Logical Models of Informational Cascades

ALEXANDRU BALTAG, ZOÉ CHRISTOFF, JENS ULRIK HANSEN, AND SONJA SMETS

ABSTRACT. In this paper, we investigate the social herding phenomenon known as *informational cascades*, in which sequential inter-agent communication might lead to epistemic failures at group level, despite availability of information that should be sufficient to track the truth. We model an example of a cascade, and check the correctness of the individual reasoning of each agent involved, using two alternative logical settings: an existing probabilistic dynamic epistemic logic, and our own novel logic for counting evidence. Based on this analysis, we conclude that cascades are not only likely to occur but are sometimes unavoidable by "rational" means: in some situations, the group's inability to track the truth is the direct consequence of each agent's rational attempt at individual truth-tracking. Moreover, our analysis shows that this is even so when rationality includes unbounded higher-order reasoning powers (about other agents' minds and about the belief-formation-and-aggregation protocol, including an awareness of the very possibility of cascades), as well as when it includes simpler, non-Bayesian forms of heuristic reasoning (such as comparing the amount of evidence pieces).

Social knowledge is what holds the complex interactions that form society together. But how reliable is social knowledge: how good is it at tracking the truth, in comparison with individual knowledge? At first sight, it may seem that groups should be better truth-trackers than the individuals composing them: the group *can "in principle" access* the information possessed by each agent, and in addition it *can access whatever follows* from combining these individual pieces of information using logic.

And indeed, in many situations, this "virtual knowledge" of a large group is much higher than the knowledge of the most expert member of the group: this is the phenomenon known as *wisdom of the crowds* [32]. Some examples of the wisdom of the crowds are explainable by the logical notion of *distributed knowledge*: the kind of group knowledge that can be realized by inter-agent communication. But most examples, typically involving no communication, are of a different, more "statistical" type, and they have been explained in Bayesian terms by Condorcet's Jury Theorem [26, 19, 25], itself based on the Law of Large Numbers. In particular, some political scientists used (variants and generalizations of) the Jury Theorem, to provide epistemic arguments in favour of deliberative democracy: this is the core of the so-called "epistemic democracy" program [26, 19]. Roughly speaking, the conclusion of this line of research is that groups are more reliable at tracking the truth than individuals and that the larger the group, the more probable it is that the majority's opinion is the right one.

However, the key word in the above paragraph is *virtual* (as in "virtual knowledge" of a group). Most explanations based on (variants of) the Jury Theorem seem to rely on a crucial condition: agents do not communicate with each other, they only *secretly* vote for their favorite answer. Their opinions are therefore taken to be completely *independent* of each other.[1] In contrast, the logical notion of distributed knowledge is tightly connected to communication: by sharing all they know, the agents can convert their virtual knowledge into actual knowledge. But without communication, how do the agents get to actualize the full epistemic potential of the group? Or do they, ever?

There seems to be a tension between the two ingredients needed for maximizing actual group knowledge: independence (of individual opinions) versus sharing (one's opinions with the group). Independence decreases when inter-agent communication is allowed, and in particular when agents are making *public and sequential* guesses or decisions. In such cases, some agents' later epistemic choices might very well be influenced by other agents' previous choices. Being influenced in this way may be perfectly justifiable on rational grounds at an individual level. After all, this is what discussion and deliberation are all about: exchanging information, so that everybody's opinions and decisions are better informed, and thus more likely to be correct. So, at first sight, it may seem that communication and rational deliberation can only be epistemically beneficial to each of the agents, and hence can only enhance the truth-tracking potential of the group. But in fact the primary consequence of communication, at the group level, is that the agents' epistemic choices become *correlated* (rather than staying independent). This correlation undermines the assumptions behind positive theoretical results (such as the Jury Theorem), and so the conclusion will also often fail. The group's (or the majority's) actual knowledge may fall way behind its "virtual knowledge": indeed, the group may end up voting unanimously for the wrong option!

Informational cascades are examples of such "social-epistemic catastrophes" that may occur as a by-product of sequential communication. By observing the epistemic decisions of the previous people in a sequence, an individual may rationally form an opinion about the information that the others might have, and this opinion may even come to outweigh her other (private) information, and thereby affect her epistemic decision. In this way, individuals in a sequence might be "ra-

[1] But see e.g. [19] for majority truth-tracking in conditions that allow for some very mild forms of communication within small subgroups.

tionally" led to ignore their own private evidence and to simply start *following the crowd*, whether the crowd is right or wrong. This is *not mindless imitation*, and it is *not due to any irrational social-psychological influence* (e.g. group pressure to conform, brainwashing, manipulation, mass hysteria etc). Rather, this is the result of rational inference based on partial information: it is not just a cascade, but an "informational" cascade. A classical example is the choice of a restaurant. Suppose an agent has some private information that restaurant A is better than restaurant B. Nevertheless, when arriving at the adjacent restaurants she sees a crowded restaurant B and an empty restaurant A, which makes her decide to opt for restaurant B. In this case our agent interprets the others' choice for B as conveying some information about which restaurant is better and this overrides her independent private information. However, it could very well be that all the people in restaurant B chose that restaurant for the exact same reason. Other examples of informational cascades include bestseller lists for books, judges voting, peer-reviewing, fashion and fads, crime etc [13].

While models of such phenomena were independently developed in [12] and [6], the term *informational cascades* is due to [12]. A probabilistic treatment of cascades, using Bayesian reasoning, can be found in [15]. Traditionally investigated by the Social Sciences, these social-informational phenomena have recently become subject of philosophical reflection, as part of the field of *Social Epistemology* [17, 18]. In particular, [21] gives an excellent philosophical discussion of informational cascades (and the more general class of "info-storms"), their triggers and their defeaters ("info-bombs"), as well as the epistemological issues raised by the existence of these social-epistemic phenomena.

In a parallel evolution, logicians have perfected new formal tools for exploring informational dynamics and agency [8], and for modeling public announcements and other forms of distributed information flow. An example is the fast-growing field of Dynamic-Epistemic Logic (DEL for short), cf. [8], [4], [14] etc. More recently, variants of DEL that focus on *multi-agent belief revision* [3, 7, 5] and on the *social dynamics of preferences* [10, 27] have been developed and used to investigate social-epistemic phenomena that are closely related to cascades: epistemic bandwagonning [22], mutual doxastic influence over social networks [31, 30], and pluralistic ignorance [29, 20].

The time seems therefore ripe for an epistemic-logical study of informational cascades. In this paper, we take the first step in this direction, by modeling "rational" cascades in a logical-computational setting based on (both probabilistic and more qualitative) versions of Dynamic-Epistemic Logic.

When the total sum of private information possessed by the members of a group is in principle enough to track the truth, but nevertheless the group's beliefs fail to do so, one might think that this is due to some kind of "irrationality" in the formation and/or aggregation of beliefs (including unsound reasoning and misinterpretation of the others' behavior, but possibly also lack of cooperation, lack of relevant communication, lack of trust etc). However, this is not always the case, as was already argued in the original paper [12]. One of the standard examples of an informational cascade (the "Urn example" which will be discussed in section 1), has been used to show that cascades can be "rational". Indeed, in such examples, the cascade does seem to be the result of correct Bayesian reasoning [15]: each agent's opinion/decision is perfectly justified, given the information that is available to her. A Bayesian model of this example is given in [15] and reproduced by us in section 1. The inescapable conclusion seems to be that, in such cases, *individual rationality may lead to group "irrationality"*.

However, what is typically absent from this standard Bayesian analysis of informational cascades is the agents' higher-order reasoning (about other agents' minds and about the whole sequential protocol in which they are participating). So one may still argue that by such higher-order reflection (and in particular, by becoming aware of the dangers inherent in the sequential deliberation protocol), "truly rational" agents might be able to avoid the formation of cascades. And indeed, in *some* cases the cascade can be prevented simply by making agents aware of the very possibility of a cascade.

In this paper, we prove that this is *not* always the case: there are situations in which no amount of higher-order reflection and meta-rationality can stop a cascade. To show this, we present in section 2 a formalization of the above-mentioned Urn example using Probabilistic Dynamic Epistemic Logic [9, 24]. This setting assumes perfectly rational agents able to reflect upon and reason about higher levels of group knowledge: indeed, epistemic logic takes into account all the levels of mutual belief/knowledge (beliefs about others' beliefs etc) about the current *state* of the world; while dynamic epistemic logic adds also all the levels of mutual belief/knowledge about the on-going *informational events* ("the protocol"). The fact that the cascade can still form proves our point: cascades cannot in general be prevented even by the use of the most perfect, idealized kind of individual rationality, one endowed with unlimited higher-level reflective powers. Informational cascades of this "super-rational" kind can be regarded as "epistemic Tragedies of the Commons": *paradoxes of (individual-versus-social) rationality*. In such contexts, a cascade can only be stopped by an external or "irrational" force, acting as *deus ex machina*: an"info-bomb", in the sense of [21]. This can be either an intervention from an outside agent (with different interests or different information that the agents engaged in the cascade), or a sudden burst of "irrationality" from one of the participating agents.

In section 3 we address another objection raised by some authors against the Bayesian analysis of cascades. They argue that real agents, although engaging in cascades, do it for non-Bayesian reasons: instead of probabilistic conditioning, they seem to use "rough-and-ready" qualitative heuristic methods, e.g. by simply *counting* the pieces of evidence in favor of one hypothesis against its alternatives. To model cascades produced by this kind of qualitative reasoning (by agents who still maintain their higher-level awareness of the other agents' minds), we introduce a new framework – a *multi-agent logic for counting evidence*. We use this setting to show that, even if we endow our formal agents only with a heuristic way of reasoning which is much less sophisticated, more intuitive and maybe more realistic than full-fledged probabilistic logic, they may still "rationally" engage in informational cascades. Hence, the above conclusion can now be now extended to a wider range of agents: *as long as the agents can count the evidence, then no matter how high or how low are their reasoning abilities* (even if they are capable of full higher-level reflection about others' minds, or dually even if they can't go beyond simple evidence counting), *their individual rationality may still lead to group "irrationality"*.

1 An Informational Cascade and its Bayesian Analysis

We will focus on a simple example that was created for studies of informational cascades in a laboratory [1, 2]. Consider two urns, respectively named U_W and U_B, where urn U_W contains two white balls and one black ball, and urn U_B contains one white ball and two black balls. One urn is randomly picked (say, using a fair coin) and placed in a room. This setup is common knowledge to a group of agents, which we will denote $a_1, a_2, ..., a_n$ but they do not know which of the two urns is in the room. The agents enter the room one at a time; first a_1, then a_2, and so on. Each agent draws one ball from the urn, looks at it, puts it back, and leaves the room. Hence, only the person in the room knows which ball she drew. After leaving the room she makes a guess as to whether it is urn U_W or U_B that is placed in the room and writes her guess on a blackboard for all the other agents to see. Therefore, each individual a_i knows the guesses of the previous people in the sequence $a_1, a_2, ..., a_n$ before entering the room herself. It is common knowledge that they will be individually rewarded if and only if their own guess is correct.

In this section we give the standard Bayesian analysis of this example, following the presentation in [15]. Let us assume that in fact urn U_B has been placed in the room. When a_1 enters and draws a ball, there is a unique simple decision rule she should apply: if she draws a white ball it is rational to make a guess for U_W, whereas if she draws a black one she should guess U_B. We validate this by calculating the probabilities. Let w_1 denote the event that a_1 draws a white ball and b_1 denote the event that she draws a black one. The proposition that it is urn U_W which is in the room will be denoted similarly by U_W and likewise for U_B. Given that it is initially equally likely that each urn is placed in the room the probability of U_W is $\frac{1}{2}$ ($P(U_W) = \frac{1}{2}$), and similarly for U_B. Observe that $P(w_1) = P(b_1) = \frac{1}{2}$. Assume now that a_1 draws a *white* ball. Then, via Bayes' rule, the posterior probability of U_W is

$$P(U_W|w_1) = \frac{P(U_W)\cdot P(w_1|U_W)}{P(w_1)} = \frac{\frac{1}{2}\cdot\frac{2}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

Hence, it is indeed rational for a_1 to guess U_W if she draws a white ball (and to guess U_B if she draws a black ball). Moreover, when leaving the room and making a guess for U_W (resp. U_B), all the other individuals can infer that she drew a white (resp. black) ball.

When a_2 enters the room after a_1, she knows the color which a_1 drew and it is obvious how she should guess if she draws a ball of the same color. If a_1 drew a white ball and a_2 draws a white ball, then a_2 should guess U_W. Formally, the probability of U_W given that both a_1 and a_2 draw white balls is

$$P(U_W|w_1,w_2) = \frac{P(U_W)\cdot P(w_1,w_2|U_W)}{P(w_1,w_2)} = \frac{\frac{1}{2}\cdot\frac{2}{3}\cdot\frac{2}{3}}{\frac{5}{18}} = \frac{4}{5}.$$

A similar reasoning applies if both drew black balls. If a_2 draws an opposite color ball of a_1, then the probabilities for U_W and U_B become equal. For simplicity we will assume that any individual faced with equal probability for U_W and U_B will guess for the urn that contains more balls of the color she saw herself: if a_1 drew a white ball and a_2 draws a black ball, a_2 will guess U_B.[2] Hence, independent of which ball a_1 draws, a_2 will always guess for the urn matching the color of her privately drawn ball. We assume that this tie-breaking rule is common knowledge among the agents too. In this way, every individual following a_2 can also infer the color of a_2's ball.

When a_3 enters, a cascade can arise. If a_1 and a_2 drew opposite color balls, a_3 is rational to guess for the urn that matches the color of the ball she draws. Nevertheless, if a_1 and a_2 drew the same color of balls (given the reasoning previously described, a_3 will know this), say both white, then no matter what color of ball a_3 draws the posterior probability of U_W will be higher than the probability of U_B (and if a_1 and a_2 both drew black balls the other way around). To check this let us calculate the probability of U_W given that a_1 and a_2 drew white balls and a_3 draws a black one:

$$P(U_W|w_1,w_2,b_3) = \frac{P(U_W)\cdot P(w_1,w_2,b_3|U_W)}{P(w_1,w_2,b_3)} = \frac{\frac{1}{2}\cdot\frac{2}{3}\cdot\frac{2}{3}\cdot\frac{1}{3}}{\frac{1}{9}} = \frac{2}{3}.$$

It is obvious that $P(U_W|w_3,w_2,w_1)$ will be even larger, thus whatever ball a_3 draws it will be rational for her to guess for U_W. Hence, if a_1 and a_2 draw the

[2] This tie-breaking rule is a simplifying assumption but it does not affect the likelihood of cascades arising. Moreover, there seems to be some empirical evidence that this is what most people do and it is also a natural tie-breaking rule if the individuals assign a small chance to the fact that other people might make errors [2].

same color of balls a cascade will start from a_3 on![3] The individuals following a_3 should therefore take a_3's guess as conveying no new information. Furthermore, everyone after a_3 will have the same information as a_3 (the information about what a_1 and a_2 drew) and their reasoning will therefore be identical to the one of a_3 and the cascade will continue.

If U_B is, as we assumed, the urn actually placed in the room and both a_1 and a_2 draw white balls (which happens with probability $\frac{1}{9}$) then a cascade leading to everyone making the *wrong* guess starts. Note, however, that if both a_1 and a_2 draw black balls (which happens with probability $\frac{4}{9}$), then a cascade still starts, but this time it will lead to everyone making the *right* guess. Thus, when a cascade happens it is four times more likely in this example that it leads to right guesses than to the wrong guesses. This already supports the claim that rational agents can be well aware of the fact that they are in a cascade without it forcing them to change their decisions. The general conclusion of this example is that even though informational cascades can look irrational from a social perspective, they are not irrational from the perspective of any individual participating in them.

The above semi-formal analysis summarizes the standard Bayesian treatment of this example, as given e.g. in [15]. However, as we mentioned in the introduction, several objections can be raised against the way this conclusion has been reached above. First of all, the example has only been partially formalized, in the sense that the public announcements of the individuals' guesses are not explicitly present in it, neither is the reasoning that lets the individuals ignore the guesses of the previous people caught in a cascade. Moreover, the Bayesian analysis given above does not formally capture the agents' full higher-order reasoning (i.e. their reasoning about the others' beliefs and about the others' higher-order reasoning about their beliefs etc). So one cannot use the above argument to completely rule out the possibility that some kind of higher-order reflection may help prevent (or break) informational cascade: it might be the case that, after realizing that they are participating in a cascade, agents may use this information to try to stop the cascade.

For all these reasons, we think it is useful to give a more complete analysis, using a model that captures both the public announcements and the full higher-order reasoning of the agents. This is precisely what we will do in the next section, in the framework of Probabilistic Dynamic Epistemic Logic [9, 24].

[3] Note that the cascade will start even if we change the tie-breaking rule of a_2 such that she randomizes her guess whenever she draws a ball contradicting the guess of a_1. In this case, if a_1 and a_2 guess for the same urn, a_3 will not know the color of a_2's ball, but she will still consider it more likely that a_2's ball matches the ball of a_1 and hence consider it more likely that the urn which they have picked is in the one in the room.

2 A Probabilistic Logical Model

In this section we will work within the framework of Probabilistic Dynamic Epistemic Logic [9, 24]. Our presentation will be based on a simplified version of the setting from [9], in which we assume that agents are *introspective as far as their own subjective probabilities are concerned* (so that an agent's subjective probability assignment does not depend on the actual state of the world but only on that world's partition cell in the agent's information partition). We also use slightly different graphic representations, which make explicit the *odds* between any two possible states (considered pairwise) according to each agent. This allows us to present directly a *comparative* treatment of the rational guess of each agent and will make obvious the similarity with the framework for "counting evidences" that we will introduce in the next section. We start with some definitions.

DEFINITION 1 (Probabilistic Epistemic State Models). A probabilistic multi-agent epistemic state model $\mathcal{M}$ is a structure $(S, \mathcal{A}, (\sim_a)_{a\in\mathcal{A}}, (P_a)_{a\in\mathcal{A}}, \Psi, \| \bullet \|)$ such that:

- S is a set of states (or "worlds");
- $\mathcal{A}$ is a set of agents;
- for each agent a, $\sim_a \subseteq S \times S$ is an equivalence relation interpreted as agent a's epistemic indistinguishability. This captures the agent's hard information about the actual state of the world;
- for each agent a, $P_a : S \to [0, 1]$ is a map that induces a probability measure on each $\sim_a$-equivalence class (i.e., we have $\sum\{P_a(s') : s' \sim_a s\} = 1$ for each $a \in \mathcal{A}$ and each $s \in S$). This captures the agent's subjective probabilistic information about the state of the world;
- Ψ is a given set of "atomic propositions", denoted by $p, q, \ldots$. Such atoms p are meant to represent *ontic "facts'* that might hold in a world.
- $\|\bullet\| : \Psi \to \mathcal{P}(S)$ is a "valuation" map, assigning to each atomic proposition $p \in \Psi$ some set of states $\|p\| \subseteq S$. Intuitively, the valuation tells us which facts hold in which worlds.

DEFINITION 2 (Relative Likelihood). The *relative likelihood* (or "odds") of a state s against a state t according to agent a, $[s : t]_a$, is defined as

$$[s : t]_a := \frac{P_a(s)}{P_a(t)}.$$

We graphically represent probabilistic epistemic state models in the following way: each state is drawn as an oval, having inside it the name of the state and the facts p that are "true" at the state (i.e. the atomic sentences p having this state in their valuation $\|p\|$); and for each agent $a \in \mathcal{A}$, we draw a-labeled arrows going from each state s towards all the states in the same a-information cell to which a attributes equal or higher odds (than to state s). Therefore, the qualitative arrows represent both the hard information (indistinguishability relation) and the probability ordering relative to an agent, pointing towards the indistinguishable states that she considers to be at least as probable. To make explicit the odds assigned by agents to states, we label these arrows with the quantitative information (followed by the agents' names in the brackets). For instance, the fact that $[s : t]_a = \frac{\alpha}{\beta}$ is encoded by an a-arrow from state s to state t labeled with the quotient $\alpha : \beta(a)$. For simplicity, we don't represent the loops relating each state to itself, since they don't convey any information that is specific to a particular model: in every model, every state is a-indistinguishable from itself and has equal odds $1 : 1$ to itself.

To illustrate probabilistic epistemic state model with odds, consider the initial situation of our urn example presented in Section 1 as pictured in Figure 1. In this initial model $\mathcal{M}_0$, it is equally probable that U_W or U_B is true (and therefore the prior odds are equal) and all agents know this. The actual state (denoted by the thicker oval) s_B satisfies the proposition U_B, while the state s_W satisfies the proposition U_W. The bidirectional arrow labeled with "1:1 (all a)" represents the fact that all agents consider both states equally probable.

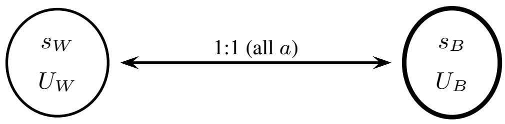


Figure 1. The initial probabilistic state model $\mathcal{M}_0$ of the urn example

DEFINITION 3 (Epistemic-probabilistic language). As in [9], the "static" language we adopt to describe these models is the epistemic-probabilistic language due to Halpern and Fagin [16]. The syntax is given by the following Backus-Naur form:

$$\varphi \ := \ p \mid \varphi \mid \varphi \wedge \varphi \mid K_a\varphi \mid \alpha_1 \cdot P_a(\varphi) + \ldots + \alpha_n \cdot P_a(\varphi) \geq \beta$$

where $p \in \Psi$ are atomic propositions, $a \in \mathcal{A}$ are agents and $\alpha_1, \ldots, \alpha_n, \beta$ stand for arbitrary rational numbers. Let us denote this language by $\mathcal{L}$.

The *semantics* is given by associating to each formula φ and each model $\mathcal{M} = (S, \mathcal{A}, (\sim_a)_{a\in\mathcal{A}}, (P_a)_{a\in\mathcal{A}})$, some *interpretation* $\|\varphi\|_{\mathcal{M}} \subseteq S$, given recursively by the obvious inductive clauses[4]. If $s \in \|\varphi\|_{\mathcal{M}} \subseteq S$, then we say that φ is *true* at

[4]It is worth noting that, when checking whether a given state s belongs to $\|\varphi\|$, every expression of the form $P_a(\psi)$ is interpreted conditionally on agent a's knowledge at s, i.e. as $P_a(\|\psi\| \cap \{s' \in S :$

state s (in model $\mathcal{M}$).

In this language, one can introduce strict inequalities, as well as equalities, as abbreviations, e.g.:

$$P_a(\varphi) > P_a(\psi) \;:=\; \neg(P_a(\psi) - P_a(\varphi) \geq 0),$$

$$P_a(\varphi) = P_a(\psi) \;:=\; (P_a(\varphi) - P_a(\psi) \geq 0) \wedge (P_a(\psi) - P_a(\varphi) \geq 0)$$

One can also define an expression saying that *an agent a assigns higher odds to φ than to ψ* (given her current information cell):

$$[\varphi : \psi]_a > 1 \;:=\; P_a(\varphi) > P_a(\psi)$$

To model the incoming of new information, we use *probabilistic event models*, as introduced by van Benthem *et alia* [9]: these are a probabilistic refinement of the notion of *event models*, which is the defining feature of Dynamic Epistemic Logic in its most widespread incarnation [4]. Here we use a simplified setting, which assumes introspection of subjective probabilities.

DEFINITION 4 (Probabilistic Event Models). A probabilistic event model $\mathcal{E}$ is a sextuple $(E, \mathcal{A}, (\sim_a)_{a\in\mathcal{A}}, (P_a)_{a\in\mathcal{A}}, \Phi, pre)$ such that:

- E is a set of possible events,
- $\mathcal{A}$ is a set of agents;
- $\sim_a \subseteq E \times E$ is an equivalence relation interpreted as agent a's epistemic indistinguishability between possible events, capturing a's hard information about the event that is currently happening;
- P_a gives a probability assignment for each agent a and each $\sim_a$-information cell. This captures some new, independent subjective probabilistic information gained by the agent during the event: when observing the current event (without using any prior information), agent a assigns probability $P_a(e)$ to the possibility that in fact e is the actual event that is currently occurring.
- Φ is a finite set of mutually inconsistent propositions (in the above probabilistic-epistemic language $\mathcal{L}$), called *preconditions*;
- pre assigns a probability distribution $pre(\bullet|\phi)$ over E for every proposition $\phi \in \Phi$. This is an "occurrence probability": $pre(e|\phi)$ expresses the prior probability that event $e \in E$ might occur in a(ny) state satisfying precondition ϕ;

$s' \sim_a s\}$. See [16], [9] for other details.

As before, the probability P_a can alternatively be expressed as probabilistic odds $[e : e']_a$ for any two events e, e' and any agent a. Our event models are drawn in the same fashion as our state models above: for each agent a, a-arrows go from a possible event e towards all the events (of a's information cell) to which a attributes equal or higher odds. As an example of an event model, consider the first observation of a ball in our urn case, as represented in the model $\mathcal{E}_1$ from Figure 2. Here a_1 draws a white ball from the urn and looks at it. According to all the other agents, two events can happen: either a_1 observes a white ball (the actual event w_1) or she observes a black one (event b_1). Moreover, only agent a_1 knows which event is the actual one. The expressions $pre(U_W) = \frac{2}{3}$ and $pre(U_B) = \frac{1}{3}$ depicted at event w_1 represents that the prior probabilities $pre(w_1 \mid U_W)$ that event w_1 occurs when U_W is satisfied is $\frac{2}{3}$ while the probability $pre(w_1 \mid U_B)$ that event w_1 happens when U_B is satisfied is $\frac{1}{3}$ (and vice versa for event b_1). The bidirectional arrow for all agents except a_1 represents the fact that agent a_1 can distinguish between the two possible events (since she *knows* that she sees a white ball), while the others cannot distinguish them and have (for now) no reason to consider one event more likely than the other, i.e., their odds are $1 : 1$.

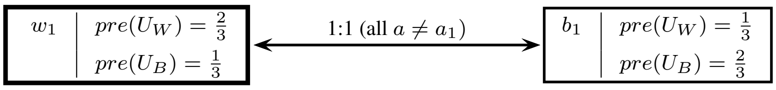


Figure 2. The probabilistic event model $\mathcal{E}_1$ of agent a_1 drawing a white ball

To model the evolution of the odds after new information is received, we now combine probabilistic epistemic state models with probabilistic event models using a notion of product update.

DEFINITION 5 (Probabilistic Product Update). Given a probabilistic epistemic state model $\mathcal{M} = (S, \mathcal{A}, (\sim_a)_{a\in\mathcal{A}}, (P_a)_{a\in\mathcal{A}}, \Psi, \| \bullet \|)$ and a probabilistic event model $\mathcal{E} = (E, \mathcal{A}, (\sim_a)_{a\in\mathcal{A}}, (P_a)_{a\in\mathcal{A}}, \Phi, pre)$, the updated state model $\mathcal{M} \otimes \mathcal{E} = (S', \mathcal{A}, (\sim'_a)_{a\in\mathcal{A}}, (P'_a)_{a\in\mathcal{A}}, \Psi', \| \bullet \|')$, is given by:

$$S' = \{(s, e) \in S \times E \mid pre(e \mid s) \neq 0\},$$

$$\Psi' = \Psi,$$

$$\|p\|' = \{(s, e) \in S' : s \in \|p\|\},$$

$$(s, e) \sim'_a (t, f) \text{ iff } s \sim_a t \text{ and } e \sim_a f,$$

$$P'_a(s, e) = \frac{P_a(s) \cdot P_a(e) \cdot pre(e \mid s)}{\sum\{P_a(t) \cdot P_a(f) \cdot pre(f \mid t) : s \sim_a t, e \sim_a f\}},$$

where we used the notation

$$pre(e \mid s) := \sum\{pre(e \mid \phi) : \phi \in \Phi \text{ such that } s \in \|\phi\|_{\mathcal{M}}\}$$

(so that $pre(e \mid s)$ is either $= pre(e|\phi_s)$ where ϕ_s is the unique precondition in Φ such that ϕ_s is true at s, or otherwise $pre(e \mid s) = 0$ if no such precondition ϕ_s exists).

This definition can be justified on Bayesian grounds: the definition of the new indistinguishability relation simply says that the agent puts together her old and new hard information[5]; while the definition of the new subjective probabilities is obtained by multiplying the old probability previously assigned to event e (obtained by applying the conditioning rule $P_a(e) = P_a(s) \cdot P_a(e \mid s) = P_a(s) \cdot pre(e \mid \phi_s)$) with the new probability independently assigned (without using any prior information) to event e during the event's occurrence, and then renormalizing to incorporate the new hard information. The reason for using multiplication is that the two probabilities of e are supposed to represent two independent pieces of probabilistic information.[6]

Again, it is possible, and even easier, to express posterior probabilities in terms of posterior relative likelihoods:

$$[(s,e) : (t,f)]_a = [s : t]_a \cdot [e : f]_a \cdot \frac{pre(e \mid s)}{pre(f \mid t)}.$$

The result of the product update of the initial state model $\mathcal{M}_0$ from Fig. 1 with the event model $\mathcal{E}_1$ of Fig. 2 is given by the new model $\mathcal{M}_0 \otimes \mathcal{E}_1$ of Fig. 3. The upper right state is the actual situation, in which U_B is true, but in which the first ball which has been observed was a white one. Agent a_1 *knows* that she observed a white ball (w_1), but she does not know which urn is the actual one, so her actual information cell consists of the upper two states, in which she considers U_W to be twice as likely as U_B. The other agents still cannot exclude any possibility.

This is going to change once the first agent announces her guess. To model this announcement we will use the standard *public announcements* of [28], where a (truthful) public announcement $!\varphi$ of a proposition φ is an event which has the effect of deleting all worlds of the initial state model that do not satisfy φ. Note that, public announcements $!\varphi$ can be defined as a special kind of probabilistic event models: take $E = \{e_{!\varphi}\}$, $\sim_a = \{(e_{!\varphi}, e_{!\varphi})\}$, $\Phi = \{\varphi\}$, $pre(e_{!\varphi} \mid \varphi) = 1$, $P_a(e_{!\varphi}) = 1$.

[5] This is the essence of the "Product Update" introduced by Baltag et alia [4], which forms the basis of most widespread versions of Dynamic Epistemic Logic.

[6] In fact, this feature is irrelevant for our analysis of cascades: no new non-trivial probabilistic information is gained by the agents during the events forming our cascade example. This is reflected in the fact that, in our analysis of cascades, we will use only event models in which the odds are $1 : 1$ between any two indistinguishable events.

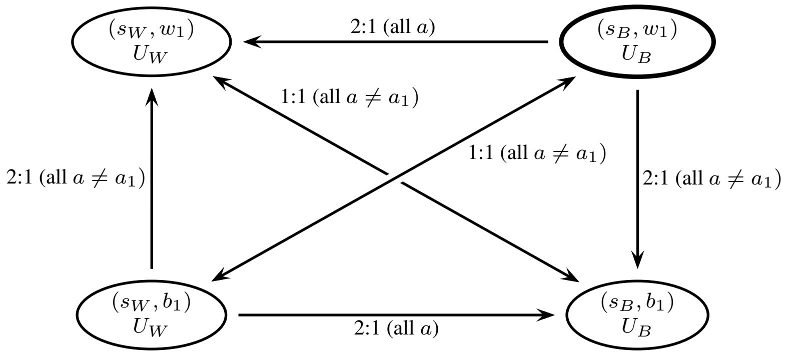


Figure 3. The updated probabilistic state model $\mathcal{M}_0 \otimes \mathcal{E}_1$ after a_1 draws a white ball

Now, after her private observation, agent a_1 publicly announces that she considers U_W to be more likely than U_B. This is a public announcement $!([U_W : U_B]_{a_1} > 1)$ of the sentence $[U_W : U_B]_{a_1} > 1$ (as defined above as an abbreviation in our language), expressing the fact that agent a_1 assigns higher odds to urn U_W than to urn U_B. Since all agents know that the only reason a_1 could consider U_W more likely than U_B is that she drew a white ball (her announcement can be truthful only in the situations in which she drew a white ball), the result is that all agents come to know this fact. This is captured by our modelling, where her announcement simply erases the states (s_W, b_1) and (s_B, b_1) and results in the new model $\mathcal{M}_1$ of Fig. 4.

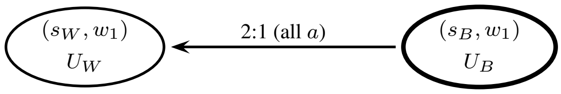


Figure 4. The updated probabilistic state model $\mathcal{M}_1$ after a_1's announcement

By repeating the above very reasoning, we know that, after another observation of a white ball by agent a_2 (the event model is as above in Fig. 2 but relative to agent a_2 instead of agent a_1) and a similar public announcement of $[U_W : U_B]_{a_2} > 1$, the resulting state model $\mathcal{M}_2$, depicted in Fig. 5, will be such that all agents now consider U_W *four* times more likely than U_B.

Let us now assume that agent a_3 enters the room and privately observes a black ball. The event model $\mathcal{E}_3$ of this action is in Figure 6, and is again similar to the earlier event model (Fig. 2) but relative to agent a_3 and this time, since a *black* ball

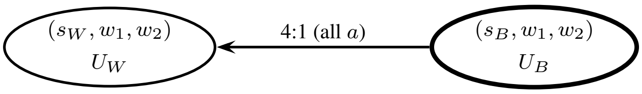


Figure 5. The updated probabilistic state model $\mathcal{M}_2$ after a_2's announcement

is observed, the actual event is b_3.

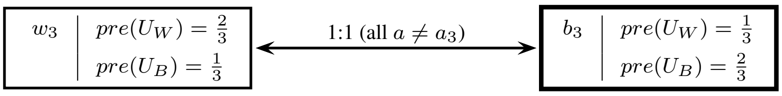


Figure 6. The probabilistic event model $\mathcal{E}_3$ of a_3 drawing a black ball

The result of a_3's observation is then given by the updated state model $\mathcal{M}_2 \otimes \mathcal{E}_3$ shown in Figure 7.

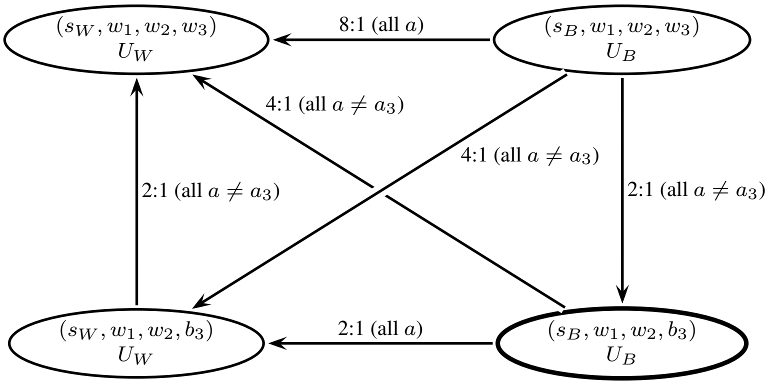


Figure 7. The probabilistic state model $\mathcal{M}_2 \otimes \mathcal{E}_3$ after a_3 draws a black ball

Since only agent a_3 knows what she has observed, her actual information cell only contains the states in which the event b_3 has happened, while all other agents cannot distinguish between the four possible situations. Moreover, agent a_3 still considers U_W more probable than U_B, *irrespective* of the result of her private observation (w_3 or b_3). So the fact that $[U_W : U_B]_{a_3} > 1$ is now *common knowledge* (since it is true at all states of the entire model). This means that announcing this fact, via a new public announcement of $[U_W : U_B]_{a_3} > 1$ will *not* delete any state: the model $\mathcal{M}_3$ after the announcement is simply *the same as before* (Fig. 7).

So the *third agent's public announcement bears no information whatsoever*: an informational cascade has been formed, even though all agents have reasoned correctly about probabilities. From now on, the situation will keep repeating itself: although the state model will keep growing, *all agents will always consider U_W more probable than U_B in all states* (irrespective of their own observations). This is shown formally by the following result.

PROPOSITION 6. *Starting in the model in Fig. 1 and following the above protocol, we have that: after $n-1$ private observations and public announcements $e_1, !([U_W : U_B]_{a_1} > 1) \dots, e_{n-1}, !([U_W : U_B]_{a_{n-1}} > 1)$ by agents $a_1, \dots, a_{n-1}$, with $n \geq 3$, $e_1 = w_1$ and $e_2 = w_2$, the new state model $\mathcal{M}_{n-1}$ will satisfy*

$$[U_W : U_B]_a > 1, \textit{ for all } a \in \mathcal{A}.$$

Proof. To show this, we prove a stronger

CLAIM: after $n-1$ private observations and announcements as above, the new state model $\mathcal{M}_{n-1}$ will satisfy

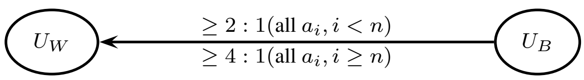
$$[U_W : U_B]_{a_i} \geq 2, \text{ for all } i < n, \text{ and}$$

$$[U_W : U_B]_{a_i} \geq 4, \text{ for all } i \geq n.$$

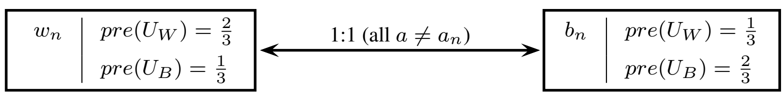
From this claim, the desired conclusion follows immediately.

PROOF OF CLAIM:

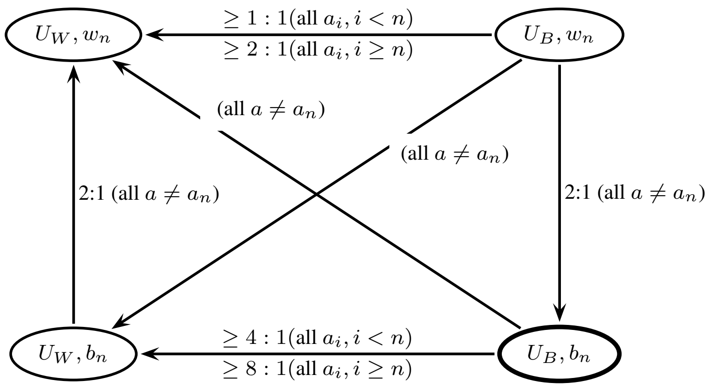
We give only a sketch of the proof, using an argument based on partial descriptions of our models. The base case $n = 3$ was already proved above. Assume the inductive hypothesis for $n-1$. By lumping together all the U_W-states in $\mathcal{M}_{n-1}$, and similarly all the U_B-states, we can represent this hypothesis via the following partial representation of $\mathcal{M}_{n-1}$:



Note that this is just a "bird's view" representation: the actual model $\mathcal{M}_{n-1}$ has 2^{n-2} states. To see what happens after one more observation e_n by agent n, take the update produce of this representation with the event model $\mathcal{E}_n$, given by:

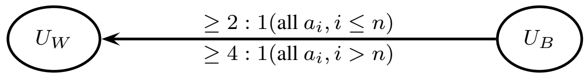


The resulting product is:



where for easier reading we skipped the numbers representing the probabilistic information associated to the diagonal arrows (numbers which are not relevant for the proof).

By lumping again together all indistinguishable U_W-states in $\mathcal{M}_{n-1}$, and similarly all the U_B-states, and reasoning by cases for agent a_n (depending on her actual observation), we obtain:



Again, this is just a bird's view: the actual model has 2^n states. But the above partial representation is enough to show that, in this model, we have $[U_W : U_B]_{a_i} \geq 2$ for all $i < n+1$, and $[U_W : U_B]_{a_i} \geq 4$ for all $i \geq n+1$. Since in particular $[U_W : U_B]_{a_n} > 1$ holds in all the states, this fact is common knowledge: so, after publicly announcing it, the model stays the same! Hence, we proved the induction step for n. ■

So, in the end, all the guesses will be wrong: the whole group will assign a higher probability to the wrong urn (U_W). Thus, we have proved that *individual Bayesian rationality with perfect higher-level reflective powers can still lead to "group irrationality"*. This shows that in some situations there simply is no higher-order information available to any of the agents to prevent them from entering the cascade; not even the information that they *are* in a cascade can help in this case. (Indeed, in our model, after the two guesses for U_W of a_1 and a_2, it is already common knowledge that a cascade has been formed!)

3 A Logical Model Based on Counting Evidence

A possible objection to the model presented in the previous section could be that it relies on the key assumption that the involved agents are perfect *Bayesian* reasoners. But many authors argue that *rationality* cannot be identified with *Bayesian* *rationality*. There are other ways of reasoning that can still be deemed rational without involving doing cumbersome Bayesian calculations. In practice, many people seem to use much simpler *"counting" heuristics*, e.g. guessing U_W when one has more pieces of evidence in favor of U_W than in favor of U_B (i.e. one knows that more white balls were drawn than black balls).

Hence, there are good reasons to look for a model of informational cascades based on counting instead of Bayesian updates. In this section we present a formalized setting of the urn example using a notion of rationality based on such a simple counting heuristic. The logical framework for this purpose is inspired by the probabilistic framework of the previous section. However, it is substantially simpler. Instead of calculating the probability of a given possible state, we will simply count the evidence in favor of this state. More precisely, we label each state with a number representing the strength of all evidence in favor of that state being the actual one. This intuition is represented in the following formal definition:

DEFINITION 7 (Counting Epistemic Models). A counting multi-agent epistemic model $\mathcal{M}$ is a structure $(S, \mathcal{A}, (\sim_a)_{a\in\mathcal{A}}, f, \Psi, \| \bullet \|)$ such that:

- S is a set of states,
- $\mathcal{A}$ is a set of agents,
- $\sim_a \subseteq S \times S$ is an equivalence relation interpreted as agent a's epistemic indistinguishability,
- $f : S \to \mathbb{N}$ is an "evidence-counting" function, assigning a natural number to each state in S,
- Ψ is a given set of atomic sentences,
- $\| \bullet \| : \Psi \to \mathcal{P}(S)$ is a valuation map.

We can now represent the initial situation of the urn example by the model of Figure 8. The two possible states s_W and s_B correspond to U_W (resp. U_B) being placed in the room. The notation U_W ; 0 at the state s_W represents that $f(s_W) = 0$ and that the atomic proposition U_W is true at s_W (and all other atomic propositions are false). The line between s_W and s_B labeled by "all a" means that the two states are indistinguishable for all agents a. Finally, the thicker line around s_B represents that s_B is the actual state.

We now turn to the issue of how to update counting epistemic models. However, first note that, at this stage there is not much that distinguish counting epistemic models from probabilistic ones. In the case the models are finite, one can simple sum the values of $f(w)$ for all states w in a given information cell and

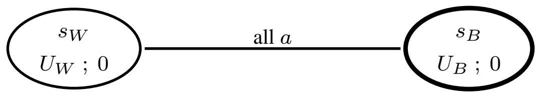


Figure 8. The initial counting model of the urn example

rescale $f(w)$ by this factor thereby obtaining a probabilistic model from a counting model. Additionally, assuming that all probabilities are rational numbers one can easily move the other way as well. In spite of this, when we move to dynamic issues, the counting framework becomes much simpler as we do not need to use multiplication together with Bayes' rule and renormalization, we can simply use addition. Here are the formal details:

DEFINITION 8 (Counting Event Models). A counting event model $\mathcal{E}$ is a quintuple $(E, \mathcal{A}, (\sim_a)_{a\in\mathcal{A}}, \Phi, pre)$ such that:

- E is a set of possible events,
- $\mathcal{A}$ is a set of agents,
- $\sim_a \subseteq E \times E$ is an equivalence relation interpreted as agent a's epistemic indistinguishability,
- Φ is a finite set of pairwise inconsistent propositions,
- $pre : E \to (\Phi \to (\mathbb{N} \cup \{\bot\}))$ is a function from E to functions from Φ to the natural numbers (extended with $\bot$)[7]. It assigns to each event $e \in E$ a function $pre(e)$, which to each proposition $\phi \in \Phi$ assigns the strength of evidence that the event e provides for ϕ.

As an example of a counting event model, the event model of the first agent drawing a white ball is shown in Figure 9. In this event model there are two events w_1 and b_1, where the actual event is w_1 (marked by the thick box). A notation like $pre(U_W) = 1$ at w_1 simply means that $pre(w_1)(U_W) = 1$.[8] Finally, the line between w_1 and b_1 labeled "all $a \neq a_1$" represents that the events w_1 and b_1 are indistinguishable for all agents a except a_1.

A counting epistemic model is updated with a counting event model in the following way:

[7]Here, $\bot$ essentially means "undefined": so it is just an auxiliary symbol used to describe the case when pre is a *partial* function.

[8]To fit the definition of counting event models properly, U_W and U_B must be pairwise inconsistent, however, this claim fits perfectly with the example where only one of the urns is placed in the room and we could simple replace U_W by $\neg U_B$.

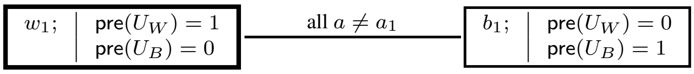


Figure 9. The counting event model of a_1 drawing a white ball

DEFINITION 9 (Counting Product Update). Given a counting epistemic model $\mathcal{M} = (S, \mathcal{A}, (\sim_a)_{a \in \mathcal{A}}, f, \Psi, \| \bullet \|)$ and a counting event model $\mathcal{E} = (E, \mathcal{A}, (\sim_a)_{a \in \mathcal{A}}, pre)$, we define the product update $\mathcal{M} \otimes \mathcal{E} = (S', \mathcal{A}, (\sim'_a)_{a \in \mathcal{A}}, f', \Psi', \| \bullet \|)$ by

$$S' = \{(s, e) \in S \times E \mid pre(s, e) \neq \bot\},$$

$$\Psi' = \Psi,$$

$$\|p\|' = \{(s, e) \in S' : s \in \|p\|\},$$

$$(s, e) \sim_a (t, f) \text{ iff } s \sim_a t \text{ and } e \sim_a f,$$

$$f'((s, e)) = f(s) + pre(s, e), \quad \text{for } (s, e) \in S',$$

where we used the notation $pre(s, e)$ to denote $pre(e)(\phi_s)$ for the unique $\phi_s \in \Phi$ such that $s \in \|\phi_s\|_{\mathcal{M}}$, if such a precondition $\phi_s \in \Phi$ exist, and otherwise we put $pre(s, e) = \bot$.

With this definition we can now calculate the product update of the models of the initial situation (Fig. 8) and the first agent drawing a white ball (Fig. 9). The resulting model is shown in Figure 10.

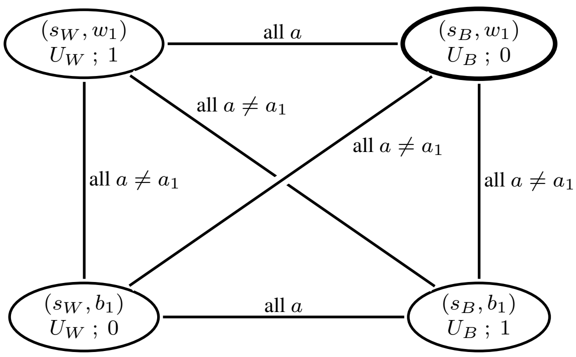


Figure 10. The updated counting model after a_1 draws a white ball

We need to say how we will represent the action that agent a_1 guesses for urn U_W. As in the probabilistic modeling we will interpret this as a public announcement. A public announcement of ϕ in the classical sense of eliminating all non-ϕ

states, is a special case of a counting event model with just one event e, $\Phi = \{\phi\}$, $\sim_a = \{(e,e)\}$ for all $a \in \mathcal{A}$, and $pre(e)(\phi) = 0$. Setting $pre(e)(\phi) = 0$ reflects the choice that we take public announcements not to provide any increase in the strength of evidence for any possible state, but only revealing hard information about which states are possible. In the urn example it is the drawing of a ball from the urn that increases the strength of evidence, whereas the guess simply convey information about the announcer's hard information about the available evidence for either U_W or U_B. Similar to the previous section, we will interpret the announcements as revealing whether their strength of evidence for U_W is smaller or larger than their strength of evidence for U_B.

We therefore require a formal language that contains formulas of the form $\phi <_a \psi$, for all formulas ϕ and ψ. The semantics of the new formula is given by:

$$\|\phi <_a \psi\|_{\mathcal{M}} = \{s \in S \mid f(a, s, \|\phi\|_{\mathcal{M}}) < f(a, s, \|\psi\|_{\mathcal{M}})\},$$

where for any given counting model $\mathcal{M} = (S, (\sim_a)_{a\in\mathcal{A}}, f, \Psi, \|\bullet\|)$ and any set of states $T \subseteq S$ we used the notation

$$f(a, s, T) := \sum\{f(t) : t \in T \text{ such that } t \sim_a s\}.$$

Now, the event that agent a_1 announces that she guesses in favor of U_W will be interpreted as a public announcement of $U_B <_{a_1} U_W$. This proposition is only true at the states (s_W, w_1) and (s_B, w_1) of the above model and thus the states (s_W, b_1) and (s_B, b_1) are removed in the resulting model shown in Figure 11.

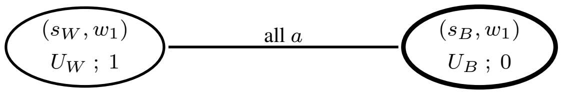


Figure 11. The counting model after a_1 publicly announces that $U_B <_{a_1} U_W$

Moreover, the event that a_2 draws a white ball can be represented by an event model identical to the one for agent a_1 drawing a white ball (Fig. 9) except that the label on the line should be changed to "all $a \neq a_2$". The updated model after the event that a_2 draws a white ball will look as shown in Figure 12. Note that in this updated model, $U_B <_{a_2} U_W$ is only true at (s_W, w_1, w_2) and (s_B, w_1, w_2), thus when a_2 announces her guess for U_W (interpreted as a public announcement of $U_B <_{a_2} U_W$) the resulting model will be the one of Figure 13.

Assuming that agent a_3 draws a black ball this can be represented by an event model almost identical to the one for agent a_1 drawing a white ball (Fig. 9). The only differences are that the label on the line should be changed to "all $a \neq a_3$)" and the actual event should be b_3. Updating the model of Figure 13 with this event will result in the model of Figure 14.

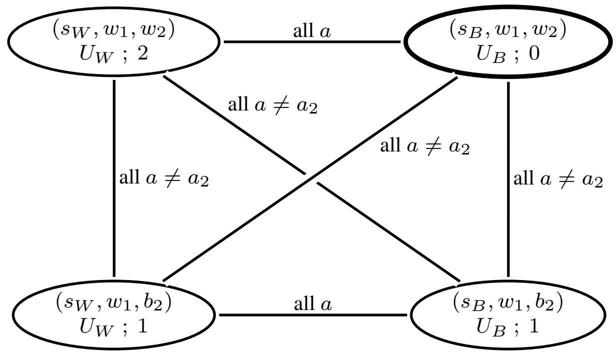


Figure 12. The updated counting model after a_2 draws a white ball

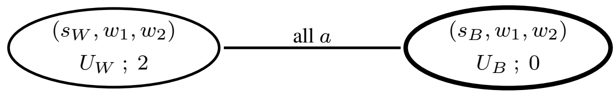


Figure 13. The counting model after a_2 publicly announces that $U_B <_{a_2} U_W$

Note that in Figure 14 the proposition $U_B <_{a_3} U_W$ is true in the entire model. Hence, agent a_3 has more evidence for U_W than U_B and thus, no states will be removed from the model when she announces her guess for U_W (a public announcement of $U_B <_{a_3} U_W$). If a_3 had drawn a white ball instead, the only thing that would have be different in the model of Figure 14 is that the actual state would be (s_B, w_1, w_2, w_3). Therefore, this would not change the fact that a_3 guesses for U_W and this announcement will remove no states from the model either. In this way, none of the following agents gain any information from learning that a_3 guessed for U_W. Subsequently whenever an agent draws a ball, she will have more evidence for U_W than for U_B. Thus, the agents will keep guessing for U_W. However, these guesses will not delete any more states. Hence, the models will keep growing exponentially reflecting the fact that no new information is revealed. In other words, *an informational cascade has started*. Formally, one can show the following result:

PROPOSITION 10. *Let $\mathcal{M}_n$ be the updated model after agent a_n draws either a white or a black ball. Then, if both a_1 and a_2 draw white balls (i.e. we are in the model of Fig. 12), then for all $n \geq 3$, $U_B <_{a_n} U_W$ will be true in all states of $\mathcal{M}_n$.*

In words: after the first two agents have drawn white balls all the following agents will all have more evidence for U_W than U_B (no matter which color ball

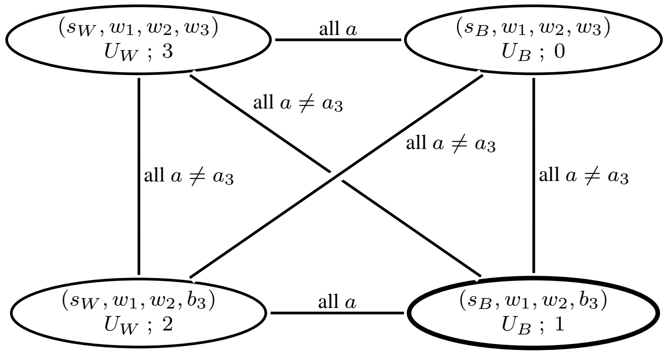


Figure 14. The updated counting model after a_3 draws a black ball

they draw) and will therefore guess for U_W, however, these guesses will be uninformative to the subsequent agents as the public announcement of $U_B <_a U_W$ will delete no worlds.

Before we can prove this proposition we need some definitions. For all $n \geq 3$, let $\mathcal{E}_n$ be the event model that agent a_n draws either a white ball (w_n) or a black ball (b_n)[9], for instance $\mathcal{E}_1$ is shown in Figure 9. Furthermore, let $\mathcal{M}_n$ denote the model obtained after updating with the event $\mathcal{E}_n$, hence $\mathcal{M}_n = \mathcal{M}_{n-1} \otimes \mathcal{E}_n$. The model $\mathcal{M}_3$ is shown in Figure 14. We will denote the domain of $\mathcal{M}_n$ by $dom(\mathcal{M}_n)$. For a proposition ϕ, we will by $f^n(\phi)$ denote $\sum\{f(s) \mid s \in \|\phi\|_{\mathcal{M}_n}\}$. Note that $f^2(U_W) = 2$, $f^3(U_W) = 5$, $f^2(U_B) = 0$, and $f^3(U_B) = 1$.

Now Proposition 2 follows from the following lemma:

LEMMA 11. *For all $n \geq 3$ the following hold:*

(i) Let $[w]_n := dom(\mathcal{M}_{n-1}) \times \{w_n\}$ and $[b]_n := dom(\mathcal{M}_{n-1}) \times \{b_n\}$. Then $[w]_n$ and $[b]_n$ are the only two information cells of agent a_n in $\mathcal{M}_n$, $dom(\mathcal{M}_n) = ([w]_n \cup [b]_n)$, and $|[w]_n| = |[b]_n| = 2^{n-2}$. Additionally, for all $k > n$, $\mathcal{M}_n$ contains only one information cell for agent a_k, namely the entire $dom(\mathcal{M}_n)$. Furthermore, U_W is true in 2^{n-3} states of $[w]_n$ and 2^{n-3} states of $[b]_n$, and similar, U_B is true in 2^{n-3} states of $[w]_n$ and 2^{n-3} states of $[b]_n$.

(ii) For $s \in [w]_n$, $f(a_n, s, U_W) = f^{n-1}(U_W) + 2^{n-3}$ and $f(a_n, s, U_B) = f^{n-1}(U_B)$. For $s \in [b]_n$, $f(a_n, s, U_W) = f^{n-1}(U_W)$ and $f(a_n, s, U_B) = f^{n-1}(U_B) + 2^{n-3}$.

[9]Which color ball a_n draws does not matter as it only affect which state will be the actual state.

(iii) $f^n(U_B) + 2^{n-2} < f^n(U_W)$.

(iv) $U_B <_{a_n} U_W$ *is true at all states of* $\mathcal{M}_n$.

Proof. The proof goes by induction on n. For $n = 3$ the statements $(i) - (iv)$ are easily seen to be true by inspecting the model $\mathcal{M}_3$ as shown in Figure 14 of section 3. We prove the induction step separately for each of the statements $(i) - (iv)$.

(i): Assume that (i) is true for n. Then, for agent a_{n+1} the model $\mathcal{M}_n$ consists of a single information cell with 2^{n-1} states where U_W is true in half of them and U_B in half of them. Considering the event model $\mathcal{E}_{n+1}$ it is easy to see that updating with this will result in the model $\mathcal{M}_{n+1}$, where there will be two information cells for agent a_{n+1} corresponding to the events w_{n+1} and b_{n+1}, i.e. $[w]_{n+1}$ and $[b]_{n+1}$, and each of these will have 2^{n-1} states. It is also easy to see that for all $k > n + 1$ there will only be one information cell for a_k. Finally, it is also easy to see that U_W will be true in 2^{n-2} states of $[w]_n$ and 2^{n-2} states of $[b]_n$ since U_w where true in 2^{n-2} states of $\mathcal{M}_n$ and likewise for U_B.

(ii): Assume that (ii) is true for n. Assume that $s \in [w]_{n+1}$. Then using (i), the definition of the product update, and (i) again, we get:

$$\begin{aligned}
f(a_{n+1}, s, U_W) &= \sum\{f(t) \mid t \in [w]_{n+1} \cap U_W\} \\
&= \sum\{f((t, w_{n+1})) \mid t \in \mathcal{M}_n, t \in U_W \text{ (in } \mathcal{M}_n)\} \\
&= \sum\{f(t) + pre(w_{n+1})(U_W) \mid t \in \mathcal{M}_n, t \in U_W \text{ (in } \mathcal{M}_n)\} \\
&= \sum\{f(t) + 1 \mid t \in \mathcal{M}_n, t \in U_W \text{ (in } \mathcal{M}_n)\} \\
&= f^n(U_W) + 2^{n-3}.
\end{aligned}$$

Similarly for U_B we get

$$\begin{aligned}
f(a_{n+1}, s, U_B) &= \sum\{f(t) \mid t \in [w]_{n+1} \cap U_B\} \\
&= \sum\{f(t) + pre(w_{n+1})(U_B) \mid t \in \mathcal{M}_n, t \in U_B \text{ (in } \mathcal{M}_n)\} \\
&= \sum\{f(t) + 0 \mid t \in \mathcal{M}_n, t \in U_B \text{ (in } \mathcal{M}_n)\} \\
&= f^n(U_B).
\end{aligned}$$

If $s \in [b]_{n+1}$ then,

- $f(a_{n+1}, s, U_W) = f^n(U_W)$ and
- $f(a_{n+1}, s, U_B) = f^n(U_B) + 2^{n-3}$,

follow by reasoning in similar manner.

(iii): Assume that $f^n(U_B) + 2^{n-2} < f^n(U_W)$. Note that from (i) and (ii) we have that

- $f^{n+1}(U_W) = 2f^n(U_W) + 2^{n-2}$,
- $f^{n+1}(U_B) = 2f^n(U_B) + 2^{n-2}$.

But, then

$$\begin{aligned} f^{n+1}(U_B) + 2^{n-1} &= 2f^n(U_B) + 2^{n-2} + 2^{n-1} \\ &= 2(f^n(U_B) + 2^{n-2}) + 2^{n-2} \\ &< 2f^n(U_W) + 2^{n-2} \\ &= f^{n+1}(U_W). \end{aligned}$$

(v): Now assume that $U_B <_{a_n} U_W$ is true at all states of $\mathcal{M}_n$. Consider agent a_{n+1}, we then want to prove that $U_B <_{a_{n+1}} U_W$ is true at all states of $\mathcal{M}_{n+1}$. That is, we need to prove that

$$f(a_{n+1}, s, U_B) < f(a_{n+1}, s, U_W),$$

for all $s \in dom(\mathcal{M}_{n+1})$. By (i) and the definition of f, we only have to consider two cases, namely when $s \in [w]_{n+1}$ and when $s \in [b]_{n+1}$. Moreover, by (ii) we just need to prove that

a) $f^n(U_B) < f^n(U_W) + 2^{n-2}$;

b) $f^n(U_B) + 2^{n-2} < f^n(U_W)$.

It is clear that $a)$ follows from $b)$ and $b)$ follows directly from (iii).

This completes the proof. ■

4 Conclusion and Further Research

We provided two logical models of the same example to argue for the somewhat counterintuitive claim that informational cascades are the direct consequence of individual rationality, even when rationality includes full higher-order reasoning or/and is non-Bayesian.

On the question whether real humans are Bayesian reasoners, a variation of the urn example was conducted as an experiment [2] to test whether people were using a simple counting heuristic instead of Bayesian update. In this variation, urn U_W contained 6 white and 1 black ball, whereas U_B contained 5 white and 2 black balls (in this way, a black ball provides more information than a white ball). Even though more decisions were consistent with Bayesian update than with a simple counting heuristic, the counting heuristic could not be neglected. For critique of the conclusion that most individuals use Bayesian updating, see [23]. In fact, these experimental results seem to rule out only the simplest counting heuristic where one compares the number of white balls against the number of black balls. However, one could argue that in this variation of the experiment another counting heuristic would be more appropriate (for instance counting one black ball several times, to reflect the higher weight of this piece of evidence) and that the experimental results are consistent with such an alternative. In general, our logic for counting evidence (presented in Section 3) can be used to capture various forms of "weighting" heuristics.

Another aim for future research is to generalize both the probabilistic logic of Section 2 and the logic for counting evidence of Section 3 to obtain a general logic of evidence that can capture both quantitative and qualitative approaches to reasoning about evidence, where an example of the latter is [11].

A deeper aim is to look at variations of our example, in order to investigate ways to *stop or prevent the cascade*. It is easy to see (using Condorcet's Jury Theorem) that, if we change the protocol to forbid all communication (thus making individual guesses private rather public), then by taking a poll at the end of the protocol, the majority vote will match the correct urn with very high probability (converging to 1 as the number of agents increases to infinity).

This proves that examples such as the one analyzed in this paper are indeed cases of "epistemic Tragedies of the Commons": situations in which *communication is actually an obstacle to group truth-tracking*. In these cases, a cascade can be stopped only in two ways: either by "irrational" actions by some of the in-group agents themselves, or else by outside intervention by an external agent with different information or different interests than the group. An example of the first solution is if some of the agents *simply disregard* the information obtained by public communication and make their guess solely on the basis of their own observations: in this way, they lower the probability that their guess is correct (which is "irrational" from their own individual perspective), but they highly increase the probability that the majority guess will be correct. An example of the second solution is if the protocol is modified (or only disrupted) by some external agent with regulative powers (the "referee" or the "government"). Such a referee can simply *forbid communication* (thus returning to the protocol in the Condorcet's Jury Theorem, which assumes independence of opinions). Or she might require *more communication*; e.g. require that the agents should announce, not only their beliefs about the urns, but also their *reasons* for these beliefs: *the evidence* supporting their beliefs. This evidence might be the number of pieces of evidence in favor of each alternative (in the case that they used the counting heuristics); or it might be the subjective probability that they assign to each alternative; or finally, it might be all their available evidence: i.e. the actual color of the ball that they observed (since all the rest of their evidence is already public). Requiring agents to share any of these forms of evidence is enough to stop the cascade in the above example.

One may thus argue that *partial communication* (sharing opinions and beliefs, but not sharing the underlying reasons is evidence) is the problem. More (and better) communication, more true deliberation based on sharing arguments and justifications (rather than simple beliefs), *may* sometimes stop the cascade. However, there are other examples, in which communicating some of the evidence is not enough: cascades can form even after each agent shares some of her private evidence with the group. It is true that a "total communication", in which everybody shares *all* their evidence, all their reasons, all the relevant facts, will be an effective way of stopping cascades (provided that the agents perfectly trust each other and they are justified to do so, i.e. nobody lies). In our toy example, this can be easily done: the relevant pieces of evidence are very few. But it is unrealistic to require such total communication in a real-life situation: the number of facts that might be of relevance is practically unlimited, and moreover it might not be clear to the agents themselves which facts are relevant and which not. So in practice this would amount to asking the agents to publicly share all their life experiences! With such a protocol, deliberation would never end, and the moment of decision would always be indefinitely postponed.

So in practice the danger remains: no matter how rational the agents are, how well-justified their beliefs are, how open they are to communication, how much time they spend sharing their arguments and presenting their evidence, there is always the possibility that all this rational deliberation will only lead the group into a cascading dead-end, far away from the truth. The only practical and sure way to prevent cascades seems to come from the existence of a significant number of "irrational" agents, who simply ignore or refuse to use the publicly available information and rely only on their own observations. Such extreme, irrational skeptics will very likely get it wrong more often than the others. But they will perform a service to society, at the cost of their own expected accuracy: due to them, in the long run society might correct its entrenched errors, evade its cascades and get better at collectively tracking the truth.

The conclusion is that communication, individual rationality and social deliberation are not absolute goods. Sometimes (and especially in Science, where the aim is the truth), it is better that some agents effectively isolate themselves and screen off some communication for some time. Allowing (and in fact encouraging) some researchers to "shut themselves up in their ivory tower" for a while, pursuing their independent thinking and tests without paying attention to the received knowledge in the field (and preferably without access to any Internet connections), may actually be beneficial for the progress of Science.

Acknowledgments.

The research of Sonja Smets and Zoé Christoff leading to these results has received funding from the European Research Council under the European Communitys Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement no. 283963.

Jens Ulrik Hansen is sponsored by the Swedish Research Council (VR) through the project "Collective Competence in Deliberative Groups: On the Epistemological Foundation of Democracy".

BIBLIOGRAPHY

[1] L.R. Anderson and C.A. Holt. Classroom games: Information cascades. *The Journal of Economic Perspectives*, 10(4):187–193, 1996.

[2] L.R. Anderson and C.A. Holt. Information cascades in the laboratory. *The American Economic Review*, 87(5):847–862, 1997.

[3] G. Aucher. A combined system for update logic and belief revision. Master's thesis, ILLC, University of Amsterdam, Amsterdam, the Netherlands, 2003.

[4] A. Baltag, L. Moss, and S. Solecki. The logic of common knowledge, public announcements and private suspicions. In *TARK Proceedings*, pages 43–56, 1998.

[5] A. Baltag and S. Smets. A qualitative theory of dynamic interactive belief revision. *Logic and the Foundations of Game and Decision Theory (LOFT7)*, (1–3):11–58, 2008.

[6] A.V. Banerjee. A simple model og herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817, 1992.

[7] J. van Benthem. Dynamic logic for belief revision. *Journal of applied and non-classical logics*, 17(2):129–155, 2007.

[8] J. van Benthem. *Logical Dynamics of Information and Interaction*. Cambridge Univ. Press, 2011.

[9] J. van Benthem, J. Gerbrandy, and B. Kooi. Dynamic update with probabilities. *Studia Logica*, 93(1), 2009.

[10] J. van Benthem and F. Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-classical Logics*, 17(2):157–182, 2007.

[11] J. van Benthem and E. Pacuit. Dynamic logics of evidence-based beliefs. *Studia Logica*, 99(1–3):61–92, 2011.

[12] S. Bikhchandani, D. Hirshleifer, and I. Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5):992–1026, 1992.

[13] S. Bikhchandani, D. Hirshleifer, and I. Welch. Learning from the behavior of others: Conformity, fads, and informational cascades. *Journal of Economic Perspectives*, 12(3):151–170, 1998.

[14] H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*. Synthese Library, Springer, 2007.

[15] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets. Reasoning about a Highly Connected World*. Cambridge University Press, 2010.

[16] R. Fagin and Halpern J.Y. Reasoning about knowledge and probability. *Journal of the ACM*, 41(2):340–367, 1993.

[17] A. Goldman. *Knowledge in a Social World*. Oxford Univ Press, 1999.

[18] A. Goldman and D. Whitcomb, editors. *Social Epistemology, Essential Readings*. Oxford University Press, 2011.

[19] R.E. Goodin. *Reflective Democracy*. Oxford Univ. Press, 2003.

[20] J.U. Hansen. A logic-based approach to pluralistic ignorance. In Jonas De Vuyst and Lorenz Demey, editors, *Future Directions for Logic – Proceedings of PhDs in Logic III*, volume 2 of *IfColog Proceedings*, pages 67–80. College Publications, 2012.

[21] V.F Hendricks, P.G Hansen, and R. Rendsvig. Infostorms. *Metaphilosophy*, 44:301–326, 2013.

[22] W. Holliday. Trust and the dynamics of testimony. In *Logic and Interactive Rationality, Dynamics Yearbook 2009*, pages 147–178. Institute for Logic, Language and Computation, Universiteit van Amsterdam, 2010.

[23] S. Huck and J. Oechssler. Informational cascades in the laboratory: Do they occur for the right reasons? *Journal of Economic Psychology*, 21(6):661–671, 2000.

[24] B.P. Kooi. Probabilistic dynamic epistemic logic. *J. of Logic, Lang. and Inf.*, 12(4):381–408, September 2003.

[25] C. List. Group knowledge and group rationality: A judgment aggregation perspective. In *Social Epistemology, Essential Readings*, pages 221–241. Oxford University Press, 2011.

[26] C. List and R.E. Goodin. Epistemic democracy: Generalizing the condorcet jury theorem. *Journal of Political Philosophy*, 9(3):277–306, 2001.

[27] F. Liu. *Changing for the Better: Preference Dynamics and Agent Diversity*. PhD thesis, University of Amsterdam, 2007.

[28] J. Plaza. Logics of public communications. In M.L. Emrich, M.S. Pfeifer, M. Hadzikadic, and Z.W. Ras, editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, pages 201–216. Oak Ridge National Laboratory, 1989.

[29] C. Proietti and E. Olsson. A ddl approach to pluralistic ignorance and collective belief. *Journal of Philosophical Logic*, 2013.

[30] J. Seligman, P. Girard, and F. Liu. Logical dynamics of belief change in the community, under submission. 2013.

[31] J. Seligman, F. Liu, and P. Girard. Logic in the community. In Mohua Banerjee and Anil Seth, editors, *Logic and Its Applications*, volume 6521 of *Lecture Notes in Computer Science*, pages 178–188. Springer Berlin Heidelberg, 2011.

[32] J. Surowiecki. *Wisdom of Crowds*. Anchor books USA, 2004.