GAMES AND LOGICS FOR INFORMATIONAL CASCADES

MSc Thesis *(Afstudeerscriptie)*

written by

Andreea Cristina Achimescu
(born 1st October, 1987 in Drobeta-Turnu Severin, Romania)

under the supervision of
Dr. Alexandru Baltag and **Dr. Joshua Sack**
and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:	**Members of the Thesis Committee:**
21st February, 2014	Prof. Dr. Maria Aloni
	Dr. Alexandru Baltag
	Prof. Dr. Johan van Benthem
	Prof. Dr. Jan van Eijk
	Prof. Dr. Vincent Hendricks
	Dr. Joshua Sack

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

ABSTRACT

Informational cascades occur when it is optimal for decision-makers to abandon their own private information in favour of inferences they make about other individuals' information. The informational cascade model, centred on the core notion of Bayesian update, has been able to explain, at least partially, many observed conformism patterns in social settings.

The aim of this thesis is to put the informational cascade model in game theoretic terms and analyse it using a new probabilistic epistemic logic. The strength of game theory lies in its mathematical apparatus that structures and identifies strategic choices. Regarding informational cascades as games of imperfect information with chance moves allows us to capture, in a natural way, the reasoning of agents engaged in an informational cascade. The strength of a logical treatment of games is, among others, the incorporation of all levels of an agent's beliefs into an analysis of optimal behaviour. This attribute is instrumental in analysing games with paradoxical collective outcomes like informational cascades. False cascades, a term that denotes people herding on the wrong decision, are paradoxical outcomes because they are intuitively inconsistent with the intentions of the individuals that generate them.

We first formalize the Urn Model, the canonical example of informational cascades, as a game of imperfect information. Next, we prove that the unique perfect Bayesian equilibrium of this game sometimes leads to false cascades. Then, we determine various changes that need to be put in place in order to ensure more socially desirable outcomes in informational cascade games. Finally, we propose a new logic, Probabilistic Logic of Communication and Change, to treat social dynamics of information games. We prove it is a sound and complete logic with respect to Bayesian Kripke structures and proceed to apply it to sequential social information flow games.

ACKNOWLEDGEMENTS

I would like to begin by thanking my supervisors, Alexandru Baltag and Joshua Sack, for their tenacity in seeing through to the completion of this thesis, for their patience and generosity in sharing insights with me and their sustained encouragement and guidance throughout the last 6 months. In particular, I would like to thank Alexandru for his support and mentorship during my time at ILLC and for communicating his passion for Logic so expressively to his students. Similarly, I would like to thank Joshua for agreeing to take on this project and help me navigate through countless technical difficulties.

I am especially indebted to Tanja Kassenaar, Ulle Endriss and Frank Veltman for giving me the means to carry out my studies in the Netherlands. I am grateful to my fellow students, who made Logic a lot of fun, and from whom I learned immensely. I am however particularly appreciative of Aybuke Ozgun, Will Frager, Facundo Carreiro, Nathaniel Forde, Paolo Galeazzi and Rasmus Rendsvig for reading my thesis and providing me with numerous insights and pieces of advice.

I wish to express my gratitude to Peter van Ham and his wonderful family, for making Amsterdam feel like home and for so readily offering me their advice and support during times of hardship. I especially wish to thank Peter for his dedication to helping me succeed and for his generous acts of kindness that in one way or another have found their way into my thesis.

Finally, I would like to thank my family for their extraordinary sacrifice over the last few years and their uncompromising dedication to helping me achieve my dreams.

Andreea Achimescu
Utrecht, February 2014

Imitation is not just the sincerest form of flattery — it's the sincerest form of learning.

George Bernard Shaw

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

INTRODUCTION

An informational cascade is, *grosso modo*, a mechanism that explains conformity in certain social situations. Suppose, as a first example that, while in a new town, you need to choose a dining restaurant. Based on your own research, you intend to go to restaurant A. However, when you arrive there, you notice that restaurant A is empty, whereas, just across the street, in restaurant B, there is a big crowd of people. At this point, if you believe that the other diners have obtained individual, but imperfect information with regard to which of the two restaurants is superior, and if you believe they have similar tastes in restaurants as you, then it might be rational for you to join them at restaurant B. This would make sense if your private information is outweighed by the public information you infer from so many other people having chosen the other venue. In this case, we say that an informational cascade has occurred (Banerjee, 1992).

Informational cascades are worth studying because they show how our intuition about social knowledge fails. The expectation is that groups are able to track the truth better than the individuals composing them, by virtue of *communication*, whereby the individual pieces of information each member of the group possesses are pooled together and analysed in a rational manner. However, informational cascades show how communication, when structured sequentially, can impair the attainment of truth. This stems from a fundamental misalignment of interests between the individual and the group, in the use of communication, which is individually rational and, at the same time, group-irrational. This feature, together with the strategic considerations that rationality engenders among members of a group, offers a natural interpretation of informational cascades within game theory.

This thesis pursues a game theoretic treatment of informational cascades, in an attempt to explain their outcomes as consequence of strategic reasoning. Moreover, a game theoretic perspective will open the door to investigating various other variants of cascade-like games and in this way broaden the applicability of informational cascades as explanatory tools to other social dynamics of information phenomena.

One class of examples is given by truth-tracking situations. These can easily be seen as cascade models where individuals' payoffs are tied to the group's performance. Later, informational cascade models will provide a platform for building a logic for games, grounded in Bayesian reasoning. We provide a sound and complete axiomatization for this logic with respect to Bayesian Kripke frames and then move on to proving the occurrence of cascades using the newly defined semantics. The logical treatment of cascades at the close of the thesis contains a formal representation of all levels of reasoning the agents in the cascade engage in, and a stepping stone in the direction of a logical treatment of imperfect information games.

The thesis is structured as follows: in Chapter 2 we define an informational cascade, give an overview of the main literature results and draw upon the recent global financial turmoil to exemplify its workings. We then present the canonical cascade example, the urn model, and close the chapter with a theorem that computes the likelihood of a false cascade occurring.

Chapter 3 deals with a classical game theoretic formalization of informational cascades. We take the canonical model presented in the previous chapter and model it as an imperfect information game with chance moves. Then we prove that the reasoning entrenched in informational cascades represents a Perfect Bayesian Nash Equilibrium (PBNE) of the cascade game. Further, we define a refinement to the PBNE that applies exclusively to urn-type games.

Chapter 4 has a two-fold aim: to generalize the cascade setting to other social dynamics of information games and to implement structural changes that incentivise agents to pursue socially optimal strategies. Our analysis shows that there is little we can do about avoiding cascades, since they emerge as deeply seated consequences of best-response behaviour in almost all of these games.

Chapter 5 presents a new logic suited for a probabilistic epistemic analysis of social dynamic of information games. The logic we present, called Probabilistic Logic of Communication and Change (PLCC), has both a sound and a weakly complete axiomatization with respect to Bayesian Kripke frames, a kind of probabilistic epistemic frames grounded in Bayesian decision theory. PLCC represents a probabilistic version of the Logic of Communication and Change introduced by van Benthem et al. (2006), which is, in turn, the epistemic interpretation of propositional dynamic logic with updates.

Chapter 6 gives a taste of the potential applications of PLCC in game theory, by proving that a false cascade ensues in the Urn Game, as soon as after the first two players' moves. The disconcerting result of this analysis is that no amount of higher level information that the agents might possess, including common knowledge of the possibility of a cascade, is sufficient to prevent it.

The conclusion of the thesis is that informational cascades are inescapably a “paradox” of rationality, an inherent tension that arises when individual rationality is applied in certain social situations. Game theory, mechanism design, logic, Bayesian reasoning or even combinations of all these are not always able to dissolve the paradox. Informational cascades remain a fundamental challenge to social rationality.

2

THE PHENOMENON OF INFORMATIONAL CASCADES

This chapter introduces the reader to the phenomenon of informational cascades. It contains an informal definition of informational cascades, followed by a real-life example of a cascade inspired from the recent global financial turmoil and finally, a standard formal analysis of informational cascades. Along the way, we review the main literature results in cascade theory and contribute to it a theorem about the likelihood of false cascades forming in the canonical cascade setting.

2.1 DEFINITION

Informational cascades provide an explanation for how groups of people quickly converge, in a rational way, to a common behavioural pattern, based on very little information. The notion was introduced by the seminal work of Bikhchandani et al. (1992), and ever since economists, social psychologists and sociologists alike have used informational cascades to explain the occurrence and evolution of the phenomena of people falling in line with the crowd (Surowiecki (2005), Bicchieri (2005), Easley and Kleinberg (2010), Hansen et al. (2013)).

Definition 1: Informational Cascades

An *informational cascade* is defined as a situation in which it is optimal for an individual, having observed the actions of other individuals, to follow the behaviour of his predecessor without regard to his own information (Bikhchandani et al., 1992).

In general, informational cascades occur in settings where:

1. People make decisions sequentially, with every person being able to observe the actions of earlier decision-makers.

   This is, in some sense, the pivotal assumption of the model. It gives way for agents to interpret the actions of others, and infer from their choices the private information the others possess.

2. There is common knowledge that the individuals in the cascade are rational.

   Rationality is interpreted here as instrumental rationality: what means to take to achieve the satisfaction of your preferences, given your beliefs and information about the probabilities that various means will achieve those ends (Kolodny and Brunero, 2013). Instrumental rationality is a standard assumption in decision theory, game theory and economics. For example, Aumann (2006) states: "A person's behaviour is rational if it is in his best interests, given his information".

3. Each person has some piece of private information that he receives prior to making a decision.

   This aptly models the numerous real-life situations of decision-making under imperfect information.

4. Individuals in the sequence do not have access to others' private information.

   Again, as above, this assumption reflects in the empirical phenomena we want to explain.

5. The space of available actions is discrete.

Interpretation of the definition

We would like to make two observations about the informational cascades definition of Bikhchandani et al. (1992). First is that *sensu stricto*, informational cascades are interpretations of certain phenomena, not the phenomena themselves. More precisely they are interpretations of situations of apparent conformity, that satisfy the five conditions laid out earlier. Observe that in the definition, the word "optimal" appears, which hints at the inherent formal interpretation, represented by standard *Bayesian reasoning*.[1] Later in this chapter we present the Bayesian analysis of informational cascades in detail.

The second observation related to the informational cascades definition is that it might be misleading to say agents in a cascade disregard their own information. A more accurate description would be that the private information of an agent caught in a cascade has no bearing on his decision. Put differently, an agent's decision does not carry any

1 By Bayesian reasoning we understand the use of Bayes' Rule for computing conditional probabilities together with the Simple Principle of Conditionalization and more, broadly, Bayesian decision theory. For more details see Talbott (2013)

information regarding his private signal to the others that observe his behaviour.

To facilitate our discussion of informational cascades, we divide them into two categories:

- *false cascades* (or reverse cascades) in which the individuals stabilize on the wrong choice,
- *true cascades* in which individuals stabilize on the correct outcome.

A defining characteristic of cascades is *fragility*, as introduced by Bikhchandani et al. (1992) and Hirshleifer and Anderson (1994). According to Easley and Kleinberg (2010) cascades are fragile if, after having persisted for a long time, can be overturned with comparatively little effort . Fragility is due to the scarcity of public information, which can sometimes be composed only of the information available to the first two individuals in a group. Due to this,the release of as little as one more individual's information is enough to break the cascade. Later in the thesis we show that the public release of the private signal of one player, during a play of the canonical game of informational cascades with a large number of players, is sufficient to reduce the likelihood of a false cascade occurring by almost a quarter. For the detailed proof, see Theorem 6.

The informational cascade model thus presents herding as the result of hyper-rational individual behaviour. However, one of its perplexing features is that individuals end up taking the wrong collective decision, despite the group having epistemic access to the truth. One may think that, perhaps, agents with higher-order reasoning powers could engage in meta-level considerations that would help avoid these kind of unfortunate situations. However, as we will show later, this is not the case with certain cascades, which survive this type of higher order reasoning. For this reason, one may find in informational cascades a striking resemblance with the Tragedy of the Commons-type of situations, where people fail to consider the social impact of their actions (Baltag et al., 2013).

2.2 A CASE FOR HOMO ECONOMICUS

Herding is a well-documented phenomenon in financial markets (Welch (1992), Avery and Zemsky (1998), Bikhchandani and Sharma (2001)). Stock market bubbles, financial speculation, bank runs are just a few instances where herding behaviour is observed. Roughly speaking, there are two types of explanations for the observed herding behaviour in the financial markets (Devenow and Welch, 1996, p. 604). There are the explanations which claim that investors are boundedly rational, relying on rules of thumb and informational processing short-cuts in order to reach a decision (Shiller, 1995). Then, there are theories that offer a rational explanation of herding in financial markets (Scharfstein and Stein, 1990), amongst which informational cascades (Bikhchandani and Sharma, 2001).

In the remainder of the section, we argue for the appropriateness of informational cascades as explanations of how rational investors may be led astray and create a market bubble. Seen as informational cascades, market bubbles illustrate the failure of disseminating information about the true value of an asset (Shiller, 2000).

In economic theory, the current mainstream explanation of investor behaviour is the Efficient Market Hypothesis (EMH), one of the tenets of neoclassical economics. This theory is a form of the wisdom of the crowds, which states that investors are making rational, independent decisions, which in aggregate, are better than any individual one. According to this theory, financial markets always get asset prices right given the available information. However, there are many empirical situations that the Efficient Market Hypothesis cannot explain, for example numerous financial market phenomena that are characterized by both fragility and wave-like dynamics, like stock market bubbles. A bubble is a type of investing phenomenon that occurs when there is "trade in high volumes at prices that are considerably at variance with intrinsic values" (Smith et al., 1993). This definition characterizes both the boom and burst phase of a bubble, that we often encounter in reality. While some economists deny the existence of bubbles, numerous historical examples have convinced many others to find new explanations for investor behaviour, divorced from the EMH.

There are two main grounds on which to attack the EMH: by challenging the assumption that investors are rational, and by challenging that investors make decisions independently. Behavioural economics takes the first route of attack and asserts that investors are irrational. The apparent irrationality of investors is due to known biases in human cognition, like the tendency to care more about small losses than small gains or the tendency to extrapolate too hastily from small samples (Krugman, 2009).

An alternative explanation for the EMH failing to describe market bubbles is the contention that people decide independently of one another. This constitutes the second route of attack, adopted by the informational cascade model. According to it, investors are hyper-rational individuals who, as part of their decision, extract the information that other people's choices reveal. Since information gathering is costly, in situations of imperfect information like financial markets, it becomes rational for individuals to appeal to the public information available. It is arguably less costly to do so than to gather additional private information.

The result of taking into consideration public information is that, after a certain stage, the private information of decision makers does not make its way into the market. Thus a self-defeating loop is created, in which an avidness for information prevents people from learning anything new. This situation seems to lend itself very well to a cascade-like explanation. Take as an example the housing market bubble that represented the ignition of the 2008 financial crisis, the worst after the Great Depression. In Shiller (2008), it is argued that the housing bubble is an informational cascade: in a sequence of perfectly rational individuals, each decided erroneously, but rationally, to invest in the housing market, based on the imperfect information they had.

Following Shiller's argument, the pre-crisis housing market was one in which houses were of a low investment value. Talking in stylized terms, there were two types of signals that individuals in the market could receive: a high signal, which indicated that houses were of high investment value and a low signal, which signalled houses were of a low investment value. Say that the high signal occurred with probability 0.6 whereas the low signal with probability 0.4. Consequently, every individual had a piece of private information that lead, 60% of the time, to the right decision.

We could look at the group of potential investors in the housing market as making decisions in a sequence. Consider a first individual in the sequence. He decided to pay a high price for a house based on receiving a high signal in a low investment value world, thereby signalling to the other people that houses are a good investment. The second investor, after having received a high signal and having observed the public action of the first investor, decided to buy a house for a high price too. At this point, whatever the signal the third person received, the information he inferred from the previous players' actions outweighed that carried by his own signal, and therefore decided to purchase a high priced house.

Any subsequent player, being able to reproduce the reasoning of the third individual, if rational, also chose to enter the housing market. Intuitively, what happened was that, as people observed others purchasing houses at higher and higher prices, they concluded that these investors' information about the market outweighed their own. This created a housing market bubble, where prices were set much above their intrinsic value. Later in the chapter we will show that the probability of this type of bubble forming is 20%, which means that one in five times, a group of individuals will reach the wrong collective decision.

2.3 THE BAYESIAN ANALYSIS OF INFORMATIONAL CASCADES

Moving towards an abstract model of informational cascades, we now present the Urn Model. This represents a canonical example of a formal representation of informational cascades. The Urn Model was, to our knowledge, first described as part of an experimental setting of Anderson and Holt (1996) and then presented, with minor variations, in subsequent cascade theory literature.

We present a narrative of the Urn Model that is close to Baltag et al. (2013): each individual in a group tries to correctly identify the proportion of black and white marbles contained in an urn, that was placed in a room by "Nature". It is common knowledge that the urn can either contain a mix denoted W of $2/3$ white marbles and $1/3$ black marbles or a mix denoted B of $1/3$ white marbles and $2/3$ black marbles, each equally likely. The agents enter the room one at a time. Upon entering, each agent draws a marble from the urn, looks at it, and puts it back. Then he makes a guess as to which mix he thinks is more probable, W or B, and writes it on a whiteboard, for all the subsequent agents entering the room to see. After everyone has entered the room, the game ends and the arbiter individually rewards the players who have guessed correctly.

Formally, consider a game with n players, $\{1, \ldots, n\}$, where every individual is rational, in the sense that he obeys by the Bayesian decision-theoretic reasoning. It is common knowledge amongst agents that there are two possible states of the world: urn W and urn B, each equally likely *ex ante* and that every agent is rational. Urn W contains a proportion of $2/3$ white marbles and $1/3$ black marbles, whereas urn B contains a proportion of $2/3$ black marbles and $1/3$ white marbles. Each agent in turn receives an independent, private signal, w or b, which has the following probability of being true, given the true state of the world:

$$P(w|W) = P(b|B) = 2/3$$
$$P(w|B) = P(b|W) = 1/3$$

Intuitively, this signal corresponds to a private, independent draw of a marble from the real urn. Note that we interchangeably make use of the following terminology for signals that have a probability of occurring higher than $\frac{1}{2}$: *high precision*, high probability or *high quality* signals. Signals with a probability of occurrence of less than $\frac{1}{2}$ we call *low precision*, low probability or *low quality* signals.

Then, agents make a prediction about the true urn based on their private signal and the publicly announced predictions of previous players in the sequence. They do not observe previous agents' private signals, only their actions. There is a *tie-breaking rule* called the *self-biased rule*, which dictates that, whenever an agent assigns the same probability to both states of the world, respectively 1/2, he chooses the action indicated by his own signal. There is common knowledge of the tie-breaking rule amongst the players. A player receives a positive payoff if and only if his guess is correct.

The urn cascade

To illustrate the concept of an informational cascade, we show the rational reasoning steps that lead agents into one, after only two moves.

Assume the real state of the world is B. Player 1 draws a marble from urn B, unbeknownst to him, looks at it and makes his guess based solely on this information (since there are no actions of previous players that he can use to extract further information from). Assume his private signal is w. Then, conditioning upon observing a white marble, player 1 infers, using Bayes' theorem, that a W world is twice as likely than a B world:

$$\begin{aligned}P(W|w) &= \frac{P(w|W)P(W)}{P(w)} = \frac{P(w|W)P(W)}{P(w|W)P(W) + P(w|B)P(B)} \\ &= \frac{\frac{2}{3} \cdot \frac{1}{2}}{\frac{1}{2}} = \frac{2}{3}\end{aligned}$$

All subsequent players are able to infer that player 1 saw a white marble, given that they all know he is rational. Consequently, player 2 makes his guess based on the private signal of player 1, w_1, and his own private draw. Assume he receives a w_2 signal. Then it becomes obvious that player 2 will predict W. However, were player 2 to receive a b_2 signal, he would have predicted B, by the tie breaking rule. In fact, player 2's action fully reflects his signal. Then, by announcing W, player 2 reveals his private signal to all subsequent players.

Now consider player 3, whose information can be divided into two categories: the information publicly available to him, constituted of player 1 and player 2's actions, (W_1 and W_2) and his own private information, either w_3 or b_3. Then player 3, having inferred from previous actions the signals w_1 and w_2 of player 1 and player 2 respectively, is now rationally bound to choose W, no matter what his private signal is. This can easily be shown by tracing player 3's Bayesian reasoning:

- If player 3 draws a w_3 then it is obvious that he will choose W.
- If player 3 draws a b_3 then, by applying his Bayesian reasoning, his subjective degree of belief in W will be higher than in B:

$$P(W \mid w_1 \wedge w_2 \wedge b_3) = \frac{P(w_1 \wedge w_2 \wedge b_3 \mid W) \cdot P(W)}{P(P(w_1 \wedge w_2 \wedge b_3)} = \frac{\frac{4}{27} \cdot \frac{1}{2}}{\frac{6}{27} \cdot \frac{1}{2}} = \frac{2}{3}$$

  whereas

$$P(B \mid w_1 \wedge w_2 \wedge b_3) = \frac{1}{3}$$

We can say that player 3's private information is "dominated" by the public information he receives from previous players. The third agents' reasoning can then be reproduced by any subsequent individual $i > 3$ in the sequence, who then understands that player 3's action carries no informational content.

Therefore, agent 4 is exactly in the same informational state as agent 3, since the information available to him is represented by the signals of the first two players and his own. Engaged in the same rational reflection as agent 3, agent 4 decides to choose W, irrespective of his own private signal.

By iterating this last step for every subsequent agent, we conclude that all players in the game will choose W. Given that we assumed the true state of the world is B, then what results is a false informational cascade, where agents' decisions stabilize on the wrong choice forever.

2.4 LIKELIHOOD OF CASCADES

In this section, we prove a limit result about the probability of false cascades arising during a game with a large number of players.[2].

> *Theorem 1: Likelihood of false cascades*
>
> If the number of players $n \to \infty$, and the probability of a low signal is p, then the likelihood of a false cascade starting during a play of the urn game is $\frac{p^2}{(2p^2-2p+1)}$.

2 Bikhchandani et al. (1992) computes such a probability but for a slightly different setting. The difference in the tie-breaking rule used in this thesis and the one in Bikhchandani et al. (1992), accounts for the difference in the calculation of the likelihood of a false cascade occurring during a play of the game. While we use the self-biased rule, they use the coin-toss rule.

Proof. Let n be the total number of agents. Let p be the probability of a low signal (not matching the real colour). Denote the private signal of player i by c_i and the guess of player i by C_i. We put k to be the largest number $0 \le k \le \frac{n}{2}$ s.t.

$$\text{colour}(c_{2k-1}) \neq \text{colour}(c_{2k})$$

This means that we have:

$$\text{colour}(c_{2i-1}) \neq \text{colour}(c_{2i}) \forall 1 \le i \le k$$

$$\text{colour}(c_{2k+1}) = \text{colour}(c_{2k+2})$$

This implies that before $2k$ there was no cascade (in fact no decisive majority). Hence:

$$\text{colour}(C_m) = \text{colour}(c_m) \forall m \le 2k$$

From this it follows that there is still no cascade at stages $2k+1$ and $2k+2$.:

$$\text{colour}(c_{2k+1}) = \text{colour}(c_{2k+2}) = \text{colour}(C_{2k+1}) = \text{colour}(C_{2k+2})$$

But now we have a decisive majority in favour of colour(c_{2k+2}), therefore a cascade forms at stage $2k+3$:

$$\text{colour}(C_{2k+3}) = \text{colour}(C_{2k+2}) = \text{colour}(C_{2h}) \forall h \ge 2k+3$$

Based on the above, we draw the following conclusions:

1. A cascade forms, in the sense that people disregard their signals, if and only if k, as defined above, has the property that $2k+3 \le n$. If so, the cascade starts at stage $2k+3$ but not earlier.

2. A cascade can only form at stages of the form $2k+3$, with $1 \le k \le \frac{n-3}{2}$

What is the probability that the cascade starts exactly at stage $2k+3$, and moreover that it is false?

$$P\Big(\bigwedge_{1\le i\le k} \text{colour}(c_{2i-1}) \neq \text{colour}(c_{2i}) \wedge \text{colour}(c_{2k+1}) = \text{colour}(c_{2k_+2})\Big) =$$

$$P\left(\bigwedge_{1\le i\le k}\Big(\Big(\text{colour}(c_{2i-1}) = \text{“high”} \wedge \text{colour}(c_{2i} = \text{“low”})\Big)\right.$$

$$\vee \Big(\text{colour}(c_{2i-1}) = \text{“low”} \wedge \text{colour}(c_{2i}) = \text{“high”} \Big)\Big)$$

$$\left.\wedge \Big(\text{colour}(c_{2k+1}) = \text{colour}(c_{2k+2}) = \text{“low”}\Big)\right)$$

$$= \Big((1-p)p + p(1-p)\Big)^k \cdot p^2$$
$$= 2^k(1-p)^k \cdot p^{k+2})$$

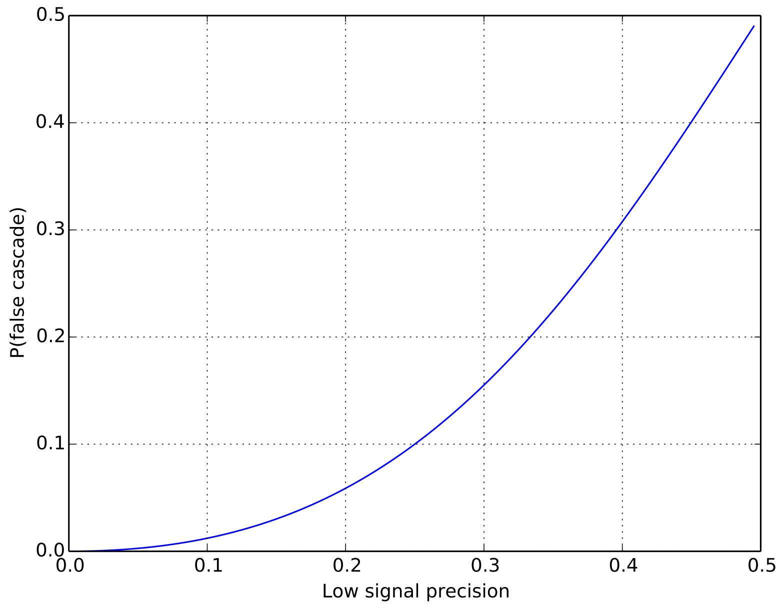


Figure 2.1: The probability of a false cascade as a function of the low precision signal p.

Therefore the probability that a false cascade forms is:

$$P(\text{false cascade forms}) = \sum_{1 \leq k \leq \frac{n-3}{2}} 2^k (1-p)^k \cdot p^{k+2}$$

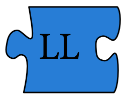
for a group of n agents. As $n \to \infty$, we have

$$\begin{aligned} lim_{n\to\infty} P(\text{a false cascade}) = &\sum_{k \in \mathbb{N}} 2^k (1-p)^k \cdot p^{k+2} \\ = &p^2 \cdot \frac{1}{1 - 2(1-p)p} \\ = &\frac{p^2}{2p^2 - 2p + 1} \end{aligned}$$

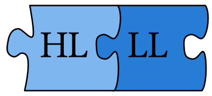
The higher the p, the higher the probability of a false cascade starting. However, recall that $p \leq \frac{1}{2}$, since p is "low". So the maximum probability of a false cascade is achieved for $p = \frac{1}{2}$, which gives $P(\text{low cascade}) = \frac{1}{2}$. A graph of the probability of a false cascade can be found in Fig. 2.1.

□

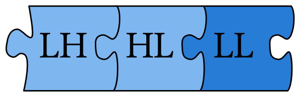
One can think of the initial segment of guesses before a cascade starts as partitioned in pairs of consecutive guesses of different colours. However, as soon as a pair whose elements have the same colour occurs in the initial segment, a cascade of the same colour starts. Some



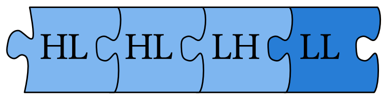
(a) A false cascade starting after the first two moves.



(b) A false cascade starting after the first 4 moves.



(c) An initial sequence of 6 moves leading up to a false cascade.



(d) An initial sequence of play leading to a false cascade after 8 moves

Figure 2.2: Sequences leading to false cascades. L denotes an action that doesn't match the state of the world while H does. The action of player i matches the colour of the private signal i received. Each piece of the puzzle represents a pair of guesses.

examples of initial sequences that lead to false cascades are included in Figure 2.2. This claim can easily be verified, if we take into account that the actions of players actually represent their signals. For a cascade to start, there must be two extra signals of one kind before a player's turn. The signal this player receives will not matter, since, if it's of the colour in minority, then that colour will remain in minority by 1. However, if the number of signals of each colour differed by one before a player's turn, the colour of that player's signal is sufficient to restore his belief in either urn to $\frac{1}{2}$. This, together with the self-biased rule, dictates that this player reveals the colour of his signal through his action to subsequent players. Therefore this player is not in a cascade.

In Chapter 2, we made an overview of the cascade literature in order to frame the dialogue on informational cascades and give context to our results. We reviewed the established definition of informational cascades and then surveyed their most salient characteristics, like fragility. We gained the intuition that informational cascades can have pernicious implications for economic welfare, by impeding the correct aggregation of mass information. Then we presented a simple-minded urn model to demonstrate the features of cascades in a Bayesian decision theoretic analysis. The original contribution of this chapter is Theorem 1, which puts a number on the likelihood of false cascades forming in urn games.

3

INFORMATIONAL CASCADES AS GAMES

This section is devoted to a classical game theoretic approach to informational cascades. The game-theoretic lenses provide the opportunity for modelling the dynamics of information inherent in informational cascades as consequences of strategic behaviour. The actions that set in motion the informational cascade, regarded as irrational in a non-strategic world, can now be explained using the apparatus of game theory. Although the association between informational cascades and games has been made in several cascade literature papers like Banerjee (1992), Hirshleifer and Anderson (1994), Hung and Plott (2001) an explicit and formal game-theoretic treatment of cascades and equilibria concepts is lacking.

We begin by defining the Urn Model as an extensive form game with imperfect information and chance moves. In effect, the Urn Model describes a sequence of individual decisions made by rational agents, based on some private information each obtains and the public information generated by previous players' actions. The extensive form reflects the sequential character of the game, the imperfect information stems from the private signals that players receive and the chance moves mimic the choice of the initial urn and the private draws taken from the chosen urn. Subsequently, we prove that the reasoning that leads agents into a cascade in the Urn Model, can be captured in game theoretic jargon as the unique Perfect Bayesian Nash Equilibrium (PBNE) of the game. In terms of our dining example in Chapter 1, the cascade-like reasoning is a PBNE because every tourist acts optimally, given his beliefs about the other tourists, and the information he has at his disposal at the time of his choice.

Structurally, the chapter is organized as follows: the first section introduces the Condorcet Jury Theorem and the definition of extensive games we use throughout the thesis. The second section contains the formalization of the Urn Model, introduced in Section 2.3 as an imperfect information game. The last section formally defines the concept PBNE for games with imperfect information and chance moves, and then proves that the reasoning of agents captured by the cascade represents a PBNE. Along the way, we define a refinement of PBNE, entitled self-biased equilibrium, in order to strengthen the notion of optimality for cascade settings.

3.1 PRELIMINARIES

3.1.1 *The Condorcet jury theorem*

The Condorcet jury theorem represents a mathematical characterization of the observation that "the many are smarter than the few" (Surowiecki, 2005). It states that majorities are much more likely than any single individual to select the "correct" of two alternatives, when there exists uncertainty about which of the two alternatives is in fact the best. More precisely, the theorem characterizes the conditions under which the majority of individuals in a group, whose size tends to infinity, is correct.

> *Theorem 2: The Condorcet Jury theorem*
>
> The probability that the majority selects the correct alternative exceeds p and approaches 1 as $n \to \infty$, whenever the following constraints are met:
>
> - Each individual in a group of n people takes an independent decision, based solely on his own private information.
> - The likelihood of each private signal to be right is $p > 1/2$.
> - The choice is binary, i.e. there are two mutually exclusive alternatives to choose from.

This technical result could be seen to vindicate the potential of groups to attain truth in situations with imperfect information and sequential moves. It stands in an antithetical position to the reasoning of agents caught in a cascade. Later in the thesis, we prove that in fact the reasoning underpinned by the Condorcet Jury Theorem is a PBNE of a structurally identical game to the Urn Model.

3.1.2 *Extensive games with imperfect information*

In the language of games, imperfect information signifies the inability of some player to distinguish between two different histories h and h' of a game, whereas incomplete information refers to the lack of information that certain players have about the structure of the game, the strategies of other players or of payoff functions, including their own.

The informational cascade setting is, naturally, one with incomplete information. This is due to the fact that players have a common prior belief about the objective payoff uncertainty and update it according to their private signal about the underlying state of the world. Each private signal initiates a belief hierarchy, which is build upon the first-order beliefs about the game fundamentals, like the posterior probability of the true state of the world, and first-order beliefs about the signals of others, given by conditioning upon the common prior.

However, Harsanyi (1967-68) showed that a game with incomplete information can be replaced by a game of imperfect information, in which Nature conducts a lottery, the outcome of which will decide which particular subgame will be played. This will be achieved by fixing the values of the unknown parameters of the original game. The result is a game of complete information with randomized moves by Nature, in which each player will receive partial information about the outcome of the lottery and about the values of the parameters. To sum up, by adding Nature as a player, we are able to transform the players' payoff uncertainty into unobservable moves by Nature, and hence, transform a game of incomplete information into a game of imperfect information.

We use a formalization of extensive games with imperfect information that is close to Kreps and Wilson (1982). Informally, an extensive form game with imperfect information and chance moves by Nature consists of:

1. a set of *players*
2. *choices* available to a player whenever it is his turn to move
3. a physical *order* of play
4. rules for determining whose *turn* it is at every point
5. the *information* each player has when it is his turn to move
6. the *payoffs* for each player
7. for every vertex assigned to Nature a *probability distribution* over the actions available at that vertex where each such distribution is independent of every other such distribution.

Definition 2: Extensive games with imperfect information and chance moves

A game is a structure

$$\mathcal{G} := \langle A, \mathcal{T}, Ag \cup \{0\}, P, (\mathcal{I}_i)_{i \in Ag}, (p_v)_{v \in V_0}, (\pi_i)_{i \in Ag} \rangle$$

where:

1. A is a set of actions. As usual, we denote by A^* the set of all finite sequences of elements of A, by $\leq$ the initial prefix relation on A^*, by $<$ its strict version and by r (the root) the empty sequence.

2. $\mathcal{T} \subset A^*$ is a set of nodes (or histories) which is closed under initial prefixes: $v \in \mathcal{T}, t \leq v \Rightarrow t \in \mathcal{T}$. We denote by $NonTerm(\mathcal{T}) := \{v \in \mathcal{T} \mid \exists t \in \mathcal{T} \text{ s.t. } v \leq t\}$ the non-terminal nodes of $\mathcal{T}$ and by $Term(\mathcal{T}) := \mathcal{T} \setminus NonTerm(\mathcal{T})$ the terminal nodes. We also denote by $A(v) := \{a \in A \mid va \in \mathcal{T}\}$ the actions available at node v.

3. Ag is a set of players. We will use a new symbol, $0 \notin Ag$ to denote Nature, as a separate kind of player.

4. $Pl : NonTerm(\mathcal{T}) \rightarrow Ag \cup \{0\}$ is a function which assigns a player to every non-terminal node. For $i \in Ag \cup \{0\}$ we put $V_i = \{v \mid i = Pl(v)\}$ to be the set of vertices at which player i needs to make a move. Finally, call $V := \cup_{i \in Ag} V_i$ the set of nodes assigned to players, except Nature.

5. For every player $i \in Ag$, $\mathcal{I}_i$ is a partition of V_i, called player i's information partition. We denote by $\sim_i$ the equivalence relation on V_i induced by this partition: $v \sim_i t$ iff $v \in I(t)$. We denote by $I(v)$ the information cell in the partition $\mathcal{I}_i$ that contains v.

6. p is a function assigning to each of Nature's nodes a probability distribution $p_v : A(v) \rightarrow [0, 1]$ over Nature's available actions at v.

7. for each player $i \in Ag$, $\pi_i : Term(\mathcal{T}) \rightarrow \mathbb{R}$ is the payoff function for player i that assigns a real-valued payoff to each terminal node.

This structure is required to satisfy the following condition:

$$v' \in I(v) \Rightarrow A(v) = A(v')$$

Name	Notation	Definition
precedence relation	$<$	$t < v$ if t is a node on the unique path from v to the root
the depth of vertex v	$d(v)$	$d(v) := \|\{y \mid y < v\}\|$
non-terminal nodes	$NonTerm(\mathcal{T})$	$\{t \in \mathcal{T} \mid \exists v \in \mathcal{T}$ s.t. $t < v\}$
terminal nodes	$Term(\mathcal{T})$	$\mathcal{T} \setminus NonTerm(\mathcal{T})$
initial node, root	r	r s.t. $\forall x \in \mathcal{T},\ r \leq x$
kth action of node v	$v(k)$	the kth element of v
predecessor of v of depth $k-1$	$v \restriction k$	the initial segment of v of length $k-1$

Table 3.1: Tree notions.

Additional notation

For a node v, we put $d(v) := |\{y \mid y < v\}|$ to be the depth of node v. Since the nodes/histories are sequences of actions, we can think of a node as a function $v : dom(v) \to A$ such that $dom(v) = \{1, \ldots, d(v) + 1\}$, is an initial segment of $\mathbb{N}^*$. Denote by $v(k)$ the kth element of v and by $v \restriction k$ the initial segment of v that ends with the $k-1$th element, where $k \geq 1$. Finally, we call $last(v) := v(2Pl(v))$, for $v \in V$, the last action taken before reaching node v.

The notation is summarized in Table 3.1.

3.2 THE URN GAME

The aim is to formalize the canonical example of social learning, the Urn Model, as a sequential (extensive form) game of imperfect information with randomized moves by Nature, called the Urn Game. This aptly describes the situation where the payoff function of each player i depends not only on the set of strategies chosen by the n players, but also on a set of random variables, represented by Nature's choices. It is assumed that all players know the joint probability distribution of these random, independent variables, but that each player only knows the realization of his own random variable (his private signal given by the draw of the marble from the urn).

Now we introduce a special class of games, called *urn games*, that exemplifies the setting described in the Urn Model in Chapter 2.

Definition 3: Urn games

An urn game is a structure

$$\mathcal{U} = \langle Ag \cup \{0\}, A, \mathcal{T}, P, (\mathcal{I}_i)_{i \in Ag}, (p_x)_{x \in V_0} \rangle$$

representing an imperfect information extensive form game such that:

- The set Ag is of the form $Ag = \{1, \ldots, n\}$, for some $n \in \mathbb{N}$, and 0 to denote the player Nature, like in the general case.
- A is a set of actions of the form $A := \cup_{k=1}^{k=2n+1} A_k$ where

  $$A_k = \begin{cases} \{W_{\frac{k-1}{2}}, B_{\frac{k-1}{2}}\} & \text{if } k \text{ is odd and } k \neq 1 \\ \{w_{\frac{k}{2}}, b_{\frac{k}{2}}\} & \text{if } k \text{ is even} \\ \{W, B\} & \text{if } k = 1 \end{cases}$$

  Notice that A_k is indexed by the turns in the game and each individual action index represents a player in Ag. For example, $A_4 = \{w_2, b_2\}$ and $A_5 = \{W_2, B_2\}$. We denote by colour(a) the colour given by the function colour: $A \to \{\text{white}, \text{black}\}$ s.t.

  $$\text{colour}(a) = \begin{cases} \text{white} & \text{if } a \in \{W_i, w_i, W\} \\ \text{black} & \text{otherwise} \end{cases}$$

- $\mathcal{T} := \{t \mid t \in \times_{i=0}^{j} A_i, \text{ for } j \leq 2n+1\}$, where $A_0 = \{\emptyset\}$. $\mathcal{T}$ represents all the possible partial and terminal histories of the game that represent the order of play.
- The player function Pl that requires players move at the even depths of nodes and Nature at the odd depths of nodes, as follows:

  $$Pl(v) = \begin{cases} \frac{d(v)}{2} & \text{if } d(v) \text{ is even} \\ 0 & \text{otherwise.} \end{cases}$$

  Notice that every agent, except Nature, only plays once and the depth of the vertex at which a player $i \neq 0$ has to move is equal to $2i$. The odd and even distinction ensures the alternation between Nature's turns and the players'. It also emphasizes the sequential character of the Urn Game, whereby each agent in turn observes a private draw from the urn, and then makes a guess. Also, notice that $V_i = \{v \in \mathcal{T} : d(v) = 2i\}$ if $i \in Ag$ and $V_0 = \{v \in \mathcal{T} : d(v) = 0$ or $d(v)$ is odd $\}$.

- For each player $i \in Ag$, the information partition $\mathcal{I}_i$ is given by:

$$I(v) = \{v' \in V_i | v(2j+1) = v'(2j+1), \\ \forall j < i \text{ and } v(2i) = v'(2i)\}$$

  This formalization ensures that agents only observe the actions of the previous players and their own signal. The signals of the previous players and the choice of world are not observed.

- A family of probability mass functions

$$(p_v : A(v) \to [0,1])_{v \in V_0}$$

  such that

$$p_v(a) = \begin{cases} \frac{2}{3} & \text{if colour}(v(1)) = \text{colour}(a) \\ \frac{1}{3} & \text{otherwise} \end{cases}$$

  where $a \in A(v)$, and $p_r : A(r) \to [0,1]$ with $a \in A(r)$ such that:

$$p_r(a) = \begin{cases} \frac{1}{2} & \text{if } a = W \\ \frac{1}{2} & \text{otherwise} \end{cases}$$

  This family of probability distributions represents Nature's "behavioural strategy", where each probability distribution is independent of the other. In our game, Nature only deals in two probability distributions: the one that assigns probabilities to states of the world, and the other that assigns probabilities to the private signals drawn by agents from the urn.

Notice that in the above definition, we left the payoffs unspecified, hence we defined a class of urn games, not only a specific game. We make the observation that urn games are games with perfect recall[1], and moreover, they are a synchronous system: there is a global clock and time is common knowledge.

1 For every player i, if $v_1, v_2 \in \mathcal{I}_i$, $a \in A(v_1)$ and $v_1 a \leq v_2$, then for every $v' \in I_i(v_2)$, $\exists v \in I_i(v_1)$ s.t. $va \leq v'$. Intuitively, perfect recall says that agents remember what they knew in the past and what actions they previously took.

Name	Notation	Definition
colour function	$\text{colour}(\cdot)$	$\text{colour}\colon A \to \{\text{white}, \text{black}\}$ s.t. $$\text{colour}(a) = \begin{cases} \text{white} & \text{if } a \in \{W_i, w_i, W\} \\ \text{black} & \text{otherwise} \end{cases}$$

Table 3.2: The colour function returns, for any action in urn games, the colour of the action: either black or white. This comes in handy, for example, when stating payoffs that depend on whether the agent guessed the correct colour of the true urn.

Definition 4: The Basic Urn Game

Define the Basic Urn Game as an extensive form game with imperfect information $\mathcal{G}^{\mathcal{U}} := \langle \mathcal{U}, \pi_i^{Ind} \rangle_{i \in Ag}$, constituted of an urn game where the payoff function is:

For each player i, the *individual* payoff function π_i, such that:

$$\pi_i^{Ind}(v) = \begin{cases} 1 & \text{if colour}(v(1)) = \text{colour}(v(2i+1)) \\ 0 & \text{otherwise} \end{cases}$$

This says that agents are rewarded if and only if the colour of their action matches the colour of the true state of the world. Notice that there is no payoff for player 0: Nature.

In the next section, we intend to explicitly and formally capture the cascade-like reasoning as an equilibrium concept of the Basic Urn Game. The aim of this is to explain cascade reasoning as a species of higher order rational inference.

3.3 THE CASCADE AS A PERFECT BAYESIAN NASH EQUILIBRIUM

Most solution concepts developed in game theory have no bite in games of incomplete information. However, the notion of *Perfect Bayesian Nash Equilibrium* captures an appropriate idea of rationality for this setting. PBNE broadens the classical concept of an equilibrium by adding a set of beliefs, on top of the usual strategy profile. The belief set represents players' probability assignments at the nodes of the game tree: it represents the beliefs of each player about where in the tree he is.

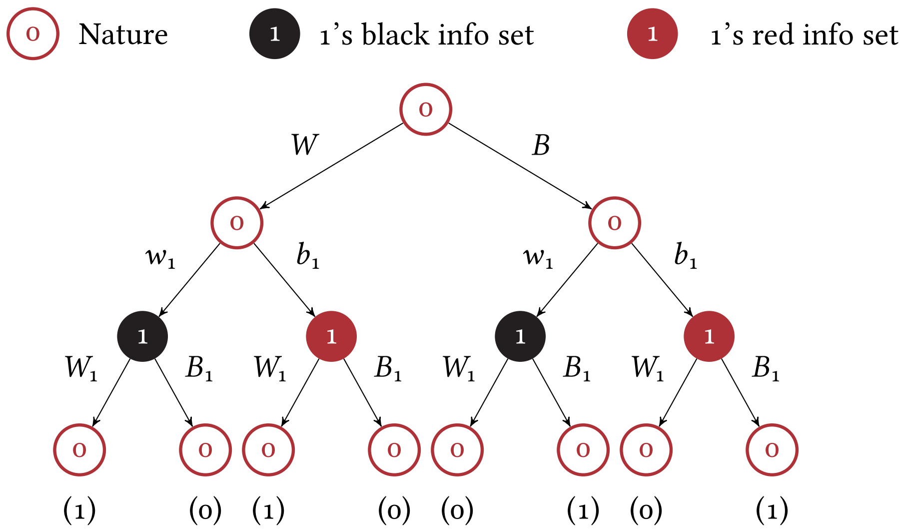


Figure 3.1: The Urn Game tree for one player. Precedence is represented graphically by arrows: one node precedes another if there is a sequence of arrows pointing from the first to the second. The game begins at the initial node, and proceeds along some path from node to immediate successor, until it reaches a terminal node. The game ends there, and each player receives the payoff associated with that terminal node. The payoffs are written in vector form, with the first element being the payoff of player 1, the second the payoff of player 2, and so on. The 0 in the hollow circles represents Nature's nodes, and the 1 in the filled circle represents player 1's nodes. Each action a is written along the branch that connects v to va.

This construction is motivated by the complication that arises in games of incomplete information, where events in one players' decision tree may correspond to actions of other players. In this sense, the optimality of an action at an information set may depend upon a history that has occurred.

3.3.1 *Definitions*

We first introduce the formal concepts needed for the definition of a Perfect Bayesian Nash Equilibrium and then proceed to define the equilibrium itself. In doing so, we tailor the definitions for extensive games of imperfect information with chance moves, by giving a formal interpretation to the common understanding in the literature of what a PBNE should entail.

Definition 5: Belief systems

A *belief system* is a function $\mu\colon V \to [0,1]$, where the restriction of μ to an information set $I \in \mathcal{I}_i$ of player i, is a probability measure $\mu_i\colon = \mu \restriction I \to [0,1]$.

Each $\mu(v)$ represents the probability that $Pl(v)$ assigns to being at node v. Consequently, μ_i summarizes the beliefs of player i about the actions of the players that moved before him in the game.

Definition 6: Behavioural strategies

A *behavioural strategy* σ_i of player i is represented by a function $\sigma_i : V_i \to (A \to [0,1])$ such that the following conditions are met:

- $\sigma_i(v)$ is a probability measure on A
- $\sigma_i(v)(a) = 0$ for $\forall a \notin A(v)$
- if $v \sim_i v'$ then $\sigma_i(v) = \sigma_i(v')$

Denote by $\sigma(v) := \sigma_i(v)\forall i\ \forall v \in V_i$. Then σ is a profile of behavioural strategies for every player in the game.

A behavioural strategy of player i is a function that assigns to each of i's information sets I_i a probability distribution over the actions in $A(I_i)$. Notice that a behavioural strategy in which every probability distribution assigns probability 1 to a single action is equivalent to a pure strategy.

Two possible interpretations of behavioural strategies

1. Behavioural strategy σ_i can be interpreted to signify the common belief of all players, except i, about the strategy that player i is going to play in the game.

2. Alternatively, behavioural strategy σ_i states that there is common knowledge of the beliefs that each player has about his own strategy, before the game starts.

Definition 7: Assessments

A pair of the form (σ, μ), where σ is a profile of behavioural strategies for each player in the game, and μ is a belief system represents an *assessment*.

Computation of expected payoff in extensive-form games

Given an extensive form game

$$\mathcal{G} = \langle Ag \cup \{0\}, A, \mathcal{T}, P, (\mathcal{I}_i)_{i \in Ag}, (\pi_i)_{i \in Ag}, (p_v)_{v \in V_0} \rangle$$

and an assessment (σ, μ), we want to compute the expected payoff of player i at info set I. We do so piecemeal, using the following constructions: first, let $Pr^\sigma(w, a)$ denote the probability that action a is played at node w, given the strategy profile σ:

$$Pr^\sigma(w, a) = \begin{cases} \sigma_{Pl(w)}(w)(a) & \text{if } Pl(w) \neq 0 \\ p_w(a) & \text{if } Pl(w) = 0 \end{cases}$$

Next, call $Pr_v^\sigma(t)$ the probability that node t will be reached given that node v was reached and given behavioural strategy σ.

$$Pr_v^\sigma(t) = \begin{cases} \prod_{d(v) < k \leq d(t)} Pr^\sigma(t \upharpoonright k, t(k)) & \text{if } v \leq t \\ 0 & \text{if } v \nleq t \end{cases}$$

Finally, call the probability that terminal node t will be reached, given the beliefs μ of player i at information set I and the behavioural strategy σ:

$$Pr_\mu^{\sigma,I}(t) := \sum_{v \in I} \mu(v) Pr_v^\sigma(t)$$

Then, the expected payoff of player i at information set I, is given by the expression below:

$$E_i^{\mu,\,I}(\sigma) = \sum_{t \in Term} Pr_\mu^{\sigma,I}(t) \pi_i(t)$$

Now we are ready to define a Perfect Bayesian Nash Equilibrium.

Definition 8: Perfect Bayesian Nash Equilibria

The assessment (σ, μ) is a perfect Bayesian equilibrium if and only if the following conditions are satisfied, with definitions adapted from Osborne (2004, p. 323):

1. *Sequential rationality*: A strategy profile σ is sequentially rational according to a belief system μ if at every information set I, the player who moves at I is behaving optimally according to the beliefs generated by μ at I. Formally, let i be the player whose information set I is. Then, σ is sequentially rational at I if:

$$E_i^{\mu,I}(\sigma_i, \sigma_{-i}) \geq E_i^{\mu,I}(\sigma_i', \sigma_{-i}) \text{ for all } \sigma_i'$$

2. *Weak consistency of beliefs with strategies:* For every information set I, reached with positive probability[2], given the strategy profile σ, the probability assigned by the belief system to each vertex t in I is given by:

$$\mu(t) = \frac{\prod_{0<k<d(t)} Pr^{\sigma}(t \upharpoonright k, t(k))}{\sum_{t' \in I} \prod_{0<k<d(t')} Pr^{\sigma}(t' \upharpoonright k, t'(k))}$$

In terms of notation introduced earlier:

$$\mu(t) = \frac{Pr_r^{\sigma}(t)}{\sum_{t' \in I} Pr_r^{\sigma}(t')}$$

where recall that r denotes the root, or the empty history.

The above conditions can be interpreted intuitively as follows: the first condition requires that agents' strategies are optimal, after any history of events. More precisely, it requires that every player's strategy be optimal "in the part of the game that follows each of her information sets, given the strategy profile and the player's belief about the history in the information set that has occurred, regardless of whether the information set is reached if the players follow their strategies" (Osborne, 2004, p. 325).

In brief, the second condition says that each player's beliefs, at nodes that accord with the strategy profile, must be correct: the probability it assigns to every vertex must be the probability with which that vertex is reached if the players follow their strategies. The difficulty arises when a player i, at his information set, needs to consider the probability of a vertex that, according to the strategy profile must have been reached with probability 0. It is stipulated, that in such a case, the player can have any beliefs at all. Then we can say that the weak consistency of beliefs condition under-determines the beliefs that players should hold in a game. We call the nodes that have a positive probability of being reached given a strategy profile *"on path"* nodes, and all the others *"off-path"* nodes.

3.3.2 *PBNE in Urn Games*

At this stage we want to formalize in game theoretic terms the reasoning that players in the Urn Model informally engage in, and check whether it is an equilibrium in the Basic Urn Game. Of the several ver-

2 Notice that given the nature of information sets, either every node in an information set I is 0, or every node is non-zero. Thus the expression "an information set is reached with positive probability" denotes an information set where every node is non-zero.

sions of Perfect Bayesian Nash Equilibrium proposed in the literature, the one we presented is the weakest, because it places no restrictions at all on the beliefs following a zero-probability event. However, given the paradoxical nature of cascade reasoning, we want to consider more stringent criteria on the consistency-of-beliefs condition, that could strengthen the equilibrium concept. We propose a new consistency criterion, which we call self-biased consistency, that applies exclusively to urn games and that strengthens the weak consistency-of-beliefs condition of PBNE. After doing so, we check which other consistency criteria presented in the PBNE literature our criterion satisfies.

Definition 9: Self-biased consistency

Given a behavioural strategy σ and a node v, we call a belief system *self-biased consistent* iff:

$$\mu(v) := \frac{Pr_r^{sb(\sigma,v)}(v)}{\sum_{v\sim_i v'} Pr_r^{sb(\sigma,v)}(v')}$$

where $sb(\sigma, v)$ is defined, for all $w \in dom(\sigma)$:

$$sb(\sigma,v)(w)(a) = \begin{cases} 1 & \text{if } \sigma(w)(a) = 0 \text{ and also } colour(a) = last(w) \text{ and } wa \leq v' \text{ for all } v' \in I(v) \\ 0 & \text{if and } \sigma(w)(a) = 0 \text{ and also } colour(a) \neq last(w) \text{ and } wa \leq v' \text{ for all } v' \in I(v) \\ \sigma_{Pl(w)}(w)(a) & \text{otherwise} \end{cases}$$

Definition 10: Self-biased equilibrium

An assessment (σ, μ) is a self-biased equilibrium if and only if it satisfies the following requirements:

1. *Sequential rationality*: σ is sequentially rational given the belief system μ.
2. *Self-biased consistency*: μ is self-biased consistent with respect to σ

Definition 9 gives us a unique belief system that is determined, at "on path" nodes by the consistency condition of PBNE, and at "off path" nodes by the belief that agents who deviated from their strategy, in fact follow their signal. This conviction is common knowledge amongst agents and represents, intuitively, a mechanism of rationalization of the actions of those players that deviate from their strategy. We wish to refer to this rule for dealing with "surprises" as *the self-biased rule* and note that it applies only to observable defectors: agents that have falsified their strategy during a play of the game. For all other players whose guesses have been consistent with their strategies, the self-biased rule is not applied.

One question that arises is how the self-biased consistency rule relates to the consistency restrictions of PBNE proposed in the literature. We present three such restrictions:

- *Structural consistency* (Kreps and Wilson, 1982): An assessment is structurally consistent if, for each information set I, there exists a strategy profile σ_I^* such that $\sigma_I^*(v) > 0,\ \forall v \in I$ and

$$\mu(v) = \frac{Pr_r^{\sigma_I^*}(v)}{\sum_{v \sim_i v'} Pr_r^{\sigma_I^*}(v')}$$

  That is, for every information set reached, the agent having to move at that information set, can find a strategy profile that would yield exactly the same beliefs as those held by the agent at that information set.

- *The reasonability criterion* (Fudenberg and Tirole, 1991): Even though this condition was developed for multi-stage games with observable actions, it constitutes an important reference point in the literature and we wish to mention it. The key condition of this criterion imposes that a player's action cannot signal private information that the player does not possess when choosing that action. Applying this to "off path" information sets, the strategy assigned to a deviator by the agent moving at that information set, should not depend upon the deviator possessing information he has no access to.

- *The strategic independence principle* (Battigalli, 1996): Information about player k's strategic behaviour is irrelevant for the probability assignments exclusively concerning player j's strategic behaviour, where $j \neq k$. In terms of deviations, it says that if player k deviates and player j does not, the beliefs about player j are updated in accordance with Bayes' rule.

We observe that all of these requirements, interpreted on extensive games with imperfect and chance moves, are fulfilled by the self-biased consistency requirement. Next, we are going to define the *Bayesian assessment* and prove that, modulo the tie-breaking rule, it represents the unique PBNE of the game.

Lemma 1

Every self-biased equilibrium is a Perfect Bayesian Nash Equilibrium.

Proof. It is straightforward to notice that the criteria for self-biased equilibrium are a refinement on the criteria of PBNE. □

Definition 11: The Bayesian assessment

We define the assessment $(\sigma^{Bayes}, \mu^{Bayes})$ by recursion on the depth of $v \in V$:

Stage 1: $v \in V_1$

$$\mu^{Bayes}(v) = \frac{p_{v\restriction 1}(v(1)) \cdot p_{v\restriction 2}(v(2))}{\sum_{v' \sim_1 v}(p_{v'\restriction 1}(v'(1)) \cdot p_{v'\restriction 2}(v'(2)))}$$

$$\sigma_1^{Bayes}(v)(a) = \begin{cases} 1 & \text{if } colour(a) = colour(last(v)) \\ 0 & \text{otherwise} \end{cases}$$

Stage i: $v \in V_i$

$$\mu^{Bayes}(v) := \frac{Pr_r^{sb(\tau,v)}(v)}{\sum_{v \sim_i v'} Pr_r^{sb(\tau,v')}(v'))}$$

where $\tau = (\sigma_1, \ldots, \sigma_{i-1})$

Next, we define

$$\mu^{Bayes}(W|v) = \sum_{v' \in I(v) s.t. v'(1) = W} \mu^{Bayes}(v')$$

which depends on the restriction of σ^{Bayes} to the predecessors of v. $\mu^{Bayes}(W|v)$ informs a decision on:

$$\sigma_i^{Bayes}(v)(a) = \begin{cases} 1 & \text{if } \mu^{Bayes}(W|v) > \frac{1}{2} \text{ and } a = W_i \\ 1 & \text{if } \mu^{Bayes}(W|v) < \frac{1}{2} \text{ and } a = B_i \\ 1 & \text{if } \mu^{Bayes}(W|v) = \frac{1}{2} \text{ and colour}(a) = \text{colour}(\text{last}(v)) \\ 0 & \text{otherwise} \end{cases}$$

Considerations on the Bayesian assessment

The strategy profile σ^{Bayes} is the one that corresponds to every player guessing the colour of the urn they believe is most probable, given their beliefs and other player's strategies. In this sense we call the individual strategy σ_i^{Bayes} the Bayesian strategy . However intuitive, the reader is invited to resist the association of the Bayesian strategy with Bayesian reasoning, since the latter concept is captured by all the other strategy profiles we consider in the thesis. In more general terms, Bayesian reasoning is equivalent, in games, to instrumental rationality. We do not, therefore, want to associate an entire species of reasoning with a unique strategy.

The belief system μ^{Bayes} is defined as to fulfil the self-biased consistency requirement. This means two things: one, that on path, μ^{Bayes} satisfies the weak consistency-of-beliefs requirement of PBNE and two, that off path μ^{Bayes} is determined by the self-biased rule, which requires that agents use a back-up strategy for forming beliefs (by applying Bayes' rule) about observable deviators in a game. This back-up strategy is, in effect, the *Condorcet pure strategy* we will introduce formally in Chapter 4, which states that players always guess the urn that matches the colour of their signal.

The choice of the Condorcet strategy, as a fall-back option for the Bayesian strategy, is motivated by two principles. The first principle is unfalsifiability. Say a player j defects from his strategy σ^{Bayes}. Then player i believes j has followed his signal. Since player i cannot observe j's signal, his new belief about j cannot be proved wrong. The second principle is rationalizibility: the Condorcet strategy ensures that players entertain a certain belief in the other's rationality. Even if their strategy is not optimal, it is at least justifiable on other grounds, like the maximization of the social welfare. We will discuss the merits of the Condorcet strategy more in depth in Chapter 4, where we prove that it is a PBNE of a different urn game.

> *Theorem 3: The PBNE of the Urn Game*
>
> The assessment $(\sigma^{Bayes}, \mu^{Bayes})$ represents a Perfect Bayesian Nash Equilibrium of the Urn Game $\mathcal{G}^{\mathcal{U}} = \langle \mathcal{U}, \pi_i^{Ind} \rangle_{i \in Ag}$.

Proof. In order to prove that $(\sigma^{Bayes}, \mu^{Bayes})$ is a Perfect Bayesian Nash Equilibrium, we will need to prove the following (see Section 3.3):

CLAIM 1 σ^{Bayes} is sequentially rational given the belief system μ^{Bayes}.

CLAIM 2 μ^{Bayes} is consistent with respect to the strategy profile σ^{Bayes}.

Proof of Claim 1. We need to check whether, at any node v of player i, the action dictated by σ^{Bayes} gives i the highest expected payoff. But this is immediate, since the expected payoff for any player i is equal to the subjective probability they assign to their guess being correct.

It only remains to pursue this intuition formally. We know that, for any σ:

$$E_i^{\mu, I}(\sigma) = \sum_{t \in Term} \sum_{v \in I} \mu(v) Pr_v^{\sigma}(t) \pi_i^{Ind}(t)$$

Since $\pi_i^{Ind}(v)$ returns 0 whenever the action of the agent does not match the state of the world, the terminal nodes t that don't satisfy the condition $colour(t(1)) = colour(t(2i+1))$ disappear from the sum. Then:

$$E_i^{\mu, I}(\sigma) = \sum_{t \in Term}^{colour(t(2i+1))=colour(t(1))} \sum_{v \in I} \mu(v) Pr_v^{\sigma}(t)$$

which is equivalent to:

$$E_i^{\mu, I}(\sigma) = \sum_{\substack{t \in Term \\ colour(t(2i+1))=colour(v(1)) \\ v \in I}} \mu(v) Pr_v^{\sigma}(t)$$

which is equivalent to:

$$E_i^{\mu, I}(\sigma) = \sum_{\substack{t \in Term \\ colour(t(2i+1))=colour(v(1)) \\ v, t\restriction(2i+1) \in I}} \mu(v) \sigma(v)(t(2i+1)) Pr_{t\restriction(2i+2)}^{\sigma}(t)$$

Observe that $\sum_{\substack{t \in Term \\ t\restriction(2i+1) \in I}} Pr_{t\restriction(2i+2)}^{\sigma}(t) = 1$ and $t(2i+1) \in \{W_i, B_i\}$. Therefore, we get that:

$$E_i^{\mu, I}(\sigma) = \sum_{\substack{v(1)=W \\ v \in I}} \mu(v) \sigma(v)(W_i) + \sum_{\substack{v(1)=B \\ v \in I}} \mu(v) \sigma(v)(B_i)$$

Now, it is easy to notice that the maximum of the expected payoff at node v $E_i^{\mu, I}(\sigma)$ is attained for σ a behavioural strategy that assigns probability 1 to $a \in A(v)$ for which $\sum \mu(v)_{colour(v(1)=colour(a))}$ is largest. In case $\sum_{\substack{v(1)=W \\ v \in I}} \mu(v) = \sum_{\substack{v(1)=B \\ v \in I}} \mu(v))$, then any strategy is optimal.

In the particular case of μ^{Bayes}, we get that σ^{Bayes} is optimal.

Intuitively, the expected payoff of a strategy for player i is nothing than the subjective belief player i assigns to the world being of the colour of his guess. Therefore, as long as the choice of action is guided by the colour of the world that is more probable than the other, players are maximizing their payoff. We proved that the strategy profile σ^{Bayes} is sequentially rational given μ^{Bayes}.

□

Proof of Claim 2. We have defined μ^{Bayes} using the definition of self-biased consistency. Therefore, it satisfies the weak consistency requirement of PBNE by definition.

□

□

The "uniqueness" of PBNE in the Basic Urn Game

We observe that all the PBNE of the Basic Urn Game agree "on path" with the above Bayesian assessment $(\sigma^{Bayes}, \mu^{Bayes})$, by virtue of the sequential rationality proof above. Moreover, the Bayesian assessment is not only a PBNE but also a self-biased equilibrium, by virtue of fulfilling a stronger consistency requirement than demanded by PBNE. We showed in Theorem 1 that cascades in the Basic Urn Game happen 20% of the time if players play according to any PBNE at "on path" nodes, that has the same tie-breaking rule as us. This leads to the conclusion that cascades are, in a sense, "unavoidable", if players are rational.

As the last result of this chapter, we would like to formally characterize the result by Bikhchandani et al. (1992), that argued the probability of a cascade forming during a play of an urn game is 1, whenever the number of players is infinite. We slightly generalize his result, by showing that the probability of a cascade happening approaches one as the number of players reaches ∞, for any Urn-type game, where all but an initial segment of players, play the Bayesian strategy, σ^{Bayes}. We will make use of this theorem in order to prove things about cascades later in the thesis.

> *Theorem 4: Certainty of cascades*
>
> The probability that a cascade starts during a play of an Urn-type game $\mathcal{G} = \langle \mathcal{U}, \pi \rangle$, approaches 1 as the number of players $n \to \infty$, whenever all but a finite initial number of players, play the Bayesian strategy, s^{Bayes}.

Proof. Let, for simplicity, n be even total number of players in an urn-type game. Let p denote the probability of a low signal (one that doesn't match the true colour of the world). The proof is based on the observations made in Theorem 1, namely that for a cascade to not occur during an urn-type game, the value of k constrained to:

$$colour(c_{2i-1}) = colour(c_{2i}) \forall i \leq k$$

must be maximal, i.e. $k = \frac{n-2}{2}$. Only then are cascades, true or false, ensured to not happen. What is the probability of such a string of guesses to occur during a play of an urn game ?

$$P\Big(\bigwedge_{1\le i\le k} colour(c_{2i-1}) \neq colour(c_{2i})\Big) =$$

$$Pr\left(\bigwedge_{1\le i\le k}\Big(\big(colour(c_{2i-1}) = \text{“high”} \wedge colour(c_{2i}) = \text{“low”}\big)\right.$$

$$\left.\vee \big(colour(c_{2i-1}) = \text{“low”} \wedge colour(c_{2i}) = \text{“high”}\big)\Big)\right)$$

$$= \big(p(1-p)\big)^k$$

$$P(\text{no cascade forms}) = (p(1-p))^{\frac{n-2}{2}}$$

for a group of n agents. As $n \to \infty$, we have

$$lim_{n\to\infty}P(\text{no cascade}) = 0$$

Therefore, the probability of a cascade occurring during a play of the game is equal to 1. □

This concludes Chapter 3, which presented informational cascades in the architecture of game theory and made use of equilibrium concepts in order to underpin the reasoning of agents caught in a cascade. This was achieved by giving a clear and formal account of the Perfect Bayesian Nash Equilibrium concept for imperfect information games with chance moves. Along the way, a refinement of PBNE for urn-type games was defined, called self-biased equilibrium, and was proved that the reasoning leading agents into cascades abides by these even stringent equilibrium requirements. The chapter also played a foundational role, by introducing background theoretical notions and preparing the ground, with Theorem 3 and Theorem 4, for assessing optimal behaviour in games that depart from the Basic Urn Game in natural ways.

We end with noting that the game model of cascades is able to capture a wide array of informational situations. For example, it can incorporate the scenario in which players do not have perfect recall, in the sense that they forget some of the things that they knew earlier. In our Basic Urn Game, we can interpret this feature as expressing the limited observational powers of agents: people in the sequence do not remember what everyone else before them did, but only the last 2 actions, for example. This is only one of several potential generalization of our informational cascade model. Others will be considered in the next chapter.

4

SOCIAL DYNAMICS OF INFORMATION GAMES

In this chapter various payoff configurations for urn games are considered. By implementing different scenarios, we generalize the cascade setting and explore, using game theoretic tools, the interactions and tensions between notions like self-interest and cooperation, and their composite influence on the likelihood of cascades. These considerations bring to life interesting patterns of information flow.

Modifications to the canonical payoff structure in informational cascade games have been considered in the literature before. The conformity and the majority payoff were treated by Hung and Plott (2001), however their focus was on experimental results. Banerjee (1992) treats modifications that reward agents based on their rank within the sequence of decisions, something that we do not consider in this thesis. One of the original contributions of this chapter is a formal game theoretic analysis of the games resulting from payoff modifications, with a focus on the impact of the changes on the social welfare of the group. Another is constituted by the proof of Theorem 6, which represents, apart from a result about optimal strategies, a quantification of the observed fragility of cascades in urn-type games. The last contribution of the chapter takes the form of a strategic game called the Hybrid Coordination game, where joint strategies are considered. This game induces a socially desirable outcome in an urn-type game, where players are rewarded both for individually and collectively finding the truth.

Structurally, the chapter is divided into two parts: the first part introduces payoff structures that depart from the Basic Urn Game payoff, in order to determine what changes need to be put in place for groups to achieve better social outcomes. The second part of the chapter contains the Hybrid Coordination Game, a strategic form game, designed to force agents to cooperate.

4.1 MODIFICATIONS OF THE BASIC URN GAME

In general, under common knowledge of rationality, different payoffs make different actions rational, and therefore, to a certain extent, players can indirectly communicate their private signals to each other via their actions. As we have seen in the Basic Urn Game, the cascade starts as soon as the communication of private signals amongst agents ends. Then a natural question to ask is how changes in payoff structures alter communication, via actions, amongst agents. This section tries to answer this question by modifying the payoff structure of the canonical Basic Urn Game, in order to check whether, and to what degree, the likelihood of cascades is dependent upon the payoff structure of the game.

Definition 12: The majority function

The majority function $maj : \mathcal{T} \setminus r \to \{W, B, 0\}$ returns the majority decision in urn-type games:

$$maj(v) = \begin{cases} W & \text{if } |\{i \mid v(2i+1) = W_i, i < j\}| > \frac{j}{2} \\ B & \text{if } |\{i \mid v(2i+1) = B_i, i < j\}| > \frac{j}{2} \\ 0 & \text{otherwise.} \end{cases}$$

where j=Pl(v)

Consider a family of urn games, with different payoff structures, which we call social dynamics of information games. The payoff structures that diverge from the Basic Urn Game reflect different empirical situations that broaden the scope of informational cascade models.

Definition 13: The Conformity Urn Game

With the *conformity payoff* subjects receive a positive payoff if and only if their guess matches that of the majority, no matter if the majority is right or wrong.

$$\pi_i^{Conf}(v) = \begin{cases} 1 & \text{if } \mathrm{colour}(v(2i+1)) = \mathrm{colour}(\mathrm{maj}(v)) \\ 0 & \text{otherwise} \end{cases}$$

The Conformity Urn Game is an urn game

$$\mathcal{G}^{Conf} = \langle \mathcal{U}, \pi_i^{Conf} \rangle_{i \in Ag \cup \{0\}}$$

that rewards agents according to the conformity payoff. Recall that $\mathcal{U}$ represents a generic urn game, introduced in Definition 3.

The conformity payoff turns the urn game into a coordination game, whereby the reward to an agent adopting an action increases in the number of other agents adopting the same action (for example the convention to drive on the right side of the road, the use of fax machines, etc).

Definition 14: The Majority Urn Game

The *majority payoff* is a payoff according to which subjects receive a positive payment if and only if the group decision (determined by the majority rule) is correct.

$$\pi_i^{Maj}(v) = \begin{cases} 1 & \text{if } maj(v) = v(1) \\ 0 & \text{otherwise} \end{cases}$$

The Majority Urn Game is an urn game

$$\mathcal{G}^M = \langle \mathcal{U}, \pi_i^{Maj} \rangle_{i \in Ag \cup \{0\}}$$

in which agents are rewarded according to the majority payoff.

Experimental results (Hung and Plott, 2001) show that the majority payoff reduces conformity among early decision-makers since each agent has an incentive to reveal his private information to subsequent players. The complete revelation of private information precludes the formation of cascades altogether. However, the conformity payoff, as the name suggests, induces pure conformity, and therefore leads to a cascade from the beginning of the game. Therefore, evidence shows that variations in payoff structures lower or rise, depending on the variation, the likelihood of cascades forming.

The majority payoff applied to the urn game incentivizes agents to take the social welfare of the group into consideration. It aptly models voting situations, in which the majority decides upon a group enforceable law, that each individual must abide by. From a mechanism design perspective, modifying the payoff of the Basic Urn Game makes sense if the aim of the designer is to maximize the learning potential of the group, or ensure a socially desirable outcome.

Definition 15: The Hybrid Urn Game

The *hybrid payoff* rewards a player if the majority is right and if he guesses correctly.

Let $\pi_i^{Hyb}(v) = y \cdot \pi_i^{Ind}(v) + x \cdot \pi_i^{Maj}(v)$

The x, y-Hybrid game is an urn game $\mathcal{G}^{Hyb} = \langle \mathcal{U}, \pi_i^{Hyb} \rangle_{i \in Ag \cup \{0\}}$ where agents are rewarded both according to the hybrid payoff function with parameters x and y.

The hybrid payoff is designed in such a way as to create conflicting interests for strategic reasoners, between individualistic and collective behaviour. This is achieved by tying individual payoffs to the performance of the majority vis-à-vis the truth. This type of payoff characterizes the adoption decisions of new technologies, where one component of the payoff is determined by the quality of the product, perceived subjectively by an individual, and the other payoff component is determined by whether the majority of the population adopts the same technology (since, by adoption of the majority, there will be more updates, extensions, new versions for that technology, and hence more satisfaction for the individual adopter).

4.2 INCIDENCE OF CASCADES

We set out to assess the likelihood of cascades formally, by determining how rational agents would play, under different payoffs. In doing so, we make use of concepts from game theory, namely PBNE, and try to show a series of positive and negative results, in various games, about two main strategies: the Bayesian strategy and the Condorcet strategy. These two strategies are of particular importance, since they represent very distinct levels of communication. While the Bayesian strategy can lead to no communication, relatively soon in the game, the Condorcet strategy is the equivalent of full communication. Since communication levels in games determine the incidence of cascades, analysing whether one of these two salient strategies, or a combination of both, will be played by rational agents, will in fact determine how likely cascades are to start during a play of the game.

Definition 16: The conformity assessment

The assessment $(\sigma^{Conf}, \mu^{Conf})$ is an assessment where:

$$\sigma^{Conf}(v)(a) = \begin{cases} 1 & \text{if } \mathrm{maj}(v) \neq 0 \text{ and } \mathrm{colour}(a) = \mathrm{colour}(\mathrm{maj}(v)) \\ 1 & \text{if } \mathrm{maj}(v) = 0 \text{ and } \mathrm{colour}(a) = \mathrm{colour}(\mathrm{last}(v)) \\ 0 & \text{otherwise} \end{cases}$$

for any v and i s.t. $Pl(v) = i$ and $a \in A(v)$. μ^{Conf} is the belief system determined by the self-biased consistency requirement introduced in Definition 10:

$$\mu^{Conf}(v) := \frac{Pr_r^{sb(\sigma^{Conf},v)}(v)}{\sum_{v\sim_i v'} Pr_r^{sb(\sigma^{Conf},v)}(v'))}$$

Intuitively, σ^{Conf} represents the behavioural equivalent of the pure strategy that requires agents to follow the majority decision of the players who moved before him. This strategy is the formal equivalent of the concept of conformity. In Theorem 5 we show that the conformity assessment represents a PBNE of the Conformity Game.

Theorem 5

The assessment $(\sigma^{Conf}, \mu^{Conf})$ is a PBNE of the Conformity Urn Game.

Proof. We have to check only one of two requirements for assessments to qualify as PBNE, namely, sequential rationality. Weak consistency-of-beliefs is satisfied by definition. The intuition is that, after the first player, who guesses the colour of his marble, all subsequent players will imitate the first. This induces unanimity in guesses throughout the entire game, and therefore, the guarantee of the conformity payoff. It is evident that no other strategy could achieve a better payoff than this. In the proof, we pursue this intuition formally:

We compute the expected utility at each node v of a player i, given her beliefs μ^{Conf}, her information set I and the payoff function π_i^{Conf}:

$$E_i^{\mu^{Conf},I}(\sigma^{Conf}) = \sum_{t\in Term} Pr_{\mu^{Conf}}^{\sigma^{Conf},I}(t)\pi_i^{Conf}(t)$$

where $Pr_{\mu^{Conf}}^{\sigma^{Conf},I}(t)$ is the probability that terminal node t is reached, given the strategy profile σ^{Conf} and the beliefs μ^{Conf} at information set I. First, notice that there will always exist a majority, established after player 1's move. Then, notice that every terminal history t that extends v, in which a player $j > i$ at node $t \upharpoonright 2j+1$ does not follow the guess of the majority, has probability 0 of occurring because $\sigma_j(t \upharpoonright 2i+1)(t(2i+1)) = 0$. Finally, notice that $\pi_i^{Conf}(t)$ for the terminal histories that have non-zero probability of occurring is always 1. Given all this, we can re-state the expected payoff of player i at information set I as:

$$E_i^{\mu^{Conf},I}(\sigma^{Conf}) = \sum_{t\in Term} Pr_{\mu^{Conf}}^{\sigma^{Conf},I}(t)$$

Since, naturally, one of the possible terminal histories must occur at the end of the game, the expected payoff of player i is:

$$E_i^{\mu^{Conf},I}(\sigma^{Conf}) = 1$$

We obtained that the expected payoff of player i, given his beliefs μ^{Conf} and strategy profile σ^{Conf} is maximum and equal to 1. Therefore, any other behavioural strategy σ_i will not yield a better payoff to player i than his strategy σ_i^{Conf}. The conformity strategy profile is sequentially rational with respect to its consistent beliefs μ^{Conf}. □

Before we move on to Theorem 6 that shows the Bayesian strategy profile is not a PBNE of the Majority Urn Game, we introduce the Condorcet assessment, which we have discussed informally until now, as part of self-biased consistency requirement.

Definition 17: The Condorcet assessment

The assessment $(\sigma^{Cond}, \mu^{Cond})$ is a self-biased consistent assessment where:

$$\sigma^{Cond}(v)(a) = \begin{cases} 1 & \text{if } colour(a) = colour(last(v)) \\ 0 & \text{otherwise} \end{cases}$$

for any v and $a \in A(v)$. μ^{Cond} is the self-biased consistent belief system introduced in Definition 10:

$$\mu(v) := \frac{Pr_r^{sb(\sigma^{Cond},v)}(v)}{\sum_{v\sim_i v'} Pr_r^{sb(\sigma^{Cond},v)}(v'))}$$

Theorem 6

The assessment $(\sigma^{Bayes}, \mu^{Bayes})$, defined in Definition 11, is *not* a PBNE of the Majority Urn Game, when the number of players $n \to \infty$.

Proof. We show that the Bayesian strategy profile is not sequentially rational, given the players' beliefs μ^{Bayes}, and therefore not a PBNE. To this end, we prove the following claim:

$$E_k^{\mu^C,I}(\sigma_k^{Cond}, \sigma_{-k}^{Bayes}) > E_k^{\mu^{Bayes},I}(\sigma_k^{Bayes}, \sigma_{-k}^{Bayes})$$

for some $k \in Ag$ and one of his information sets I, where μ^C and μ^{Bayes} represent the self-biased consistent belief systems with respect to strategy profile $(\sigma_k^{Cond}, \sigma_{-k}^{Bayes})$, respectively σ^{Bayes}. We begin by noticing that $\pi^{Maj}(t) = 0$ if $maj(t) \neq t(1)$). Then expected payoff of player k at information set I is equal to:

$$E_k^{\mu,I}(\sigma) = \sum_{t \in Term(\mathcal{T})}^{maj(t)=t(1)} Pr_\mu^{\sigma,I}(t)$$

which can be restated more intuitively as

$$E_k^{\mu,I}(\sigma) = P(\text{majority is right} \mid \sigma, \mu), \ \forall k \in Ag$$

Given that every player k seeks to maximize his expected payoff, every player therefore seeks to maximize the social welfare of the group, represented by the probability that the majority will be right. The assumption that $n \to \infty$ allows to set the probability of a cascade forming during the game to 1 by Theorem 4. This means terminal histories contain either a false cascade and therefore a majority that is wrong, or a true cascade and therefore a majority that is correct. Writing TC for true cascade and FC for false cascade, this means:

$$P(\text{majority is right}) = P(\text{TC forms})$$

Therefore, the claim can now be restated as:

$$P(\text{TC} \mid (\sigma_k^{Cond}, \sigma_k^{Bayes}), \mu^C) > P(\text{TC} \mid (\sigma_k^{Bayes}, \sigma_k^{Bayes}), \mu^{Bayes})$$

Given that $P(\text{TC}) = 1 - P(\text{FC})$, we can state the claim as:

$$P(\text{FC} \mid (\sigma_k^{Cond}, \sigma_k^{Bayes}), \mu^C) < P(\text{FC} \mid (\sigma_k^{Bayes}, \sigma_k^{Bayes}), \mu^{Bayes})$$

Take player k to have his turn after a cascade has started and consider the decision at his node v. Player k is better advised to play the Condorcet rather than the Bayesian strategy, since the former yields a lower probability of a false cascade happening in the game than the latter. This is what we show in the rest of the proof, by calculating a series of probabilities.

- The probability that a false cascade has formed before player k's turn can be easily approximated by the probability that a false cascade happens in the entire game. By Theorem 1, this is equal to $\frac{1}{5}$.
- The probability that a false cascade formed before k will continue after k's move, given that k plays the Condorcet strategy, is

$$\underbrace{\frac{1}{5}}_{pr\ that\ false\ cascade\ forms\ before\ k} \cdot \underbrace{\frac{1}{3}}_{pr\ that\ k's\ signal\ is\ low\ quality}$$

- The probability that a false cascade will form at some point after k, given that k broke a false cascade started before k and k plays the Condorcet strategy, is:

$$\underbrace{\frac{1}{5}}_{pr\ false\ cas\ before\ k} \cdot \underbrace{\frac{2}{3}}_{pr\ k's\ signal\ is\ high} \cdot \underbrace{\frac{1}{3}}_{pr\ (k+1)'s\ signal\ is\ low}$$

together with

$$\underbrace{\frac{1}{5}}_{pr\ FC\ before\ k} \cdot \underbrace{\frac{2}{3}}_{pr\ k's\ signal\ is\ high} \cdot \underbrace{\frac{2}{3}}_{pr\ (k+1)'s\ signal\ is\ high} \cdot \underbrace{\frac{1}{5}}_{pr\ FC\ after\ k}$$

- The probability that a false cascade will form after k given that a true cascade was broken by k and k plays the Condorcet strategy is:

$$\underbrace{\frac{4}{5}}_{pr\ TC\ before\ k} \cdot \underbrace{\frac{1}{3}}_{pr\ k's\ signal\ is\ low} \cdot \underbrace{\frac{1}{3}}_{pr\ (k+1)'s\ signal\ is\ low} \cdot \underbrace{\frac{1}{5}}_{pr\ FC\ after\ k}$$

Now we are ready to compute the probability of a false cascade starting after player k, given that he plays the Condorcet strategy and everyone else is plying the Bayesian strategy :

$$\begin{aligned} P(\text{FC} \mid \sigma_k^{Cond}, \sigma_{-k}^{Bayes}) = & P(\text{FC continues} \mid \text{FC before } k) \\ & \cdot P(\text{FC before } k) \\ & + P(\text{FC after } k \mid \text{FC broken by } k) \\ & \cdot P(\text{FC broken by } k) \\ & + P(\text{FC after } k \mid \text{TC before } k) \\ & \cdot P(\text{TC before } k) \end{aligned}$$

By plugging in the numbers computed earlier, we get that:

$$\begin{aligned} P(\text{FC} \mid \sigma_k^{Cond}, \sigma_{-k}^{Bayes}) &= \frac{1}{5} \cdot \frac{1}{3} + \frac{1}{5} \cdot \frac{2}{3} \cdot \frac{1}{3} + \frac{1}{5} \cdot \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{5} + \frac{4}{5} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{5} \\ &= \frac{11}{75} \end{aligned}$$

By applying Theorem 1 for $p = \frac{1}{3}$, we get that

$$P(\text{FC forms} \mid (\sigma_k^{Bayes}, \sigma_{-k}^{Bayes})) = \frac{1}{5}$$

and since $\frac{11}{75} < \frac{1}{5}$, the Condorcet strategy for player k yields a higher expected payoff than the Bayesian strategy, given that everyone else plays the Bayesian strategy. We proved that σ^{Bayes} is not sequentially rational given μ^{Bayes}.

□

If we calculate the percentage difference in likelihood of cascades, it turns out that player k, by making his private signal public, reduces the probability of a false cascade by a little more than a quarter. The interesting conclusion to be drawn from Theorem 6 is that it takes very little to reverse a cascade, even in a population dominated by conformists. One deviator, known as revealing his private signal, is sufficient to induce a sudden change of actions in the direction of the true state of the world. This testifies to the fragility of cascades that allow single agents to break and (or) start a new pattern of conformism.

We now move to the next result of this section. We show that the Condorcet strategy profile together with its consistent belief system is a Perfect Bayesian Nash Equilibrium of the Majority Urn game. The interesting part of the proof is showing that the Condorcet strategy is a best response to everyone else choosing Condorcet. As before, we only consider the two salient strategy profiles Condorcet and Bayesian. In the language of mechanism design, we could say that the imposition of π^{Maj} induces a Bayesian incentive compatible mechanism for the revelation of private information.

> *Theorem 7*
>
> The assessment $(\sigma^{Cond}, \mu^{Cond})$ represents a Perfect Bayesian Nash Equilibrium of the Majority Urn Game.

Proof. In what follows, we only prove that σ^{Cond} is sequentially rational given the set of beliefs μ^{Cond}, as proving weak consistency-of-beliefs is evident. For this purpose, we proceed to prove the following claim:

$$E^{\mu^{Cond},I}(\sigma_i^{Cond}, \sigma_{-i}^{Cond}) = 1$$

where

$$E^{\mu^{Cond},I}(\sigma_i^{Cond}, \sigma_{-i}^{Cond}) = P(\text{majority is right} \mid \sigma_i^{Cond}, \sigma_{-i}^{Cond})$$

as proved in Theorem 6.

CASE 1 Assume n is large.

By applying the Condorcet Jury Theorem (See 2) to the Majority Urn Game, we get that the probability of a majority of people being right approaches 1, as $n \to \infty$.

$$\lim_{n\to\infty} P(\text{majority is right} \mid \sigma_i^{Cond}, \sigma_{-i}^{Cond}) = 1$$

In fact, given the precision of our high quality signal, $\frac{2}{3}$, the probability that the majority gets it right becomes approximatively 1, for relatively low n's. So, for large n's, we approximate the probability that the majority is right, given σ_i^{Cond}, by 1. Therefore, we can say that no strategy will do better than Condorcet since:

$$E(\sigma_i^{Bayes}, \sigma_{-i}^{Cond}) = P(\text{majority is right} \mid \sigma_i^{Bayes}, \sigma_{-i}^{Cond}) < 1$$

CASE 2 Assume n is small.

Here, we can further assume that player i acts as if his move would make a difference to the outcome of the game. For otherwise, he would be indifferent between choosing among his strategies.

Player i can only make a difference, if at the end of the game, all the other players would have chosen W_i and B_i in equal numbers. In this case, both his Bayesian and Condorcet strategies will advise him to follow his signal. Therefore, the argument is that, if player i believes he has a say in the game, no matter where he is in the sequence, then he must believe that by the end of the game, the scores are tied. In that case then, both the Bayesian and the Condorcet strategies tell him the same thing: to follow his signal.

Therefore, a rational agent i, ignorant of whether he will have made a difference by the end of the game, will always choose to follow his signal, since following his signal is as good as choosing according to his belief in the right urn in case he doesn't make a difference, and better in case he does get to have the swing vote, in retrospect.

□

The last results of the section concern showing that in the Hybrid Game, the Bayesian strategy and its accompanying belief system persists as a PBNE, for certain values of x and y. As expected, this result shows the mitigating effect on the likelihood of cascades occurring in the game, that the addition of a payoff component based on the performance of the majority. However, we prove last that the Condorcet strategy is never a PBNE, no matter the values of x and y, as long as $x \neq 0$. The Hybrid Game thus represents a middle ground between the Basic Urn Game and the Majority Urn Game in terms of cascade likelihood.

Theorem 8

> For n large enough, and for x, y such that $y > 4x/10$, then the assessment $(\sigma^{Bayes}, \mu^{Bayes})$ is a PBNE of the x, y-Hybrid Urn Game with n players.

Proof. As before, we only need to check the sequential rationality condition of a PBNE. More specifically, we are looking for the x and y values for which the Bayesian strategy is sequentially rational, given the beliefs μ^{Bayes}. We begin by computing player i's expected payoff of the Bayesian strategy, $E_i(\sigma_i^{Bayes}, \sigma_{-i}^{Bayes})$. Given that $n \to \infty$ and every player plays σ_i^{Bayes}, by Theorem 4 we get that the likelihood of a cascade forming during the game is 1. Therefore, there will always be a majority guess, and that guess will be correct if and only if the cascade that forms is a true cascade. By Theorem 1, the probability of a true cascade forming during a play of the game is $\frac{4}{5}$ and it is precisely the probability that the majority is right. Therefore, the expectation of the collective payoff component x is determined by the probability of a true cascade forming and is, consequently, equal to $\frac{4}{5}x$. Moreover, given that n is very large, cascades tend to ensue relatively early in the game, so we can safely approximate the probability that any individual is right by the probability of a true cascade forming. This gives us the following expression for the expected payoff of player i:

$$
\begin{aligned}
E(\sigma_i^{Bayes}, \sigma_{-i}^{Bayes}) &= P(\text{true cascade})x + P(\text{true cascade})y \\
&= \frac{4}{5}(x + y)
\end{aligned}
$$

It remains to show that the expected payoff of playing the Bayesian strategy cannot be surpassed by playing any other strategy. We argue that we only need to look at only one other candidate strategy, namely the only other that seems reasonable given the payoff structure: the Condorcet strategy.

As computed in Theorem 6, we know that the likelihood of a cascade forming when one player plays the Condorcet strategy and all the others play the Bayesian strategy is:

$$
P(\text{ false cascade } \mid (\sigma_i^{Cond}, \sigma_{-i}^{Bayes})) = \frac{11}{75}
$$

However, the probability of a cascade forming during a play of this game is, by Theorem 4, also 1. Then, following the same reasoning as above, we obtain that:

$$
\begin{aligned}
P(\text{majority is right} \mid (\sigma_i^{Cond}, \sigma_{-i}^{Bayes})) &= P(\text{true cascade} \mid (\sigma_i^{Cond}, \sigma_{-i}^{Bayes})) \\
&= 1 - \frac{11}{75} \\
&= \frac{64}{75}
\end{aligned}
$$

and hence the expected payoff of playing the Condorcet strategy is:

$$\begin{aligned} E(\sigma_i^{Cond}, \sigma_{-i}^{Bayes}) &= P(\text{true cascade})x + P(\text{i's signal is correct})y \\ &= \frac{11}{75}x + \frac{2}{3}y \end{aligned}$$

It remains to calculate for which values of x and y:

$$\frac{4(x+y)}{5} > \frac{2y}{3} + \frac{64x}{75}$$

After calculations, we obtain that the Bayesian strategy is sequentially rational whenever $y > \frac{4x}{10}$. Therefore, we showed that the Bayesian assessment is a PBNE, whenever $y > \frac{4x}{10}$. □

We make the observation that, in the Hybrid Urn Game, the Condorcet assessment is never a PBNE, for all values x, y with $y \neq 0$, since, for a player i, it gives a lower expected payoff than the Bayesian strategy. In the next section, we try to rectify this sub-optimal result by modifying the structure of the game.

As summarized in Table 4.1, in this section we have shown that:

- The conformity assessment is a PBNE when the payoff rewards agents based on the ability to conform to the majority's decision.
- The Bayesian assessment ceases to be a PBNE, when the payoff rewards agents solely based on the group outcome.
- The Condorcet assessment is a PBNE, when an agents' payoff is entirely dependant on the group outcome.
- The Bayesian assessment is a PBNE, whenever agents are rewarded, in certain ratios, both based on their individual performance and that of the group.

The inevitable conclusion that follows is that, whenever an incentive is present that is not exclusively based on the group attainment of truth, optimal behaviour leads to cascades.

4.3 THE HYBRID COORDINATION GAME

As we've seen in the previous section, under our most general payoff considered, the Hybrid Urn Game is still very vulnerable to cascade behaviour (we proved the Bayesian assessment represents a PBNE and the Condorcet assessment doesn't). In this section, we modify some of the structural assumptions of the Hybrid Urn Game in order to direct the optimal behaviour of players towards more socially desirable

Assessment	Conformity Urn G	Majority Urn G	Hybrid Urn G
Condorcet		✓	✗
Bayesian		✗	✓
Conformity	✓		

Table 4.1: PBNE status for different Urn Games and assessment types. Empty combinations are conjectured to be a ✗, but not proven in this thesis.

outcomes. In particular, we want to create the effects of a pre-play deliberation stage in which players can coordinate on their *strategies*, prior to them knowing their private information (including their turn in the game). This handily generalizes the cascade setting, in order to accommodate more interesting and life-like situations, such as the deliberations of a jury in a trial (Osborne, 2004, p. 301). Given the symmetry of each player's information, it's only natural to look at situations in which they *all* need to agree, as a grand coalition, on a joint strategy profile. Deviations from an agreed upon joint strategy will be penalized.

The pure joint strategies available to a group are just the Cartesian product of the individual pure strategies of the group members. We assume, for simplicity of exposition, that there are only two salient individual pure strategies each agent can "rationally" choose from: the Condorcet and the Bayesian strategies, introduced below, together with the definition of a pure strategy:

Definition 18: Pure strategies

A pure strategy of player i in an extensive game is a function $s_i : V_i \to A$, such that:

- $s_i(v) \in A(v)$
- if $v \sim_i v'$ then $s(v) = s(v')$

We denote a pure strategy profile by $s = (s_1, \ldots, s_n)$.

A pure strategy can be seen as a behavioural strategy in which every distribution assigns probability 1 to a single action.

Definition 19: The Bayesian Pure Strategy

The *Bayesian pure strategy* dictates that every player i chooses, at each of his nodes v, the action with the highest probability of being true, according to his beliefs μ_i.

$$s_i^{Bayes}(v) = \begin{cases} W_i & \text{if } \mu_i(W) > \frac{1}{2} \\ B_i & \text{if } \mu_i(B) > \frac{1}{2} \\ a \in A(v) \text{ s.t. colour}(a) = \text{colour(last}(v)) & \text{otherwise} \end{cases}$$

where $\mu_i(W) = \sum_{t \in I}^{s.t.t(1)=W} \mu_i(t)$.

This definition reflects the self-biased tie-breaking rule assumed in the Urn Model.

Definition 20: The Condorcet Pure Strategy

The *Condorcet strategy* dictates that players never take into consideration any information other than the one conveyed by their own private signal:

$$s_i^{Cond}(v) = \begin{cases} W_i & \text{if } last(v) = w_i \\ B_i & \text{otherwise} \end{cases}$$

Of all the possible joint strategies that the grand coalition can agree upon, three are of particular interest, and therefore we limit the agents to considering only these in the deliberation period. These salient joint strategies, which we call protocols, are described below in detail:

Definition 21: The Condorcet Protocol

Under the Condorcet protocol s^{Cond}, every player i plays the Condorcet strategy s_i^{Cond}. This is the strategy according to which one guesses the colour of the marble one draws from the urn. Formally, $s^{Cond} = (s_1^{Cond}, s_2^{Cond}, \ldots, s_n^{Cond})$.

Definition 22: The Bayesian Protocol

Under the Bayesian protocol s^{Bayes}, every player i plays the Bayesian strategy s_i^{Bayes}. This requires that every player guesses the urn for which the posterior probability, after having observed his signal and the actions of players before him, is higher. Formally, $s^{Bayes} = (s_1^{Bayes}, s_2^{Bayes}, \ldots, s_n^{Bayes})$.

Definition 23: The Hybrid Protocol

The Hybrid protocol prescribes that the first $\frac{n}{m}$ people in the sequence(where $\frac{n}{m}$ is odd and m is chosen such that $\frac{n}{m} \in \mathbb{N}$)

guess the colour of their marble, and the rest each guess what the majority of the first $\frac{n}{m}$ people have guessed. Let $m \geq 2$ be fixed.

$$\underbrace{1 \quad 2 \ldots \frac{n}{m}}_{Condorcet} \mid \overbrace{\frac{n}{m} + 1 \ldots n}^{imitation}$$

No player knows, a priori, what his place in the sequence will be. Formally, $s^{Hyb} = (s_1^{Cond}, \ldots, s_{\frac{n}{m}}^{Cond}, s_{\frac{n}{m}}^{Bayes}, \ldots, s_n^{Bayes})$.

Our goal is to design a game that simulates a pre-play deliberation period, before the Hybrid Urn Game starts, where players can coordinate on the joint strategy that maximizes the social welfare of the group[1]. First, we prepare the ground by introducing the definition of *extensive form games with simultaneous moves.* Then, we define the Hybrid Coordination Game, which introduces a new turn into urn games, at which all agents choose simultaneously which of three pure strategies to play for the rest of the game. The choices of each player are publicly revealed when the turn ends, at which point an urn game starts. The payoffs of the Hybrid Coordination Game are the hybrid payoffs with two twists: a penalty for those players who no not follow the strategy they chose in the first turn of the game and a penalty for lack of coordination amongst agents.

Definition 24: Extensive form games with simultaneous moves

(Adapted to imperfect information games from Osborne (2004, p. 206). An extensive form game with imperfect information and simultaneous moves is a structure

$$\mathcal{G} := \langle A, \mathcal{T}, Ag \cup \{0\}, Pl^*, (\mathcal{I}_i)_{i \in Ag}, (\pi_i)_{i \in Ag}, (p_v)_{v \in V_0} \rangle$$

where all the elements are as defined in Definition 3.1.2, except:

- The player function $Pl^* : NonTerm(\mathcal{T}) \rightarrow \mathcal{P}(Ag \cup \{0\})$ that assigns a set of players to every sequence of nodes pertaining to information sets of agents.

This definition generalizes the player function to allow more than one player to move at each node.

1 The social welfare of a group is the sum of the payoffs of the individuals in the group.

Definition 25: The Hybrid Coordination Game

The Hybrid Coordination Game represents an extensive form game with simultaneous moves

$$C = \langle Ag \cup \{0\}, A, \mathcal{T}, Pl^*, (\mathcal{I}_i)_{i \in Ag}, (p_x)_{x \in V_0} \rangle$$

such that:

- The set Ag is of the form $Ag = \{1, \ldots, n\}$, for some $n \in \mathbb{N}$, and 0 denotes the player Nature, like in the general case.
- A is a set of actions of the form $A := \cup_{k=1}^{k=2n+1} A_k \cup A_{Ag}$ where:

  $A_{Ag} = \{\text{Hybrid}, \text{Bayes}, \text{Cond}\}$ and

  $$A_k = \begin{cases} \{W_{\frac{k-1}{2}}, B_{\frac{k-1}{2}}\} & \text{if } k \text{ is odd and } k \neq 1 \\ \{w_{\frac{k}{2}}, b_{\frac{k}{2}}\} & \text{if } k \text{ is even} \\ \{W, B\} & \text{if } k = 1 \end{cases}$$

- $\mathcal{T} := \{t \mid t \in A_0 \times A_{Ag} \times A_1 \ldots A_j, \text{ for } j \leq 2n+1\}$ $\mathcal{T}$ represents all the possible partial and terminal histories of the game that depict the order of play.
- The player function Pl^* that requires players to move at the even depths of nodes and Nature at the odd depths of nodes, as follows:

  $$Pl^*(v) = \begin{cases} Ag & \text{if } d(v) \text{ is } 0 \\ \frac{d(v)-1}{2} & \text{if } d(v) \text{ is odd} \\ 0 & \text{otherwise} \end{cases}$$

- For each player $i \in Ag$, the information partition $\mathcal{I}_i$ is given by:

  $$I(v) = \{v' \in V_i | v(1) = v'(1), v(2j+2) = v'(2j+2), \\ \forall j < i \text{ and } v(2i+1) = v'(2i+1)\}$$

  where $v(1) = (v_1(1), v_2(1), \ldots, v_n(1))$ is the sequence of choices made on the first turn, by each player. This formalization ensures that agents observe everyone's choice of

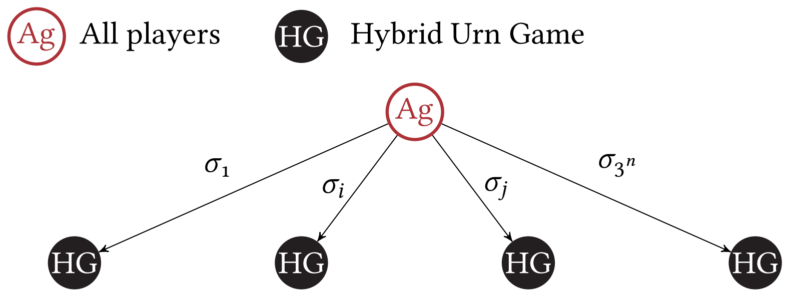


Figure 4.1: A graphical representation of the Hybrid Coordination game.

strategy, the actions of the previous players and their own signal. The signals of the previous players and the choice of world are not observed.

- A family of probability mass functions

$$(p_v : A(v) \to [0, 1])_{v \in V_0}$$

as before

$$p_v(a) = \begin{cases} \frac{2}{3} & \text{if colour}(v(1)) = \text{colour}(a) \\ \frac{1}{3} & \text{otherwise} \end{cases}$$

where $a \in A(v)$, and $p_r : A(r) \to [0, 1]$ with $a \in A(r)$ such that:

$$p_r(a) = \begin{cases} \frac{1}{2} & \text{if } a = W \\ \frac{1}{2} & \text{otherwise} \end{cases}$$

- The payoff function of player i, π_i^{Coor}, is represented by:

$$\pi_i^{Coor} = \begin{cases} \pi_i^{Hyb}(v) & \text{if } v_i(1) = v_j(1) \forall i, j \in Ag \text{ and} \\ & v_i(1)(v \upharpoonright 2i+2) = v(2i+2) \\ 0 & \text{otherwise} \end{cases}$$

Clarification on the first move

At the first turn of the Hybrid Coordination Game, all players except Nature make a move. Each agent can choose from a set of three actions they can perform at this turn: $\{\text{Hybrid}, \text{Bayes}, \text{Cond}\}$. The action Hybrid states that if i's place in the sequence of the Hybrid Game is lower than $\frac{n}{m}$, then player i plays the s_i^{Cond} strategy, and if i's place in the sequence of the Hybrid Game is higher than $\frac{n}{m}$, then player i plays the s_i^{Bayes} strategy. The Bayes action represents the choice of player i

player $-i$		Hybrid	Bayes	Cond
	Hybrid	$x + \frac{(3m-1)}{3m}y$ $x + \frac{(3m-1)}{3m}y$	0 0	0 0
	Bayes	0 0	$\frac{4(x+y)}{5}$ $\frac{4(x+y)}{5}$	0 0
	Cond	0 0	0 0	$x+\frac{2}{3}y$ $x+\frac{2}{3}y$

Figure 4.2: Hybrid Coordination Game Payoff Matrix. We collapsed the n-player game into a two-person game in which each agent plays against all the others taken collectively. If $m \geq 2$ and $y \neq 0$ then out of the three Nash Equilibria, $(Hybrid, Hybrid)$ yields the highest expected individual payoff.

s_i^{Bayes} in the Hybrid Game and similarly, the Cond action represents the choice of s_i^{Cond}.

The Hybrid Coordination Game thus defined could be visualized as a strategic, one-off, game in which players commit to strategies in the Hybrid Urn Game. For simplicity, we constructed the payoff matrix of the Hybrid Coordination game bi-dimensionally, describing player i's payoffs against the individual payoffs of all the other players in the game. This is represented by Figure 4.2. We note that formally, this does not constitute a strategic game proper, since its payoffs are given in expectation form, and not deterministic form, as usual.

The strategy profile in which every player chooses the same strategy, Bayesian, Condorcet or Hybrid and then plays according to it represent PBNE of the Hybrid Coordination Game. Before we show this, we set to calculate the individual expected payoffs given in Figure 4.2 of playing one of three protocols.

(1) The individual expected payoff for s^{Hyb}

Let $E_i(s^{Hyb})$ be the expected payoff of player i when the group is playing according to the Hybrid Protocol.

$$E_i(s^{Hyb}) = \overbrace{P\left(i \leq \frac{n}{m}\right)}^{\frac{1}{m}} \cdot E_i(s^{Hyb} \mid i \leq \frac{n}{m}) \\ + \underbrace{P\left(i > \frac{n}{m}\right)}_{\frac{m-1}{m}} \cdot E(\pi_i^{Coor})(s^{Hyb} \mid i > \frac{n}{m})$$

The next step is to apply the limit to the expression obtained:

$$\lim_{n\to\infty} E(\pi_i^{Coor})(s^{Hyb}) = \lim_{n\to\infty}\left(\frac{1}{m} \cdot E_i(s^{Hyb} \mid i \leq \frac{n}{m}\right) \tag{4.1}$$

$$+ \lim_{n\to\infty}\left(\frac{(m-1)}{m} \cdot E_i(s^{Hyb} \mid i > \frac{n}{m})\right) \tag{4.2}$$

We manipulate the limit operator for the first component 4.1 and obtain:

$$\begin{aligned}
(4.1) &= \frac{1}{m} \cdot \left(\lim_{n\to\infty} E_i(s^{Hyb} \mid i \leq \frac{n}{m})\right) \\
&= \frac{1}{m}(\lim_{n\to\infty}\left(x \cdot P(\text{ maj is right } \mid s^{Hyb}\right) \\
&\quad + \lim_{n\to\infty}\left(y \cdot P(\text{guess of i is correct} \mid s_i^{Cond})\right)) \\
&= \frac{1}{m}[x \cdot \lim_{n\to\infty} P(\text{ maj is right } \mid s^{Hyb}) \\
&\quad + y \cdot \lim_{n\to\infty} P(sg_i \text{ is high})] \\
&= \frac{1}{m} \cdot x + \frac{1}{m} \cdot y \cdot \frac{2}{3} \\
&= \frac{1}{m}\left(x + \frac{2y}{3}\right)
\end{aligned}$$

The last two lines follow by an application of Theorem 2:

$$\lim_{n\to\infty} P(\text{maj is right} \mid s^{Hyb}) = \\ \lim_{n\to\infty} P(\text{maj of the first } \frac{n}{m} \text{ is right} \mid s^{Hyb}) = 1$$

and

$$\lim_{n\to\infty} P(sg_i \text{ is high} \mid i \leq \frac{n}{m}) = \lim_{n\to\infty} \frac{2}{3} = \frac{2}{3}$$

Similarly, for the second component of the equation, Eqn. 4.2, we obtain:

$$\begin{aligned}
(4.2) &= \lim_{n\to\infty} P(i > \frac{n}{m}) \cdot E_i(s^{Hyb} \mid i > \frac{n}{m}) \\
&= \frac{m-1}{m} \cdot \left(x \cdot \lim_{n\to\infty} P(\text{ maj is right } \mid s^{Hyb})\right) \\
&\quad + \frac{m-1}{m} \cdot \left(y \cdot \lim_{n\to\infty} P(\text{guess of i is correct} \mid s_i^{Bayes}))\right) \\
&= \frac{m-1}{m}(x \cdot 1 + y \cdot 1)
\end{aligned}$$

The last line follows by an application of Theorem 2:

$$\begin{gathered}
\lim_{n\to\infty} P(\text{ maj is right } \mid s^{Hyb}) = \\
\lim_{n\to\infty} P(\text{guess of i is correct} \mid s_i^{Bayes})) = \\
\lim_{n\to\infty} P(\text{ maj of the first } \frac{n}{m} \text{ is right } \mid s^{Hyb}) = 1 \text{ (by Thm. 2)}
\end{gathered}$$

Bringing together our components (4.2) and (4.1), we obtain the desired result:

$$\begin{aligned}
E_i(s^{Hyb}) &= \frac{1}{m}\left(x + \frac{2y}{3}\right) + \frac{m-1}{m}(x \cdot 1 + y \cdot 1) \\
&= x + \left(\frac{3m-1}{3m}\right) \cdot y
\end{aligned}$$

(2) *The individual expected payoff of* s^{Cond}

Let $E_i(s^{Cond})$ be the expected payoff of player i when the group is playing according to the Condorcet Protocol. We need to calculate: $\lim_{n\to\infty} E_i(s^{Cond})$.

$$\begin{aligned}
\lim_{n\to\infty} E_i(s^{Cond}) &= x \cdot \lim_{n\to\infty} P(\text{ maj is right } \mid s^{Cond}) \\
&\quad + y \cdot \lim_{n\to\infty} P(\text{guess of i is correct} \mid s_i^{Bayes}) \\
&= x \cdot 1 + y \cdot \frac{2}{3}
\end{aligned}$$

(3) The individual expected payoff of s^{Bayes}

Given that $n \to \infty$, the likelihood of a cascade forming is 1, see Bikhchandani et al. (1992). This together with Theorem 1, gives us:

$$E_i(s^{Bayes}) = \frac{4}{5}(x+y)$$

In the remainder of this section we will assess which protocol gives the highest expected payoff for the individuals in the game.

Definition 26: Protocol Dominance

A protocol A is said to dominate another protocol B, if every individual in the group derives a higher expected utility from the group playing according to protocol A than from the group playing according to protocol B.

We prove three theorems that offer an order of dominance among these protocols:

Theorem 9

The Hybrid Protocol strongly dominates the Condorcet Protocol, for all payoff functions $(\pi_i^{Coor})_{i\in\mathcal{N}}$ with $x \geq 0,\ m > 1$ *and* $y \neq 0$ as $n \to \infty$.

Proof. We want to find out for what values of $x,\ y$ and m the Hybrid protocol dominates the Condorcet protocol:

$$\lim_{n\to\infty} E_i(s^{Hyb}) > \lim_{n\to\infty} E_i(s^{Cond})$$

Plugging in the values computed above, we obtain:

$$x + \frac{3m-1}{3m} \cdot y > x + \frac{2y}{3} \Leftrightarrow \frac{3m-1}{3m} > \frac{2}{3} \Leftrightarrow m > 1,\ y \neq 0$$

□

Theorem 10

The Condorcet Protocol strongly dominates the Bayesian Protocol, for all payoff functions $(\pi_i)_{i\in\mathcal{N}}$ with $x \geq 0$ and $y \neq 0$, $y < \frac{3}{2}x$ as $n \to \infty$.

Proof. We need to show, in a manner akin to the previous proof, for which values of x and y the Condorcet protocol is better than the Bayesian protocol:

$$\lim_{n\to\infty} E_i(s^{Cond}) > \lim_{n\to\infty} E_i(s^{Bayes})$$

We know from previous calculations that:

$$E_i(s^{Cond}) = x + \frac{2}{3}y$$

$$E_i(s^{Bayes}) = \frac{4}{5}(x+y)$$

Then, after some straightforward manipulation, we obtain:

$$x + \frac{2}{3}y > \frac{4}{5}(x+y) \Leftrightarrow y < \frac{3}{2}x$$

□

Theorem 11

The Hybrid protocol strongly dominates the Bayesian protocol for all payoff functions $(\pi_i)_{i\in N}$ with $x, y \geq 0$, and $m \geq 2$ as $n \to \infty$.

Proof. We need to find for what values of x, y and m the Hybrid protocol dominates the Bayesian protocol:

$$\lim_{n\to\infty} E_i(s^{Hyb}) > \lim_{n\to\infty} E_i(s^{Bayes})$$

After plugging in the values, we get that:

$$x + \frac{3m-1}{3m}y > \frac{4}{5}(x+y) \text{ if and only if } m > 2$$

□

Theorem 12

The Condorcet Protocol is weakly dominant for all payoff functions $(\pi_i)_{i\in N}$ with $y = 0$ and $x \neq 0$ as $n \to \infty$.

Proof. Given that $y = 0$, the expected utilities for the different protocols become:

$$\begin{aligned} E_i(s^{Bayes}) &= \frac{4}{5}x \\ E_i(s^{Cond}) &= x \\ E_i(s^{Hyb}) &= x \end{aligned}$$

The result is then immediate. □

Putting all of these results together, we get that, when $n \to \infty$:

$$E_i(s^{Hyb}) > E_i(s^{Cond}) > E_i(s^{Bayes})$$

for $y < 1.5x,\ y \neq 0,\ m \geq 2$ and that:

$$E_i(s^{Hyb}) = E_i(s^{Cond}) > E_i(s^{Bayes})$$

for $y = 0$.

Given that players receive a payoff of 0 whenever they deviate from their chosen strategy or their strategy is not uniform, it is straightforward to show that each of the three protocols represents a PBNE. However, of the individual expected payoffs of playing each of these protocols, the Hybrid protocol yields the maximum. Then, by structurally changing the game and the payoff function to punish players from deviating from the group strategy, we succeeded in directing players towards a more socially desirable outcome. In conclusion, in the first part of this chapter we looked at alternative payoff structures for urn games. In the second part of the chapter we took a mechanism design perspective on the Hybrid Urn Game, and created a new strategic form game that forces cooperation at the level of the group, with the aim of maximizing its social welfare.

5

PROBABILISTIC LOGIC OF COMMUNICATION AND CHANGE

In this chapter we introduce a new epistemic dynamic logic, whose static language is Probabilistic Epistemic PDL (PE-PDL), a probabilistic extension to Epistemic PDL, given in (van Benthem et al., 2006). What results is a powerful logic, called Probabilistic Logic of Communication and Change (PLCC) that unifies, under its framework, arbitrary levels of mutual knowledge, including common knowledge and a multi-agent Bayesian update mechanism that accommodates various social-epistemic scenarios. The strength of this logic comes from its combination of probabilistic features and higher levels of mutual knowledge. We provide a sound and complete axiomatization for PLCC.

5.1 PROBABILISTIC LOGIC OF COMMUNICATION AND CHANGE

5.1.1 *Language of PLCC*

Let At be a set of atomic propositions and Ag a set of basic agents $Ag = \{a, b, c \ldots\}$. Consider the PLCC language $\mathcal{L}_{\text{PLCC}}$ given by the following Backus Naur form:

$$\phi ::= p \mid p \wedge \neg p \mid \neg\phi \mid \phi \wedge \psi \mid [\pi]\phi \mid [e]\phi \mid \alpha_1 \cdot P_a(\phi_1) + \ldots + \alpha_n \cdot P_a(\phi_n) \geq \beta$$
$$\pi ::= a \mid \pi;\pi \mid \pi \cup \pi \mid \pi^* \mid \phi?$$

where $p \in At$ is an atom proposition, $\alpha_1, \ldots, \alpha_n, \beta$ are rational numbers, e is an event from the domain of a probabilistic update model A, and $a \in Ag$ is an agent.

The static underlying language of PLCC is the language that results from using the same syntactic constructions as in PLCC, except for $[e]\phi$. We call this probabilistic epistemic PDL, which we abbreviate by PE-PDL.

In the epistemic interpretation we are pursuing, $[a]\phi$ corresponds, in natural language paraphrase, to "agent a knows that ϕ". However, by making use of constructors, we can create more complex levels of knowledge (Parikh, 2003). One extreme state of knowledge is represented by a proposition ϕ being common knowledge amongst a group of agents $A \subset Ag$ which is captured as $[A^*]\phi$.

Weaker levels of knowledge are illustrated, for example, by statements of the kind "a and b know that ϕ" , which appears as $[a \cup b]\phi$. Another example is given by the composition constructor, that can express notions like "agent a knows that agent b knows that ϕ" via the expression $[a; b]\phi$. The language of PLCC is very expressive and can accommodate for more esoteric notions of group knowledge, however there is currently no salient interpretation for them. Lastly, the unusual character of the test operator allows us to construct a modality$[\phi?]$ out of any formula ϕ . Intuitively, this modality accesses the present state if the present state satisfies ϕ.

We abbreviate: $\neg[\pi]\neg\phi$ by $\langle\pi\rangle\phi$, $\neg(\neg\phi_1 \wedge \neg\phi_2)$ by $(\phi_1 \vee \phi_2)$, $\neg(\phi_1 \wedge \phi_2)$ by $(\phi_1 \rightarrow \phi_2)$, $(\phi_1 \rightarrow \phi_2) \wedge (\phi_2 \rightarrow \phi_1)$ by $\phi_1 \leftrightarrow \phi_2$. On top of these, we add other obvious abbreviations for formulas involving probabilities such as:

- $\sum_{j=1}^{j=n} \alpha_j P_a(\phi_j)$ for $\alpha_1 \cdot P_a(\phi_1) + \ldots + \alpha_n \cdot P_a(\phi_n) \geq \beta$. We call formulas of this form a-probability formulas.

- $\sum_{j=1}^{j=n} \alpha_j P_a(\phi_j) \leq \beta$ for $\sum_{j=1}^{j=n} -\alpha_j P_a(\phi_j) \geq -\beta$

- $\sum_{j=1}^{j=n} \alpha_j P_a(\phi_j) \geq \sum_{i=1}^{i=m} \beta_i P_b(\psi_i)$ for

$$\sum_{j=1}^{j=n} \alpha_j P_a(\phi_j) + \sum_{i=1}^{i=m} -\beta_i P_b(\psi_i) \geq 0$$

- $\sum_{j=1}^{j=n} \alpha_j P_a(\phi_j) > \beta$ for

$$\left(\sum_{j=1}^{j=n} \alpha_j P_a(\phi_j) \geq \beta\right) \wedge \neg\left(\sum_{j=1}^{j=n} \alpha_j P_a(\phi_j) \leq \beta\right)$$

- $\sum_{j=1}^{j=n} \alpha_j P_a(\phi_j) < \beta$ for

$$\left(\sum_{j=1}^{j=n} \alpha_j P_a(\phi_j) \leq \beta\right) \wedge \neg\left(\sum_{j=1}^{j=n} \alpha_j P_a(\phi_j) \geq \beta\right)$$

- $\sum_{j=1}^{j=n} \alpha_j P_a(\phi_j) = \beta$ for

$$\left(\sum_{j=1}^{j=n} \alpha_j P_a(\phi_j) \geq \beta\right) \wedge \left(\sum_{j=1}^{j=n} \alpha_j P_a(\phi_j) \leq \beta\right)$$

5.1.2 *Semantics of $\mathcal{L}_{PLCC}$*

Definition 27: Bayesian Kripke models

Take as given a set of agents Ag and a set of literals At. A Bayesian Kripke model is a quadruple $M = (S, \sim, \mu_a, V)_{a \in Ag}$ where:

- S is a finite non-empty set of states.
- $\sim$ is a set of equivalence relations $\sim_a$ on S for agent a.
- A function $\mu_a : S \to (S \to [0, 1])$, for each agent $a \in Ag$ such that the following conditions are met:

  SDP if $s \sim_a v$ then $\mu_a(s)(s') = \mu_a(v)(s') \forall s' \in S$ (Fagin and Halpern, 1994), corresponding to axiom W10

  CONS $\mu_a(s)(v) = 0$ if $s \nsim_a v$ (Fagin and Halpern, 1994)

  CAUT $s \nsim_a v$ if $\mu_a(s)(v) = 0$ (van Eijck and Schwarzentruber, 2014)

  PROB for every $s \in S$ $\sum_{t \in S} \mu_a(s)(t) = 1$ corresponding to axiom W2

  To improve legibility we will denote an expression of the form $\mu_a(s)(s')$ by $\mu_a^s(s')$.

- V is a function that assigns a set of states in S to each propositional variable in At.

Following van Benthem et al. (2006), before describing the update mechanism, we generalize our event model by allowing factual change. Through the feature of substitution functions we are able to reset the propositional valuation of the epistemic model anterior to the event being observed. The generalization to factual change is driven by the natural interpretation of the kinds of events we usually model in games. These do not typically only involve informational change, but also factual changes, as in the case of players' actions.

Definition 28: Substitutions (van Benthem et al., 2006)

A substitution function $\sigma \colon At \to \mathcal{L}_{\mathrm{PLCC}}$ is a function that maps all but a finite number of propositional atoms into themselves. Call the set $\{p \in At \mid \sigma(p) \neq p\}$ the domain of σ and denote it by $dom(\sigma)$. Let $sub_{\mathcal{L}_{\mathrm{PLCC}}}$ denote the set of all such possible substitution functions and ϵ the identity substitution.

Definition 29: Event Models

The event model for a finite set of agents Ag with the $\mathcal{L}_{\text{PLCC}}$ language, is the quintuple $A = (E, \sim, PRE, pre_a, sub)_{a \in Ag}$ where:

- E is a non-empty finite set of events.
- $\sim$ is a set of equivalence relations $\sim_a$ for each agent $a \in Ag$.
- $PRE : E \to \mathcal{P}(\mathcal{L}_{\text{PLCC}})$ such that $PRE(e)$ is finite set of pairwise inconsistent sentences called preconditions for the occurrence of event e such that $\Phi := \bigcup_{e \in E} PRE(e)$ is a finite set of pairwise inconsistent formulas. Further, denote by $pre_e = \bigvee PRE(e)$.
- $pre_a : \Phi \to (E \to [0, 1])$ assigns to each precondition $\phi \in \Phi$ a subjective probability distribution over E, such that $pre_a(\phi)(e) = 0$ iff $\phi \notin PRE(e)$. We abbreviate $pre_a(\phi)(e)$ by $pre_a(e|\phi)$ in order to improve legibility.
- $sub : E \to sub_{\mathcal{L}_{\text{PLCC}}}$ assigns a substitution function to each event in E.

Discussion on the modelling of preconditions

According to Definition 29, the function *pre* is treated differently than in the standard probabilistic DEL framework (van Benthem, Gerbrandy and Kooi, 2009, p. 77). The first contrast with the standard probabilistic event models is represented by the addition of the function *PRE* (Baltag et al., 1999). Function PRE represents the set of objective qualitative conditions that give the preconditions for an event happening, and are objective in the sense that agents share this common language when thinking about events. These *PRE's* can be recovered in the standard definition as: $PRE(e) = \{\phi \in \Phi | pre(e|\phi) > 0\}$.

The second point of divergence with the standard definition is the function $pre_a(\phi)$ that represents the subjective probability distribution that agent a assigns to events, given a certain precondition ϕ. Our *pre* is the subjective version of the objective occurrence probabilities in the probabilistic DEL framework. The condition imposed on *pre* ensures that agents cannot be wrong about an event being possible at a given world; however the actual assignment of probabilities to possible events is left to the discretion of each agent. In a sense, our event models have absorbed the observation probabilities of van Benthem, Gerbrandy and Kooi (2009). In our view the divergence of objective and subjective probabilities does not have grounds in the Bayesian framework.

Definition 30: Product Update Models

Given a static Bayesian Kripke model $M = (S, (\sim_a)_{a \in Ag}, \mu, V)$ and an event model $A = (E, \sim, \Phi, pre_a, sub)$, we say that the result of executing A in M, is the product update model $M \otimes A = (S \otimes E, \sim, \mu, V)$ where:

- $S \otimes E = \{(s, e) \mid s \in S, e \in E, M, s \models pre_e\}$. We denote by $pre_a(e|s)$ the value of $pre_a(e|\phi_s)$, where ϕ_s is the element of Φ that is satisfied in M, s. If no such ϕ exists then $pre_a(e|s) = 0$.
- $(s, e) \sim_a (s', e')$ iff $s \sim_a s'$ and $e \sim_a e'$.
- Let $D := \sum_{(s',e') \sim_a (w,g)} (\mu_a^w(s') \cdot pre_a(e'|s'))$. Then we have that

$$\mu_a^{(w,g)}(s, e) := \begin{cases} \frac{\mu_a^w(s) \cdot pre_a(e|s)}{D} & \text{if } (s, e) \sim_a (w, g) \\ 0 & \text{otherwise} \end{cases}$$

- $V(p) = \{(s, e) \mid M, s \models sub(e)(p)\}$

Justification of the update rule

We claim that our product update rule is fully grounded in Bayesian reasoning. By inspection, it can be noticed that the product update rule is nothing else than an application of Bayes' theorem. If we regard pairs of the form (s, e) as "outcomes", then event $\bar{s}$ could be interpreted as the set of outcomes $\{(t, e)|t = s\}$ and event $\bar{e}$ can be interpreted as the set of outcomes $\{(s, f)|f = e\}$. Then the product update $\mu_a^{(w,g)}(s, e)$ is nothing than the posterior probability player a assigns to $\bar{s} \wedge \bar{e}$, after conditionalizing on her information set, where this is put in standard Bayesian terms as: $w(a) \wedge g(a)$, where $w(a) := \{w'|w' \sim_a w\}$ and $g(a) := \{g'|g' \sim_a g\}$.

Definition 31: Semantics of PLCC

The semantics of $\mathcal{L}_{\text{PLCC}}$ is an extension of the semantics for epistemic logic. We assign truth values to formulas in $\mathcal{L}_{\text{PLCC}}$ at a state s in the semantic structure M and we write $(M, s) \models \phi$ if the formula ϕ is true at a state s in the Bayesian Kripke structure M:

$$
\begin{array}{ll}
M, s \models p & \text{iff } s \in V(p) \\
M, s \models \neg\phi & \text{iff } M, s \not\models \phi \\
M, s \models \phi \wedge \psi & \text{iff } M, s \models \phi \text{ and } M, s \models \psi \\
M, s \models [a]\phi & \text{iff for all } t \in S: \text{ if } s \sim_a, t \text{ then } M, t \models \phi \\
M, s \models [e]\phi & \text{iff } M, s \models pre_e \text{ then } M \times A, (s, e) \models \phi, \\
& \quad \text{e an event in action model A} \\
M, s \models [\pi]\phi & \text{iff for all } t \in S: \text{ if } sR_\pi t \text{ then } M, t \models \phi, \\
& \quad \forall \pi \text{ a complex agent} \\
M, s \models \sum_{j=1}^{j=n} \alpha_j P_a(\phi_j) & \text{iff } \sum_{j=1}^{j=n} \alpha_j \cdot \mu_a^s(S(\phi_j)) \geq \beta
\end{array}
$$

where $\mu_a^s(S(\phi_j))$ is an abbreviation for $\sum_{s' \in S, s' \models \phi_j} \mu_a^s(s')$.

The binary relations R_π mentioned above are built, for each agent π, starting from $(\sim_a)_{a \in Ag}$ and using the following rules:

$$
\begin{aligned}
R_a &= \sim_a \\
R_{\pi_1 \cup \pi_2} &= R_{\pi_1} \cup R_{\pi_2} \\
R_{\pi_1 ; \pi_2} &= R_{\pi_1} ; R_{\pi_2} \\
R_{\pi^*} &= (R_\pi)^* \\
sR_{\phi?}t &\Leftrightarrow s = t \text{ and } s \models \phi
\end{aligned}
$$

where ; is the set-theoretic operation of concatenation and $(R_\pi)^*$ represents the reflexive-transitive closure of R_π.

We write $\models \phi$ if $M, s \models \phi$ for every pointed Bayesian Kripke model M, s.

5.1.3 *Proof System of $\mathcal{L}_{PLCC}$*

In what follows, we describe a complete axiomatization for our dynamic probabilistic PDL language $\mathcal{L}_{\text{PLCC}}$ with change of facts.

We begin with the static axiomatic system which contains the axioms of PDL together with the axioms of (Fagin et al., 1990). We compartment them into five parts, pertaining to propositional, knowledge, inequalities, probabilities and programs reasoning as portrayed below. The set of axiom schemas that results, together with the restriction of these to the language of PE-PDL, represents the proof system of PE-PDL.

Inference Rules for PLCC

K1 All instances of propositional tautologies

R1 From ϕ and $\phi \to \psi$ infer ψ(modus ponens)

R2. From ϕ infer $[\pi]\phi$

Axioms for reasoning about knowledge

For any basic program a, we have that:

K3. $[a]\phi \to \phi$

K4. $[a]\phi \to [a][a]\phi$

K5. $\neg[a]\phi \to [a]\neg[a]\phi$

Axioms for programs

1. $[\pi](\phi \to \psi) \to ([\pi]\phi \to [\pi]\psi)$
2. $[\pi_1;\pi_2]\phi \leftrightarrow [\pi_1][\pi_2]\phi$
3. $[\pi_1 \cup \pi_2]\phi \leftrightarrow [\pi_1]\phi \wedge [\pi_2]\phi$
4. $[\pi^*]\phi \leftrightarrow (\phi \wedge [\pi][\pi^*]\phi)$
5. $[\pi^*](\phi \to [\pi]\phi) \to (\phi \to [\pi^*]\phi)$
6. $[\phi?]\psi \leftrightarrow (\phi \to \psi)$

Axioms for reasoning about linear inequalities

I1. $\alpha_1 \cdot P_a(\phi_1) + \ldots + \alpha_n \cdot P_a(\phi_n) \geq \beta \leftrightarrow \alpha_1 \cdot P_a(\phi_1) + \ldots + \alpha_n \cdot P_a(\phi_n) \geq \beta + 0 \cdot P(\phi_{n+1})$

I2. $\alpha_1 \cdot P_a(\phi_1) + \ldots + \alpha_n \cdot P_a(\phi_n) \geq \beta \to \alpha_{j_1} \cdot P_a(\phi_1) + \ldots + \alpha_{j_n} \cdot P_a(\phi_n) \geq \beta$ where $j_1 \ldots j_n$ is a permutation of $1, \ldots, n$

I3. $(\alpha_1 \cdot P_a(\phi_1) + \ldots + \alpha_n \cdot P_a(\phi_n) \geq \beta) \wedge$
$(\alpha'_1 \cdot P_a(\phi_1) + \ldots + \alpha'_n \cdot P_a(\phi_n) \geq \beta')$
$\to ((\alpha_1 + \alpha'_1) \cdot P_a(\phi_1) + \ldots + (\alpha_n + \alpha'_n) \cdot P_a(\phi_n) \geq (\beta + \beta'))$

I4. $(\alpha_1 \cdot P_a(\phi_1) + \ldots + \alpha_n \cdot P_a(\phi_n) \geq \beta)$
$\leftrightarrow (d\alpha_1 \cdot P_a(\phi_1) + \ldots + d\alpha_n \cdot P_a(\phi_n) \geq d\beta)$ if $d > 0$

I5. $(t \geq b) \vee (t \leq b)$ if t is a term

I6. $(t \geq b) \to (t > c)$ if t is a term and $b > c$

Axioms for reasoning about probabilities

W1. $P_a(\phi) \geq 0$

W2. $P_a(true) = 1$

W3. $P(\phi \wedge \psi) + P(\phi \wedge \neg\psi) = P_a(\phi)$

W4. $P_a(\phi) = P_a(\psi)$ if $\phi \leftrightarrow \psi$ is a propositional tautology.

W5. $P_a(false) = 0$

W7. $[a]\phi \leftrightarrow P_a(\phi) \geq 1$

W10. $w \rightarrow [a]w$, for any w an a-probability formula.

On top of these, we add a series of reduction axioms for product update, that convey the effects of informational events. These allow the translation of any sentence from the dynamic language PLCC into an equivalent sentence in the underlying static language PE-PDL. The set of static axioms together with the reduction axioms represent the proof system of PLCC.

The reduction axioms for update models

$$\begin{aligned}
&[e]p \leftrightarrow (\mathrm{pre}_e \rightarrow sub(e)(p))\\
&[e]\neg\phi \leftrightarrow (\mathrm{pre}_e \rightarrow \neg[e]\phi)\\
&[e](\phi \wedge \psi) \leftrightarrow ([e]\phi \wedge [e]\psi)\\
&[e_i][\pi]\phi \leftrightarrow \bigwedge_{i=0}^{n-1}[T_{ij}(\pi)][e_j]\phi)\\
&A \leftrightarrow \big(pre(e) \rightarrow (C \geq D)\big)
\end{aligned}$$

where the letters in the last line stand for

$$\begin{aligned}
A &= [e]\left(\sum_{1 \leq h \leq k} \alpha_h \cdot P_a(\psi_h) \geq \beta\right)\\
C &= \sum_{\substack{1 \leq h \leq k\\ \phi_i \in \Phi\\ f \sim_a e}} \alpha_h \cdot \mathrm{pre}_a(f \mid \phi_i) \cdot P_a(\phi_i \wedge [f]\psi_h)\\
D &= \sum_{\substack{\phi_i \in \Phi\\ f \sim_a e}} \beta \cdot \mathrm{pre}_a(f \mid \phi_i) \cdot P_a(\phi_i)
\end{aligned}$$

and we used the following definition.

Definition 32: T_{ij} programme transformers

$$\begin{aligned}
T_{ij}(a) &= \begin{cases} ?\mathrm{pre}(e_i); a & \text{if } e_i R(a) e_j \\ ?\perp & \text{otherwise} \end{cases} \\
T_{ij}(?\phi) &= \begin{cases} ?(\mathrm{pre}(e_i) \wedge [e_i]\phi) & \text{if } i = j \\ ?\perp & \text{otherwise} \end{cases} \\
T_{ij}(\pi_1; \pi_2) &= \bigcup_{k=0}^{n-1} \big(T_{ik}(\pi_1); T_{kj}(\pi_2)\big) \\
T_{ij}(\pi_1 \cup \pi_2) &= T_{ij}(\pi_1) \cup T_{ij}(\pi_2) \\
T_{ij}(\pi^*) &= K_{ijn}(\pi)
\end{aligned}$$

where $K_{ijn}(\pi)$ is given by Definition 33 and n denotes the number of events in E.

Definition 33: $K_{ijk}(\pi)$ path transformers

$K_{ijk}(\pi)$ is defined by recursion on k, as follows:

$$\begin{aligned}
K_{ij0}(\pi) &= \begin{cases} ?(p \vee \neg p) \cup T_{ij}(\pi) & \text{if } i = j \\ T_{ij}(\pi) & \text{otherwise} \end{cases} \\
K_{ij(k+1)}(\pi) &=
\end{aligned}$$

$$\begin{cases} (K_{kkk}(\pi))^* & \text{if } i = k = j \\ (K_{kkk}(\pi))^* ; K_{kjk}(\pi) & \text{if } i = k \neq j \\ K_{ikk}(\pi); (K_{kkk}(\pi))^* & \text{if } i \neq k = j \\ K_{ijk}(\pi) \cup \big(K_{ikk}(\pi); (K_{kkk}(\pi))^* ; K_{kjk}(\pi)\big) & \text{otherwise } (i \neq k \neq j) \end{cases}$$

As usual, both axioms and inference rules are to be understood as schemata in which the formulas above stand for any well-formed formulas. The reduction axioms show that the static language PE-PDL is rich enough to pre-encode the dynamic language PLCC.

Theorem 13: Soundness

The proof system $\mathcal{L}_{PLCC}$ is sound with respect to Bayesian Kripke structures if and only if, for $\phi \in \mathcal{L}_{PLCC}$:

$$\vdash \phi \text{ implies } \models \phi$$

Induction on the length of the proof. It is sufficient to prove that every axiom is sound and each inference rule preserves truth. This is a routine proof, so we will only check the soundness of the most difficult reduction axiom:

$$A \leftrightarrow \Big(pre(e) \to (C \geq D)\Big)$$

where the letters stand for:

$$A = [e]\left(\sum_{1\leq h\leq k} \alpha_h \cdot P_a(\psi_h) \geq \beta\right)$$

$$C = \sum_{\substack{1\leq h\leq k\\ \phi_i\in\Phi\\ f\sim_a e}} \alpha_h \cdot \mathrm{pre}_a(f \mid \phi_i) \cdot P_a(\phi_i \wedge [f]\psi_h)$$

$$D = \sum_{\substack{\phi_i\in\Phi\\ f\sim_a e}} \beta \cdot \mathrm{pre}_a(f \mid \phi_i) \cdot P_a(\phi_i)$$

Take an arbitrary Bayesian Kripke model M and a state s, such that:

$$M, s \models [e]\left(\sum_{1\leq h\leq k} \alpha_h \cdot P_a(\psi_h) \geq \beta\right)$$

$$\overbrace{\Leftrightarrow}^{\text{by semantic definition}}$$

$$\sum_{1\leq h\leq k} \alpha_h \sum_{\substack{(s',e')\sim_a(s,e)\\ M\otimes A,(s',e')\models\psi_h}} \mu_a^{(s,e)}(s',e') \geq \beta$$

$$\overbrace{\Leftrightarrow}^{\text{by the product update rule}}$$

$$\sum_{1\leq h\leq k} \alpha_h \frac{\sum_{\substack{(s',e')\sim_a(s,e)\\ M\otimes A,(s',e')\models\psi_h}} \mu_a^s(s')pre(e'|s')}{\sum_{(w,f)\sim_a(s,e)} \mu_a^s(w)pre(f|w)} \geq \beta$$

$$\sum_{1\leq h\leq k} \alpha_h \sum_{\substack{(s',e')\sim_a(s,e)\\ M\otimes A,(s',e')\models\psi_h}} \mu_a^{(s,e)}(s',e') \geq \beta$$

$$\overbrace{\Leftrightarrow}^{\text{by the product update rule}}$$

$$\sum_{1\leq h\leq k} \alpha_h \frac{\sum_{\substack{(s',e')\sim_a(s,e)\\ M\otimes A,(s',e')\models\psi_h}} \mu_a^s(s')pre(e'|s')}{\sum_{(w,f)\sim_a(s,e)} \mu_a^s(w)pre(f|w)} \geq \beta$$

$$\overbrace{\Leftrightarrow}^{\text{re-aranging the terms}}$$

$$\sum_{1\leq h\leq k} \alpha_h \sum_{\substack{(s',e')\sim_a(s,e)\\ M\otimes A,(s',e')\models\psi_h}} \mu_a^s(s')pre(e'|s') \geq \beta \sum_{(w,f)\sim_a(s,e)} \mu_a^s(w)pre(f|w)$$

$$\overset{\text{by grouping worlds according to the preconditions they satisfy, for every } f \sim_a e}{\overbrace{\Leftrightarrow}}$$

$$\sum_{1\leq h\leq k} \alpha_h \sum_{\substack{f\sim_a e\\ \phi_i\in\Phi}} pre_a(f|\phi_i) \sum_{\substack{s'\sim_a s\\ s'\models[f]\psi_h\\ s'\models\phi_i}} \mu_a^s(s') \geq \beta \sum_{\substack{f\sim_a e\\ \phi_i\in\Phi}} pre_a(f|\phi_i) \sum_{w\models\phi_i} \mu_a^s(w)$$

$$\overset{\text{by semantic definition}}{\overbrace{\Leftrightarrow}}$$

$$M,s \models \sum_{1\leq h\leq k} \alpha_h \sum_{\substack{f\sim_a e\\ \phi_i\in\Phi}} pre_a(f|\phi_i)\cdot P_a(\phi_i \wedge [f]\psi_h) \geq \beta \sum_{\substack{f\sim_a e\\ \phi_i\in\Phi}} pre_a(f|\phi_i)P_a(\phi_i)$$

$$\overset{\text{re-grouping the sums}}{\overbrace{\Leftrightarrow}}$$

$$M,s \models \sum_{\substack{1\leq h\leq k\\ f\sim_a e\\ \phi_i\in\Phi}} \alpha_h pre_a(f|\phi_i)\cdot P_a(\phi_i \wedge [f]\psi_h) \geq \sum_{\substack{f\sim_a e\\ \phi_i\in\Phi}} \beta pre_a(f|\phi_i)P_a(\phi_i)$$

Reintroducing the notation for terms, we obtained that:

$$M,s \models \big(pre(e) \to (C \geq D)\big)$$

and thus proved that the reduction axiom is sound.

□

5.2 COMPLETENESS OF PLCC

We first prove the completeness of the static language PE-PDL, which we call PE-PDL, and then argue, via the reduction axioms, that every $\mathcal{L}_{\mathrm{PLCC}}$-formula can be translated into a $\mathcal{L}_{\mathrm{PLCC}}$-formula.

5.2.1 *Completeness of the static language PE-PDL*

The following definitions have been adapted from Blackburn et al. (2001) to include the test operator.

Definition 34: Fischer-Ladner closure

Let X be a set of formulas. Then X is *Fischer-Ladner closed* if it is closed under subformulas and satisfies the following additional constraints:

(i) If $[\pi_1;\pi_2]\phi \in X$ then $[\pi_1][\pi_2]\phi \in X$

(ii) If $[\pi_1 \cup \pi_2]\phi \in X$ then $[\pi_1]\phi \wedge [\pi_2]\phi \in X$

(iii) If $[\pi^*]\phi \in X$ then $[\pi][\pi^*]\phi \in X$

(iv) If $[\phi?]\psi \in X$ then $\phi \to \psi \in X$.

(v) If $w \in X$ then $[a]w \in X$, for w an a-probability formula.

If Σ is any set of formulas then $\text{FL}(\Sigma)$ (the *Fischer-Ladner closure* of Σ) is the smallest set of formulas containing Σ that is Fischer-Ladner closed.

Given a formula ϕ, we define $\sim \phi$ as the following formula

$$\sim \phi = \begin{cases} \psi & \text{if } \phi \text{ is of the form } \neg\psi \\ \neg\phi & \text{otherwise.} \end{cases}$$

A set of formulas X is *closed under single negations* if $\sim \phi$ belongs to X whenever $\phi \in X$.

We define $\neg\text{FL}(\Sigma)$, the *closure of* Σ, as the smallest set containing Σ which is Fischer-Ladner closed and closed under single negations.

Definition 35: Atoms

Let Σ be a set of formulas. A set of formulas A is an *atom* over Σ if it is a maximal consistent subset of $\neg\text{FL}(\Sigma)$. That is, A is an atom over Σ if $A \subseteq \neg\text{FL}(\Sigma)$, if A is consistent, and if $A \subset B \subseteq \neg\text{FL}(\Sigma)$ then B is inconsistent. $At(\Sigma)$ is the set of all atoms over Σ.

The following lemma (except for item 6) closely follows Blackburn et al. (2001, lemma 4.81).

Lemma 2

Let Σ be any set of formulas, and A any element of $At(\Sigma)$. Then

1. For all $\phi \in \neg\text{FL}(\Sigma)$: exactly one of ϕ and $\sim \phi$ is in A
2. For all $\phi \vee \psi \in \neg\text{FL}(\Sigma)$: $\phi \vee \psi \in A$ iff $\phi \in A$ or $\psi \in A$
3. For all $\langle \pi_1; \pi_2 \rangle\phi \in \neg\text{FL}(\Sigma)$: $\langle \pi_1; \pi_2 \rangle\phi \in A$ iff $\langle \pi_1 \rangle\langle \pi_2 \rangle\phi \in A$
4. For all $\langle \pi_1 \cup \pi_2 \rangle\phi \in \neg\text{FL}(\Sigma)$: $\langle \pi_1 \cup \pi_2 \rangle\phi \in A$ iff $\langle \pi_1 \rangle\phi \in A$ or $\langle \pi_2 \rangle\phi \in A$
5. For all $\langle \pi^* \rangle\phi \in \neg\text{FL}(\Sigma)$: $\langle \pi^* \rangle\phi \in A$ iff $\phi \in A$ or $\langle \pi \rangle\langle \pi^* \rangle\phi \in A$
6. For all $[\phi?]\psi \in \neg FL(\Sigma) : [\phi?]\psi \in A$ iff $\phi \to \psi$.

Now it is time to define the canonical model over Σ.

Definition 36: Canonical model over Σ

Let Σ be a finite set of formulas. The canonical model over Σ is the triple $(At(\Sigma), \{S_\pi^\Sigma\}_{\pi\in\Pi}, V^\Sigma)$ where for all the propositional variables p, $V^\Sigma(p) = \{A \in At(\Sigma) \mid p \in A\}$ and for all atoms $A, B \in At(\Sigma)$ and all programs π,

$$AS_\pi B \text{ if } \phi_A S_\pi[\pi]\phi_B \text{ is consistent.}$$

where ϕ_A is defined as the conjunction of all formulas that belong to A.

Definition 37: Regular model over Σ

Let Σ be a finite set of formulas. For all basic programs a, define R_a^Σ as:

$$AR_aB \text{ if } \forall\phi \in \neg FL(\Sigma), [\pi]\phi \in A \text{ iff } [\pi]\phi \in B$$

For the complex programs, inductively define the PDL relations such that $\forall\pi, \pi_1, \pi_2$, we have:

$$\begin{aligned} R_{\pi_1\cup\pi_2} &= R_{\pi_1} \cup R_{\pi_2} \\ R_{\pi_1;\pi_2} &= R_{\pi_1}; R_{\pi_2} \\ R_{\pi^*} &= (R_\pi)^* \\ AR_{\phi?}B \Leftrightarrow A &= B \text{ and } \phi \in A \end{aligned}$$

Finally, define $\mathcal{R}$, the regular model over Σ, to be

$$\mathcal{R} = (At(\Sigma), \{R_\pi^\Sigma\}_{\pi\in\Pi}, V^\Sigma)$$

where V^Σ is the canonical valuation.

Given the regular model $\mathcal{R} = (At(\Sigma), \{R_\pi^\Sigma\}_{\pi\in\Pi}, V^\Sigma)$, our goal is to define a probability assignment

$$\mu^\Sigma : Ag \to \Big(At\big(\Sigma\big) \to (At(\Sigma) \to [0,1])\Big)$$

s.t. if we consider the Bayesian Kripke structure

$$\mathcal{M} = (At(\Sigma), \{R_\pi^\Sigma\}_{\pi\in\Pi}, \mu^\Sigma, V^\Sigma)$$

then for every state $A \in At(\Sigma)$ and every $\psi \in \neg\mathrm{FL}(\Sigma)$ we have $(\mathcal{M}, A) \models \psi$ iff $\psi \in A$.

Lemma 3

For any $a \in Ag$, there exists a probability function μ_i, such that all a-probability formulas $w \in \neg FL(\Sigma)$ can be true together.

Proof. Using only propositional reasoning, we can show that:

$$\vdash \psi \leftrightarrow \bigvee_{\{A \in At(\Sigma) | \psi \in A\}} \phi_A, \text{ for all } \psi \in \neg \mathrm{FL}(\Sigma) \tag{5.1}$$

$$\vdash \phi_A \rightarrow \neg \phi_B, \text{ for any } A, B \in At(\Sigma), A \neq B \tag{5.2}$$

Using these observations and Axioms W1-W5, we can show that

$$P_a(\psi) = \sum_{\{A \in At(\Sigma) | \psi \in A\}} P_a(\phi_A)$$

is provable in PE-PDL. Using this fact, together with I1 and I3, we can show that an a-probability formula $\psi \in \neg \mathrm{FL}(\Sigma)$ is provably equivalent to a formula of the form

$$\sum_{A \in At(\Sigma)} c_A P_a(\phi_A) \geq b$$

for some appropriate coefficients c_A. Let $\mathcal{K}_a(A)$ be the set of atoms B such that $\nvdash \phi_A \rightarrow P_a(\phi_B) = 0$. Now, in the presence of W7, we can show that if $A \in At(\Sigma)$ and $A' \notin \mathcal{K}_i(A)$, then

$$\vdash \phi_A \rightarrow (\mu_i(\phi_{A'}) = 0)$$

Now, fix an agent a and a state $A \in At(\Sigma)$. We describe a set of linear equalities and inequalities corresponding to a and s, over variables of the form $x_{iAA'}$, for $A' \in At(\Sigma)$. We can think of $x_{aAA'}$ as representing $\mu_a^A(A')$, that is the probability of state A' under agent a's probability distribution at state A. We have one inequality corresponding to every a-probability formula $\psi \in \neg \mathrm{FL}(\Sigma)$. Assume that ψ is equivalent to

$$\sum_{A' \in At(\Sigma)} c_{A'} P_a(\phi_{A'}) \geq b$$

Notice that exactly one of ψ and $\neg\psi$ is in A. If $\psi \in A$, then the corresponding inequality is

$$\sum_{A' \in At(\Sigma)} c_{A'} x_{aAA'} \geq b$$

If $\neg\psi \in A$, then the corresponding inequality is

$$\sum_{A' \in At(\Sigma)} c_{A'} x_{aAA'} < b$$

Further, due to W7, we have the following equalities:

$$x_{aAA'} = 0$$

for $A' \notin \mathcal{K}_a(A)$ and

$$x_{aAA'} > 0$$

for $A' \in A \setminus \mathcal{K}_a(A)$. Finally, we have the equality

$$\sum_{A' \in At(\Sigma)} x_{aAA'} = 1$$

As shown in Fagin et al. (1990, Thm. 2.2), since ϕ_A is consistent, this set of linear equalities and inequalities has a solution $x^*_{a,A,A'}$, for $A' \in At(\Sigma)$. Set $\mu_{a,A}(A) = x^*_{a,A,A'}$. This is the probability assignment μ that we are looking for. Before we proceed to the truth lemma, we only need to make sure that our model $\mathcal{M}$, thus constructed, satisfies the SDP condition, corresponding to the introduction of Axiom W10 in the logic. This can be easily checked by inspecting of the definition of R_a. Given this, we can assume, without loss of generality, that if AR_aA' then $\mu_{a,A} = \mu_{a,A'}$, since we have that the definition of $\mu_{a,A}$ depends only on the a-probability formulas and their negations at state A. □

Before we prove the truth lemma, we need to establish two important results: an existence lemma for S_π and a theorem which states that $S_\pi \subseteq R_\pi$.

> *Lemma 4: The Existence Lemma for S_π*
>
> Let A be an atom and let $\langle\pi\rangle\phi$ be a formula in $\neg FL(\Sigma)$. Then $\langle\pi\rangle\phi \in A$ iff there is a B such that $AS_\pi B$ and $\phi \in B$.

Proof. Following the strategy laid out in Blackburn et al. (2001, p. 244), we set out to construct an appropriate atom B by forcing choices. We begin by enumerating the formulas in the finite set $FL(\Sigma)$ as $\sigma_1, \ldots, \sigma_m$ and define B_0 to be $\{\phi\}$. Suppose as an inductive hypothesis that B_n is defined such that $\phi_A \wedge \langle\pi\rangle\phi_{B_n}$ is consistent (where $1 \leq n \leq m$). We get that

$$\vdash \langle\pi\rangle\phi_B \leftrightarrow \langle\pi\rangle\left((\phi_B \wedge \sigma_{n+1}) \vee (\phi_B \wedge \neg\sigma_{n+1})\right)$$

and thus

$$\vdash \langle\pi\rangle\phi_B \leftrightarrow (\langle\pi\rangle(\phi_B \wedge \sigma_{n+1})) \vee (\langle\pi\rangle(\phi_B \wedge \neg\sigma_{n+1}))$$

Therefore, either for $B' = B \cup \{\sigma_{n+1}\}$ or for $B' = B \cup \{\neg\sigma_{n+1}\}$, we have that $\phi_A \wedge \langle\phi_{B'}\rangle$ is consistent. Choose B_{n+1} to be this consistent expansion, and let B_m be B. Then B is the atom we want. □

Lemma 5: Lemma for basic programs

For all programs $a \in Ag$, $S_a \subseteq R_a$.

Proof. We need to show that, if AS_aB, then AR_aB, for all $A, B \in At(\Sigma)$. We begin by noting that, since $\phi_A \wedge \langle a \rangle \phi_B$ is consistent, then there exists a maximally consistent set (MCS) Γ such that $\phi_A \wedge \langle \phi_B \rangle \in \Gamma$. Note that A is the maximal consistent subset of $\neg FL(\Sigma)$ that extends to Γ: $A = \Gamma \cap \neg FL(\Sigma)$. Since $\phi_A \wedge \langle a \rangle \phi_B \in \Gamma$ then $\langle a \rangle \phi_B \in \Gamma$ too. So, there exists a Δ, a maximally consistent set, such that $\Gamma \sim_a \Delta$, where $\sim_a$ is the canonical relation, defined by $A \sim_a B$ iff for all formulas ϕ, $\phi \in A$ implies $\langle a \rangle \phi \in B$. Let $B = \Delta \cap \neg FL(\Sigma)$. Then, we have that $A \sim_a B$. We can show that, by the standard results on canonical models, we have that if the logic includes the S5 axioms, then $\sim_a$ is an equivalence relation.

We prove the following claim:

$$T \sim_a U \leftrightarrow \forall \phi, w \ \langle a \rangle \phi \in T \text{ iff } \langle a \rangle \phi \in U$$

Proof.

"⇒" Suppose $T \sim_a U$ and $\langle a \rangle \phi \in T$. Then, by the definition of $\sim_a$, we have that $\langle a \rangle \langle a \rangle \phi \in U$. By Axiom $K4$, we have that $\langle a \rangle \phi \in U$. The other direction follows from the symmetry of $\sim_a$.

"⇐" Suppose that $\forall \phi, \langle a \rangle \phi \in T$ iff $\langle a \rangle \phi \in U$. Let $\psi \in U$. We need to show that $\langle a \rangle \psi \in T$. From $\psi \in U$ and Axiom K3, we then have that $\langle a \rangle \phi \in T$.

□

Therefore, we proved then that if AS_aB then AR_aB. □

Lemma 6

If Σ is finite, then $\neg FL(\Sigma)$ is finite.

Proof. We skip the proof of this theorem, as it is a straightforward proof by induction. □

Lemma 7

For all programs π, we have that $S_{\pi^*} \subseteq (S_\pi)^*$.

Proof. Identical to the proof of Lemma 4.87 in (Blackburn et al., 2001, p. 244). □

Theorem 14

For all programs π, $S_\pi \subseteq R_\pi$.

Proof by induction on the complexity of π.

BASE CASE: $\pi = a$ is given by the Lemma for basic programs.

INDUCTIVE HYPOTHESIS: Assume the claim holds for all programs of complexity lower than π. Now we try to show it for π

Case 1: π is of the form $\pi_1; \pi_2$. Suppose $AS_{\pi_1;\pi_2}B$, that is, $\phi_A \wedge \langle \pi_1; \pi_2 \rangle \phi_B$ is consistent. It follows, by Axiom 3, that $\phi_A \wedge \langle \pi_1 \rangle \langle \pi_2 \rangle \phi_B$ is consistent. By the IH, we get that $AR_{\pi_1}C$ and $CR_{\pi_2}B$. It follows immediately that $AR_{\pi_1;\pi_2}B$.

Case 2: π is of the form $\pi_1 \cup \pi_2$. Similar to Case 1. Omitted here.

Case 3: π is of the form π^*. Suppose $AS_{\pi^*}B$, that is, $\phi_A \wedge \langle \pi^* \rangle \phi_B$ is consistent. Since $S_{\pi^*} \subseteq (S_\pi)^*$, we get that there exists a chain $A = C_0 S_\pi C_1 \ldots C_k = B$, such that, for every pair $C_i C_{i+1}$, by the IH, if $C_i S_\pi C_{i+1}$ then $C_i R_\pi C_{i+1}$. But then $AR_{\pi^*}B$.

Case 4: π is of the form $\phi?$. Assume $\phi \in \neg FL(\Sigma)$. Suppose $AS_{\phi?}B$. Then $\phi_A \wedge [\phi?]\phi_B$ is consistent. From Axiom 7, using propositional reasoning, we get that $\langle \phi? \rangle \psi \leftrightarrow (\phi \wedge \psi)$. It follows that

$$\phi_A \wedge (\phi \wedge \psi_B) \text{ is consistent}$$

However, it's easy to notice that for any two atoms are mutually exclusive, therefore $\vdash \phi_A \rightarrow \neg \phi_B \forall A \neq B$. We can conclude then that $A = B$. Finally, since $\phi_A \wedge \phi$ is consistent and $\phi \in \neg FL(\Sigma)$, we conclude that $\phi \in A$.

□

Before we prove the truth lemma, we need to establish an existence lemma as follows:

Lemma 8: Existence Lemma

Let A and B be atoms in $At(\Sigma)$ and let $[\pi]\phi \in \neg FL(\Sigma)$. Then if $[\pi]\psi \in A$ and $AR_\pi B$ then $\phi \in B$.

Proof: induction on the complexity of π.

BASE CASE: π is a basic program a

We need to show that if AR_aB and $[a]\phi \in A$, then $[a]\phi \in B$. It immediately follows that if $[a]\phi \in A$ and AR_aB then, by the definition of R_a we get $[a]\phi \in B$. By this and the transitivity axiom $[a]\phi \to \phi$, it follows that $\phi \in B$.

INDUCTIVE STEP: Assume the claim holds for all π of a certain complexity and lower.

Case 1: ϕ is of the form $\pi_1 \cup \pi_2$. We have that $AR_{\pi_1 \cup \pi_2}B$ and $[\pi_1 \cup \pi_2]\phi \in A$. By Ax. (iv), we have that $[\pi_1]\phi, [\pi_2]\phi \in A$. Since $AR_{\pi_1 \cup \pi_2}B$ then we have that either $AR_{\pi_1}B$ or $AR_{\pi_2}B$. Applying the IH, we get that in either case $\phi \in B$.

Case 2: π is of the form $\pi_1; \pi_2$. Similar to Case 1. Will omit here.

Case 3: π is of the form $\phi?$. We have that $[\phi?]\psi \in A$ and $AR_{\phi?}B$ and we need to show that $\psi \in B$. By Axiom 7, then $\phi \to \psi \in A$. Further, from $AR_{\phi?}B$, we get that $A = B$ and $\phi \in A$. By an application of modus ponens, we get that $\psi \in A$.

Case 4: π is of the form $[\pi^*]$. In order to prove this, it suffices to show that:

> *Theorem 15*
>
> $\forall\phi$ such that $[\pi^*] \in \neg FL(\Sigma)$, if $[\pi^*]\phi \in A$ and $AR_{\pi^*}B$ then $[\pi^*]\phi \in B$.

Proof by induction on the length of the path k from A to B: $A = C_0 R_\pi C_1 \ldots R_\pi C_k = B$.

Proof.

BASE CASE: the length of the path $k = 1$. We know that $AR_\pi B$ and $[\pi^*]\phi \in B$. By Axiom 5 and clause 5 of the FL closure, we get that $[\pi][\pi^*]\phi \in A$. Applying the $IH_{\exists Lemma}$, we get that $[\pi^*]\phi \in B$.

INDUCTIVE HYPOTHESIS: Assume the claim holds for all lengths lower than k and try to prove it for k.

We have that $[\pi^*]\phi \in A$ and

$$A = C_0 R_\pi C_1 \ldots C_{k-1} R_\pi C_k = B$$

By the IH, given that $AR_{\pi^*}C_{k-1}$, we have that $[\pi^*]\phi \in C_{k-1}$. By Axiom 5 and clause 5 of the FL closure, $[\pi][\pi^*]\phi \in C_{k-1}$. Since we also have that $C_{k-1}R_\pi B$, by the $IH_{\exists Lemma}$, we get have $[\pi^*]\phi \in B$.

□

□

Theorem 16: Truth Lemma

Let $\mathcal{R}$ be a regular PE-PDL model over Σ. For all atoms A and all $\phi \in \neg\text{FL}(\Sigma)$, $\mathcal{R}, A \models \phi$ iff $\phi \in A$.

Proof: Induction on the number of connectives.

BASE CASE Follows immediately from the definition of the canonical valuation over Σ.

INDUCTIVE STEP

CASE 1: The Boolean case.

It follows immediately from Lemma 2 above, and from Lemma 3, in case of a-probability formulas w.

CASE 2: ϕ is of the form $[a]\psi$

"$\Rightarrow$" $\mathcal{R}, A \models [a]\psi$ then $[a]\psi \in A$.

Proof. Define $\Delta = \{[a]\chi | [a]\chi \in A\} \cup \{\neg[a]\chi | \neg[a]\chi \in A\}$. Then $\Delta \cup \{\neg\psi\}$ is inconsistent. For suppose otherwise. Then $\Delta \cup \{\neg\psi\}$ could be expanded to a MCS $B \in At(\Sigma)$. We have that by construction AR_aB. If $\neg\psi \in B$ then by the IH I get that $B \models \neg\psi$. Since $\mathcal{R}, A \models [a]\psi$ and $AR_aB \Rightarrow B \models \psi$. Contradiction. Then $\vdash \phi_\Delta \rightarrow \psi$. By $R_2 \Rightarrow \vdash [a](\phi_\Delta \rightarrow \psi)$ and by Ax. 4 and Ax. 5 we get that $\phi_\Delta \rightarrow [a]\phi_\Delta$. By $K_2 \Rightarrow \vdash \phi_\Delta \rightarrow [a]\psi$. This, together with the fact that $[a]\psi \in \neg\text{FL}(\Sigma)$ and the fact that for $\forall\phi \in \text{FL}(\Sigma)$, either ϕ or its negation is in $A \Rightarrow [a]\psi \in A$. □

"$\Leftarrow$" if $[a]\psi \in A$ then $\mathcal{R}, A \models [a]\psi$.

Proof. Consider $B \in At(\Sigma)$ s.t. AR_aB. Then $[a]\phi \in A \Leftrightarrow [a]\phi \in B$. This together with the assumption imply that $[a]\psi \in B$. By Ax. K3 we know that $[a]\psi \rightarrow \psi$ and since $[a]\psi \in \neg\text{FL}(\Sigma)$ then $\psi \in \neg\text{FL}(\Sigma)$, we get that $\psi \in B$. By the IH $\Rightarrow B \models \psi$. This holds for any B s.t. AR_aB. $\Rightarrow A \models [a]\psi$. □

CASE 3: $\phi = [\pi_1 \cup \pi_2]\psi$

"$\Rightarrow$" if $\mathcal{R}, A \models [\pi_1 \cup \pi_2]\psi$ then $[\pi_1 \cup \pi_2]\psi \in A$.

Proof. $A \models [\pi_1 \cup \pi_2]\psi \Rightarrow \forall B$ s.t. $AR_{\pi_1}B$ or $AR_{\pi_2}B$ then $B \models \psi \Rightarrow A \models [\pi_1]\psi$ and $A \models [\pi_2]\psi$. By IH we get that $[\pi_1]\psi, [\pi_2]\psi \in A$. By Ax. 4, we have that $[\pi_1]\psi \wedge [\pi_2]\psi \Leftrightarrow [\pi_1 \cup \pi_2]\psi$. Since $[\pi_1 \cup \pi_2]\psi \in \neg\mathrm{FL}(\Sigma) \Rightarrow [\pi_1 \cup \pi_2]\psi \in A$. □

"$\Leftarrow$" if $[\pi_1 \cup \pi_2]\psi \in A$ then $A \models [\pi_1 \cup \pi_2]\psi$.

Proof. Consider $B \in At(\Sigma)$ s.t. $AR_{\pi_1 \cup \pi_2}B$. This means that $AR_{\pi_1}B$ or $AR_{\pi_2}B$. Now $[\pi_1 \cup \pi_2]\psi \in A$. By Ax. 4 $[\pi_1 \cup \pi_2]\psi \leftrightarrow [\pi_1]\psi \wedge [\pi_2]\psi \Rightarrow [\pi_1]\psi, [\pi_2]\psi \in A \Rightarrow$ by IH $A \models [\pi_1]\psi \wedge [\pi_2]\psi$. Since B is s.t. $AR_{\pi_1}B$ or $AR_{\pi_2}B$ then $B \models \phi \Rightarrow A \models [\pi_1 \cup \pi_2]\psi$ □

CASE 4: ϕ is of the form $[\pi_1; \pi_2]\psi$

"$\Rightarrow$" $\mathcal{R}, A \models [\pi_1; \pi_2]\psi$ then $[\pi_1; \pi_2]\psi \in A$

Proof. From $\mathcal{R}, A \models [\pi_1; \pi_2]\psi$ we get that for $\forall C, \forall B \in At(\Sigma)$ s.t. $AR_{\pi_1}C$ and $CR_{\pi_2}B$, then $B \models \psi$. It follows that $C \models [\pi_2]\psi$ for any C s.t. $AR_{\pi_1}C \Rightarrow A \models [\pi_1][\pi_2]\psi$. By IH $\Rightarrow [\pi_1][\pi_2]\psi \in A$. By Ax. 3 $\Rightarrow [\pi_1; \pi_2]\psi \in A$, since $[\pi_1; \pi_2]\psi \in \neg\mathrm{FL}(\Sigma)$. □

"$\Leftarrow$" if $[\pi_1; \pi_2]\psi \in A \Rightarrow A \models [\pi_1; \pi_2]\psi$

Proof. We need to show that if $[\pi_1; \pi_2]\psi \in A$ then for $\forall C, B \in At(\Sigma)$ s.t. $AR_{\pi_1}C$ and $CR_{\pi_2}B$ then $B \models \psi$. By Ax. 3 and by clause 1, if $[\pi_1; \pi_2]\psi \in A \Rightarrow [\pi_1][\pi_2]\psi \in A$. By IH $\Rightarrow A \models [\pi_1][\pi_2]\psi \Rightarrow$ for $\forall C$ s.t. $AR_{\pi_1}C$ we have $\mathcal{R}, C \models [\pi_2]\psi \Rightarrow \forall B$ s.t. $CR_{\pi_2}B \Rightarrow B \models \psi \Rightarrow A \models [\pi_1; \pi_2]\psi$ □

CASE 5: ϕ is of the form $\phi?$

"$\Rightarrow$" Assume $A \models [\phi?]\psi$. We need to show that $[\phi?]\psi \in A$. By Axiom 6, $A \models \phi \rightarrow \psi$. By the IH, we get that $\phi \rightarrow \psi \in A$. By Axiom 7, $[\phi?]\psi \in A$.

"$\Leftarrow$"Assume $[\phi?]\psi \in A$. Show that $A \models [\phi?]\psi$. By Axiom 7, we get that $\phi \rightarrow \psi \in A$. By the IH, we have that $A \models \phi \rightarrow \psi$. By Axiom 7,we get that $A \models [\phi?]\psi$.

CASE 6: ϕ is of the form $[\pi^*]\psi$

"$\Leftarrow$" if $[\pi^*]\psi \in A \Rightarrow A \models [\pi^*]\psi$.

From the Existence Lemma and an application of Axiom 5, it follows immediately that $\forall \phi s.t.[\pi]\phi \in \neg FL(\Sigma)$, if $AR_{\pi^*}B$ and $[\pi^*]\phi \in A$ then $\phi \in B$.

"$\Rightarrow$" if $A \models [\pi^*]\psi \Rightarrow [\pi^*]\psi \in A$.

Proof. By contraposition, we need to prove that if $\neg[\pi^*]\phi \in A$ then $A \not\models [\pi^*]\psi$. By the Existence Lemma for S_{i^*}, we have that if $\neg[\pi^*]\phi \in A$ then $\exists B \in At(\Sigma)$ such that $AS_{\pi^*}B$, then $\neg\phi \in B$. But we have shown that $S_{\pi^*} \subseteq R_{\pi^*}$, therefore we have that $AR_{\pi^*}B$. By the $IH_{\text{Truth Lemma}}$ and the fact that $\neg\phi \in B$ we have that $B \models \neg\phi$. Therefore $A \not\models [\pi^*]\phi$. □

CASE 7 : ϕ is of the form $\sum_{j=1}^{k} \alpha_j P_a(\phi_j) \geq \beta$. By Lemma 3, $\sum_{j=1}^{k} \alpha_j P_a(\phi_j) \geq \beta \in A$ if and only if $\sum_{j=1}^{k} \sum_{B \in At, \vdash B \to \phi_j} \alpha_j \mu_{a,A}(B) \geq \beta$ if and only if $\mathcal{R}, A \models \sum_{j=1}^{k} \sum_{B \in At, \vdash B \to \phi_j} \alpha_j \mu_{a,A}(B) \geq \beta$ if and only if $\mathcal{R}, A \models \sum_{j=1}^{k} \alpha_j P_a(\phi_j) \geq \beta$.

This concludes the proof of the Truth Lemma.

□

Theorem 17: Weak completeness of PE-PDL

PE-PDL is weakly complete with respect to the class of all Bayesian Kripke frames.

The reduction axioms presented in Section 5.1.3 determine a translation procedure, for reducing the $\mathcal{L}_{\text{PLCC}}$-formulas into $\mathcal{L}_{\text{PE-PDL}}$ formulas.

Definition 38: Translation

The function t takes a formula from the language of $\mathcal{L}_{\text{PLCC}}$ and yields a formula in the language of $\mathcal{L}_{\text{PE-PDL}}$.

$$
\begin{aligned}
t(T) &= T & r(a) &= a \\
t(p) &= p & r(B) &= B \\
t(\neg\phi) &= \neg t(\phi) & r(?\phi) &= ?t(\phi) \\
t(\phi_1 \wedge \phi_2) &= t(\phi_1) \wedge t(\phi_2) & r(\pi_1;\pi_2) &= r(\pi_1); r(\pi_2) \\
t([\pi]\phi) &= [r(\pi)]t(\phi) & r(\pi_1 \cup \pi_2) &= r(\pi_1) \cup r(\pi_2) \\
t([e]T) &= T & r(\pi^*) &= (r(\pi))^* \\
t([e]p) &= t(\text{pre}(e)) \to t(\text{sub}(e)(p)) \\
t([e]\neg\phi) &= t(\text{pre}(e)) \to \neg t([e]\phi) \\
t([e](\phi_1 \wedge \phi_2)) &= t([e]\phi_1) \wedge t([e]\phi_2) \\
t([e_i][\pi]\phi) &= \bigwedge_{j=0}^{n-1} [T_{ij}(r(\pi))]t([e_j]\phi) \\
t([e][e']\phi) &= t([e]t([e']\phi)) \\
t\left(\sum_{1\le h\le k} \alpha_h \cdot P_a(\psi_h) \ge \beta\right) &= \sum_{1\le h\le k} \alpha_h \cdot P_a(t(\psi_h)) \ge \beta
\end{aligned}
$$

$$t(A) = t(pre(e)) \to t(C \ge D)$$

where the letters in the last line stand for

$$
\begin{aligned}
A &= [e]\left(\sum_{1\le h\le k} \alpha_h \cdot P_a(\psi_h) \ge \beta\right) \\
C &= \sum_{\substack{1\le h\le k \\ \phi_i\in\Phi \\ f\sim_a e}} \alpha_h \cdot \text{pre}_a(f \mid \phi_i) \cdot P_a(t(\phi_i \wedge [f]\psi_h)) \\
D &= \sum_{\substack{\phi_i\in\Phi \\ f\sim_a e}} \beta \cdot \text{pre}_a(f \mid \phi_i) \cdot P_a(t(\phi_i))
\end{aligned}
$$

Finally, we have that:

Theorem 18: Completeness of PLCC

For any ϕ, a formula of the language $\mathcal{L}_{\text{PLCC}}$, we have that:

$$\models \phi \text{ iff } \vdash \phi$$

Proof. Given the completeness of the static language PE-PDL $\mathcal{L}_{\text{PE-PDL}}$, and the translation procedure above, which ensures every formula in the language of PLCC is equivalent to a formula in the language of PEPDL, the result follows immediately. □

In conclusion, Chapter 5 developed the probabilistic version of the Logic of Communication and Change, a powerful epistemic logic that we believe could become well-suited for modelling social dynamics of information games. PLCC aspires to becoming a candidate for the standard logic for imperfect information games for the reason that it can express diverse notions of group knowledge, especially common knowledge alongside probabilistic statements that are revised by the principle of Bayesian conditionalization. The main result of this chapter is the proof that PLCC is a sound and weakly complete logic with respect to Bayesian Kripke structures.

6

THE LOGIC APPROACH TO CASCADES

This section presents a logical formulation of informational cascades, both semantically and syntactically. The semantic approach will dispel the argument according to which the Bayesian analysis of the cascade (see Chapter 2) fails to take into consideration the higher order beliefs of players. These beliefs encode the agents' reflection on the overall game and on the possibility of a cascade formation. On the other hand, statements that describe informational cascades are quite naturally expressed in a logical syntax. More generally, the syntactic approach can spot hidden assumptions behind ordinary economic or game-theoretic reasoning, and the axioms and inference rules can also help to analyse such assumptions.

Using Bayesian Kripke structures, we present a semantic treatment of cascades. This construction has several advantages. The first is that it represents a quantitative belief system that tracks each agent's observational learning, as the game unfolds. The use of probabilistic epistemic structures allows the incorporation of all levels of an agents' beliefs and knowledge about the current state of the world into the model. This includes taking into account all the players' meta-considerations regarding the game. Moreover, it allows for a clear delineation between the knowledge of an agent before and after an action has been observed. This brings forth a transparent analysis of the consequences information has on agents' knowledge and beliefs. Logical approaches to cascades have been considered in the literature before Baltag et al. (2013), Rendsvig (2013), in the framework of Dynamic Epistemic Logic. Our presentation will be carried out in the Bayesian Kripke semantics for Probabilistic Logic of Communication and Change introduced in Chapter 5.

6.1 THE SEMANTIC PROOF OF THE CASCADE

This section presents a PLCC semantic proof of the cascade formation in the Basic Urn Game. The strategy consists of simulating a play of the game where the first two players receive a low precision signal, and show, using Bayesian Kripke models, that under common knowledge of rationality, every subsequent player $i \geq 3$ will follow the first two, no matter their own private signal.

Consider a set of agents $Ag = \{1, \ldots, n\}$ together with 0 denoting Nature, and a set of propositional atoms

$$At = \{\bar{W}, \bar{B}\} \cup \{\bar{W}_i, \bar{B}_i\}_{i \in Ag} \cup \{\bar{w}_i, \bar{b}_i\}_{i \in Ag}$$

These propositional atoms allow us, through the mechanism of substitutions or postconditions, to track the progression of the game. The idea is to set all propositional atoms as false at all worlds in the initial state model M_0. Then, as agents take turns to play, the event models underlying this progression execute change of facts corresponding to the actions announced during the game.

A play of the game represents, in semantic notions, a succession of product updates of an initial Bayesian Kripke model, after event models that alternate between being private announcements of signals and public announcements of guesses. We will denote the events that model announcements of player i's private signal by A_{2i-1} and the events that model the guess of player i by A_{2i}. We encode the assumption of common knowledge of rationality into the modelling of the events, more precisely in the preconditions, that constrain agents to perform only "rational" moves. Essentially, in this setting common knowledge of rationality is a constraint on the event models.

We begin with a Bayesian Kripke model denoted by M_0, which represents the informational state of the agents, after Nature has picked the state of the world:

$$M_0 = (S_0, \sim_i, \mu_i, V)_{i \in Ag}$$

where:

- $S_0 = \{W, B\}$.
- $\sim_i$ is an equivalence relation on S for agent i: $W \sim_i B \; \forall i \in Ag$.
- $\mu_i : S_0 \to [0, 1]$ such that $\mu_i(s) = \frac{1}{2} \; \forall s \in S$.
- V assigns a set of states in S to each propositional variable in At as follows:
  - ⋆ $V(\bar{W}) = \{W\}$ and $V(\bar{B}) = \{B\}$
  - ⋆ $V(\bar{W}_i) = \emptyset$, $V(\bar{B}_i) = \emptyset$,, for all $i \in Ag$
  - ⋆ $V(\bar{w}_i) = \emptyset$, $V(\bar{b}_i) = \emptyset$,, for all $i \in Ag$

We graphically represent Bayesian Kripke models as follows: the worlds are depicted by rectangles, which contain the atoms true at that world. These atoms, put together in a sequence, label the world they're in. Worlds are connected by arrows, which represent equivalence relations. Every arrow is labelled by the agents who cannot distinguish between the worlds connected by the arrow. We do not denote reflexive arrows in our models, since they are always assumed to be there for any model represented. Given the SDP condition, we have that within every agent's information set, his probability assignments at each world are the same. Therefore, we represent only one probability assignment per world per agent, next to the name of the world. In model M_0 of Figure 6.1, the actual state B is represented by the bold-font rectangle, whereas the label on the arrows designates the agents that cannot distinguish between the two states of the world, W and B. The probabilities that each player assigns to the worlds are represented on the side of each rectangle, preceded by the players that hold these beliefs.

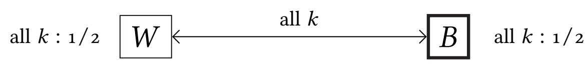


Figure 6.1: The initial state model M_0, after Nature picks the state of the world B. This action is not observable by any of the agents $k \in Ag$. All players give each world equal chances of being true.

Nature's pick of player 1's private signal

Next in the sequence of play, Nature reveals the private signal of player 1. This turn can be modelled using a probabilistic event model $A_1 = (E_1, \sim_i, PRE, pre_i, sub_1)_{i \in Ag}$ in which:

- $E_1 = \{w_1, b_1\}$
- $\sim_i$ such that $w_1 \sim_i b_1 \forall i \neq 1$
- $PRE(e) = \{\bar{W}, \bar{B}\}$ for all $e \in E_1$. Then $\Phi = \{\bar{W}, \bar{B}\}$.
- $pre_i \colon \Phi \to (E_1 \to [0,1])$ such that $pre_i(w_1|\bar{W}) = pre_i(b_1|\bar{B}) = \frac{2}{3}$ and $pre_i(b_1|\bar{W}) = pre_i(w_1|\bar{B}) = \frac{1}{3}$ for all $i \neq 0$
- $sub_1(w_1) = \sigma_w$ and $sub_1(b_1) = \sigma_b$, where $dom(\sigma_w) = \{\bar{w}_1\}$, $dom(\sigma_b) = \{\bar{b}_1\}$, and $\sigma_w(\bar{w}_1) = \neg\bar{w}_1$, $\sigma_b(\bar{b}_1) = \neg\bar{b}_1$

The event model A_1 is depicted graphically in Figure 6.2. We abuse notation by dropping the agent indexation of the *pre* function in the drawings because it is assumed that all players in the game share the same subjective beliefs regarding the probabilistic preconditions.

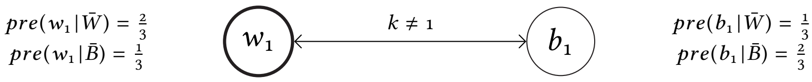


Figure 6.2: The event model A_1 represented by player 1 drawing a ball from the urn. Player 1 sees a w_1, however none of the other players can distinguish between the two events at this stage.

The result of updating M_0 with event A_1 is the new state model $M_1 = M_0 \otimes A_1 = (S_1, \sim_i, \mu_i, V)_{i \in Ag}$, represented in Figure 6.3. Observe that agent 1 knows she has drawn a white marble w_1, while not being able to discern the true urn B. All the other players in the game remain ignorant with regard to player 1's private draw, and can therefore exclude no world.

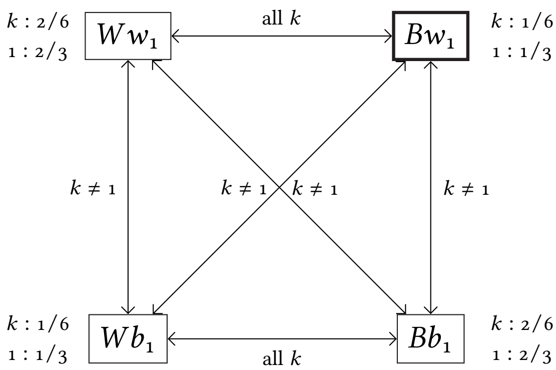


Figure 6.3: The product update model M_1 resulting after the private announcement of player 1's signal.

Given the new state model M_1, we can calculate the probabilities that each agent gives to the new states, using the probability update formula introduced in Section 5.1.2. For example, we compute the revised probability assignment of player 1, since any other players' informational state does not change as a result of event E_1. We drop the world indexation in the product update rule in order to improve legibility, since by the SDP condition, the probability assignment of an agent is the same at every world within the same information set. For example, applying the product update rule

$$\mu_i(s,e) = \frac{\mu_i(s) \cdot pre_i(e|s)}{\sum_{(s',e') \in S_1} \mu_i(s') \cdot pre_i(e'|s')}$$

we have that

$$\mu_1(W, w_1) = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

This represents, intuitively, the probability player 1 assigns to the state of the world being W, given he received the private signal w_1. The other probabilities are computed in the same way and included in Figure 6.3.

Player 1's action

We formalize player 1's action as an event model A_2 with two events, W_1 and B_1, unrelated to each other by player arrows since they all know what action is being taken. Each action changes facts according to a substitution function. Formally we have $A_2 = (E_2, \sim_i, PRE, pre_i, sub_2)_{i \in Ag}$ in which:

- $E_2 = \{W_1, B_1\}$
- $\sim_i$ such that $W_1 \nsim_i B_1 \forall i$
- $PRE(e) = \{\phi_W, \phi_B\}$ for all $e \in E_2$, such that:

$$\phi_W^1 := P_1(\bar{W}) > P_1(\bar{B}) \vee \Big(P_1(\bar{W}) = P_1(\bar{B}) \wedge w_1\Big)$$

$$\phi_B^1 := P_1(\bar{W}) < P_1(\bar{B}) \vee \Big(P_1(\bar{W}) = P_1(\bar{B}) \wedge b_1\Big)$$

- $pre_i : \Phi \to (E_2 \to [0, 1])$ such that $pre_i(W_1|\phi_W^1) = pre_i(B_1|\phi_B^1) = 1$ and $pre_i(W_1|\phi_B^1) = pre_i(B_1|\phi_W^1) = 0$ for all i.
- $sub_2(W_1) = \sigma_{W_1}$ and $sub_2(B_1) = \sigma_{B_1}$, where $dom(\sigma_{W_1}) = \{\bar{W}_1\}$, $dom(\sigma_{B_1}) = \{\bar{B}_1\}$, and $\sigma_{W_1}(\bar{W}_1) = \neg\bar{W}_1$, $\sigma_{B_1}(\bar{B}_1) = \neg\bar{B}_1$.

We depict event model A_2 graphically in Figure 6.4 The *pre* function

Figure 6.4: The event model A_2 represented by player 1's announcement of his guess. Everyone can distinguish between these two events, hence the model has no outgoing arrows.

encodes the common knowledge of rationality and tie-breaking rule assumptions: agent 1 only announces W_1 if he either believes W to be more likely than B or he believes them to be equally probable but his private signal was w_1. In the previous paragraph, we computed player 1's subjective probability that W is the true state of the world, $\mu_1(W, w_1) = \frac{2}{3}$. Therefore player 1 will choose W_1. This event gives rise to the update model M_2. The product update model will still have four worlds, but the two worlds where W_1 is true will be unrelated (for all players) to the two worlds in which B_1 is true: it is common knowledge in which of these two zones the players are. Assuming the real world is Bw_1W_1, we can thus disregard the B_1 worlds as irrelevant (inaccessible, impossible). More precisely, the 4-world model with actual world Bw_1W_1 is *bisimilar* to the the 2-world model (having only the W_1 worlds) with the same actual world Bw_1W_1. So we can just delete the B_1 worlds, obtaining the model in Figure 6.5. From now on, we abuse notation and refer to the bisimilar model as being the product update model itself.

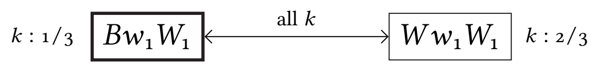


Figure 6.5: The product update model M_2 after player 1's announcement. The consequence of imposing common knowledge of rationality on the current model is the deletion of the worlds at which b_1 is true. All players, knowing that player 1 is rational, are able to deduce that player 1 saw a white ball.

Nature's pick of player 2's private signal

Assume that Nature sends player 2 the signal w_2. This private communication is represented by the event model

$$A_3 = (E_3, \sim_i, \Phi, pre, sub)_{i \in Ag}$$

where, the elements are, *mutatis mutandis*, as for A_1. This private announcement is modelled as an event model and is represented in Figure 6.4

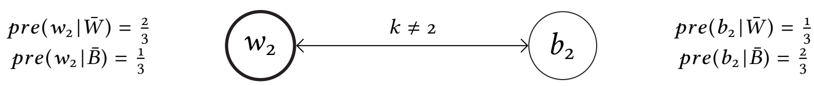


Figure 6.6: The event model A_3 represented by player 2 drawing a ball from the urn. Player 2 sees a w_2, however none of the other players can distinguish between the two events at this stage.

Again, as above, we model the effect of this event using a product update model $M_3 = M_2 \otimes A_3$ in Figure 6.7.

Player 2's action

The action of player 2 can be encoded as an event model A_4, which is depicted in Figure 6.8, where :

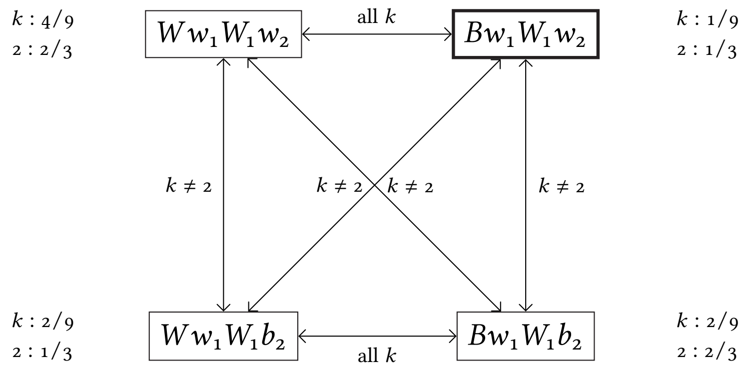


Figure 6.7: The product update model M_3 resulting after the private announcement of player 2's signal.

$$\phi^2_W := \big(P_2(\bar{W}) > P_2(\bar{B})\big) \vee \big(P_2(\bar{W}) = P_2(\bar{B}) \wedge w_2\big)$$
$$\phi^2_B := \big(P_2(\bar{W}) < P_2(\bar{B})\big) \vee \big(P_2(\bar{W}) = P_2(\bar{B}) \wedge b_2\big)$$

and in general, let:

$$\phi^i_W := \big(P_i(\bar{W}) > P_i(\bar{B})\big) \vee \big(P_i(\bar{W}) = P_i(\bar{B}) \wedge w_i\big)$$
$$\phi^i_B := \big(P_i(\bar{W}) < P_i(\bar{B})\big) \vee \big(P_i(\bar{W}) = P_i(\bar{B}) \wedge b_i\big)$$

Figure 6.8: The event model A_4 represented by player 2's announcement of his guess. Everyone can distinguish between these two events

The preconditions are designed to make player 2 choose action W_2, given that the subjective belief he attaches to the world being W is given by $\mu_2(W) = \frac{4}{5} > \frac{1}{2}$. This public announcement will give rise to the product update model M_4. Graphically this is represented in Figure 6.9:

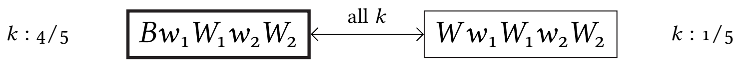


Figure 6.9: The product update model M_4 after player 2's public announcement. The consequence of imposing common knowledge of rationality on the current model is the deletion of the worlds at which b_2 is true.

Nature's pick of player 3's private signal

Assume, further, that Nature chooses a black ball b_3 for player 3. As expected, we model the private communication using an event model A_5, represented in Figure 6.10.

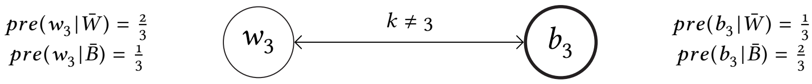


Figure 6.10: The event model A_5 represented by player 3 drawing a black ball from the urn. None of the players, except 3, can distinguish between the two events at this stage.

The update model that follows this event, denoted by M_5, is given by Figure 6.11. In the pictorial representation, you can find the subjective probability player 3 assigns to the world being W, given the observed actions of previous players and his own private signal.

Player 3's action

As dictated by rationality, player 3 will choose action W_3, as depicted in Figure 6.12. Moreover, if Nature had chosen a white ball w_3 then the subjective probability that player 3 assigned to the world being white would have been higher than $\frac{2}{3}$. Therefore, no matter what signal player 3 had received, he would have chosen W_3.

As we have gotten used to by now, we are going to construct a product update model M_6, as a result of the public communication of action W_3. In this new model, no worlds will be deleted, since no player except player 3 can distinguish between b_3 and w_3, based solely on the assumption of common knowledge of rationality. For player 3, as we have argued, it is consistent with rationality to choose action W_3, both in the case that he receives a w_3 and b_3. This means this agent has entered into a false cascade, and others cannot infer his private signal though his choice.

The same reasoning can be applied to any subsequent player, who will rationally choose W_i, regardless of his private signal. This is due to the failure of extracting any extra information for any players, except player 1, player 2 and himself. Thus, as argued before, agents will enter a cascade, in which every player $i > 2$ imitates his predecessor.

We set to prove this claim formally by induction on the number of players. First, we encode the initial conditions of this example into one sentence called *initial*. This says that it is common knowledge that either $\overline{W}$ or $\overline{B}$ holds, that nobody knows which of them is true

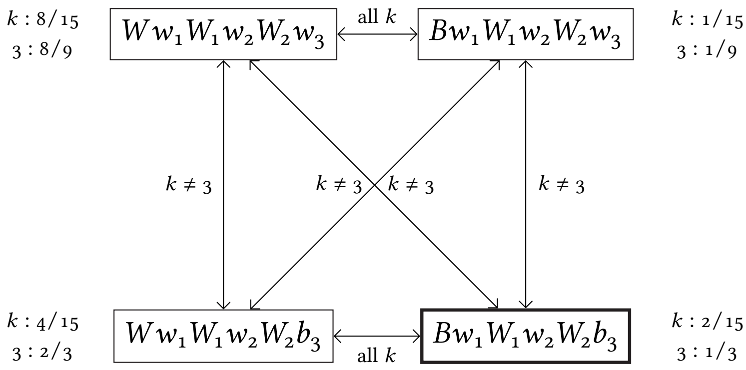


Figure 6.11: The product update model M_5 resulting after the private announcement of player 3's signal.

Figure 6.12: The event model A_6 represented by player 3's announcement of his guess. Everyone can distinguish between these two events

and that everybody assigns probability 0.5 to each possibility. More precisely,

$$\text{initial} := \Big((\overline{W} \vee \overline{B}) \wedge \bigwedge_{i \in Ag} (\neg[i]\overline{W} \wedge \neg[i]\overline{B}) \wedge P_i(\overline{W}) = 0.5)\Big)$$

Proposition 1

For all $i \geq 3$, all $e_2 \in E_2$ and $e_4 \in E_4$, all $5 \leq j \leq 2i - 1$, and all $e_5 \in E_5, e_6 \in E_6 \ldots, e_{2i-1} \in E_{2i-1}$, we have that:

$$[\cup^*_{i \in Ag}]\text{initial} \Rightarrow [w_1][e_2][w_2][e_4][e_5] \ldots [e_{2i-1}]$$
$$\Big(\bigwedge_{j<i} \big(P_j(\overline{W}) \geq 2 \cdot P_j(\overline{B})\big) \wedge \bigwedge_{j \geq i} \big(P_j(\overline{W}) \geq 4 \cdot P_j(\overline{B})\big)\Big)$$

is a valid formula.

We set to prove this in induction on the number of players. Note that by the completeness result, it follows that all these formulas should also be provable in the axiomatic system.

Proof by induction on the number of players.

BASE CASE The case $n = 3$ was already proved above.

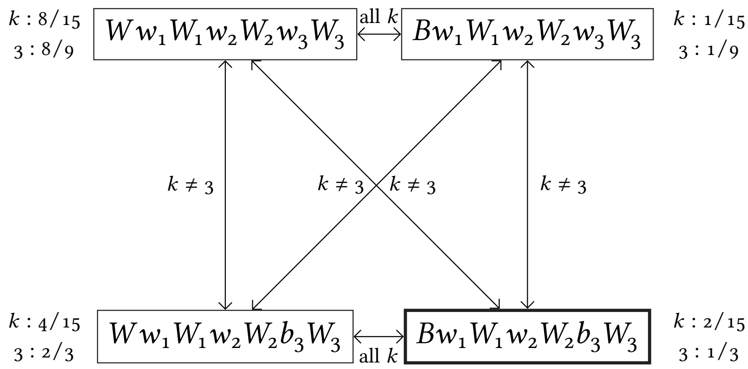


Figure 6.13: The product update model M_6 resulting after the private announcement of player 3's signal.

INDUCTIVE HYPOTHESIS Assume the proposition holds for all $i \leq n-1$ and try to prove it holds for n. In particular the inductive hypothesis holds for $n-1$, and therefore get state model M_{2n-3}, which we will represent partially, by lumping together all the W-worlds and, respectively, B-worlds as presented in Figure 6.14

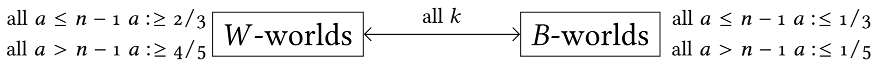


Figure 6.14: The state model M_{2n-3}, representing the beliefs of players after player $n-1$ has seen his signal. The probabilities express the sentence in the proposition, in terms of actual probability assignments. For example, $P_j(\overline{W}) \geq 2 \cdot P_j(\overline{B})$ is equivalent to saying that $\mu_j(\overline{W}) \geq \frac{2}{3}$

.

Next, player $n-1$ will publicly announce $W_{n-1}!$, as demanded by his beliefs. This announcement will not change the informational state of any agent, since no one except $n-1$ can infer anything about the private signal of player $n-1$. Therefore, the new model M_{2n-2} will be identical to M_{2n-3} in terms of beliefs of players. This is so because at every world in the model M_{2n-3}, the sentence $P_{n-1}(W) > \frac{1}{2}$ is true, and therefore common knowledge. The next event is represented by the private announcement of player n's signal, which is reflected in event model A_{2n-1} depicted in Figure 6.15.

The new product update model that results from M_{2n-2} and A_{2n-1} is presented graphically in Figure 6.16

Applying the technique of lumping together W-worlds and respectively B-worlds, we end up with a model of the form:

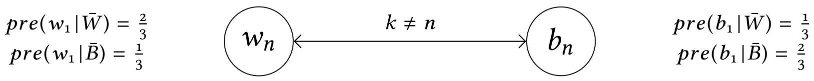


Figure 6.15: The event model A_{2n-1} represented by player n drawing a ball from the urn. None of the other players can distinguish between the two events at this stage.

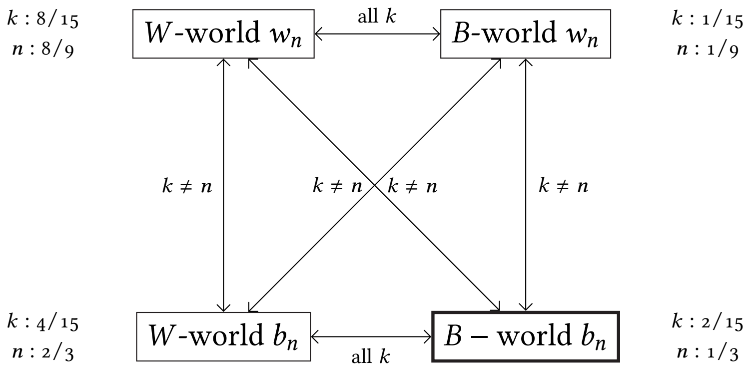
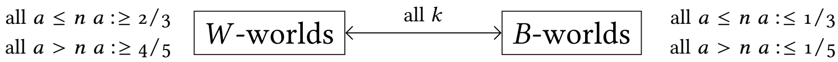


Figure 6.16: The product update model M_{2n-1} resulting after the private announcement of player n's signal.



Therefore, model M_{2n-1} satisfies

$$P_j(\overline{W}) \geq 2 \cdot P_j(\overline{B}) j \leq n, \text{ and}$$

$$P_j(\overline{W}) \geq 4 \cdot P_j(\overline{B}) \text{ for all } j > n$$

Hence we proved the induction step for n.

□

In conclusion, in Chapter 6 we used a logical formalism to capture the meta-level reasoning of players about game cascades. The conclusion of this analysis is that informational cascades can be seen as an infelicity of the social dynamics of information engendered by strategic considerations, rather than a failure of individual rational cognisance. This claim is endorsed by the observation that from player 3 onwards, all players were aware that they were in a cascade. What is more, there was common knowledge that a cascade ensued, and yet, these higher order reflective powers that agents were endowed with failed to dissuade any of them from changing their strategy.

7

CONCLUSION AND DIRECTIONS FOR FURTHER WORK

7.1 SYNTHESIS

In this thesis, we looked at situations where "rational individual action, in pursuit of well-defined preferences, leads to outcomes undesirable to individuals and surprising, given their intention" (Granovetter, 1978, p. 15).

Essentially, the thesis shows how hard it is to avoid false cascades. The logical chapter and its application show that no amount of higher-level rationality, theory of the mind, perfect Bayesian reasoning and common knowledge (including common knowledge of the protocol, of everybody's rationality and of everybody's awareness of the possibility of cascades) is enough to avoid the cascade. The game theory chapters show that the strategy leading to cascades is a PBNE, and indeed an unique equilibrium as far as its on path prescriptions are concerned.

The chapters on other games are attempts at social mechanism design aimed at avoiding the cascade, by changing the payoffs to favour collective rationality, or by imposing team-play (joint choice of strategy) and forbidding or punishing individual deviations from the group decision. By and large, most of these attempts fail: some kind of cascading behaviour is produced by best response actions in most of these variations. The only exception is constituted by truth-tracking games, which reward agents if and only if the group attains the truth.

A condensed view of the thesis' contributions can be divided according to their nature. Part of our contribution is exegetical. We interpreted and formalized one of the main solution concepts, namely Perfect Bayesian Nash Equilibrium, for a new type of games: imperfect information games with chance moves. Another part of our contribution is explanatory: we developed a formal definition of informational cascades, based on a fine-grained conception of cascade behaviour, as generated by strategic considerations. From this perspective, informational cascades can now be described as sequential games with imperfect information and chance moves in which the Bayesian strategy is a PBNE. We compressed cascade reasoning into a strategy and showed that the likelihood of cascades depends on the structure of rewards the group shares at the end of the game. This dependence is indirect, connected by the notion of optimality. Finally, a part of the contribution is conceptual. We put forth a proposal for a logic of imperfect information games. While there are a few open problems in the way of achieving a viable logic for games, the start seems promising.

7.2 DIRECTIONS FOR FURTHER WORK

The developments of this thesis could be expanded in various directions

DIRECTION 1 In this thesis we talked, in fact, about one particular kind of externality, called informational externality. It is the very familiar concept that refers to the informational benefit an agent enjoys from observing the actions of some other agents. We showed before that the information which can be inferred from someone else's action constituted the motor for the propagation of conformity in the informational cascade models presented.

In contrast, a future direction could be considering an alternative explanation for rational herd behaviour: payoff externalities. These denote situations in which the payoff of an agent is an increasing function in the number of other agents adopting the same action. Some examples are the convention of driving on the right side of the road or the adoption of the fax technology. In both of these cases, the more people adopt a particular behaviour or technology, the more benefit everyone derives from them. Payoff externalities are sometimes called network effects, because an agent can incur an explicit benefit by aligning his behaviour with the behaviour of others.

A potential focus could be characterizing the reasoning of connected agents, in a way that places emphasis on the influence individuals have on their network neighbours. This perspective is motivated by the observation that most social informational processes take place at a local, rather than global level: when considering whether to adopt a technology, for example, one often turns to his circle of friends and acquaintances, and less to statistics regarding what the overall population is doing.

DIRECTION 2 This thesis was mainly concerned with informational cascades and their logical formalization. However, a fruitful direction could be the formalization, in the same logical language, of other social epistemic phenomena, like pluralistic ignorance. With regard to other urn games, a logical formalization can only be done by moving to a "protocol" version of the logic (in the style explored in van Benthem, Gerbrandy, Hoshi and Pacuit (2009)), since if one player's payoff depends also on all the other players' future moves, then the assumption of common knowledge of rationality becomes a constraint on the whole future history of events, so a "protocol" constraint rather than a one-event constraint.

DIRECTION 3 We can further generalize the informational cascade setting. For example, we would want to endogenize the order in the sequence of choices. This seems like a natural extension of the cascade model presented in this thesis. Another would be, as hinted earlier in Chapter 3, to change the informational structure of the game. For example, consider cases where players do not observe the order of the decisions made before them in the sequence, but just the number of people opting for each choice (like in the restaurant example). Another possible change would be to limit the observational powers of the players to the last couple of decisions.

BIBLIOGRAPHY

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

Aumann, R. J.: 2006, Autobiography, *Nobel Prize in Economics documents 2005-3*, Nobel Prize Committee.

Avery, C. and Zemsky, P.: 1998, Multidimensional Uncertainty and Herd Behavior in Financial Markets, *The American Economic Review* **88**(4), 724–748.

Baltag, A., Christoff, Z., Hansen, J. U. and Smets, S.: 2013, *Logical Models of Informational Cascades*, Logic Across the University: Foundations and Applications, College Publications, pp. 405–432.

Baltag, A., Moss, L. S. and Solecki, S.: 1999, The logic of public announcements, common knowledge, and private suspicions, *Technical report*.

Baltag, A., Smets, S. and Zvesper, J. A.: 2009, Keep 'hoping' for rationality: A solution to the backward induction paradox.

Banerjee, A. V.: 1992, A Simple Model of Herd Behavior, *The Quarterly Journal of Economics* **107**(3), 797–817.

Battigalli, P.: 1996, Strategic independence and perfect bayesian equilibria, *Journal of Economic Theory* **70**(1), 201 – 234.

Bicchieri, C.: 2005, The grammar of society: The nature and dynamics of social norms.

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

Bikhchandani, S. and Sharma, S.: 2001, Herd behavior in financial markets.

Blackburn, P., de Rijke, M. and Venema, Y.: 2001, *Modal Logic*, Cambridge University Press, New York, NY, USA.

Devenow, A. and Welch, I.: 1996, Rational herding in financial economics, *European Economic Review* **40**(3-5), 603–615.

Diamond, D. and Dybvig, P.: 1983, Bank runs, deposit insurance, and liquidity, *The Journal of Political Economy* pp. 401–419.

Easley, D. and Kleinberg, J.: 2010, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*, Cambridge University Press, New York, NY, USA.

Fagin, R. and Halpern, J. Y.: 1994, Reasoning about knowledge and probability, *Journal of the ACM* **41**, 340–367.

Fagin, R., Halpern, J. Y. and Megiddo, N.: 1990, A logic for reasoning about probabilities, *Inf. Comput.* **87**(1-2), 78–128.

Fudenberg, D. and Tirole, J.: 1991, Perfect bayesian equilibrium and sequential equilibrium, *Journal of Economic Theory* **53**(2), 236–260.

Granovetter, M.: 1978, Threshold models of collective behavior, *The American Journal of Sociology* **83**(6), 1420–1443.

Hansen, P. G., Hendricks, V. F. and Rendsvig, R. K.: 2013, Infostorms, *Metaphilosophy* **44**(3), 301–326.

Harsanyi, J. C.: 1967-68, Games with Incomplete Information Played by 'Bayesian' Players, *Management Science* **14**, 159–182, 320–324, 486–502.

Hirshleifer, D. and Anderson, J. E.: 1994, *The Blind Leading the Blind: Social Influence, Fads and Informational Cascades: for Volume in Honor of Gary Becker*, Anderson Graduate School of Management, UCLA.

Hung, A. A. and Plott, C. R.: 2001, Information cascades: Replication and an extension to majority rule and conformity-rewarding institutions, *American Economic Review* **91**(5), 1508–1520.

Kolodny, N. and Brunero, J.: 2013, Instrumental rationality, *in* E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, fall 2013 edn.

Kreps, D. and Wilson, R.: 1982, Sequential equilibria, *Econometrica* **50**, 863–894.

Krugman, P.: 2009, How did economists get it so wrong?, *The New York Times* .

Osborne, M.: 2004, *An introduction to game theory*, Oxford Univ. Press.

Parikh, R.: 2003, Levels of knowledge, games, and group action, *Research in Economics 57, 267*, p. 281.

Rendsvig, R. K.: 2013, Aggregated beliefs and informational cascades, *in* D. Grossi, O. Roy and H. Huang (eds), *Logic, Rationality, and Interaction*, Vol. 8196 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 337–341.

Rogers, E. M.: 2003, *Diffusion of innovations*, 5th edn, Free Press, New York, NY [u.a.].

Scharfstein, D. S. and Stein, J. C.: 1990, Herd behavior and investment, *American Economic Review* **80**(3), 465–79.

Shiller, R.: 2000, *Irrational exuberance*, Princeton Univ. Press, Princeton, NJ.

Shiller, R.: 2008, How a bubble stayed under the radar, *New York Times* .

Shiller, R. J.: 1995, Conversation, information, and herd behavior, *Cowles Foundation Discussion Papers 1092*, Cowles Foundation for Research in Economics, Yale University.

Smith, V. L., King, R. R., Williams, A. W. and Boening, M. V.: 1993, The robustness of bubbles and crashes in experimental stock markets, *Non Linear Dynamics and Evolutionary Economics*, Oxford Press.

Surowiecki, J.: 2005, *The Wisdom of Crowds*, Anchor.

Talbott, W.: 2013, Bayesian epistemology, *in* E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, fall 2013 edn.

van Benthem, J., Gerbrandy, J., Hoshi, T. and Pacuit, E.: 2009, Merging frameworks for interaction, *Journal of Philosophical Logic* **38**(5), 491–526.

van Benthem, J., Gerbrandy, J. and Kooi, B. P.: 2009, Dynamic update with probabilities, *Studia Logica* **93**(1), 67–96.

van Benthem, J., van Eijck, J. and Kooi, B.: 2006, Logics of communication and change, *Information and Computation* **204**(11), 1620 – 1662.

van Eijck, J. and Schwarzentruber, F.: 2014, Epistemic probability logic simplified, *Manuscript* .

Welch, I.: 1992, Sequential Sales, Learning, and Cascades, *The Journal of Finance* **47**(2), 695–732.

INDEX