Journal of Mathematical Economics 59 (2015) 47–57

Contents lists available at ScienceDirect

Journal of Mathematical Economics

journal homepage: www.elsevier.com/locate/jmateco

Multiagent belief revision☆

Antoine Billot [a,b], Jean-Christophe Vergnaud [c,*], Bernard Walliser [d]

[a] LEMMA-Paris 2, Paris, France
[b] IUF, Paris, France
[c] CES-CNRS-Paris 1, Paris, France
[d] PSE, Paris, France

ARTICLE INFO

Article history:
Received 30 January 2014
Received in revised form
17 April 2015
Accepted 5 May 2015
Available online 14 May 2015

Keywords:
Belief
Belief revision
Epistemic logics
Game

ABSTRACT

An original epistemic framework is proposed for the modeling of beliefs and messages within a multiagent belief setting. This framework enables public, private and secret messages as well, even when the latter contains errors. A revising rule—*i.e.* the *product rule*—is introduced in pure epistemic terms in order to be applied to all structures and message. Since any syntactic structure can be expressed through various semantic ones, an equivalence principle is given by use of the semantic notion of *bisimilarity*. Thereafter, a robustness result proves that, for a given prior structure, bisimilar messages yield bisimilar posterior structures (Theorem 1). In syntax, the beliefs revised thanks to the product rule are then shown to be unique (Theorem 2). Finally, an equivalence theorem is established between the product rule and the Belief-Message Inference axiom (Theorem 3).

1. Introduction

Belief revision has been intensively studied in epistemic logics. However, if epistemic models are well-established in a multiagent setting, there is no convincing general modeling of the effect of a message onto prior beliefs in this framework, in particular when this message is secret and/or when it contains *false* information on the world or, more subtly, on the information received by another agent. Yet, a lot of economic and strategic situations involve erroneous information or 'secret communication'. The purpose of this paper is then to propose a framework allowing to revise consistently all prior beliefs by means of a generalized rule called the *product rule* (PR). This rule works as follows: first, a message is designed as a standard belief structure where possible worlds and possibility domains translate what is learned by agents and what is learned about what is learned; second, posterior worlds are built up by combining prior worlds with message worlds; third, at each posterior world, agents believe as possible all combinations of prior and message worlds that are compatible.

1.1. An example: the Rothschild 'myth'

Nathan Mayer Rothschild is associated with one of the mythical cases of 'insider trading' entailing erroneous information: he is said to have used his early knowledge of an English victory at Waterloo to fool people on the Stock Exchange and consequently make a vast fortune. A version of this story can be found in Morton (1962): *A Rothschild agent jumped into a boat at Ostend to bring a dispatch. When he received it, Nathan Rothschild let his eye fly over the lead paragraphs (claiming Wellington's victory). Then he proceeded to the Stock Exchange. He did not invest. He sold. He dumped consols. Consols dropped still more. 'Rothschild knows,' the whisper rippled through the Change. 'Waterloo is lost.' Nathan kept on selling and bought a giant parcel for a song.* Let us now consider an imaginary version of this story involving Nathan Mayer Rothschild (denoted R) and the British Prime Minister, Robert Banks Jenkinson (denoted J), which leaves out the strategic aspect to focus onto beliefs and revision. This version is developed along the two standard approaches in epistemics—namely, first in syntax and second in semantics.

For Rothschild as for Jenkinson, it does matter whether 'Wellington wins'—a proposition formally denoted p—or 'Wellington does not win'—*i.e.*, $\neg p$. Before any message, they both have no definite belief about the outcome of the battle and their prior beliefs (*priors* for brief, denoted $\mathbf{B}_i$) can then be formally expressed as follows: $\neg \mathbf{B}_i p \wedge \neg \mathbf{B}_i \neg p$ with $i \in \{R, J\}$ – meaning that they do not believe p and they do not believe $\neg p$ neither – and it is common

☆ We thank A. Baltag, G. Bonanno, P. Fleckinger, I. Gilboa, J. Lang, J.M. Tallon, J. van Benthem and three referees for helpful remarks and comments on earlier version of this paper.

* Corresponding author.
E-mail addresses: billot@u-paris2.fr (A. Billot), vergnaud@univ-paris1.fr (J.-C. Vergnaud), walliser@mail.enpc.fr (B. Walliser).

http://dx.doi.org/10.1016/j.jmateco.2015.05.004
0304-4068/

belief. Now, Rothschild is *secretly* informed by a spy while Jenkinson is just informed through the official channel and ignores that a spy came informing Rothschild. The content of the spy-message – 'Wellington wins' – is captured by another proposition $\overline{m}$ while the content of the official-message – 'the outcome of the battle is still uncertain' – is captured by a proposition $\overline{m}'$. (Notice that the language in which the messages are represented differs from the language in which the facts of the matter are represented.) To formally describe the diffusion protocol of the message, the operator $\overline{\mathbf{B}}_i$ is introduced depicting what 'agent i learns'. At the first level, the information displayed can be captured by $\overline{\mathbf{B}}_R\overline{m} \wedge \overline{\mathbf{B}}_J\overline{m}' \wedge \neg\overline{\mathbf{B}}_J\overline{m}$. At a second level, Jenkinson erroneously learns that Rothschild has only learned the official message: $\overline{\mathbf{B}}_J\overline{\mathbf{B}}_R\overline{m}' \wedge \neg\overline{\mathbf{B}}_J\overline{\mathbf{B}}_R\overline{m}$, while Rothschild learns correctly that Jenkinson has only learned the official message: $\overline{\mathbf{B}}_R\left(\overline{\mathbf{B}}_J\overline{m}' \wedge \neg\overline{\mathbf{B}}_J\overline{m}\right)$. In addition, Rothschild learns that Jenkinson learns erroneously that he has only learned the official message: $\overline{\mathbf{B}}_R\overline{\mathbf{B}}_J\left(\overline{\mathbf{B}}_R\overline{m}' \wedge \neg\overline{\mathbf{B}}_R\overline{m}\right)$. Posterior beliefs (*posteriors* for brief, denoted $\mathbf{B}_i^*$) are then deduced from a logical revision of priors according to what agents have learned. For Rothschild, learning $\overline{m}$ removes his initial doubt: hence, $\mathbf{B}_R^*p$, while for Jenkinson, learning $\overline{m}'$ has no particular effect on his first-order posteriors: hence, $\neg\mathbf{B}_J^*p \wedge \neg\mathbf{B}_J^*\neg p$. Rothschild being able to simulate Jenkinson's reasoning, his second-order posteriors are then given by $\mathbf{B}_R^*\left(\neg\mathbf{B}_J^*p \wedge \neg\mathbf{B}_J^*\neg p\right)$. In return, Jenkinson can also simulate Rothschild's reasoning but, since he did not learn correctly what Rothschild has learned, he believes that Rothschild's posteriors remain: $\mathbf{B}_J^*\left(\neg\mathbf{B}_R^*p \wedge \neg\mathbf{B}_R^*\neg p\right)$. Hence, Jenkinson makes an obvious mistake when considering Rothschild's final beliefs. The posteriors can be lengthened towards the higher level: since Rothschild has learned that Jenkinson erroneously learned what he has learned himself, Rothschild believes that Jenkinson believes that his posteriors remain: $\mathbf{B}_R^*\mathbf{B}_J^*\left(\neg\mathbf{B}_R^*p \wedge \neg\mathbf{B}_R^*\neg p\right)$, etc.

Let now turn to a semantic interpretation of this story. For priors, a standard partitional model (Aumann, 1976) with two possible worlds – one for p and one for $\neg p$ – is sufficient to represent how Rothschild and Jenkinson both ignore the state of nature (the outcome of the battle). At each world, they both regard these two worlds as possible (see Fig. 1):

Similarly, the secret spy-message can be captured through a two-world structure: an actual world where the correct message is $\overline{m}$—'Wellington wins'—and a second world where no information is available—corresponding to $\overline{m}'$. In the actual world, Rothschild believes this first world possible while Jenkinson believes possible the second one at which, besides, both regard this world as possible (see Fig. 2):

To deduce the semantical structure representing posteriors, PR works as follows (see Fig. 3). Firstly, the true actual world proceeds from the combination of the true prior world with the true world of the message—more generally, posterior worlds result from the combination of prior worlds with message worlds. In the true posterior world, Rothschild believes possible any combination of the prior worlds that he thought possible (*i.e.*, $\mathbf{w}$, w_1) with compatible message worlds that he has learned to be possible (*i.e.*, $\overline{\mathbf{w}}$). Given the content of the message $\overline{m}$, the two worlds w_1 and $\overline{\mathbf{w}}$ are incompatible and, therefore, the only posterior world Rothschild believes possible is $(\mathbf{w}, \overline{\mathbf{w}})$. The same process holds for Jenkinson but, since the world $\overline{w}_1$ that he has learned to be possible is compatible with both $\mathbf{w}$ and w_1, there are two posterior worlds that he believes possible: $(\mathbf{w}, \overline{w}_1)$ and $(w_1, \overline{w}_1)$.

Let us now display how PR is relevant to the revision process syntactically presented above. Recall that Jenkinson finally believes that it is a common belief that Rothschild and he still doubt the outcome of the battle. This corresponds to the two possible worlds $(\mathbf{w}, \overline{w}_1)$ and $(w_1, \overline{w}_1)$. In return, Rothschild knows the outcome and Jenkinson's beliefs as well. This corresponds to the actual world $(\mathbf{w}, \overline{\mathbf{w}})$. Hence, three worlds are definitely necessary to describe posteriors in this case.

1.2. Background and motivation

In epistemic logics, Alchourron et al. (1985) can be seen as the first attempt to design and axiomatize belief revision. In their static set-up, a unique agent is endowed with set-theoretic priors and she is assumed to receive a message defined as an event of this space. Messages are generally meant to clarify agents' beliefs but they could refute them as well (Fagin et al., 1995). More recently, in a dynamic set-up this time, Baltag et al. (1998) characterize a message by its content—*i.e.*, its meaning—and its diffusion protocol—*i.e.*, a specification of the agents who receive the content of the message and the knowledge of all agents about its scattering. Messages are then semantically represented by a structure of possible worlds. To explicit the revision process, two sorts of worlds need to be distinguished: prior worlds and message ones. A method is then given for mixing them up in order to obtain posterior worlds (see also Baltag and Moss, 2004; Aucher, 2009, 2011, and Baltag et al., 2008). Incidentally, all these contributions focus on nonpartitional possibility correspondences.

In interactive epistemology, Aumann (1976) defines a framework where agents' beliefs are based on two possible sources of information: probabilistic common priors and private information partitions over the same set of possible worlds. The existence of a Bayesian revision process is suggested but no precise method is made explicit about the way a prior partition evolves into a posterior one. As for Geanakoplos and Polemarchakis (1982), they consider a dynamic process about public communication connected with partition transformations. At each world, agents revise their prior partitions by intersection with the message partition. This corresponds to the standard *intersection rule* (IR) where priors and message are both expressed over the same set of worlds. However, IR is not theoretically founded. Especially, there exists no current general axiomatization despite two interesting attempts we discuss now. First, Board (2004) considers infinitely many possibility correspondences – each being associated with a proposition that behaves as a message content – but does not account for a general structure to catch on the diffusion protocol. Second, Bonanno (2005) suggests considering three possibility correspondences – one for priors, one for messages and the latter for posteriors – but, in return, does only relate to a single agent.

In the Rothschild–Jenkinson example, as shown above, it is impossible to apply IR because the number of posterior worlds is larger than that of prior worlds, even though, previously – that is for public or private messages – IR has been successfully applied. The difficulty to use IR in our case is apparently based upon the fact that a secret message can involve persisting errors and these errors entail nonpartitional belief structures (to be considered each time the Truth axiom does no longer hold). However, Geanakoplos (1989) introduces a definition of equivalence between information structures and shows that there always exists a partitional structure which is equivalent to a nonpartitional one. Therefore, the Rothschild–Jenkinson secret message example can be dealt with IR in a partitional framework. Yet, this equivalence relation is founded *in terms of decision*—that is, two structures are equivalent if they lead to the same decision according to the Expected Utility (EU) criterion. Here, we only consider a pure epistemic framework where agents' preferences are not specified. In our framework, the equivalence relation that associates two semantical structures corresponding to identical beliefs is given by a purely epistemic notion; *i.e.*, *bisimilarity*. Furthermore, under bisimilarity, Geanakplos's conclusion may not hold: there does not always exist a partitional structure that is bisimilar to a nonpartitional one—this is precisely the case for the posterior structure in the Rothschild–Jenkinson example. In conclusion, IR is not an universal rule for revision, especially in the presence of a failure of the Truth axiom.