Independent opinions? On the causal foundations of belief formation and jury theorems

Dietrich, Franz and Spiekermann, Kai

CNRS, Paris, France, University of East Anglia, Norwich, U.K., London School of Economics, London, U.K.

July 2012

Independent Opinions?

On the causal foundations of belief formation and jury theorems

Franz Dietrich
CNRS & UEA
&
Kai Spiekermann
LSE

Abstract

It is often claimed that opinions are more likely to be correct if they are held independently by many individuals. But what does it mean to hold *independent* opinions? To clarify this condition, we distinguish four notions of probabilistic opinion independence. Which notion applies depends on environmental factors such as commonly perceived evidence, or, more formally, on the causal network in which people interact and form their opinions. In a general theorem, we identify conditions on this network that guarantee opinion independence in each sense. Our results have implications for 'wisdom of crowds' arguments, as we illustrate by providing old and new jury theorems.

1   Introduction

What exactly does it mean to have independent opinions or beliefs on a factual question? And under what circumstances can we expect opinions of individuals to be independent from each other? The concept of opinion independence is not well understood, even though it is crucial in social epistemology.

Opinion independence is of central importance in the context of 'wisdom of crowds' arguments and, more formally, jury theorems. The claim that 'crowds are wise' stems from the idea that decisions based on many opinions are more likely to be correct than decisions based on a few or just one opinion (e.g., Surowiecki 2004). Much of the trust in the judgment of large electorates, for instance, comes from the idea that a judgment is likely to be correct if it is approved by many voters. Similarly, in a court case a single witness may well be mistaken, but twenty witnesses who all say the same may not, so the intuition. But these arguments assume that the group aggregates (sufficiently) *independent* individual opinions. Opinion dependence can undermine the 'wisdom of crowds': if, for instance, most individuals blindly follow the same 'opinion leader', then the majority is no 'wiser' than that opinion leader. However, exaggerated influence of opinion leaders is only one form of harmful opinion dependence; other popular forms are information cascades and systematic biases.

---

[1] The formal results presented in this paper are the work of Franz Dietrich.

In threatening the 'wisdom of crowds', opinion dependence ultimately threatens the epistemic superiority of democratic decision-making bodies and institutions. The importance of opinion independence for 'wisdom of crowds' arguments becomes explicit in *jury theorems*. Indeed, such theorems typically make an 'independence' assumption (as well as an assumption of voter 'competence'), and they then conclude that 'crowds are wiser' in the technical sense that larger groups are more likely to be correct in majority than smaller groups or single individuals.

But opinion independence is not only relevant in the aggregative context of the 'wisdom of crowds' and jury theorems. It also matters, for instance, when studying the effect of group deliberation on opinions, regardless of whether any voting or aggregation takes place. And finally, it is important to carefully define the notion of opinion independence for the sake of conceptual clarity, quite apart from its relevance elsewhere.

In a first step, we need to distinguish between causal and probabilistic dependence. The former states that the opinions causally affect each other, the latter that they display probabilistic correlations. Causation is clearly distinct from correlation; yet the two notions are intertwined.[2] Indeed, whether and how opinions are probabilistically dependent is very much determined by causal interconnections between individuals and their environment. This paper considers four possible notions of probabilistic independence (Section 2). It determines in a general theorem which notion applies, depending on the underlying causal relations between individuals and other relevant factors (Section 3). As an application, we finally present two jury theorems with different independence conditions (Section 4).

Several technical treatments of opinion independence can be found in the literature on jury theorems; see for instance Boland (1989), Boland *et al.* (1989), Ladha (1992, 1993, 1995), Berg (1993), Dietrich and List (2004), Dietrich (2008), Kaniovski (2010) and Dietrich and Spiekermann (2010) (and for jury theorems more generally see for instance Grofman et al. 1983, Nitzan and Paroush 1984, List and Goodin 2001, Bovens and Rabinowicz 2006, Romeijn and Atkinson 2011, Hawthorne unpublished). But opinion dependence has also been discussed less formally in political philosophy, especially with reference to epistemic and deliberative democracy. Contributions include Grofman and Feld (1988), Estlund et al. (1989), Estlund (1994, 2008), Anderson (2006), Vermeule (2009, ch. 1) and Spiekermann and Goodin (2012). We approach the problem from a causal angle, which is crucial for gaining a deeper understanding of independence. To do so, we draw on causal networks. Causal network reasoning has been employed before in the context of jury theorems (see Dietrich and List 2004, Dietrich 2008, and Dietrich and Spiekermann 2010), but a general account of probabilistic independence in terms of causal interactions is still missing.

Our causal approach highlights the effect of social practices and institutions on opinion independence, and as such on the epistemic quality of these practices and institutions. In assuming an external standard for epistemic quality and correctness

---

[2]For instance, two phenomena that do not causally affect each other can be correlated because they have a common cause. Here, we do not commit ourselves to any specific account of causation and our formal (network-theoretic) representation of causation is compatible with different metaphysical accounts.

we are in line with the correspondence theory of truth and Alvin Goldman's influential "veritism" approach in social epistemology (Goldman 1999, 59, 79ff.; 2004).

2 Four different independence conditions and their causal motivations

We assume that some individuals, labeled $i = 1, 2, 3, ...$, must form opinions on a given issue.[3] This might arise in the context of deciding between two alternatives, such as: convict or acquit the defendant in a court trial, predict that global warming will continue or that it will not, and so on. The opinions may, for instance, serve as votes in a formal voting procedure, or as inputs or outputs of group deliberation.

In the simple base-line case, our model involves
- an *opinion* of each voter, which can take only two possible values (e.g., 'guilty' or 'innocent', or 'yes' or 'no', or '1' or '0'),
- the *state (of the world)*, which represents the objectively correct opinion and which therefore can take the same two values, and
- the *correctness notion,* according to which an opinion is either 'correct' or 'incorrect' depending on whether or not it matches the state.

Since our framework is probabilistic, phenomena – such as opinions and the state – are outcomes of random variables (with an underlying probability function denoted Pr). We thus formally consider a random variable $\mathbf{x}$ generating the state of the world, and random variables $\mathbf{o}_1$, $\mathbf{o}_2$, ... generating the opinions of individuals 1, 2, ... Note our convention of using bold face letters to denote random variables. In our mentioned base-line case, the state $\mathbf{x}$ and the opinions $\mathbf{o}_1$, $\mathbf{o}_2$, ... all range over the same binary set (e.g., the set {'yes', 'no'} or $\{1, 0\}$), where an opinion $\mathbf{o}_i$ is thought of as being correct if $\mathbf{o}_i = \mathbf{x}$. All our illustrations will follow this binary base-line case (as does the literature on jury theorems). But our model is much more general:
- The opinions $\mathbf{o}_1$, $\mathbf{o}_2$, ... need not be binary and might, for instance, be sets of believed propositions (belief sets or judgment sets), numerical estimates (say, of temperature), or even degrees of belief.
- The state $\mathbf{x}$ need not be binary and might, for instance, be a set of true propositions, the true temperature, or an objective probability or probability function.[4]
- The notion of correctness of opinions need not be binary, i.e., there may be multiple degrees of objective correctness (rightness, goodness etc.[5]) For instance,

---

[3]The total number of individuals does not matter for us. Technically, it may even be taken to be countably infinite, as in jury theorems.

[4]In fact, the state and the opinions may be two different kinds of objects, i.e., range over a different sets – which is philosophically natural since an opinion differs conceptually from its truth-making fact, i.e., the state. For instance, the opinion could be 'guilty' or 'innocent', and the state could be the fact of what exactly the defendant did on 7 January 2010.

[5]Different notions of objectivity are compatible with that assumption, as long as the fact about the correct opinion is not determined by the actual opinions of people. This excludes procedural notions of quality of opinion, where an opinion is correct *if and because* it matches the opinion that arose collectively by applying an appropriate procedure. It also excludes constructivist notions of quality of opinion where an opinion's quality is constitutively determined by the opinions of the agents.

a temperature estimate (the opinion) is correct to a degree given by its proximity to true temperature (the state); and, in a presumably more controversial application, subjective probabilities (the opinion) are correct to the extent to which they resemble the objective or 'rational' probabilities (the state).[6]

In general, the variables $\mathbf{o}_1, \mathbf{o}_2, \ldots$ thus range over an arbitrary set of possible opinions, and the state $\mathbf{x}$ ranges over an arbitrary set of possible states. We do not yet formalize the correctness notion, as it is not needed to analyse opinion dependence, but only to state jury theorems.[7]

The simplest independence assumption one might come up with refrains from conditionalizing on any background information:

Unconditional Independence (UI). The opinions $\mathbf{o}_1$, $\mathbf{o}_2$, ... are unconditionally independent.

Counterexamples to UI are easily constructed. In short, since opinions are typically (indeed, hopefully) correlated with the state, they are usually correlated with each other. To see, for instance, why the first two individuals (jurors) in our court trial example presumably hold positively correlated rather than independent opinions, note that a 'guilty' opinion of juror 1 typically raises the probability of a 'guilty' opinion of juror 2 – formally, $\Pr(\mathbf{o}_2 = \text{'guilty'}|\mathbf{o}_1 = \text{'guilty'}) > \Pr(\mathbf{o}_2 = \text{'guilty'})$ – because juror 1's 'guilty' opinion raises the probability that the state is that of a guilty defendant (assuming juror 1 is competent), which in turn raises the probability that juror 2 holds a 'guilty' opinion (assuming juror 2 is competent). This is a clear violation of UI, which would have required that opinions are of no informational relevance to each other.

Note that this argument implicitly assumes that the state $\mathbf{x}$ has a causal effect on each opinion, as indicated in Figure 1. In this (and all following) plots, only
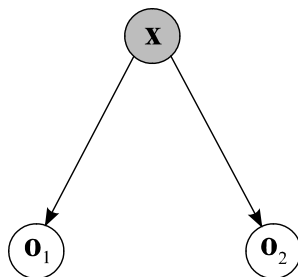

Figure 1: The state is a direct cause of the opinions.

two opinions are shown for simplicity. The arrows represent causal relationships,

---

[6]The literature on jury theorems rarely considers more than two opinions and has perhaps never considered more than two correctness levels. List and Goodin (2001) use many opinions but only two correctness levels (where exactly one opinion is correct and all others are incorrect without further refinement).

[7]A formalization could include a set S of possible 'correctness levels' (e.g., the set {'correct', 'incorrect'} for a binary correctness notion) and a function mapping each opinion-state pair (o, x) to the correctness level (in S) of opinion o in state x.

4

pointing from the causing variable to the affected variable. In Figure 1 the opinions are probabilistically dependent 'through' their common cause $\mathbf{x}$.

That UI is easily violated should not surprise scholars familiar with the Condorcet Jury Theorem, given that this theorem does *not* assume that opinions are unconditionally independent but (usually) that they are state-conditionally independent. What is more surprising is that UI *does* hold in some circumstances, but we postpone this issue for now and turn to the more classical state-conditional notion of independence:

State-Conditional Independence (SI). The opinions $\mathbf{o}_1$, $\mathbf{o}_2$, ... are independent conditional on the state $\mathbf{x}$.[8]

This conditional notion of independence is the basis of Condorcet's classical jury theorem (e.g., Grofman et al., 1983), which can be summarized as follows (see Section 4 for details). Suppose a group performs a majority vote between two alternatives of which exactly one is correct. The correct alternative corresponds to our state $\mathbf{x}$, and the votes to our opinions $\mathbf{o}_1$, $\mathbf{o}_2$, ... Condorcet's jury theorem states that if SI holds and moreover in each state voters are (homogeneously) more often right than wrong, then the probability of a correct majority outcome increases in (odd) group size and converges to one.[9]

State-Conditional Independence says that once we know the state of the world the opinions do not anymore bear any information on each other. The earlier objection to UI – namely that the opinions of some of the people tell us something about what the state is likely to be, and hence about what other people are likely to believe – does not work against SI because we cannot learn anything new about the state if we have already conditionalized on it. The state plays the role of a common cause of the opinions. If the state is indeed the only common cause, SI is in line with Reichenbach's famous *Common Cause Principle*, which is often understood roughly as follows: if a common cause fully explains the correlations between certain events, then these events are independent conditional on the common cause (Reichenbach 1956, 159-60). In slightly more general terms:

---

[8]Conditional independence is defined like independence but with probabilities replaced by conditional probabilities. More precisely, as long as $\mathbf{x}$ is discrete (e.g., binary), independence conditional on $\mathbf{x}$ by definition means that for every value x which $\mathbf{x}$ may take (with positive probability) there is independence under the conditional probability function $\Pr(\cdot|x)$. Without discreteness restriction, the opinions are independent conditional on $\mathbf{x}$ if they are independent under the conditional probability measure $\Pr(\cdot|x)$ for all values x that the random variable $\mathbf{x}$ may take, except possibly from a set of values of $\mathbf{x}$ that occurs with zero probability. (The clause 'except...' appears for technical reasons related to the general mathematical definition of conditional probabilities, which takes care of the case in which $\mathbf{x}$ takes some or even all of its values with zero probability. We spare the reader the technicalities.)

[9]Some statements of Condorcet's classical jury theorem use an *unconditional* independence condition, but only in a quite different sense. *Either* the state is not modeled as a random variable but is assumed to take a given value (in which case the framework is different since all probabilities are implicitly posterior probabilities given the state, so that the independence is implicitly state-conditional). *Or* the assumption of unconditional independence is applied not to the opinions themselves but to the events of correct, i.e., state-matching opinions (which makes the assumption more plausible, although a convincing causal-network-theoretic justification seems hard to give).

Common Cause Principle (stated informally). Phenomena which do not causally affect each other are probabilistically independent conditional on their common causes.

While the Common Cause Principle at first sight supports SI, it can be turned against SI once we consider other causal networks in which **x** is not the only common cause of the opinions. Consider for instance the network in Figure 2. Here the
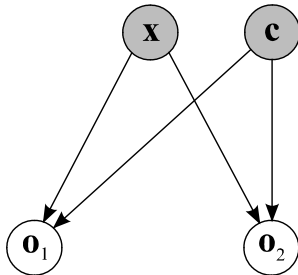

Figure 2: Multiple direct common causes of opinions.

opinions have two common causes, the state **x** and another cause **c**, which could be a factor like weather or room temperature. SI can now fail in much the same way as UI. Suppose for instance that the variable weather has an effect on each juror in a court trial: the sunnier weather is, the more the jurors see the good in the defendant, and hence the more they are inclined to form the opinion that the defendant is innocent. Now, even after having conditionalized on the state of the world that the defendant is innocent, the opinions of the jurors are informative on each another, this time through the common cause of weather; for instance, an opinion 'innocent' by the first juror increases the probability that weather is sunny and hence the probability that the second juror has the opinion 'innocent' too. In other words, the opinions are not state-conditionally independent but state-conditionally positively correlated, namely through the other common cause ('weather') on which SI fails to conditionalize.

This discussion suggests replacing SI by a notion of independence which conditionalizes on all common causes of the opinions. By doing so we 'control' for all factors that causally affect more than one opinion, eliminating the dependence induced by such common factors.

To state such a condition formally, let us extend the formal framework, which so far consists just of the state **x** and the opinions o_1, o_2, ... Now we consider these and any number of additional random variables (representing phenomena which are directly or indirectly causally related to the opinions), and we consider a causal network over the variables. Formally, a causal network over some variables is a so-called *directed acyclic graph* over these variables, that is, a set of directed arrows between pairs of variables (representing causal relevance) such that there is no directed cycle of arrows.[10] Figures 1 or 2 were examples of how the network might look like. Figure 3 is yet another example. Here, the state causally affects a variable **c**, which is interpretable as evidence (e.g., fingerprints, witness reports, etc.) and which in turn

_____
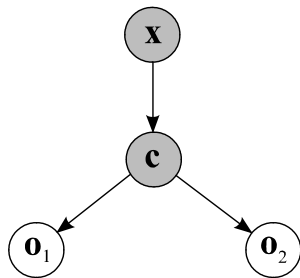[10]For thorough discussions of causal networks, see Pearl 2000, ch. 1.

Figure 3: The state is an indirect common cause of opinions.

influences each opinion. Individuals are thus affected by the state only indirectly, via the 'trace' which the state leaves in the form of c. The additional variables (such as the variable c in Figures 2 and 3) may be binary or multi-valued. For instance, the variable weather may take the values 'sunny', 'cloudy', 'rainy' and so on; and the variable 'body of evidence' may take several forms as well. Some variables (such as room temperature) might even range over a continuum of values.

In the causal network, a variable a is said to be a *direct cause* of another b (and b a *direct effect* of a) if there is an arrow pointing from a towards b ('$a \rightarrow b$'). Further, a is a *cause* of b (and b an *effect* of a) if there is a directed path from a to b, i.e., a sequence of two or more variables starting with a and ending with b such that each of these variables (except from the last one) directly causes the next one. For instance, in Figure 3 the state x directly causes c, and indirectly causes the opinions. Generally, when we use the verb "cause" we refer only to causal contribution; no sufficiency or necessity is implied.[11]

A variable is a *common cause* (*effect*) of some variables if it is a cause (effect) of each of them. By a 'common cause' simpliciter we mean a common cause of (two or more) opinions. In all figures, such common causes are shown in gray. While in Figures 1-3 all causes of opinions are common causes, Figure 4 contains four *private causes* of opinions; they causally affect just one opinion. Note also that in Figure 4 some of the causes of opinions (namely, c_2, c_4 and c_6) are non-evidential: they are not related to the state. Although this might be viewed as 'irrational', individuals are often influenced by non-evidential causes such as room temperature (a *common* non-evidential cause) or the quality of one's sleep last night (a *private* non-evidential cause).

Let us write χ (Greek 'chi') for the family of all common causes. In Figure 1 χ consists just of the state x; in Figures 2 and 3 of x and c; and in Figure 4 of x, c_3 and c_4. In general, χ is a compound random variable with as many parts as there are common causes of opinions.[12] We are now ready to state a new independence assumption, which is a direct application of the Common Cause Principle:

Common-Cause-Conditional Independence (CI). The opinions o_1, o_2, ... are

---

[11] More precisely, each variable (i.e., the probabilities of its values) is affected by its direct causes. Our network is a macroscopic simplification of the world and we do not commit ourselves to whether the world is fundamentally governed by probabilistic or deterministic processes.

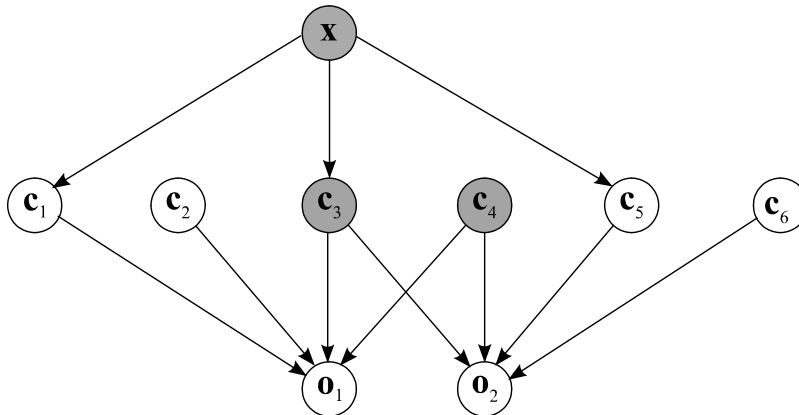[12] The range of χ is the Cartesian product of the ranges of the common causes of opinions.

Figure 4: Private and common causes of opinions.

independent conditional on the common causes $\boldsymbol{\chi}$.

This independence assumption may seem the most appealing one. It is backed by the Common Cause Principle and more generally by probabilistic theories of causality. With CI, the independence of opinions is guaranteed as long as the opinions do not causally affect each other. It has, however, a weakness in the context of jury theorems and 'wisdom of crowds' arguments. The problem with CI is not so much that it is not sufficiently justified – CI is perhaps the most justifiable independence assumption – but rather that CI (like UI) is a premise which does not easily lend itself to arguments that 'crowds are wise(r)'. Let us now explain this subtle point informally; in Section 4 we work it out more formally.

It is important to first realize that what matters ultimately in a jury theorem is not independence of the *opinions* but rather independence of the *events of correct opinions*. The typical reasoning is that a group whose members are independently more likely to get it right will quite probably get it right in majority. This reasoning involves independence of the events of holding correct opinions, not independence of the opinions (or votes) *simpliciter*.[13] Now, independence of the opinions *implies* independence of the correct opinion events once we have conditionalized on the state $\mathbf{x}$. It is easy to see why: conditional on the state being x, if the opinions $\mathbf{o}_1$, $\mathbf{o}_2$, ... are independent then so are the events that $\mathbf{o}_1$ matches x, $\mathbf{o}_2$ matches x, and so on. In other words, the assumption of State-Conditional Independence implies what is needed, namely (conditional) independence of the correctness events. Similarly, if it so happens that the state $\mathbf{x}$ features among the common causes $\boldsymbol{\chi}$ – as it does indeed in all of the above Figures 1-4 – Common-Cause-Conditional Independence also implies (conditional) independence of the correctness events. But there are plausible situations in which the state $\mathbf{x}$ is not a common cause. Figure 5 is one such case. Here the common cause $\mathbf{c}$ affects both the state and the opinions. As a plausible

---

[13]The kind of aggregative conclusion that independence of opinions *simpliciter* lends itself to is different and less relevant. One might for instance reason that a group of jurors who are independently more likely to express the opinion 'innocent' will quite probably hold this opinion in majority. But what matters is less the probability of an 'innocent' majority than that of a *correct* majority.
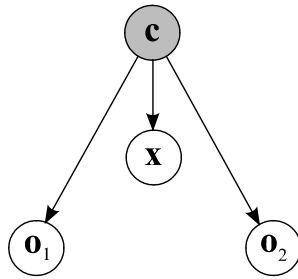
8

Figure 5: The state is not a cause of opinions.

example, imagine a homicide case in which the jurors learn that the defendant has bought cyanide (represented by $\mathbf{c}$). This fact is a common cause of the opinions of the jurors, who take murder (and guilt) to be more likely if the defendant has bought cyanide in advance. Since having bought cyanide facilitates poisoning, the purchase causally affects not just the opinions but also whether the murder takes place; hence the network of Figure 5. Note that state $\mathbf{x}$ is not a cause of any opinions.

Whenever the state $\mathbf{x}$ is not a common cause, CI does not conditionalize on it, and therefore does not lend itself to jury-theorem-type arguments about the probability of majority correctness. So we have to add the state into the conditionalisation, just as Condorcet's jury theorem conditionalizes on the state by using SI rather than UI. To sum up, we have to conditionalize on all common causes plus the state. But what does this mean? Following Dietrich (2008) and Dietrich and Spiekermann (2010), the *decision problem* faced by the group can be conceptualized as being a description of two things:

- the fact to find out about, conceptualized as the state of the world;
- the circumstances (environment) in which people form opinions, conceptualized as the common causes influencing the opinions.

By conditionalising on the decision problem, we include the state by default (thus making sure that not only the opinions but as a consequence also the events of correct opinions are independent).

Formally, let us write $\boldsymbol{\pi}$ for the decision problem defined as a family containing the state $\mathbf{x}$ and all common causes. Clearly, the problem $\boldsymbol{\pi}$ reduces to the common causes $\boldsymbol{\chi}$ if the state is a common cause (as in Figures 1-4). In general $\boldsymbol{\pi}$ is isomorphic to the state-circumstances pair $(\mathbf{x}, \boldsymbol{\chi})$.[14] We are now in a position to state our final independence condition:

Problem-Conditional Independence (PI). The opinions $\mathbf{o}_1$, $\mathbf{o}_2$, ... are independent conditional on the problem $\boldsymbol{\pi}$.

This assumption is put to work in a jury theorem presented in Section 4.

---

[14]This pair contains the state twice if the state is among the common causes, but such a redundancy poses no problem.

9

3 The causal foundation of each independence condition: a general theorem

While the last section has given informal causal motivations for the four independence conditions, this section turns to a formal result. The result gives us precise sufficient (and in fact essentially necessary) conditions on causal interconnections for each independence condition to hold. Given this result, once we know the individuals' causal environment we can infer which kinds of opinion independence should (not) be assumed. And if a social planner can design the environment, he can do so to induce the kind of independence he aims for.

To be able to infer probabilistic features from causal interconnections, one must of course assume that probabilities are compatible with the causal network. What such compatibility amounts to has been settled precisely in the theory of causal (and Bayesian) networks (e.g., Pearl 2000). Formally, probabilities (more precisely: the joint probability distribution of the variables) are *compatible* with the causal network if the so-called *(Parental) Markov Condition* holds: any variable in the network is independent of its non-effects[15] conditional on its direct causes. For instance, in Figure 1 opinion $\mathbf{o}_1$ is independent of opinion $\mathbf{o}_2$ conditional on the direct cause $\mathbf{x}$; in Figure 2, $\mathbf{o}_1$ is independent of $\mathbf{o}_2$ given its direct causes $\mathbf{x}$ and $\mathbf{c}$; in Figure 3, $\mathbf{o}_1$ is independent of both $\mathbf{o}_2$ and $\mathbf{x}$ conditional on the only direct cause $\mathbf{c}$; and so on. Note the importance of causal independence between the opinions for (probabilistic) opinion independence: if $\mathbf{o}_1$ had a causal effect on $\mathbf{o}_2$ then the Markov Condition would not imply that $\mathbf{o}_1$ is conditionally independent of $\mathbf{o}_2$.

The following theorem gives causal conditions for our last two independence conditions; the first two conditions are dealt with by a corollary below.

Theorem 1. Suppose probabilities are compatible with the causal network, and no opinion is a cause of any other opinion. Then:
  (a) Common-Cause-Conditional Independence holds;
  (b) Problem-Conditional Independence holds if the state is not a common effect of any opinions or private causes thereof.

Part (a) is an instance of the Principle of Common Cause and as such should come without surprise to specialists.[16] Part (b) settles the question of how the state should (not) be causally related to the opinions in order for independence to be preserved after conditionalising also on the state (in addition to the common causes). The condition stated in part (b) requires that the state is not a common effect of variables each of which is or privately causes a different opinion.[17]

---

[15] With the non-effects of a variable $\mathbf{a}$ we mean the variables which are not effects of $\mathbf{a}$ (and differ from $\mathbf{a}$).

[16] We nonetheless present a formal proof of part (a), given that standard renderings of the Common Cause Principle are often less general than our application in that they focus on (in)dependence between *only two* random variables and often assume that there is *only one* common cause. We allow several opinions and common causes.

[17] Equivalently, none of the following three cases obtains: (i) two (or more) opinions cause the state (as in Figure 6a); (ii) private causes of two (or more) opinions cause the state (as in Figure 6b); (iii) an opinion and a private cause of a different opinion cause the state.

Figure 6 gives counterexamples in which the state **x** *is* such a common effect. In
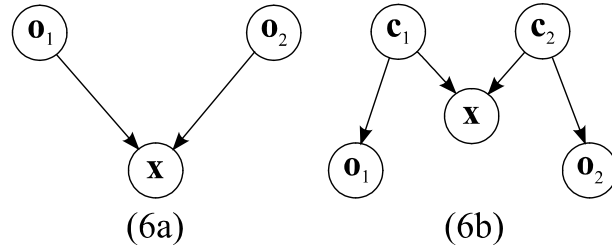

(6a) (6b)

Figure 6: Violations of the condition for Problem-Conditional Independence.

6a we see a causal setup where the state is a common effect of the opinions. A causal structure like 6a arises if the opinions *influence* the state. For instance, the prediction of a bank run might cause the bank run. Though interesting and sometimes very real, such cases violate one of the core assumptions of many theories of social epistemology (at least among those committed to a veritistic approach): the assumption that an external fact determines correctness. 'Self-fulfilling prophecies' are ruled out.

In Figure 6b, by contrast, the state is a common effect of private causes of opinions. To show the relevance of such a setup, we need a more complex example. Suppose an intelligence agency observes two different subjects at different ends of the town. The agency knows from reliable sources that if and only if both subject 1 leaves the house at noon (c_1) and subject 2 leaves the house at noon (c_2), the two subjects will have a conspiratorial meeting (**x**). One agent observes subject 1, another subject 2, and for security reasons they cannot directly communicate with each other. Both agents form opinions on whether the meeting will take place. Each agent's opinion is influenced only by his own observation (either c_1 or c_2), so that these two causes influence both the opinions and the state. This example shows that 6b is a plausible causal setup, but, again, the condition in clause (b) of Theorem 1 is violated and Problem-Conditional Independence should not be assumed. Indeed, PI is intuitively violated: conditional on the state, we can infer something about an agent's opinion if we learn about other opinions. For instance, if we know that the conspiratorial meeting does not take place (we conditionalize on **x** being 'no meeting') and we learn that agent 1 believes that the meeting will take place, we can infer that subject 1 has left the house. But since there is no meeting, we also infer that subject 2 stays at home and that agent 2 holds the corresponding opinion. We have learned something about 2's opinion from 1's opinion, a violation of PI. A plausible notion of opinion independence in cases like 6a and 6b must not conditionalize on the state. Therefore, only Common-Cause-Conditional Independence and hence – somewhat surprisingly, as there are no common causes – Unconditional Independence hold.

Although Theorem 1 seems to deal only with two of our independence conditions (CI and PI), an immediate corollary of part (a) gives us causal conditions for our other two independence conditions:

Corollary. Suppose probabilities are compatible with the causal network, and no opinion is a cause of any other opinion. Then:

- State-Conditional Independence holds if only the state is a common cause;
- Unconditional Independence holds if there are no common causes at all.

Notice how strong the causal conditions for SI and UI are. Among the above figures, only Figure 1 satisfies the condition for SI that the state is the only common cause, and only Figures 6a and 6b satisfy the 'no common cause' condition for UI. It might surprise that there exist plausible causal interconnections for which UI holds. Figure 6b is such a plausible network, as discussed above.

4 Application to jury theorems

We have mentioned informally that those of our independence assumptions which conditionalize (at least) on the state – namely, SI and PI – can be used in jury theorems, i.e., formal 'wisdom of crowds' arguments. The present section substantiates this claim by stating two simple jury theorems, namely Condorcet's classical jury theorem (which is based on SI) and a new jury theorem (which is based on PI).

To state jury theorems, we first need to enrich our formal framework by an additional ingredient: the notion of correctness of opinions. As before, for simplicity it is best to assume that there are only two correctness levels, 'correct' and 'incorrect', two states and two possible opinions, such that exactly one opinion is correct (though generalizations are compatible with our formal framework).

Formally, for each state x let some opinion o^x be specified as the 'correct' opinion in state x. We write R_i for the event that i's opinion is correct, i.e., the event that $\mathbf{o}_i = o^{\mathbf{x}}$. Jury theorems are concerned with the event that a majority of the group is correct, i.e., that more than half of the correctness events $R_1, ..., R_n$ hold. The question is how this probability of 'group wisdom' depends on the group size n (which we assume to always be odd, to avoid ties under majority voting[18]).

Jury theorems typically assume that the correctness events $R_1, R_2, ...$ are independent in some sense. However, causal reasoning of the sort presented above leads to independence *of the opinions* $\mathbf{o}_1, \mathbf{o}_2, ...$, rather than the correctness events. A key issue therefore is whether opinion independence (in one of our four senses) implies correctness independence (in the same sense).

The answer is negative for UI and CI. For instance, perfectly independent opinions (satisfying UI) may lead to correlated correctness events. For instance, consider independent opinions taking the value 1 or 0 with equal probabilities: $\Pr(\mathbf{o}_i = 1) = \Pr(\mathbf{o}_i = 0) = \frac{1}{2}$ for all agents i. Let the state be fully determined by the first two opinion such that it is 1 if $\mathbf{o}_1 = \mathbf{o}_2 = 1$ and 0 otherwise. (This could be the case in causal environments like those of Figure 6.) Then R_1 and R_2 are negatively correlated

---

[18]As usual, our two theorems can be generalized to possibly even group size n by assuming that ties are broken by tossing a fair coin.

rather than independent: $\Pr(R_1 \cap R_2) < \Pr(R_1)\Pr(R_2)$. The reason is that

$$
\begin{aligned}
\Pr(R_1 \cap R_2) &= \Pr(\mathbf{o}_1 = \mathbf{o}_2 = 1) + \Pr(\mathbf{o}_1 = \mathbf{o}_2 = 0) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}, \\
\Pr(R_1) &= \Pr(\mathbf{o}_1 = \mathbf{o}_2 = 1) + \Pr(\mathbf{o}_1 = \mathbf{o}_2 = 0) + \Pr(\mathbf{o}_1 = 0 \neq \mathbf{o}_2) \\
&= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}, \\
\Pr(R_2) &= \Pr(\mathbf{o}_1 = \mathbf{o}_2 = 1) + \Pr(\mathbf{o}_1 = \mathbf{o}_2 = 0) + \Pr(\mathbf{o}_2 = 0 \neq \mathbf{o}_1) \\
&= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4},
\end{aligned}
$$

where of course $\frac{1}{2} < \frac{3}{4} \times \frac{3}{4} = \frac{9}{16}$.

By contrast, the answer is positive for SI and PI, and more generally for any form of independence which conditionalizes at least on the state. Let us state this fact formally (the simple proof is trivial):

Proposition 1. For any family $\boldsymbol{\rho}$ of random variables containing $\mathbf{x}$ (e.g., for $\boldsymbol{\rho} = \mathbf{x}$ or $\boldsymbol{\rho} = \boldsymbol{\pi}$), if the opinions $\mathbf{o}_1, \mathbf{o}_2, ...$ are independent conditional on $\boldsymbol{\rho}$, then so are the correctness events $R_1, R_2, ...$

This fact is a key to the two jury theorems to be stated now. The first theorem combines SI with Condorcet's classical competence assumption:

Classical Competence. There is a parameter p in $(\frac{1}{2}, 1)$ such that for all states x the conditional correctness probability $\Pr(R_i|x)$ is p for all individuals i.[19]

This assumption states that in each given state the voters all have the same correctness probability of more than $\frac{1}{2}$ and less than 1.[20] Combined with SI (and Proposition 1), we may thus compare the correctness events $R_1, R_2, ...$ with independent tosses of the same coin biased towards the truth. This metaphor provides an intuition for the classical jury theorem:

Theorem 2. Under the assumptions SI and Classical Competence, the probability of a correct majority opinion strictly increases in group size and converges to one.

We now turn to a new jury theorem based on the assumption of problem- (rather than state-) conditional independence. For each value $\boldsymbol{\pi}$ of the problem $\boldsymbol{\pi}$ we consider the voter's problem-specific correctness probability, $\Pr(R_i|\pi)$, which is interpretable either as a measure of how 'easy' this problem is (for agents like i) or as the voter's competence *on problem* π (see Dietrich 2008 and Dietrich and Spiekermann 2010).

---

[19]This condition (and the resulting Theorem 2) can be generalized by allowing p to depend on the state x.

[20]Our statement of Classical Competence assumes that the conditional probability $\Pr(R_i|x)$ is defined for all states x, i.e., that each state x occurs with non-zero probability. This excludes that the state has a continuous distribution. Analogously, our statement of Easy/Hard Dichotomy below assumes that all problems π occur with non-zero probability. Both conditions could easily be stated without this restriction, drawing on the generalized definition of conditional probabilities.

Intuitively, when the problem is easy, the evidence and the circumstances are truth-conducive. Conversely, when the problem is hard, the evidence is misleading. For an extreme example of misleading evidence, assume all witnesses in a court trial lie, falsely claiming the defendant is innocent, so that the problem-specific probability of the correct 'guilty' opinion approaches zero. In the real world some problems π are undeniably 'hard' in the sense that $\Pr(R_i|\pi) < \frac{1}{2}$ – just consider the previous court example. As another example of a hard problem due to misleading evidence, suppose a team of doctors is asked to assess whether a woman is pregnant. They perform a pregnancy test, which gives a negative result, while in fact the woman is pregnant. As for more realistic cases, imagine voting on a factual question related to a complex and controversial public policy issue, such as the use of genetically modified crops, bank bailouts, energy policy, etc. Sometimes such issues are dominated by misleading evidence, rendering the problem hard.

The existence of 'hard' problems need not undermine the plausibility of Classical Competence, since hard problems might be less frequent than easy problems, so that *on average* over all problems a voter might still be more often right than wrong for each state x – so that $\Pr(R_i|x)$ exceeds $\frac{1}{2}$ for all states x. But it would be inappropriate to assume that the *problem*-specific (rather than state-specific) correctness probability $\Pr(R_i|\pi)$ exceeds $\frac{1}{2}$ for all problems π, since a specific problem may be 'hard', as explained.

We now introduce a problem-specific variant of the Classical Competence assumption which does *not* state that the problem-specific correctness probability $\Pr(R_i|\pi)$ always equals some fixed parameter p in $(\frac{1}{2}, 1)$. Rather, we allow this probability to be below $\frac{1}{2}$ for some 'hard' problems π. Apart from this, we keep our assumption as simple as and as close to the classical assumption. Indeed, just as in the classical assumption the competence parameter p does not depend on the voter or the state, so in the following assumption all voters are equally competent, all 'easy' problems are equally easy, and all 'hard' problems are equally hard:

Easy/Hard Dichotomy. There is a parameter p in $(\frac{1}{2}, 1)$ such that for all problems $\boldsymbol{\pi}$ the conditional correctness probability $\Pr(R_i|\boldsymbol{\pi})$ is
- either p for all individuals i (we then call π an '*easy*' problem)
- or $1 - p$ for all individuals i (we then call π a '*hard*' problem).[21]

Note that this assumption is silent on how many problems are easy or hard. In the extreme case that all problems are easy, we obtain Classical Competence – but this case is by no means plausible. In most cases, however, the following jury theorem reaches a substantially different conclusion than the classical theorem, since the probability of a correct majority does not converge to one but to the proportion (probability) of easy problems:

Theorem 3. Under the assumptions PI and Easy/Hard Dichotomy, the probability of a correct majority opinion converges to $\Pr(\boldsymbol{\pi}$ is easy$)$ as the group size increases, and is

---

[21]This condition (and the resulting Theorem 3) could be generalized by allowing p to depend on the state x contained in π.

- strictly increasing if $\Pr(\boldsymbol{\pi}$ is easy$) > \Pr(\boldsymbol{\pi}$ is hard$)$,
- strictly decreasing if $\Pr(\boldsymbol{\pi}$ is easy$) < \Pr(\boldsymbol{\pi}$ is hard$)$,
- constant if $\Pr(\boldsymbol{\pi}$ is easy$) = \Pr(\boldsymbol{\pi}$ is hard$)$ $(= \frac{1}{2})$.

By the first bullet point, 'larger groups are wiser' as long as the problem is more often easy than difficult. The latter assumption about the problem can be defended by a stability argument: if more problems are hard than easy, i.e., if the voter is more often wrong than right (and if as usual only two opinions, say, 'yes' or 'no', are possible), then as soon as a voter realizes this (through observing her frequent failures) the voter can systematically reverse each opinion, thereby making herself more often right than wrong.

The fragility of 'wisdom of crowds' arguments is illustrated by the fact that Theorem 3 would not survive the following tempting generalization of the assumption of Easy/Hard Dichotomy. Suppose instead of requiring that the correctness probability on a problem is either $p \in (\frac{1}{2}, 1)$ (for an easy problem) or $1 - p \in (0, \frac{1}{2})$ (for a hard problem) one merely requires that this probability is either $p \in (\frac{1}{2}, 1)$ or $q \in (0, \frac{1}{2})$ *where q may differ from* $1 - p$, i.e., where p and q may have different distance to the midpoint $\frac{1}{2}$. In such cases the bullet points of Theorem 3 fail to follow, so that larger crowds may be *less* 'wise', even if most problems are easy.

5   Conclusion

The widespread conceptual confusion about the notion of opinion independence has hindered progress in assessing whether and when 'crowds are wise', and it explains the disagreements in social epistemology about whether jury theorems have any relevance. Understanding causal interrelations is indispensable for a proper analysis of opinion independence.

Our paper distinguishes between four notions of independence, as summarized in Table 1. The table highlights our two dimensions of categorization. To make a notion

Table 1: Different notions of independence.

		explicit state-conditionalization?	
		yes	**no**
explicit common-	**yes**	PI	CI
cause-conditionalization?	**no**	SI	UI

of independence realistic, the conditionalization has to include the common causes; to make it suitable for jury theorems – i.e., formal 'wisdom of crowds' arguments – the conditionalization has to include the state of the world, as illustrated in Section 4 by two jury theorems.

State-Conditional Independence is the commonly used notion in orthodox statements of Condorcet's jury theorem. Common-Cause-Conditional Independence is most generally defensible from the perspective of the theory of causal networks: it applies *always* as long as the opinions do not causally affect each other. Problem-Conditional Independence resembles Common-Cause-Conditional Independence, ex-

cept that it conditionalizes on the state of the world even when the state does not feature among the common causes. Unconditional Independence is the simplest form of independence. If the literature ignores this condition, it is probably because most scholars take it to be obviously false; but surprisingly we find plausible causal setups in which this independence assumption is justified.

Theorem 1 and its corollary give formal causal-network-theoretic foundations for each of the four independence assumptions. These results suggest that the causal conditions for the classical State-Conditional Independence assumption are quite special and of limited real-world significance. One will usually have to go beyond classical independence to make *sound* arguments in support of the 'wisdom of crowds'.

Much further work is needed to develop the causal approach. We leave it as a future challenge to develop new jury theorems for the aggregation of non-binary opinions, such as judgment sets or degrees of belief.

References

Anderson, E. (2006), 'The Epistemology of Democracy', *Episteme* 3(1-2), 8-22.

Berg, S. (1993), 'Condorcet's jury theorem, dependency among voters:', *Social Choice and Welfare* 10, 87–95.

Boland, P. J. (1989), 'Majority systems and the Condorcet jury theorem', *Journal of the Royal Statistical Society.* Series D (The Statistician) 38(3), 181–189.

Boland, P. J.; Proschan, F. & Tong, Y. (1989), 'Modelling dependence in simple and indirect majority systems', *Journal of Applied Probability* 26(1), 81–88.

Bovens, L. & Rabinowicz, W. (2006), 'Democratic Answers to Complex Questions - an Epistemic Perpective', *Synthese* 150, 131–153.

Dietrich, F. (2008), 'The premises of Condorcet's jury theorem are not simultaneously justified', *Episteme* 58(1), 56–73.

Dietrich, F. & List, C. (2004), 'A model of jury decisions where all jurors have the same evidence', *Synthese* 142, 175–202.

Dietrich, F. and Spiekermann, K. (2010), 'Epistemic Democracy with Defensible Premises', Unpublished manuscript.

Estlund, D. M. (1994), 'Opinion leaders, independence, and Condorcet's Jury Theorem', *Theory and Decision* 36(2), 131–162.

Estlund, D. M. (2008), *Democratic Authority: A Philosophical Framework*, Princeton University Press, Princeton.

Estlund, D.; Jeremy, W.; Grofman, B. & Feld, S. L. (1989), 'Democratic Theory and the Public Interest: Condorcet and Rousseau Revisited', *American Political Science Review* 83(4), 1317–1340.

Goldman, A. I. (2004), 'Group Knowledge versus Group Rationality: Two Approaches to Social Epistemology', *Episteme* 1, 11–22.

Goldman, A. I. (1999), *Knowledge in a Social World*, Clarendon, Oxford.

Grofman, B. & Feld, S. L. (1988), 'Rousseau's General Will: A Condorcetian Perspective', *American Political Science Review* 82(2), 567-576.

Grofman, B.; Owen, G. & Feld, S. L. (1983), 'Thirteen Theorems in Search of the Truth', *Theory and Decision* 15, 261–278.

Hawthorne, J. (unpublished) Voting in Search of the Public Good: the Probabilistic Logic of Majority Judgments, University of Oklahoma.

Kaniovski, S. (2010), 'Aggregation of correlated votes and Condorcet's Jury Theorem', *Theory and Decision* 69(3), 453-468.

Ladha, K. K. (1995), 'Information pooling through majority-rule voting: Condorcet's jury theorem with correlated votes', *Journal of Economic Behavior & Organization* 26(3), 353–372.

Ladha, K. K. (1993), 'Condorcet's jury theorem in light of de Finetti's theorem', *Social Choice and Welfare* 10(1), 69–85.

Ladha, K. K. (1992), 'The Condorcet Jury Theorem, Free Speech, and Correlated Votes', *American Journal of Political Science* 36(3), 617–634.

List, C. & Goodin, R. E. (2001), 'Epistemic Democracy: Generalizing the Condorcet Jury Theorem', *Journal of Political Philosophy* 9(3), 277–306.

Nitzan, S. & Paroush, J. (1984), 'The significance of independent decisions in uncertain dichotomous choice situations', *Theory and Decision* 17(1), 47–60.

Pearl, J. (2000), *Causality: models, reasoning and inference*, Cambridge University Press, Cambridge.

Reichenbach, H. (1956), *The direction of time*, University of California Press, Berkeley.

Romeijn, J. & Atkinson D. (2011), 'A Condorcet jury theorem for unknown juror competence', *Politics, Philosophy, and Economics,* 10(3), 237–262.

Spiekermann, K., Goodin, R. E. (2012), 'Courts of Many Minds', *British Journal of Political Science,* 12, 555-571.

Surowiecki, J. (2004) *The Wisdom of Crowds*, New York, Doubleday/Anchor

Vermeule, A. (2009), *Law and the Limits of Reason*, Oxford University Press, Oxford.

A   Appendix: Proofs

We now prove our main result (Theorem 1) and the two jury theorems presented as applications (Theorems 2 and 3).

Proof of Theorem 1. Assume that probabilities are compatible with causal interconnections (in the sense of the Parental Markov Condition) and no opinion is a cause of another opinion. We first prove part (a) and then part (b). The informal idea in both proofs is that dependence between opinions can only arise if information can travel along a path in the network without the path being 'blocked' by the variables on which one conditionalizes. The formal definition of 'blocking' (or 'd-separating') and the theorem whereby such blocking implies conditional independence are borrowed from the theory of causal networks, where they play a central role. Throughout we write $\mathcal{C}$ for the set of common causes.

Proof of part (a). We have to show that the opinions are independent conditional on $\mathcal{C}$. By the 'blocking theorem' in the theory of causal networks (e.g., Pearl, 2000, Theorem 1.2.4) it suffices to show that $\mathcal{C}$ blocks every path from an opinion to another

opinion, in the usual technical sense that for any such path

(i) *either* the path contains a chain '$\mathbf{a} \to \mathbf{b} \to \mathbf{c}$' or fork '$\mathbf{a} \leftarrow \mathbf{b} \to \mathbf{c}$' such that $\mathbf{b}$ is in $\mathcal{C}$

(ii) *or* the path contains a collider '$\mathbf{a} \to \mathbf{b} \leftarrow \mathbf{c}$' such that $\mathbf{b}$ is not in $\mathcal{C}$ and does not cause any variable in $\mathcal{C}$.

To show this, consider any path from an opinion $\mathbf{o}_i$ to another opinion $\mathbf{o}_j$. Call the path $(\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_m)$, where $m \, (\geq 2)$ is the number of variables in the path and $\mathbf{a}_1 = \mathbf{o}_i$ and $\mathbf{a}_m = \mathbf{o}_j$. By definition of a path, any two neighbors $\mathbf{a}_t, \mathbf{a}_{t+1}$ are connected by an arrow, of the form '$\mathbf{a}_t \to \mathbf{a}_{t+1}$' or '$\mathbf{a}_t \leftarrow \mathbf{a}_{t+1}$'.

Case 1: the arrow between $\mathbf{a}_1$ and $\mathbf{a}_2$ points towards $\mathbf{a}_2$ ('$\mathbf{a}_1 \to \mathbf{a}_2$'). It is impossible that between all neighboring variables $\mathbf{a}_t, \mathbf{a}_{t+1}$ the arrow points towards $\mathbf{a}_{t+1}$, since otherwise $\mathbf{o}_i$ would be a cause of $\mathbf{o}_j$. Let $\mathbf{a}_t$ be the earliest variable in the path such that an arrow points from $\mathbf{a}_{t+1}$ to $\mathbf{a}_t$. Notice the collider '$\mathbf{a}_{t-1} \to \mathbf{a}_t \leftarrow \mathbf{a}_{t+1}$'. It is impossible that $\mathbf{a}_t$ is in $\mathcal{C}$ or causes a variable in $\mathcal{C}$, since otherwise $\mathbf{o}_i$ (which causes $\mathbf{a}_t$) would cause other opinions. Therefore $\mathcal{C}$ blocks the path via clause (ii).

Case 2: the arrow between $\mathbf{a}_1$ and $\mathbf{a}_2$ points towards $\mathbf{a}_1$ ('$\mathbf{a}_1 \leftarrow \mathbf{a}_2$'). It is impossible that between all neighboring variables $\mathbf{a}_t, \mathbf{a}_{t+1}$ the arrow points towards $\mathbf{a}_t$, since otherwise $\mathbf{o}_j$ would be a cause of $\mathbf{o}_i$. Let $\mathbf{a}_t$ be the earliest variable in the path such that an arrow points from $\mathbf{a}_t$ to $\mathbf{a}_{t+1}$. Notice the fork '$\mathbf{a}_{t-1} \leftarrow \mathbf{a}_t \to \mathbf{a}_{t+1}$'.

Subcase 2.1: $\mathbf{a}_t \in \mathcal{C}$. Then $\mathcal{C}$ blocks the path via clause (i).

Subcase 2.2: $\mathbf{a}_t \notin \mathcal{C}$. Then $\mathbf{a}_t$ (which already causes $\mathbf{o}_i$) cannot also cause $\mathbf{o}_j$. So we do not have a chain '$\mathbf{a}_t \to \mathbf{a}_{t+1} \to ... \to \mathbf{a}_m$'. Choose $\mathbf{a}_s$ as the earliest variable among $\mathbf{a}_t, \mathbf{a}_{t+1}, ..., \mathbf{a}_{m-1}$ such that the arrow between $\mathbf{a}_s$ and $\mathbf{a}_{s+1}$ points towards $\mathbf{a}_s$. Note the collider '$\mathbf{a}_{s-1} \to \mathbf{a}_s \leftarrow \mathbf{a}_{s+1}$'. The variable $\mathbf{a}_s$ neither belongs to $\mathcal{C}$ nor causes a member of $\mathcal{C}$, since otherwise the variable $\mathbf{a}_t$ (which causes $\mathbf{a}_s$) would belong to $\mathcal{C}$. Therefore $\mathcal{C}$ blocks the path via clause (ii).

Proof of part (b). Now suppose that $\mathbf{x}$ is not a common effect of any opinions or private causes thereof. We have to show that the opinions are independent conditional on $\{\mathbf{x}\} \cup \mathcal{C}$. Again by the 'blocking theorem' (e.g., Pearl, 2000, Theorem 1.2.4), it suffices to show that $\{\mathbf{x}\} \cup \mathcal{C}$ blocks every path from an opinion to another opinion, i.e., that for every such path

(i*) *either* the path contains a chain '$\mathbf{a} \to \mathbf{b} \to \mathbf{c}$' or fork '$\mathbf{a} \leftarrow \mathbf{b} \to \mathbf{c}$' such that $\mathbf{b}$ is in $\{\mathbf{x}\} \cup \mathcal{C}$

(ii*) *or* the path contains a collider '$\mathbf{a} \to \mathbf{b} \leftarrow \mathbf{c}$' such that $\mathbf{b}$ is not in $\{\mathbf{x}\} \cup \mathcal{C}$ and does not cause any variable in $\{\mathbf{x}\} \cup \mathcal{C}$.

Consider any path $(\mathbf{a}_1, ..., \mathbf{a}_m)$ from an opinion $\mathbf{o}_i \, (= \mathbf{a}_1)$ to another opinion $\mathbf{o}_j \, (= \mathbf{a}_m)$.

Case 1: the arrow between $\mathbf{a}_1$ and $\mathbf{a}_2$ points towards $\mathbf{a}_2$ ('$\mathbf{a}_1 \to \mathbf{a}_2$'). Construct a collider '$\mathbf{a}_{t-1} \to \mathbf{a}_t \leftarrow \mathbf{a}_{t+1}$' as in Case 1 of part (a). Again, $\mathbf{a}_t$ neither is in $\mathcal{C}$ nor causes a variable in $\mathcal{C}$.

18

Subcase 1.1: $\mathbf{a}_t$ neither is $\mathbf{x}$ nor causes $\mathbf{x}$. Then, in summary, $\mathbf{a}_t$ neither belongs to $\{\mathbf{x}\} \cup \mathcal{C}$ nor causes a variable in $\{\mathbf{x}\} \cup \mathcal{C}$. So, $\{\mathbf{x}\} \cup \mathcal{C}$ blocks the path via clause (ii*).

Subcase 1.2: $\mathbf{a}_t$ is $\mathbf{x}$ or causes $\mathbf{x}$. We cannot have arrows '$\mathbf{a}_t \leftarrow \mathbf{a}_{t+1} \leftarrow ... \leftarrow \mathbf{a}_m$', since otherwise $\mathbf{a}_t$, and hence $\mathbf{x}$, would be a common effect of the opinions $\mathbf{a}_1 \ (= \mathbf{o}_i)$ and $\mathbf{a}_m \ (= \mathbf{o}_j)$. Let $\mathbf{a}_s$ be the earliest variable among $\mathbf{a}_t, \mathbf{a}_{t+1}, ..., \mathbf{a}_{m-1}$ with an arrow '$\mathbf{a}_s \rightarrow \mathbf{a}_{s+1}$'. Note the fork '$\mathbf{a}_{s-1} \leftarrow \mathbf{a}_s \rightarrow \mathbf{a}_{s+1}$' (see Figure A.1).
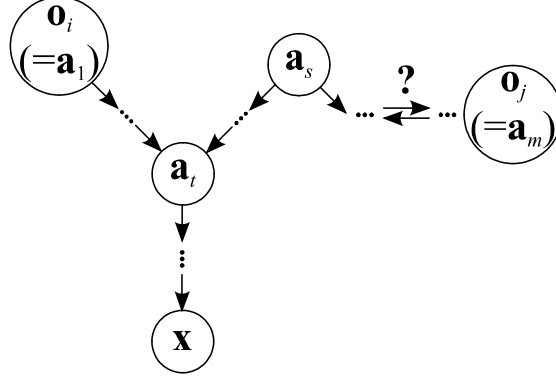

Figure A.1: The path in Subcase 1.2.

Subsubcase 1.2.1: $\mathbf{a}_s \in \mathcal{C}$. Then we are done as $\{\mathbf{x}\} \cup \mathcal{C}$ blocks the path via clause (i*).

Subsubcase 1.2.2: $\mathbf{a}_s \notin \mathcal{C}$. Then we do not have arrows '$\mathbf{a}_s \rightarrow \mathbf{a}_{s+1} \rightarrow ... \rightarrow \mathbf{a}_m$', since firstly $\mathbf{a}_s$ does not commonly cause the opinion $\mathbf{a}_m \ (= \mathbf{o}_j)$ and another opinion because $\mathbf{a}_s \notin \mathcal{C}$, and secondly $\mathbf{a}_s$ does not *privately* cause the opinion $\mathbf{a}_m$ because otherwise $\mathbf{a}_t$, and hence $\mathbf{x}$, would be a common effect of two variables (namely $\mathbf{o}_i$ and $\mathbf{a}_s$) which are or privately cause distinct opinions. Given that we do not have arrows '$\mathbf{a}_s \rightarrow \mathbf{a}_{s+1} \rightarrow ... \rightarrow \mathbf{a}_m$', there must be a variable $\mathbf{a}_r$ among $\mathbf{a}_{s+1}, ..., \mathbf{a}_{m-1}$ with a collider '$\mathbf{a}_{r-1} \rightarrow \mathbf{a}_r \leftarrow \mathbf{a}_{r+1}$'. Let $\mathbf{a}_r$ be the *last* such variable among $\mathbf{a}_{s+1}, ..., \mathbf{a}_{m-1}$. Note that we either have arrows '$\mathbf{a}_r \leftarrow ... \leftarrow \mathbf{a}_m$' or arrows '$\mathbf{a}_r \leftarrow ... \leftarrow \mathbf{a}_p \rightarrow ... \rightarrow \mathbf{a}_m$' for some $r < p < m$ (see Figure A.2). We may assume without loss of generality that in the second case a_p does not belong to $\mathcal{C}$, since otherwise $\{\mathbf{x}\} \cup \mathcal{C}$ would block the path via clause (i*). Now,

(*) a_r does not belong to $\mathcal{C}$ (hence, does not cause a member of $\mathcal{C}$),

In the case of '$\mathbf{a}_r \leftarrow ... \leftarrow \mathbf{a}_m$' this is because otherwise the opinion $\mathbf{a}_m \ (= \mathbf{o}_j)$ would cause another opinion, and in the case of '$\mathbf{a}_r \leftarrow ... \leftarrow \mathbf{a}_p \rightarrow ... \rightarrow \mathbf{a}_m$' it is because otherwise $\mathbf{a}_p$ would belong to $\mathcal{C}$. Notice further that $\mathbf{a}_r$ is an effect either of the opinion $\mathbf{a}_m = \mathbf{o}_j$ (in the case of '$\mathbf{a}_r \leftarrow ... \leftarrow \mathbf{a}_m$') or of a private cause of this opinion (in case of '$\mathbf{a}_r \leftarrow ... \leftarrow \mathbf{a}_p \rightarrow ... \rightarrow \mathbf{a}_m$'). From this it follows that

(**) $\mathbf{a}_r$ is not $\mathbf{x}$ and does not cause $\mathbf{x}$,
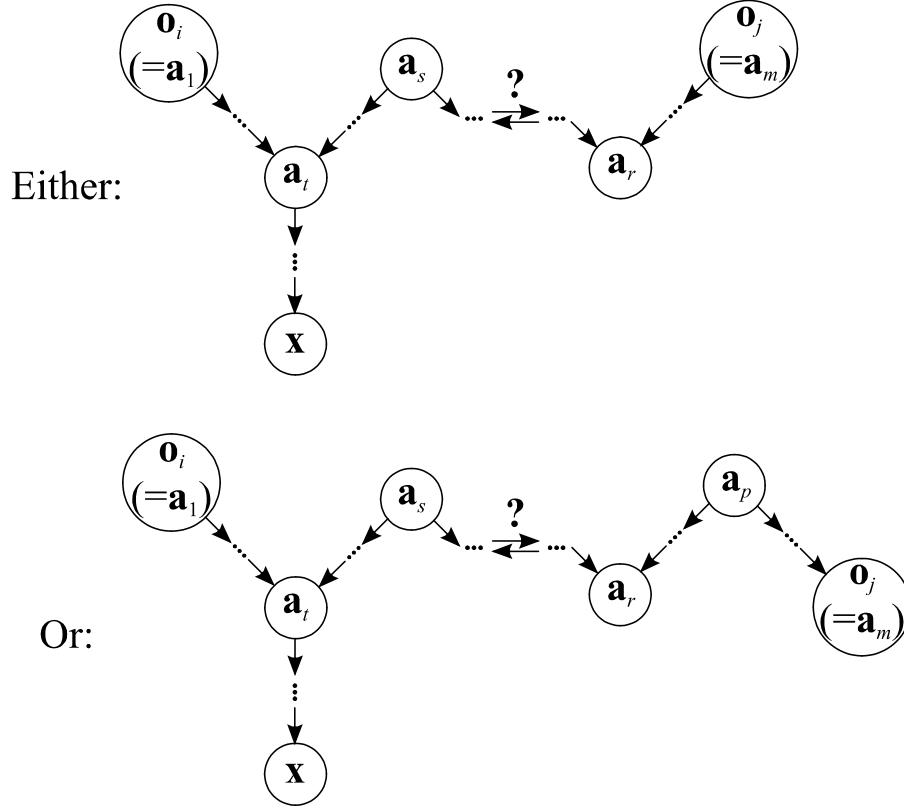
19

Figure A.2: The path in Subsubcase 1.2.2.

as otherwise $\mathbf{x}$ would be a common effect of on the one hand (via $\mathbf{a}_r$) the opinion $\mathbf{o}_j$ or a private cause thereof, and on the other hand (via $\mathbf{a}_t$) the opinion $\mathbf{o}_i$ (= $\mathbf{a}_1$). By (*) and (**), $\mathbf{a}_r$ neither belongs to $\{\mathbf{x}\} \cup \mathcal{C}$ nor causes a member of $\{\mathbf{x}\} \cup \mathcal{C}$. So, the path is blocked via clause (ii*).

Case 2: the arrow between $\mathbf{a}_1$ and $\mathbf{a}_2$ points towards $\mathbf{a}_1$ ('$\mathbf{a}_1 \leftarrow \mathbf{a}_2$'). Construct the fork '$\mathbf{a}_{t-1} \leftarrow \mathbf{a}_t \rightarrow \mathbf{a}_{t+1}$' as in Case 2 of part (a).

Subcase 2.1: $\mathbf{a}_t \in \{\mathbf{x}\} \cup \mathcal{C}$. Then $\mathcal{C}$ blocks the path via clause (i*).

Subcase 2.2: $\mathbf{a}_t \notin \{\mathbf{x}\} \cup \mathcal{C}$. Since in particular $\mathbf{a}_t \notin \mathcal{C}$, we can construct a collider '$\mathbf{a}_{s-1} \rightarrow \mathbf{a}_s \leftarrow \mathbf{a}_{s+1}$' as in Subcase 2.2 of part (a) (see Figure A.3), and again $\mathbf{a}_s$ neither belongs to $\mathcal{C}$ nor causes a member of $\mathcal{C}$.
Subsubcase 2.2.1: $\mathbf{a}_s$ neither is nor causes $\mathbf{x}$. Then, in summary, $\mathbf{a}_s$ neither is in nor causes a variable of $\{\mathbf{x}\} \cup \mathcal{C}$, and hence $\{\mathbf{x}\} \cup \mathcal{C}$ blocks the path via clause (ii*).

Subsubcase 2.2.2: $\mathbf{a}_s$ is or causes $\mathbf{x}$. We cannot have arrows '$\mathbf{a}_s \leftarrow \mathbf{a}_{s+1} \leftarrow \dots \leftarrow \mathbf{a}_m$', since otherwise $\mathbf{x}$ would be a common effect (via $\mathbf{a}_s$) of the opinion $\mathbf{o}_j$ (= $\mathbf{a}_m$) and the private cause $\mathbf{a}_t$ of the opinion $\mathbf{o}_1$ (= $\mathbf{a}_1$). If we have arrows '$\mathbf{a}_s \leftarrow \dots \leftarrow \mathbf{a}_q \rightarrow \dots \rightarrow \mathbf{a}_m$' for some $s < q < m$, then $\mathbf{a}_q$ must be in $\mathcal{C}$ since otherwise $\mathbf{a}_q$ would be a
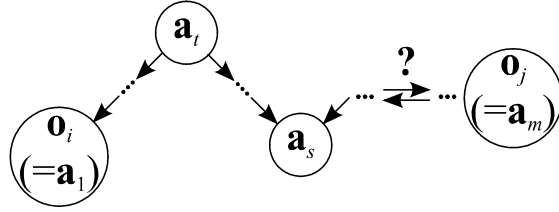
Figure A.3: The path in Subcase 2.2.

private cause of the opinion $\mathbf{a}_m$ ($= \mathbf{o}_j$) so that $\mathbf{x}$ would be a common effect of the private causes $\mathbf{a}_q$ and $\mathbf{a}_t$; and since '$\mathbf{a}_{q-1} \leftarrow \mathbf{a}_q \rightarrow \mathbf{a}_{q+1}$' is a fork with $\mathbf{a}_q \in \mathcal{C}$ we are done by clause (i*). Now assume the remaining case that we neither have arrows '$\mathbf{a}_s \leftarrow ... \leftarrow \mathbf{a}_m$' nor arrows '$\mathbf{a}_s \leftarrow ... \leftarrow \mathbf{a}_q \rightarrow ... \rightarrow \mathbf{a}_m$'. There must be a variable $\mathbf{a}_r$ among $\mathbf{a}_{s+1}, ..., \mathbf{a}_{m-1}$ such that we have a collider '$\mathbf{a}_{r-1} \rightarrow \mathbf{a}_r \leftarrow \mathbf{a}_{r+1}$'. Choose $\mathbf{a}_r$ to be the *latest* variable among $\mathbf{a}_{s+1}, ..., \mathbf{a}_{m-1}$ with the collider property. Note that we
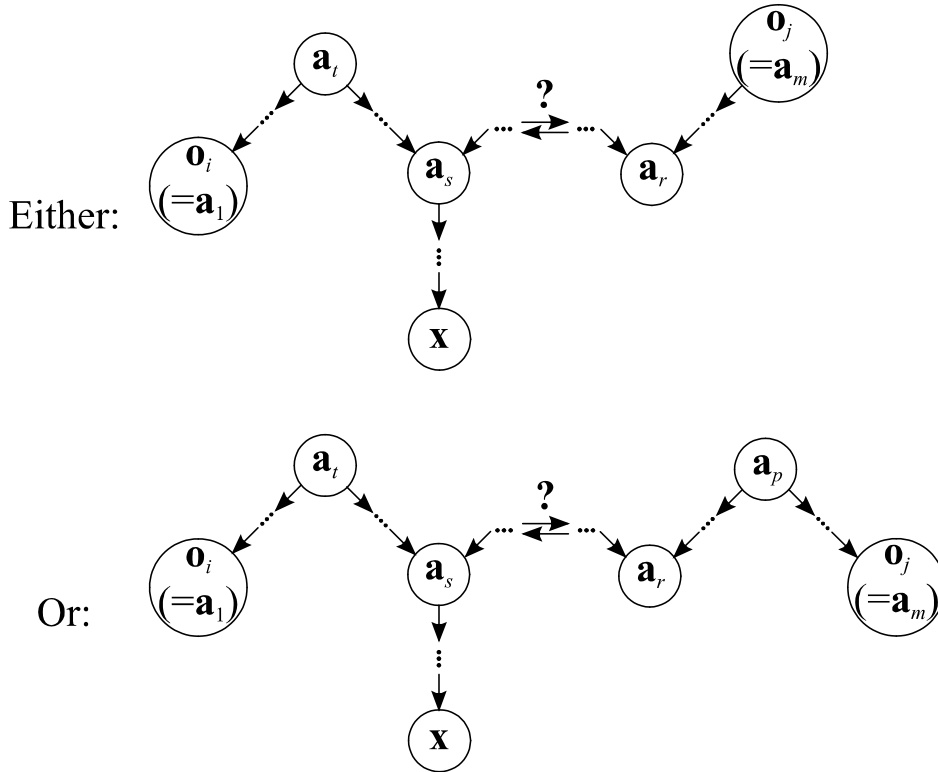

Figure A.4: The path in Subsubcase 2.2.2.

either have arrows '$\mathbf{a}_r \leftarrow ... \leftarrow \mathbf{a}_m$' or arrows '$\mathbf{a}_r \leftarrow ... \leftarrow \mathbf{a}_p \rightarrow ... \rightarrow \mathbf{a}_m$' for some $r < p < m$ (see Figure A.4). We may assume without loss of generality that in the second case $\mathbf{a}_p \notin \mathcal{C}$, as otherwise we would be done by clause (i*). Then

(***) a_r does not belong to $\mathcal{C}$ (so, does not cause a member of $\mathcal{C}$).

21

In the case of arrows '$\mathbf{a}_r \leftarrow ... \leftarrow \mathbf{a}_m$' the reason for (***) is that otherwise the opinion $\mathbf{a}_m$ $(= \mathbf{o}_j)$ would cause another opinion; and in the case of arrows '$\mathbf{a}_r \leftarrow ... \leftarrow \mathbf{a}_p \rightarrow ... \rightarrow \mathbf{a}_m$' the reason is that otherwise $\mathbf{a}_p$ would belong to $\mathcal{C}$. Note further that $\mathbf{a}_r$ is caused either by the opinion $\mathbf{a}_m = \mathbf{o}_j$ (in the case of arrows '$\mathbf{a}_r \leftarrow ... \leftarrow \mathbf{a}_m$') or by a private cause of this opinion (in case of arrows '$\mathbf{a}_r \leftarrow ... \leftarrow \mathbf{a}_p \rightarrow ... \rightarrow \mathbf{a}_m$'). It follows that

(****) $\mathbf{a}_r$ is not $\mathbf{x}$ and does not cause $\mathbf{x}$,

since otherwise $\mathbf{x}$ would be a common effect firstly (via $\mathbf{a}_r$) of the opinion $\mathbf{o}_j$ or a private cause thereof, and secondly (via $\mathbf{a}_s$) of the private cause $\mathbf{a}_t$ of the opinion $\mathbf{o}_i$ $(= \mathbf{a}_1)$. By (***) and (****), $\mathbf{a}_r$ does not belong to or cause a member of $\{\mathbf{x}\} \cup \mathcal{C}$. So, the path is blocked via clause (ii*). ∎

Although the classical jury theorem is of course well-known, we give a proof of it because of our more general framework (which allows for more than two states and opinions).

Proof of Theorem 2. Assume SI and Classical Competence. Let M_n be the event that a majority of the opinions is correct in the group of size n. For simplicity, we assume that the set of possible states is countable and each possible state occurs with non-zero probability. (The proof generalizes easily to an arbitrary set of possible states, by essentially replacing sums by Lebesgue integrals.)

Conditional on any possible state x, the events $R_1, ..., R_n$ have independent probabilities of p; so the probability that exactly k of these events hold is $\binom{n}{k}p^k(1-p)^{n-k}$, whence the probability that a majority of the events hold is

$$\beta_{n,p} \equiv \Pr(M_n|\boldsymbol{x}) = \sum_{k=(n+1)/2}^{n} \binom{n}{k}p^k(1-p)^{n-k}.$$

The unconditional probability of a correct majority can be expressed as

$$\begin{aligned}
\Pr(M_n) &= \sum_x \Pr(M_n|x)\Pr(x) \\
&= \sum_x \beta_{n,p}\Pr(x) = \beta_{n,p}\sum_x \Pr(x) = \beta_{n,p}.
\end{aligned}$$

So, it suffices to show that $\beta_{n,p}$ is strictly increasing in (odd) n and converging to 1.

First, $\beta_{n,p}$ is strictly increasing in (odd) n because $p \in (\frac{1}{2}, 1)$ and the coefficient $\beta_{n,p}$ satisfies the following well-known recursion formula:

$$\beta_{n+2,p} = \beta_{n,p} + (2p-1)\binom{n}{\frac{n+1}{2}}[p(1-p)]^{\frac{n+1}{2}}$$

(e.g., Grofman *et al.* 1983).

22

Second, $\beta_{n,p} \to 1$ as $n \to \infty$, by the following argument. Recall that $\beta_{n,p}$ is the probability that the sum of n independent and identically distributed Bernoulli-variables (which are 1 with probability $p > \frac{1}{2}$) belongs to the interval $\left(\frac{n}{2}, n\right]$. Equivalently, $\beta_{n,p}$ is the probability that $\frac{1}{n}$ times this sum, i.e., the *correctness frequency*, belongs to the interval $(\frac{1}{2}, 1]$. This probability converges to 1 because by the law of large numbers the correctness frequency converges to p $(\in (\frac{1}{2}, 1))$ with probability one. ■

Finally, we prove our new jury theorem.

Proof of Theorem 3. Assume PI and Easy/Hard Dichotomy, and let p be the value in $(\frac{1}{2}, 1)$ specified in the latter assumption. Again, let M_n be the probability of a majority for the correct opinion in the group of (odd) size n. For simplicity, let there be only countably many possible problems, where each problem occurs with positive probability. (The proof generalizes easily to an arbitrary set of possible problems.) Let Π_{easy} resp. Π_{hard} be the set of easy resp. hard problems. Let p_{easy} be the probability that the problem is easy (i.e., belongs to Π_{easy}).

1. We first calculate the probability of majority correctness *conditional* on the problem.

Conditional on any given problem π in Π_{easy}, the events $R_1, ..., R_n$ have independent probabilities of p, so that, just as in the proof of Theorem 2, the probability that a majority of these events hold is given by the coefficient

$$\Pr(M_n | \boldsymbol{\pi}) = \beta_{n,p} \equiv \sum_{k=(n+1)/2}^{n} \binom{n}{k} p^k (1-p)^{n-k}.$$

By contrast, conditional on any given problem π in Π_{hard}, the events $R_1, ..., R_n$ have independent probabilities of $1-p$, and the events of *incorrect* opinions $\overline{R_1}, ..., \overline{R_n}$ have independent probabilities of p; so the probability of a majority of *incorrect* opinions is $\beta_{n,p}$, and hence, the probability of a majority of *correct* opinions is one minus that probability, i.e.,

$$\Pr(M_n | \pi) = 1 - \beta_{n,p}.$$

2. We now calculate the *unconditional* probability of majority correctness. First, we write

$$
\begin{aligned}
\Pr(M_n) &= \sum_{\pi} \Pr(M_n | \pi) \Pr(\pi) \\
&= \sum_{\pi \in \Pi_{\text{easy}}} \Pr(M_n | \pi) \Pr(\pi) + \sum_{\pi \in \Pi_{\text{hard}}} \Pr(M_n | \pi) \Pr(\pi).
\end{aligned}
$$

23

By part 1, it follows that

$$
\begin{aligned}
\Pr(M_n) &= \sum_{\pi \in \Pi_{\text{easy}}} \beta_{n,p} \Pr(\pi) + \sum_{\pi \in \Pi_{\text{hard}}} (1 - \beta_{n,p}) \Pr(\pi) \\
&= \beta_{n,p} \sum_{\pi \in \Pi_{\text{easy}}} \Pr(\pi) + (1 - \beta_{n,p}) \sum_{\pi \in \Pi_{\text{hard}}} \Pr(\pi) \\
&= \beta_{n,p} p_{\text{easy}} + (1 - \beta_{n,p})(1 - p_{\text{easy}}) \\
&= \beta_{n,p} p_{\text{easy}} + 1 - \beta_{n,p} - p_{\text{easy}} + \beta_{n,p} p_{\text{easy}} \\
&= \beta_{n,p}(2p_{\text{easy}} - 1) + 1 - p_{\text{easy}}.
\end{aligned}
$$

3. We can finally prove the theorem's conclusions. In the last expression for $\Pr(M_n)$, the only term which depends on n is $\beta_{n,p}$. As in the proof of Theorem 2, $\beta_{n,p}$ is strictly increasing in (odd) n and converges to 1. First, since $\beta_{n,p} \to 1$, we have

$$
\Pr(M_n) \to 1 \times (2p_{\text{easy}} - 1) + 1 - p_{\text{easy}} = p_{\text{easy}}.
$$

Second, since $\beta_{n,p}$ is strictly increasing, $\Pr(M_n)$ is strictly increasing if $2p_{\text{easy}} - 1 > 0$, i.e., $p_{\text{easy}} > \frac{1}{2}$, or equivalently $\Pr(\boldsymbol{\pi} \text{ is easy}) > \Pr(\boldsymbol{\pi} \text{ is hard})$. Analogously, $\Pr(M_n)$ is strictly decreasing if $2p_{\text{easy}} - 1 < 0$, i.e., $\Pr(\boldsymbol{\pi} \text{ is easy}) < \Pr(\boldsymbol{\pi} \text{ is hard})$. And $\Pr(M_n)$ is constant if $2p_{\text{easy}} - 1 = 0$, i.e., $\Pr(\boldsymbol{\pi} \text{ is easy}) = \Pr(\boldsymbol{\pi} \text{ is hard})$. $\blacksquare$