of the purposes of an AF is to take into account all available evidence—if we did not have the goal of considering all available evidence (or at least of most), then there would be no need for a AF in the first place. Thus, the desideratum of Non-Dictatorship is a necessary feature of an AF.

6.3 Unanimity

It is a frequently appealed to intuition that if all modes of evidence confirmation order one hypothesis over another, then the AF should do the same, and if it does not, the AF is flawed. The following toy example illustrates the Unanimity desideratum. Suppose our three hypotheses are:

H_1: The global climate is warming.
H_2: The global climate is neither warming nor cooling.
H_3: The global climate is cooling.

And suppose we have three modes of evidence:

i: Atlantic ocean temperature measurements
j: Arctic ice mass measurements
k: Atmospheric CO_2 concentration measurements

Further suppose that all three modes of evidence confirmation order H_1 over H_2 and H_2 over H_3. More precisely, using the notation introduced in Sect. 4 (and the definition of the strict ordering relation in footnote 6), suppose we have the following confirmation orderings:

$$H_1 \succ_i H_2 \succ_i H_3$$
$$H_1 \succ_j H_2 \succ_j H_3$$
$$H_1 \succ_k H_2 \succ_k H_3$$

In such a situation it is intuitively compelling to demand of any AF that upon amalgamation of evidence from i, j, and k, it *must* confirmation order H_1 over H_2 and H_2 over H_3. That is, given the above confirmation orderings it is reasonable to demand of the AF that

$$H_1 \succ H_2 \succ H_3$$

Unanimity, then, is a general (but as I note below, not exceptionless) desideratum of AFs.

Despite the intuitive appeal of Unanimity, epistemic modesty requires us to recognize that Unanimity is fallible; as I have elsewhere argued, multimodal evidence can be concordant for a hypothesis which is later deemed false (Stegenga 2009). To illustrate the fallibility of Unanimity, consider another example. Suppose you are a hospital's chief of medicine, pondering a patient's survival, and you have available two modes of evidence: verbal reports from Dr. Blue, and verbal reports from Dr. Green. Let your hypotheses be:

H_1: the patient will live longer than one week
H_2: the patient will die within one week

Let your modes of evidence be:

i: verbal reports from Dr. Blue
j: verbal reports from Dr. Green

Dr. Blue tells you that she is giving the patient drug X, because drug X is known to help such patients; for Dr. Blue, as for you, $p(H_1|e_i) > p(H_1)$, and suppose that $p(H_1|e_i) - p(H_1) > p(H_2|e_i) - p(H_2)$, so in the above notation: $H_1 \succ_i H_2$. Dr. Green tells you that he is giving the patient drug Y, because drug Y is known to help such patients; for Dr. Green, as for you, $p(H_1|e_j) > p(H_1)$, and suppose that $p(H_1|e_j) - p(H_1) > p(H_2|e_j) - p(H_2)$, so in the above notation: $H_1 \succ_j H_2$. However, you know that administering drug X and drug Y together will be lethally damaging to the patient's kidney; you decide that, although $p(H_1|e_i) - p(H_1) > p(H_2|e_i) - p(H_2)$ and $p(H_1|e_j) - p(H_1) > p(H_2|e_j) - p(H_2)$, $p(H_1|e_i \& e_j) - p(H_1) < p(H_2|e_i \& e_j) - p(H_2)$, and in the above notation: $H_2 \succ H_1$. In other words, the amalgamated evidence has the opposite rank-ordering of hypotheses than do both individual modes of evidence. Unanimity fails in this case, for a seemingly good reason.[12]

[12] Although perhaps even in this case Unamity is not violated, since one's knowledge of the interaction between the two drugs was itself presumably gained by evidence from a third mode.

Thus, Unanimity should not be construed as a necessary requirement of rationality, since it is occasionally violated by seemingly truth-conducive AFs. Nevertheless, although it is not exceptionless, the 'robustness' intuition illustrated by the global-warming example has been frequently defended as a useful and relatively general heuristic for scientific reasoning.[13] The same reasons which justify the robustness intuition support the status of Unanimity as a generally desirable feature of an AF.

6.4 Independence of irrelevant alternatives

The Independence of Irrelevant Alternatives axiom might, perhaps, be construed as the most troublesome of the axioms used in the impossibility theorem for confirmation. Recall what the axiom stipulates: an AF confirmation ordering of hypotheses $\{H_1 \ldots H_n\}$ should only depend on the confirmation orderings, by individual modes, of $\{H_1 \ldots H_n\}$. Thus, the axiom has both the Independence and Ordinality features as it does in the case of amalgamating preferences. I will discuss each in turn.

6.4.1 Independence

The Independence aspect of the axiom in the case of evidence amalgamation is formally analogous to the case of preference amalgamation, and it is just as intuitively compelling, as the following example illustrates. Suppose an AF has confirmation ordered a heliocentric model of the solar system (H_h) over an epicyclic model (H_e), and an epicyclic model over an eccentric model (H_c); the confirmation ordering would then be:

$$H_h \succ H_e \succ H_c \tag{1'}$$

We might then include another hypothesis in our rank-ordering—the blue cheese (H_b) model of the solar system, for example—and determine that the AF confirmation ordering of the blue cheese model is greater than the AF confirmation ordering of an epicyclic model of the solar system, and so the following confirmation ordering is possible:

$$H_h \succ H_b \succ H_e \succ H_c \tag{2'}$$

But merely including the blue cheese model hypothesis in the ordering of astronomical hypotheses should not alter the relative confirmation orderings of the heliocentric model, the epicyclic model, and the eccentric model; thus the following confirmation ordering would be prohibited:

$$H_c \succ H_b \succ H_e \succ H_h \tag{3'}$$

In (2′), but not in (3′), Independence is satisfied.

[13] Many philosophers of science have claimed that concordant multimodal evidence is useful, including Cartwright (1983), Howson and Urbach (1989), Staley (2004), and Weber (2005).

It is possible to construct cases in which, at least at first glance, it seems reasonable to relax Independence. Consider the following example. The biodiversity of an ecosystem has long been known to be correlated with the bioproductivity of an ecosystem. The direction of causality, however, is not known. Suppose our two hypotheses are:

H_1: Biodiversity causes bioproductivity
H_2: Bioproductivity causes biodiversity

The variables 'biodiversity' and 'bioproductivity' are coarse-grained macro-level hypotheses, but suppose that for each macro-level hypothesis we hypothesize one micro-level causal mechanism, which I will call, respectively, $H_{1'}$ and $H_{2'}$. Given all available evidence, $H_{1'} \succ H_{2'}$ (e.g. suppose $p(H_{1'}) = 0.6$ and $p(H_{2'}) = 0.4$). But then suppose a second micro-level causal mechanism is proposed for H_1, which I will call $H_{1''}$, and suppose that in the absence of any new evidence, the rank-orderings of the macro-level hypotheses do not change. Whatever credence is devoted to $H_{1''}$ must come from $H_{1'}$. If $H_{1'}$ were equally plausible as $H_{1''}$ (e.g. if $p(H_{1'}) = 0.3$ and $p(H_{1''}) = 0.3$), then the plausibility of $H_{1'}$ should go down when $H_{1''}$ is introduced, and it follows that $H_{2'} \succ H_{1'}$. In other words the AF confirmation ordering of the micro-level causal hypotheses $H_{1'}$ and $H_{2'}$ was reversed, merely by introducing a competitor micro-level causal hypothesis (while the rank-ordering of the macro-level hypotheses did not change). Independence is violated.

Something has gone wrong. To the extent that one finds the Independence axiom a generally compelling desideratum, one ought to be troubled by this example. Perhaps it is unrealistic to assume that, when assigning credences to micro-level hypotheses, we have exhausted the space of possibilities, and thus we ought not distribute our full credence across the presently available micro-level hypotheses. After all, we have a long history of new micro-level hypotheses being introduced to account for macro-level relations. Thus, holding $p(H_{1'}) = 0.6$ and $p(H_{2'}) = 0.4$ in the example above is perhaps unwarranted, given a non-negligible expectation that $H_{1'}$ and $H_{2'}$ do not exhaust the space of possibilities for micro-level hypotheses. If our prior probabilities in the micro-level hypotheses reflected the possibility of future competitor micro-level hypotheses being introduced, then warranted violations of Independence illustrated by the above example might be rare indeed.

6.4.2 Ordinality

The Ordinality aspect of the axiom might strike some as odd. Ordinality limits the information regarding the support that evidence provides to a hypothesis which is made available to the AF, by ruling out information stronger than the comparative confirmation ordering of multiple hypotheses. More specifically, Ordinality rules out absolute measures of confirmation. This is also true for the Ordered Output axiom, which states that an output of an AF is also a confirmation order. These might seem unduly restrictive, since at first glance we have absolute measures of confirmation for hypotheses: probabilities are measured on an absolute scale. If we were to include such information in an AF, then the above impossibility theorem simply would not apply, since the Independence axiom would not be satisfied (and nor perhaps the Ordered Output axiom). Thus, such reasoning might go, the impossibility theorem for confirmation presented here is based on an artificial limitation of information. We can and should relax this limitation for real cases of confirmation of multiple hypotheses by multimodal evidence, since we often have absolute measures of confirmation; it would follow that the theorem presented here is of limited scope since its domain of application is limited to those situations in which absolute measures of confirmation are not available. So one might think.

It is true, of course, that there are classes of evidence–hypothesis relations such that determining a precise measure of support that the evidence provides to the hypothesis is possible. If evidence e is deductively entailed by hypothesis H, then the likelihood of the evidence, $p(e|H)$, can be trivially determined. If the opposite of evidence e is deductively entailed by hypothesis H, then $p(e|H)$ can be trivially determined. If H specifies a particular chance set-up (as in classic examples such as drawing colored balls from an urn), and e is a particular outcome of this chance set-up, and the outcome space is exhausted and well-defined by objective probabilities, then $p(e|H)$ can be trivially determined. In these cases, knowing $p(e|H)$ is necessary (but not sufficient) to determine the degree to which e supports H by the most plausible Bayesian measures of support, which include:

(i) the difference measure: $c_d(H, e) = p(H|e) - p(H)$
(ii) the ratio measure: $c_r(H, e) = p(H|e)/p(H)$
(iii) the likelihood-ratio measure: $c_l(H, e) = p(e|H)/p(e| \sim H)$
(iv) the log ratio measure: $c_{log-r}(H, e) = \log[p(H|e)/p(H)]$
(v) the log likelihood measure: $c_{log-l}(H, e) = \log[p(e|H)/p(e| \sim H)]$[14]

Confirmation measures (i), (ii), and (iv) require knowing $p(e|H)$, since the posterior, $p(H|e)$, is a term in the measures and, by Bayes' theorem, the likelihood is required to determine the posterior for any real case of confirmation in science; confirmation measures (iii) and (v) require knowing $p(e|H)$ since the term appears directly in the measures. For those classes of evidence–hypothesis relations in which a likelihood can be determined, a necessary condition for determining the absolute measure of support that the evidence provides to the hypothesis, measurable on an interval scale, is met. Other necessary conditions must be met for each of the confirmation measures; for example, to determine the support that the evidence provides to the hypothesis using (i), we would need to know $p(H)$—a notorious headache. But if these other necessary conditions are met, then in these classes of evidence–hypothesis relations the absolute measure of support that the evidence provides to the hypothesis can be determined. In such cases a stipulated limitation to an ordinal ranking of confirmations would be unduly restrictive. Ordinality, in these classes of evidence–hypothesis relations, is, for good reasons, not satisfied.

However, there is a class of evidence–hypothesis relations such that determining an absolute measure of support that the evidence provides to the hypothesis is impossible. Elsewhere I argue that this class of evidence–hypothesis relations is large, at least in the empirical sciences. In a nutshell, the argument is as follows. As discussed above, all plausible absolute measures of confirmation depend on knowing the likelihood, $p(e|H)$. However, there are numerous features of evidence (such as the quality,

14 On these measures, see e.g. Fitelson (1999).

relevance, and salience of the evidence; the theoretical plausibility of the evidence; patterns in the evidence; and concordance with other evidence) that must be weighed and variably prioritized when assessing evidence, and there are numerous equally rational yet contradictory ways to do so; thus the probability of the evidence under the assumption that a particular hypothesis is true—that is, the likelihood—is, at least in many cases in the empirical sciences, indeterminate. In this class of evidence–hypothesis relations, then, absolute measures of confirmation are impossible to determine. But for this class of evidence-hypothesis relations (or at least a significant subset), ordinal rankings of confirmation (what I above call confirmation orderings) may still be possible.

I am not the first to note the difficulty with determining likelihoods. In Earman's critical examination of a Bayesian account of the Duhem–Quine problem, he writes: ". . . while much of the attention on the Bayesian version of the problem has focused on the assignments of prior probabilities, the assignments of likelihoods involves equally daunting difficulties" (1992). Similarly, Glymour claims that determinate likelihoods are possible only in rare circumstances, and in most empirical science "no such immediate and natural alternative distribution of degree of belief is available" (1980). For much of empirical science we cannot determine precise, absolute measures of confirmation on an interval scale. At least in such cases it is reasonable to limit an AF to ordinal rankings of comparative confirmation orderings among multiple competing hypotheses.[15]

One might worry that the considerations raised against the possibility of absolute measures of confirmation are equally problematic for determining the ordinality of confirmation among multiple competing hypotheses. After all, in Sect. 4 confirmation ordering was explicated in terms of absolute measures of confirmation (likelihoodist and Bayesian measures). However, absolute measures are inherently richer in information than ordinal measures: one can derive ordinal measures from absolute measures, but not vice versa. So there is 'something more' needed to justify an absolute measure over than an ordinal measure, and the arguments above are directed at that 'something more'. Another way of putting this is: despite the complexities of evidential assessment, when we have evidence for competing hypotheses, we often can *at least* know that H_1 is confirmed more than H_2, but we usually cannot know the precise values of these confirmations. Comparative confirmation is not necessarily derived from (or reducible to) respective absolute measures of confirmation. My argument justifies skepticism in absolute measures of confirmation for a class of evidence-hypothesis relations, but not wholesale skepticism that evidence cannot sort out better from worse confirmed hypotheses.

For many empirical situations it is reasonable to limit an AF to ordinal rankings of hypotheses, since for many empirical situations anything beyond ordinal rankings of hypotheses—for example, measures of confirmations on absolute scales—is

15 An anonymous reviewer notes that this problem might be mitigated by including the relevant background assumptions when assessing the degree of confirmation. Elsewhere I argue that this will help only in limited circumstances. For a Bayesian treatment of auxiliary hypotheses see Fitelson and Waterman (2005), Strevens (2001), Fitelson and Waterman (2007), and Rowbottom (2010).

impossible, and it is reasonable to limit our methods to what is possible.[16] In these cases, the Ordinality aspect of the Independence of Irrelevant Alternatives axiom is reasonably satisfied, as is the Ordered Output axiom. And, though arguing for the point here would take me far afield, such cases are ubiquitous in science.

6.5 Summary

In this section I have discussed the four axioms required for the impossibility theorem for amalgamating evidence. Although sometimes a truth-conducive AF might not satisfy one or more of the axioms, nonetheless the axioms should be considered either generally desirable features of an AF (as I argued is the case for Unrestricted Domain, Non-Dictatorship, and, usually, Independence and Unanimity), or features of an AF which we are often stuck with (as I argued is the case with Ordinality). Moreover, the occasional mismatch between the truth-conduciveness of an AF and its satisfaction of the desiderata further strengthens the analogy between amalgamating evidence and amalgamating preferences, because there are similar occasional mismatches in the preference case.

7 Conclusion

I have argued that there is an analogy between amalgamating individuals' preferences into a group decision—a subject which has been developing technically for the past several decades—and amalgamating multimodal evidence into a confirmation ordering of competing hypotheses. Just as the former faces Arrow's famous impossibility theorem, I here prove that amalgamating multimodal evidence into a confirmation ordering of competing hypotheses faces an analogous impossibility theorem. The axioms of the theorem are either generally desirable features of an AF or represent the best that is realistically available to us (although the axioms are not exceptionless requirements of rationality)—this too is another parallel in the analogy between amalgamating preferences and amalgamating evidence. Besides proving the impossibility theorem itself, this paper more generally shows that at least some of the technical

[16] I thank an anonymous reviewer for noting an ambiguity in the strength of my original defense of ordinality. To clarify I distinguish between two claims. I will call the first the Weak Ordinality Thesis (WOT), which urges that since we cannot assume that we always have absolute measures of confirmation it is better to assume that for particular cases we only have a confirmation ordering. I will call the second the Strong Ordinality Thesis (SOT), which urges that since we cannot assume that we always have absolute measures of confirmation, the only thing we can know we always have is a confirmation ordering. For a *general* defense of the ordinality axiom, SOT, but not WOT, would suffice; however, WOT, but not SOT, is supported by the present argument. SOT depends on the additional assumption that whenever we lack absolute measures of confirmation we can at most have a confirmation ordering. If, on the other hand, we can have confirmation information stronger than an ordering but weaker than an absolute measure of confirmation, then SOT is false and the ordinality axiom would need an independent argument for a general justification. I am not aware of a general way to satisfy the antecedent of the preceding conditional. A modest view, perhaps, would be to claim that there is no general fact about the strength of information available to determine the strength of particular confirmation relations.

results in social choice theory have plausible analogues for confirmation theory. The potential to utilize these richly-developed results from social choice theory is, I hope, promising.

The general conclusion of the argument presented here should not be seen as overly pessimistic. The argument is simply based on a piece of logic best construed as a no-go theorem which directs our attention to assessing the general plausibility of its axioms. In any situation in which the axioms are satisfied, the theorem applies. I have argued for the frequent (but not universal) applicability of the axioms in empirical science, but a thorough defense of this is a task for another time.

Acknowledgements I am grateful to Franz Dietrich for extensive assistance with the notation and proof, and to Samir Okasha, Nancy Cartwright, and two anonymous reviewers for detailed feedback. This paper has also benefited from discussion with Bruce Glymour, William Harper, Christian List, Tarun Menon, Boaz Miller, Wayne Myrvold, Daniel Steel, Katie Steele, and Eran Tal.

Appendix: Proof

I include a formal proof of the theorem to make the formal analogy with preference amalgamation explicit. I rely on the notation and definitions introduced in Sect. 4, and a few additional pieces of terminology are used in the proof: A hypothesis H is 'top' in a confirmation relation if it is ranked (strictly) above all other hypotheses. A hypothesis H is 'bottom' in a confirmation relation if it is ranked (strictly) below all other hypotheses. A hypothesis H is 'extreme' in a confirmation relation if it is top or bottom. Finally, the formal statement of the AF desiderata are as follows (with the informal statement repeated in parentheses for ease of reference):

Unanimity (U)
For all profiles $(\succcurlyeq_1, \ldots, \succcurlyeq_n)$ and every pair of hypotheses H_1, H_2 in $\mathbf{H}$, if $\mathrm{H}_1 \succ_i \mathrm{H}_2$ for all modes i, then $\mathrm{H}_1 \succ \mathrm{H}_2$. (Informally: If all modes prefer one hypothesis over another, then the AF must do the same.)

Independence of Irrelevant Alternatives (I)
For every pair of hypotheses $\mathrm{H}_1,\ \mathrm{H}_2$ in $\mathbf{H}$ and every pair of profiles $(\succcurlyeq_1, \ldots, \succcurlyeq_n)$, $(\succcurlyeq'_1, \ldots, \succcurlyeq'_n)$, if [for all modes i the relations $\succcurlyeq_i$ and $\succcurlyeq'_i$ coincide on $\{\mathrm{H}_1,\ \mathrm{H}_2\}$] then $\succcurlyeq$ and $\succcurlyeq'$ coincide on $\{\mathrm{H}_1,\ \mathrm{H}_2\}$. Two relations coincide on $\{\mathrm{H}_1,\ \mathrm{H}_2\}$ when they either both rank H_1 strictly over H_2 or both rank H_2 strictly over H_1 or both rank H_1 and H_2 equally. (Informally: The way two hypotheses are ranked relative to each other by an AF depends only on how the individual modes rank these two hypotheses relative to each other, and not on how they rank them relative to other hypotheses.)

Ordered Output (O)
AF generates an order (i.e., a transitive, reflexive and connected relation) for every profile.

Non-Dictatorship (D)
There is no mode i such that $\mathrm{AF}(\succcurlyeq_1, \ldots, \succcurlyeq_n) = \succcurlyeq_i$ for all profiles $(\succcurlyeq_1, \ldots, \succcurlyeq_n)$. (Informally: No mode is dictatorial.)

Theorem *There exists no amalgamation function satisfying U, I, O, and D.*

The proof of the impossibility theorem for confirmation follows a strategy of a proof of Arrow's Theorem by Geanakoplos (2005), and proceeds in four steps. The strategy is to assume U, I and O, and then to prove the violation of D, i.e., the existence of a dictator.

Step 1 If a hypothesis H_1 is extreme in every confirmation order of a profile, then it is also extreme in the AF output.

Proof Assume the contrary: suppose hypothesis H_1 is extreme in all orders of profile $(\succcurlyeq_1, \ldots, \succcurlyeq_n)$ but not extreme in the output relation. Then there exist hypotheses H_2 and H_3 such that $H_2 \succ H_1$ and $H_1 \succ H_3$. So, by transitivity, $H_2 \succ H_3$. We may assume without loss of generality that all modes i have $H_3 \succ_i H_2$. Indeed, if this were not the case we could modify the profile so that it becomes true while retaining how each mode ranks H_1 relative to any other hypothesis; by (I), this modification would not affect how H_1 is ranked compared to any other hypothesis in the AF output. By (U), $H_3 \succ H_2$, in contradiction to $H_2 \succ H_3$. □

Step 2 For every hypothesis H there exists a mode i which is 'pivotal' for H; that is, i is the earliest mode such that H is top in the AF output relation for every profile in which all modes up to i have H top and all following modes have H bottom.

Proof Let H be an arbitrary hypothesis. Consider any fixed profile $(\succcurlyeq_1, \ldots, \succcurlyeq_n)$ in which H is bottom in every confirmation order. I will call this "profile 0": see Table 1 for a graphical representation of profiles. By (U), H must be bottom in $\mathrm{AF}(\succcurlyeq_1, \ldots, \succcurlyeq_n)$. For every mode j let $\succcurlyeq'_j$ be the confirmation order in which H is top and any two other hypotheses are ranked just as in $\succcurlyeq_j$. Define i as the *earliest* mode with the property that H is not bottom in $\mathrm{AF}(\succcurlyeq'_1, \ldots, \succcurlyeq'_i, \succcurlyeq_{i+1}, \ldots, \succcurlyeq_n)$. There exists such an i because, by (U), H is top in $\mathrm{AF}(\succcurlyeq'_1, \ldots, \succcurlyeq'_n)$. Call the profile $(\succcurlyeq'_1, \ldots, \succcurlyeq_i, \succcurlyeq_{i+1}, \ldots, \succcurlyeq_n)$ "profile I", and the profile $(\succcurlyeq'_1, \ldots, \succcurlyeq'_i, \succcurlyeq_{i+1}, \ldots, \succcurlyeq_n)$ "profile II" (see Table 1). By Step 1, since H is extreme and H is not bottom in $\mathrm{AF}(\succcurlyeq'_1, \ldots, \succcurlyeq'_i, \succcurlyeq_{i+1}, \ldots, \succcurlyeq_n)$, H is top in $\mathrm{AF}(\succcurlyeq'_1, \ldots, \succcurlyeq'_i, \succcurlyeq_{i+1}, \ldots, \succcurlyeq_n)$. Finally, i is pivotal for H because, by (I), H is top in the AF relation for *every* profile which shares with $(\succcurlyeq'_1, \ldots, \succcurlyeq'_i, \succcurlyeq_{i+1}, \ldots, \succcurlyeq_n)$ the property that all modes up to i have H top and all following modes have H bottom. □

Step 3 For every hypothesis H, there is a mode i which is dictatorial over any pair of hypotheses other than H.

Proof Consider any hypothesis H, and let i be the mode which is pivotal for H (i must exist, by Step 2). Consider any hypotheses H_1 and H_2 distinct from H, and let $(\succcurlyeq_1, \ldots, \succcurlyeq_n)$ be an arbitrary profile. Without loss of generality H_1 and H_2 are distinct and $H_1 \succ_i H_2$ (if $H_2 \succ_i H_1$ then simply exchange the roles of H_1 and H_2 in the argument). By (I) we may assume that $H_1 \succ_i H \succ_i H_2$ (since the way the output relation ranks H_1 and H_2 relative to each other is not affected by how these hypotheses compare to H by mode i). We may also assume without loss of generality that H is top in all modes earlier than i and bottom in all

Table 1 Profiles constructed in proof

Mode	Profile			
	0	I	II	III
1	$\ldots \succ_1$ H	H $\succ_1 \ldots$	H $\succ_1 \ldots$	H $\succ_1 \ldots$
2	$\ldots \succ_2$ H	H $\succ_2 \ldots$	H $\succ_2 \ldots$	H $\succ_2 \ldots$
$\ldots$				
i	$\ldots \succ_i$ H	$\ldots \succ_i$ H	H $\succ_i \ldots$	$\ldots$H$_1 \succ_i \ldots$H$\ldots \succ_i$ H$_2\ldots$
$i+1$	$\ldots \succ_{i+1}$ H	$\ldots \succ_{i+1}$ H	$\ldots \succ_{i+1}$ H	$\ldots \succ_{i+1}$ H
$\ldots$				
n	$\ldots \succ_n$ H	$\ldots \succ_n$ H	$\ldots \succ_n$ H	$\ldots \succ_n$ H
AF	$\ldots \succ$ H	$\ldots \succ$ H	H $\succ \ldots$	H$_1 \succ$ H and H $\succ$ H$_2$, so H$_1 \succ$ H$_2$

Each cell in the table means that the hypotheses listed in the cell have a confirmation ordering, by the mode listed in the left-most column, indicated with the indexed ordering symbol $\succ_i$, the hypothesis variables, and the ellipsis. Thus in the first cell of Profile 0, "$\ldots \succ_1$ H" means "hypothesis H is the least confirmed hypothesis ('bottom'), by mode 1"

modes following i, again because by (I) we can change the rank ordering of H as long as the respective ordering of H$_1$ and H$_2$ do not change. The profile we are considering I will call "profile III". I have already shown that H is bottom in the AF output of profile I and that H is top in the AF output of profile II (since that was how we defined i which we use in the current step). In profile III all H $-$ H$_1$ pairs are ordered as they were in profile I, so, since H is bottom in AF(profile I), in profile III it must be that H$_1$ $\succ$ H, by (I). In profile III all H $-$ H$_2$ pairs are ordered as they were in profile II, so, since H is top in AF(profile II), it must be that in profile III, H $\succ$ H$_2$, by (I). So, by transitivity, H$_1$ $\succ$ H$_2$. To summarize this step: we assumed H$_1$ $\succ_i$ H (which by (I) should not affect the AF ordering of the H$_1$ $-$ H$_2$ pair), and we assumed H$_1$ $\succ_i$ H$_2$ and showed H$_1$ $\succ$ H$_2$. ◻

Step 4 Some mode is dictatorial over *every* pair of distinct hypotheses H$_1$ and H$_2$.

Proof By Step 3, for every hypothesis H there is a mode i which is dictatorial over all hypotheses other than H; call this mode iH. I must show that iH is the same for all H. Suppose for a contradiction that H and H' are hypotheses such that iH and iH' are distinct. Pick any hypothesis H" distinct from H and from H'. Consider a profile $(\succcurlyeq_1, \ldots, \succcurlyeq_n)$ in which H' $\succ_{iH}$ H" and H" $\succ_{iH'}$ H and H $\succ_{iH''}$ H' (this is possible because iH, iH' and iH'' are not all the same mode). Then, by the local dictatorship properties of iH, iH' and iH'' demonstrated in step 3, H' $\succ$ H" and H" $\succ$ H and H $\succ$ H'. This cycle violates transitivity: a contradiction. This completes Step 4, and Step 4 completes the proof. ◻

Table 1 visually depicts the Profiles constructed in the proof.

References

Arrow, K. (1950). A difficulty in the concept of social welfare. *Journal of Political Economy, 58*(4), 328–346.

Arrow, K. (1951). *Social choice and individual values*. New York: John Wiley and Sons.

Baumann, P. (2005). Theory choice and the intransitivity of 'is a better theory than'. *Philosophy of Science, 72*(1), 231–240.

Black, D. (1958). *The theory of committees and elections*. New York: Cambridge University Press.

Carnap, R. (1947). On the application of inductive logic. *Philosophy and Phenomenological Research, 8*, 133–148.

Cartwright, N. (1983). *How the laws of physics lie*. Oxford: Oxford University Press.

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

Earman, J. (1992) *Bayes or bust? A critical examination of Bayesian confirmation theory*. Cambridge: MIT Press.

Fitelson, B. (1999). The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science, 66*, S362–S378.

Fitelson, B. (2007). Likelihoodism, Bayesianism, and relational confirmation. *Synthese, 156*, 473–489.

Fitelson, B., & Waterman, A. (2005). Bayesian confirmation and auxiliary hypotheses revisited: A reply to Strevens. *The British Journal for the Philosophy of Science, 56*, 293–302.

Fitelson, B., & Waterman, A. (2007). Comparative Bayesian confirmation and the Quine-Duhem problem: A rejoinder to Strevens. *The British Journal for the Philosophy of Science, 58*, 333–338.

Geanakoplos, J. (2005). Three brief proofs of arrow's impossibility theorem. *Economic Theory, 26*, 211–215.

Glymour, C. (1980). *Theory and evidence*. Princeton: Princeton University Press.

Hacking, I. (1965). *The logic of statistical inference*. New York: Cambridge University Press.

Hawthorne J. (forthcoming). Confirmation theory. In P. S. Bandyopadhyay & M. Forster (Eds.), *Philosophy of statistics: Handbook of the philosophy of science* (Vol. 7). Oxford: Elsevier

Howson, C., & Urbach, P. (1989). *Scientific reasoning: The Bayesian approach*. La Salle, IL: Open Court.

Keeley, B. (2002). Making sense of the senses: Individuating modalities in humans and other animals. *The Journal of Philosophy, 99*(1), 5–28.

Leitgeb, H., & Segerberg, K. (2007). Dynamic doxastic logic: Why, how, and where to?. *Synthese, 155*, 167–190.

MacKay, A. F. (1980). *Arrow's theorem: The paradox of social choice*. New Haven: Yale University Press.

Mayo, D. (1996). *Error and the growth of experimental knowledge*. Chicago: University of Chicago Press.

Okasha, S. (2011). Theory choice and social choice: Kuhn versus arrow. *Mind, 477*, 83–115.

Royall, R. (1997). *Statistical evidence: A likelihood paradigm*. New York: Chapman & Hall.

Rowbottom, D. (2010). Corroboration and auxiliary hypotheses: Duhem's thesis revisited. *Synthese, 177*, 139–149.

Sen, A. (1970). *Collective choice and social welfare*. San Francisco: Holden-Day.

Sober, E. (2008). *Evidence and evolution: The logic behind the science*. New York: Cambridge University Press.

Staley, K. (2004). Robust evidence and secure evidence claims. *Philosophy of Science, 71*, 467–488.

Stegenga, J. (2009). Robustness, discordance, and relevance. *Philosophy of Science, 76*, 650–661.

Strevens, M. (2001). The Bayesian treatment of auxiliary hypotheses. *The British Journal for the Philosophy of Science, 52*, 515–537.

Weber, M. (2005). *Philosophy of experimental biology*. Cambridge: Cambridge University Press.