

# Multi-Agent Justification Logic: Communication and Evidence Elimination

Bryan Renne  
Faculty of Philosophy  
University of Groningen  
Oude Boteringestraat 52  
9712 GL Groningen  
The Netherlands  
<http://bryan.renne.org/>

March 16, 2009

## Abstract

This paper presents a logic that combines *Dynamic Epistemic Logic*, a framework for reasoning about multi-agent belief dynamics, with a new multi-agent version of *Justification Logic*, a framework for reasoning about evidence and justification. This novel combination incorporates a new kind of *evidence elimination* that cleanly meshes with the multi-agent belief dynamics of Dynamic Epistemic Logic. The resulting system may be used to reason about belief dynamics and evidence elimination within groups of interacting rational agents.

## 1 Introduction

Suppose my friend Anne, who lives in Atlanta, just sent a group email addressed to me and to a mutual friend Charlie, who lives in Columbus. This email, which we will refer to as  $x_1$ , is as follows.

To: Bryan, Charlie  
From: Anne  
Date: November 21, 2008 at 3:00pm EST  
Subject: Possible Job in Columbus

Hi, Bryan and Charlie. Good news! I had an interview today, and I think I am going to be offered a great job in Columbus with XYZ Corp. If I get the job, then I will move to Columbus. Charlie, are you interested in having a new roommate?

Email  $x_1$

Charlie, who always seems to be checking his email, responds to the group shortly thereafter with email  $x_2$ .

To: Anne, Bryan  
From: Charlie  
Date: November 21, 2008 at 3:10pm EST  
Subject: Re: Possible Job in Columbus

Anne, this is great news! Of course I would like to have you as a roommate! As you know, I have an unused room in my apartment, and it is yours if you move to Columbus.

— Anne wrote on November 21, 2008 at 3:00pm EST: —  
Hi, Bryan and Charlie. Good news! I had an interview today, and I think I am going to be offered a great job in Columbus with XYZ Corp. If I get the job, then I will move to Columbus. Charlie, are you interested in having a new roommate?

Email  $x_2$

The next day, Charlie and I hear back from Anne with email  $x_3$ .

To: Bryan, Charlie  
From: Anne  
Date: November 22, 2008 at 10:00am EST  
Subject: I got the job!

Bryan, Charlie: I got the job!

Email  $x_3$

Charlie, very happy for his friend Anne, responds with email  $x_4$ .

To: Anne, Bryan  
From: Charlie  
Date: November 22, 2008 at 10:10am EST  
Subject: Re: I got the job!

Anne, I am so happy to hear that you got the job from XYZ Corp and that you will be moving to Columbus! I will start preparing the room for you today. Give me a call so we can figure out the details.

— Anne wrote on November 21, 2008 at 10:00am EST: —  
Bryan, Charlie: I got the job!

Email  $x_4$

But later in the day, I receive the private email  $x_5$  from Anne.

To: Bryan  
 From: Anne  
 Date: November 22, 2008 at 3:00pm EST  
 Subject: About the job

Bryan, XYZ Corp just called to tell me that the recent slump in the economy is really hurting their business. As a result, the senior management decided today to cut a number of jobs and to freeze all new hires. So they rescinded my job offer! I am really disappointed. I have not yet told Charlie because he is so excited about me moving to Columbus. Ugh!

Email  $x_5$

This sequence of emails— $x_1, x_2, x_3, x_4, x_5$ —together convey quite a bit of information. For example, after I read email  $x_1$ , I have evidence supporting the statement “if Anne gets the job, then Anne moves to Columbus.” Let us write this statement symbolically as  $J \supset M$ . Here we used  $J$  for “job” and  $M$  for “moves”.

Further, let us write  $t :_i \varphi$  to mean “individual  $i$  has evidence  $t$  that  $\varphi$ .” So after I have read email  $x_1$ , we have that

$$x_1 :_B (J \supset M) .$$

That is, “Bryan has evidence  $x_1$  that ‘if Anne gets the job, then Anne moves to Columbus’.” And after I read Charlie’s response  $x_2$ , we have that

$$x_2 :_B (x_1 :_C (J \supset M)) .$$

That is, “Bryan has evidence  $x_2$  that Charlie has evidence  $x_1$  that ‘if Anne gets the job, then Anne moves to Columbus’.” After all, Charlie’s reply  $x_2$  to Anne’s original email  $x_1$  is evidence that Charlie read Anne’s original email. So Charlie certainly has evidence  $x_1$  that  $J \supset M$ . But once I have read Charlie’s reply  $x_2$ , I have evidence  $x_2$  of that statement (that Charlie has evidence  $x_1$  that  $J \supset M$ ).

Once I read Anne’s second email  $x_3$ , I have evidence that Anne got the job; that is,  $x_3 :_B J$ . And for similar reasons as before, after I read Charlie’s reply  $x_4$  to Anne’s second email  $x_3$ , we have that

$$x_4 :_B (x_3 :_C J) .$$

In words: “Bryan has evidence  $x_4$  that Charlie has evidence  $x_3$  that Anne gets the job.”

But after reading Anne’s emails  $x_1$  and  $x_3$ , we can draw the conclusion that Anne will move to Columbus. After all, Anne said in her email  $x_1$  that she would move to Columbus if she gets the job, and we just learned in email  $x_3$  that she indeed got the job. So once I have read these two emails, I can combine my evidence  $x_1$  (about Anne’s plan to move to Columbus if she gets the job) with my evidence  $x_3$  (about the fact that Anne got the job) to draw the conclusion that Anne will move to Columbus. Writing this combination of evidence as  $x_1 \cdot_J x_3$ , where the subscript  $J$  names the specific part of evidence  $x_3$  that is to be combined with evidence  $x_1$  in order to reach the desired conclusion, we have that

$$(x_1 \cdot_J x_3) :_B M .$$

That is, “Bryan has evidence  $x_1 \cdot_J x_3$  that Anne moves to Columbus.” As Charlie’s emails  $x_2$  and  $x_4$  indicate that he too has read Anne’s emails  $x_1$  and  $x_3$ , then Charlie can himself draw this same conclusion using the same reasoning. We therefore have that

$$(x_1 \cdot_J x_3) :_C M .$$

So we see that both Charlie and I have good reason for thinking that Anne will move to Columbus based on the explicit way in which we each combined evidence  $x_1$  (that Anne would move if she gets the job) with evidence  $x_3$  (that Anne did in fact get the job). Further, since I read Charlie’s emails  $x_2$  and  $x_4$ , then I myself can see that Charlie reasoned as did I. After all, his emails  $x_2$  and  $x_4$  provide me with evidence of this very fact. Let us see why.

First, we see that Charlie’s reasoning can be described by the formula

$$(x_1 :_C (J \supset M)) \supset ((x_3 :_C J) \supset (x_1 \cdot_J x_3) :_C M) . \quad (1)$$

In words: “if Charlie has evidence  $x_1$  that ‘if Anne gets the job, then Anne moves to Columbus’, then if Charlie [also] has evidence  $x_3$  that Anne gets the job, then Charlie [concludes that he] has evidence  $x_1 \cdot_J x_3$  that Anne moves to Columbus.”

Second, we observe that formula (1) says that Charlie constructs the evidence  $x_1 \cdot_J x_3$  that  $M$  from both his evidence  $x_1$  that  $J \supset M$  and his evidence  $x_3$  that  $J$ . But then the construction of the evidence  $x_1 \cdot_J x_3$  from evidence  $x_1$  and  $x_3$  is just a description of Charlie’s use of a logically valid rule of inference called *Modus Ponens*:

$$\frac{J \supset M \quad J}{M} , \quad (2)$$

which may be read, “from assumptions  $J \supset M$  and  $J$ , conclude  $M$ .” So Charlie built his evidence  $x_1 \cdot_J x_3$  that  $M$  from his original evidence  $x_1$  that  $J \supset M$  and  $x_3$  that  $J$  according to the logically valid rule of Modus Ponens. Note that the subscript  $J$  in the combined evidence  $x_1 \cdot_J x_3$  indicates the important role the formula  $J$  plays as the antecedent of the implication  $J \supset M$  in the application (2) of Modus Ponens.

Third, I know that the formula (1) that describes Charlie’s reasoning is logically valid and hence true: if we let  $t$  denote the text of the previous paragraph, then  $t$  is my evidence that formula (1) is true. Written symbolically, we have that

$$t :_B [(x_1 :_C (J \supset M)) \supset ((x_3 :_C J) \supset (x_1 \cdot_J x_3) :_C M)] ,$$

which we may read as, “Bryan has evidence  $t$  that (1) [is true].”

Taken together, I have evidence  $t$  that justifies the formula (1) describing Charlie’s reasoning, I have evidence  $x_2$  that justifies the statement  $x_1 :_C (J \supset M)$ , and I have evidence  $x_4$  that justifies the statement  $x_3 :_C J$ . So if we combine my evidence  $t$  with my evidence  $x_2$ , then we have that

$$(t \cdot_{x_1 :_C (J \supset M)} x_2) :_B ((x_3 :_C J) \supset (x_1 \cdot_J x_3) :_C M) .$$

Here the subscript formula indicates the part of evidence  $x_2$  that was combined with evidence  $t$  in order to conclude that “if Charlie has evidence  $x_3$  that Anne gets the job, then Charlie has evidence  $x_1 \cdot_J x_3$  that Anne is moving to Columbus.” If we then combine my evidence  $t \cdot_{x_1 :_C (J \supset M)} x_2$  with my evidence  $x_4$ , then we have that

$$\left( (t \cdot_{x_1 :_C (J \supset M)} x_2) \cdot_{x_3 :_C J} x_4 \right) :_B \left( (x_1 \cdot_J x_3) :_C M \right) .$$

Again, the subscript  $x_3 :_C J$  indicates the part of evidence  $x_4$  that was combined with evidence  $t \cdot_{(x_1 :_C (J \supset M))} x_2$  in order to conclude that “Charlie has evidence  $x_1 \cdot_J x_3$  that Anne is moving to Columbus.” Conclusion: I have evidence

$$(t \cdot_{x_1 :_C (J \supset M)} x_2) \cdot_{x_3 :_C J} x_4 \tag{3}$$

that Charlie has evidence  $x_1 \cdot_J x_3$  that Anne is moving to Columbus. Let us see why this conclusion makes sense. First,  $t$  is my evidence of the validity of a two-step reasoning process described by (1) that Charlie can go through to conclude that  $x_1 \cdot_J x_3$  is evidence that Anne is moving to Columbus (the process: first he establishes that  $x_1$  is evidence that  $J \supset M$  and then he establishes that  $x_3$  is evidence that  $J$ ). But I have evidence  $x_2$  that Charlie has evidence  $x_1$  that  $J \supset M$ ; put another way, my evidence  $x_2$  says that Charlie has completed the first of the two-step reasoning process. Further, I have evidence  $x_4$  that Charlie has evidence  $x_3$  that  $J$ ; in words, my evidence  $x_4$  says that Charlie has (also) completed the second of the two-step reasoning process. But my evidence (10) was built in two steps by first combining  $t$  with  $x_2$  and then combining the result with  $x_4$ . So this is why (10) is my evidence that Charlie has evidence  $x_1 \cdot_J x_3$  that Anne is moving to Columbus.

But now Anne’s sad final email  $x_5$  adds a new twist to the story. In this email, Anne tells me privately that the information conveyed in her email  $x_3$ , which said that she accepted the job, is now defunct. So Anne’s message  $x_5$  is an *elimination* of my evidence  $x_3$  that Anne got the job, in that her later message  $x_5$  causes me to set aside the evidence her earlier message  $x_3$  provided to support the assertion that she got the job. Accordingly, after I have read her message  $x_5$ , then we have that

$$\neg(x_3 :_B J) .$$

That is, “Bryan does not have evidence  $x_3$  that Anne gets the job.” And since my evidence  $x_3$  that Anne got the job is no longer valid, then I should also abandon the conclusion that Anne is moving to Columbus because I came to this conclusion on the basis of my evidence  $x_1 \cdot_J x_3$ , something that was predicated on my evidence  $x_3$  that she got the job. So after I read Anne’s message  $x_5$ , we have not only that  $\neg(x_3 :_B J)$  but also that

$$\neg((x_1 \cdot_J x_3) :_B M) .$$

That is, “Bryan does not have evidence  $x_1 \cdot_J x_3$  that Anne moves to Columbus.” Again, this is because evidence  $x_1 \cdot_J x_3$ , a piece of evidence that is built by combining the part of evidence  $x_3$  that Anne got the job with the evidence  $x_1$ , cannot be a reason for the assertion that Anne is moving to Columbus because, sadly, Anne did not get the job.

Now let us see what happens with Charlie as a result of Anne’s message  $x_5$  about XYZ Corp rescinding the job offer. First,  $x_5$  was sent privately to me, so Charlie did not receive this message. Second, in  $x_5$ , Anne writes that she has not yet told Charlie about XYZ Corp rescinding the job offer. So after I read message  $x_5$ , I am led to think that Charlie still thinks that the evidence conveyed by messages  $x_1$  and  $x_3$  is valid. This is because, from my perspective, Charlie has no reason to think that evidence  $x_3$  should be eliminated. So even after I read message  $x_5$ , we still have that

$$(x_2 :_B (x_1 :_C (J \supset M))) \wedge x_4 :_B (x_3 :_C J) .$$

That is, “Bryan (still) has evidence  $x_2$  that Charlie has evidence  $x_1$  that  $J \supset M$ , and Bryan (still) has evidence  $x_4$  that Charlie has evidence  $x_3$  that  $J$ .” But then I can reason as before to conclude that Charlie must still think that Anne is moving to Columbus on the basis of his combined evidence  $x_2$  and  $x_4$ ; that is, we still have that

$$((t \cdot_{x_1 :_C (J \supset M)} x_2) \cdot_{x_3 :_C J} x_4) :_B ((x_1 \cdot_J x_3) :_C M) .$$

So while I myself have had evidence  $x_3$  eliminated, and so I do not think that  $x_1 \cdot_J x_3$  is good evidence that Anne is moving to Columbus, I can still see why it is reasonable for Charlie to reach the opposite conclusion. After all, he did not get Anne’s final email  $x_5$ .

## 1.1 About This Paper

This paper studies issues of evidence and belief within groups of interacting rational individuals. As our email example suggests, we will be interested in how communications can affect the beliefs and evidence held by the individuals in question. Communications may be public (like Anne’s first email  $x_1$ ), may be private (like Anne’s last email  $x_5$ ), or may involve subtle mixtures of privacy and misdirection (unlike any of the emails in our example). Our task in this paper will be to provide a general account of how to reason about the way these kinds of communications affect the beliefs and evidence held by the rational individuals who are a party to the communication.

Like the example above, we will restrict our evidence changes to a kind of *evidence elimination*, wherein a piece of evidence that may have justified a certain conclusion is made so that it does not justify that conclusion. We saw the result of such an elimination in the email example, where Anne’s final message  $x_5$  eliminated my evidence  $x_3$  that she got the job.

In what follows, we will present a theory for reasoning about belief change and evidence elimination within groups of interacting rational individuals. We will refer to these individuals using a simple term: *agents*. Our theory builds upon the work in two fast-growing areas of applied logic: *Dynamic Epistemic Logic* and *Justification Logic*. So let us say a bit more about each of these areas.

*Dynamic Epistemic Logic* is a framework for reasoning about multi-agent belief dynamics [6, 18, 19]. This framework grew out of the early work by Gerbrandy and Groeneveld [10, 11] on *public announcements*. (Public announcements were actually studied almost ten years

earlier by Plaza [16], but his work did not become well-known until much later.) In a series of influential papers by Baltag, Moss, and Solecki (BMS) [3, 4, 5], the work on public announcements was generalized to a broader class of communicative updates called *epistemic programs*. The name *epistemic programs* suggestively describes the flexibility that these communicative actions provide in specifying arbitrarily complicated mixtures of public and private communication, even allowing for the possibility of deceit and misdirection within the communication. With the help of a number prominent contributors (including van Benthem, van Ditmarsch, van Eijck, van der Hoek, Kooi, Pacuit, and Smets, among others), the work has grown very quickly into a broader discipline now called *Dynamic Epistemic Logic*.

*Justification Logic* is a framework for reasoning about evidence and justification for rational agents [1, 8, 14, 17]. This work originated in the proof-theoretic studies of Gödel, who sought an exact provability semantics for the modal logic **S4** [12]. Artemov later discovered the *Logic of Proofs* as this long-sought connection between **S4** and Gödel’s intended **S4** provability semantics [2], and a number of authors (including Artemov, Fitting, Iemhoff, Krupski, Kuznets, Milnikel, and others) have since grown the study of the Logic of Proofs into a broader research project—*Justification Logic*—whose purpose is to investigate a wide-ranging family of logics of evidence and justification for rational agents.

In this paper, we present a system that combines Dynamic Epistemic Logic with a new multi-agent Justification Logic. Our system includes the full range of multi-agent belief dynamics from Dynamic Epistemic Logic along with a new operation for multi-agent evidence elimination. After proving the correctness (soundness and completeness) of the system with respect to its intended semantics, we will show how the system can be used to formalize the reasoning we presented in the email example above. So let us begin by introducing the syntax of our system.

## 2 Syntax

### 2.1 The Modal and Evidence Languages

The language of our framework describes the beliefs and evidence held by each of a finite number of individuals that we call *agents*.

**Definition 2.1.** An *agent set* is a finite nonempty set. We refer to the elements of an agent set as *agents*.

**Definition 2.2** ( $\text{ML}^A$ ). Let  $A$  be an agent set. The *modal language (for  $A$ )*, written  $\text{ML}^A$ , consists of the *formulas*  $\varphi$  formed by the following grammar.

$$\begin{aligned} \varphi ::= & p_k \mid \perp \mid \top \mid \varphi_1 \star \varphi_2 \mid \neg\varphi \mid B_i\varphi \\ & k \in \mathbb{N}, \star \in \{\supset, \wedge, \vee, \equiv\}, i \in A \end{aligned}$$

The  $p_k$ ’s make up the set of *propositional letters*.  $\perp$  is the propositional constant for falsity and  $\top$  is the propositional constant for truth. To reduce the amount of horizontal space

needed to write the above grammar, we used the symbol  $\star$  as a metavariable ranging over the binary logical connectives  $\supset$  (for implication),  $\wedge$  (for conjunction),  $\vee$  (for disjunction), and  $\equiv$  (for equivalence). We will use the symbol  $\star$  in this way throughout this paper. The formula  $B_i\varphi$  is assigned the informal reading “[agent]  $i$  believes that  $\varphi$ .”

While the modal language allows us to describe the beliefs held by each of our agents, it does not provide a means of expressing evidence or justification as to why an agent may hold a given belief. To illustrate: observe that while a formula of the form  $B_i\varphi \supset B_i\psi$  expresses the statement that agent  $i$  will believe  $\psi$  in case he also believes  $\varphi$ —thereby connecting his belief of one thing with his belief of another—this modal formula does not provide a reason as to why this connection holds. This is certainly something we would like for our language to express if we are to say that it provides a notion of evidence or justification.

Taking cues from the first work on a language of multi-agent Justification Logic [20], we will introduce our *evidence language*, a new multi-agent version of Justification Logic that remedies the above-mentioned deficiency of modal logic in the same way that Justification Logic remedies this deficiency in general. Namely, we introduce structured syntactical objects called *terms* that are used to represent specific pieces of evidence. If  $t$  is a term and  $\varphi$  is a formula, then we write the formula

$$t :_i \varphi$$

to say that “agent  $i$  has evidence  $t$  that  $\varphi$ .” To see how this allows us to solve the above-mentioned deficiency with modal logic, consider the implication  $(t :_i \varphi) \supset (s_t :_i \psi)$ , where  $s_t$  is a term built by combining  $t$  with other terms using certain term-building operations we will define later. This implication tells us that agent  $i$  connects his evidence  $s_t$  for  $\psi$  with his evidence  $t$  for  $\varphi$  using a specific sequence of reasoning steps that is encoded within the structure of  $s_t$ . This eliminates the above-mentioned deficiency because agent  $i$  now has an explicit connection between  $\psi$  and  $\varphi$ . As an example: our email scenario had Charlie combine his evidence  $x_1$  with his evidence  $x_3$  to form the evidence  $x_1 \cdot_J x_3$  that Anne is moving to Columbus. The structure of the term  $x_1 \cdot_J x_3$  explicitly encodes the reasoning that Charlie underwent in drawing this conclusion; in particular, Charlie used the rule of Modus Ponens to conclude  $M$  (“Anne moves to Columbus”) from the  $x_1$ -evidenced assumption  $J \supset M$  (“If Anne gets the job, then Anne moves to Columbus”) and the  $x_3$ -evidenced assumption  $J$  (“Anne gets the job”). Thus we see that the term  $x_1 \cdot_J x_3$  represents this particular application of Modus Ponens. (The formula  $J$  appears as a subscript in  $x_1 \cdot_J x_3$  because  $J$  plays the key role of the antecedent of  $J \supset M$  in this application of Modus Ponens.)

Our evidence language also allows agents to hold a weaker connection between a piece of evidence and an assertion: we write  $t \gg_i \varphi$  to say that “agent  $i$  has evidence  $t$  relevant to  $\varphi$ .” Intuitively, the formula  $t \gg_i \varphi$  tells us that agent  $i$  thinks it reasonable to take evidence  $t$  into account when considering the truth of  $\varphi$ , but evidence  $t$  is not necessarily strong enough to compel agent  $i$  to believe  $\varphi$ . To use a slight modification of Fitting’s example [9]: if I see an article in a tabloid that says Professor Smartypants, a famous intellectual, has been named editor-in-chief of *The New York Times*, then I would certainly regard this evidence as relevant when considering the truth of the assertion “Professor Smartypants is editor-in-chief

of *The New York Times*,” but I definitely would not believe this assertion until I read it in *The New York Times* itself.

Accordingly, the meaning of the formula  $t \gg_i \varphi$  is to be contrasted with the meaning of the formula  $t :_i \varphi$  in the following way: the latter is a stronger assertion that says that agent  $i$  not only thinks it reasonable to take  $t$  into account when considering  $\varphi$  but also that he believes  $\varphi$ . Put more compactly,  $t :_i \varphi$  says that agent  $i$  believes  $\varphi$  on the basis of evidence  $t$ . Continuing the example about Professor Smartypants: if I now read in *The New York Times* that Professor Smartypants has been named editor-in-chief of *The New York Times*, then not only will I regard this evidence as relevant when considering the truth of the assertion “Professor Smartypants is editor-in-chief of *The New York Times*” but I will also believe that this assertion is true.

So we see that there is to be a connection between the meaning of  $t \gg_i \varphi$  and the meaning of  $t :_i \varphi$ ; in particular, we will later stipulate (both semantically and axiomatically) that

$$(t :_i \varphi) \equiv (t \gg_i \varphi) \wedge B_i \varphi .$$

That is, “agent  $i$  has evidence  $t$  that  $\varphi$  if and only if agent  $i$  has evidence  $t$  relevant to  $\varphi$  and agent  $i$  believes that  $\varphi$ .” With this said, let us now define our multi-agent *evidence language*.

**Definition 2.3** ( $\text{EL}^A$ ). The *evidence language (for  $A$ )*, written  $\text{EL}^A$ , consists of the *terms*  $t$  and the *formulas*  $\varphi$  formed by the following grammar.

$$\begin{aligned} t & ::= c_k \mid x_k \mid t_1 \cdot_{\varphi} t_2 \mid t_1 + t_2 \mid !t \\ \varphi & ::= p_k \mid \perp \mid \top \mid \varphi_1 \star \varphi_2 \mid \neg \varphi \mid B_i \varphi \mid t \gg_i \varphi \mid t :_i \varphi \\ & \quad k \in \mathbb{N}, \star \in \{\supset, \wedge, \vee, \equiv\}, i \in A \end{aligned}$$

A term of the form  $c_k$  is called a *constant* and a term of the form  $x_k$  is called a *variable*. To say that a term  $t$  is *variable-free* means that no non-subscript occurrence of a term in  $t$  is a variable. (Examples:  $c_0 \cdot_{x_1 :_i p_5} c_2$  is variable-free, whereas  $c_0 \cdot_{x_1 :_i p_5} x_2$  is not.) A formula of the form  $t \gg_i \varphi$  is called a *relevance assertion* and is assigned the informal reading “[agent]  $i$  has evidence  $t$  relevant to  $\varphi$ .” A formula of the form  $t :_i \varphi$  is called an *evidence assertion* and is assigned the informal reading “[agent]  $i$  has evidence  $t$  that  $\varphi$ .”

The operations of term formation listed in the grammar of Definition 2.3 are used to build more complicated terms out of simpler terms. As we described above, our intention is that each operation will correspond to a logical operation of evidence combination. We already saw that the operation  $\cdot_{\varphi}$  is used to indicate Modus Ponens (with antecedent  $\varphi$ ). For convenience, let us indicate the intended meanings for each of these operations.

- If  $t$  is evidence that  $\varphi \supset \psi$  and  $s$  is evidence that  $\varphi$ , then  $t \cdot_{\varphi} s$  is evidence that  $\psi$ . So  $\cdot_{\varphi}$  is just the operation of Modus Ponens (with antecedent  $\varphi$ ).
- If  $t_1$  or  $t_2$  is evidence that  $\varphi$ , then  $t_1 + t_2$  is evidence that  $\varphi$ . So  $+$  is a monotonic combination of evidence, in that  $t_1 + t_2$  is evidence for all of those things that one or more of  $t_1$  or  $t_2$  evidences.

- If  $t$  evidences  $\varphi$  for agent  $i$ , then  $!t$  evidences  $t{:}_i\varphi$  for agent  $i$ . So  $!$  is an evidence checker, in that  $!t$  is agent  $i$ 's verification that  $t$  is his evidence for  $\varphi$  (in case  $t$  really is his evidence for  $\varphi$ ).

Finally, it will be useful to be able to refer to the set of terms of a given language.

**Notation 2.4.** For a language  $\mathfrak{L}$ , we write  $\mathcal{T}(\mathfrak{L})$  to denote the set of all terms in  $\mathfrak{L}$ . Examples:  $\mathcal{T}(\text{ML}^A)$  denotes the empty set and  $\mathcal{T}(\text{EL}^A)$  denotes the set of all terms that can be formed using the grammar in Definition 2.3.

## 2.2 Evidence Labelings

We will soon introduce our *update language*, which extends our evidence language by adding syntax to express belief dynamics and evidence elimination. But we first need some machinery in order to describe both the semantics of this language and also how it is that this language describes its many communicative updates. As in Dynamic Epistemic Logic [6, 18, 19], both the semantics and the internal structure of the communicative updates will be based on Kripke frames.

**Definition 2.5.** Let  $A$  be an agent set. A *Kripke frame (for  $A$ )* is a tuple  $F = (W, R)$  satisfying each of the following.

- $W$  is a nonempty set. We refer to the elements of  $W$  *worlds (in  $F$ )* and write  $\Gamma \in F$  to mean that  $\Gamma$  is a world in  $F$ .
- $R : A \rightarrow 2^{W \times W}$  is a function mapping each agent  $i \in A$  to a binary relation  $R_i$  on  $W$ .<sup>1</sup>

To say that the Kripke frame  $(W, R)$  is *finite* means that  $W$  is finite. To say that the Kripke frame  $(W, R)$  satisfies a property of binary relations (such as transitivity) means that  $R_i$  satisfies this property for each  $i \in A$ .<sup>2</sup>

Kripke frames are used to describe agent uncertainty. The set of worlds simply enumerates all possible states of affairs that might obtain. And whenever we have that  $\Gamma R_i \Delta$ , then this tells us that in case  $\Gamma$  turns out to be the actual state of affairs, then agent  $i$  will consider it possible that the actual state of affairs is  $\Delta$ . In this way, if  $\Omega$  is the actual state of affairs, then the more worlds  $\Delta$  such that  $\Omega R_i \Delta$  holds, the less agent  $i$  is certain as to which state of affairs is the real one. Note that if  $\Omega$  is the actual state of affairs and yet we do not have that  $\Omega R_i \Omega$ , then agent  $i$  will not even consider the actual state of affairs as a possibility.

By adding certain features to a Kripke frame, we will obtain a model for interpreting formulas in our update language. But by adding other features to a finite Kripke frame, we will build the internal structure of our *update modals*, which our formal language will use to

---

<sup>1</sup>A *binary relation (on  $W$ )* is just a subset of  $W \times W$ . If  $R_i$  is a binary relation on  $W$ , we write  $\Gamma R_i \Delta$  to mean that  $(\Gamma, \Delta) \in R_i$ .

<sup>2</sup>To say that a binary relation  $R$  on  $W$  is *transitive* means that  $\Gamma R \Delta$  and  $\Delta R \Omega$  together imply that  $\Gamma R \Omega$  for each  $\Gamma, \Delta, \Omega \in W$ .

express what happens after an occurrence of a given communicative update whose meaning is defined by the way in which the added features fit within the structure of the underlying finite Kripke frame.

This is perhaps a bit vague right now, but it will be made clear soon enough. For now, we simply mention that one of the key features we must add to a (finite) Kripke frame, both for semantic and for syntactic purposes, is an *evidence labeling*.

**Definition 2.6.** Let  $F = (W, R)$  be a Kripke frame for an agent set  $A$  and let  $\mathcal{L}$  be a language. An *evidence labeling (in  $\mathcal{L}$  on  $F$ )* is a function

$$\mathcal{A} : A \rightarrow (\mathcal{T}(\mathcal{L}) \times \mathcal{L} \rightarrow 2^W)$$

that maps each agent  $i \in A$  to a function  $\mathcal{A}_i : \mathcal{T}(\mathcal{L}) \times \mathcal{L} \rightarrow 2^W$  that itself maps each term-formula pair  $(t, \varphi) \in \mathcal{T}(\mathcal{L}) \times \mathcal{L}$  to a set  $\mathcal{A}_i(t, \varphi)$  of worlds in  $F$ . We adopt the following notation and terminology with respect to an evidence labeling  $\mathcal{A}$  in  $\mathcal{L}$  on  $F$ .

- For each  $i \in A$ , we define  $\text{dom}(\mathcal{A}_i)$  as the set of all term-formula pairs  $(t, \varphi)$  such that  $\mathcal{A}_i(t, \varphi)$  is a nonempty set.<sup>3</sup> We call  $\text{dom}(\mathcal{A}_i)$  the *domain* of  $\mathcal{A}_i$ .<sup>4</sup>
- To say that  $\mathcal{A}$  is *finite* means that for each  $i \in A$ , we have that  $\text{dom}(\mathcal{A}_i)$  is finite and that the set  $\mathcal{A}_i(t, \varphi)$  is finite for each  $(t, \varphi) \in \text{dom}(\mathcal{A}_i)$ .

For semantic purposes, we will use an evidence labeling to describe evidence relevancy; that is, we will let  $\mathcal{A}_i(t, \varphi)$  be the set of worlds at which the relevance assertion  $t \gg_i \varphi$  is true.

But for syntactic purposes, we will use a finite evidence labeling to describe evidence elimination. The idea is that the finite Kripke frame  $F$  on which the finite evidence labeling  $\mathcal{A}$  is defined represents the set of all possible evidence elimination that might occur. In particular, for each world  $w \in F$ , the elimination  $\mathcal{A}(w)$  that  $\mathcal{A}$  associates with  $w$  is given as follows: whenever  $t \gg_i \varphi$  is a formula such that  $w \in \mathcal{A}_i(t, \varphi)$ , then the elimination  $\mathcal{A}(w)$  will cause agent  $i$  to eliminate the evidence  $t$  that  $\varphi$ , making it so that  $t \gg_i \varphi$  is false.

We will eventually wrap a finite Kripke frame  $F$ , a finite evidence labeling  $\mathcal{A}$ , and one more finite component all together into a single symbol  $U$ . This will form our communicative update.<sup>5</sup> We will then use the pair  $(U, u)$  to describe a communicative update that initiates

<sup>3</sup>Thus  $\text{dom}(\mathcal{A}_i) := \{ (t, \varphi) \in \mathcal{T}(\mathcal{L}) \times \mathcal{L} : (\exists S \in 2^W)[((t, \varphi), S) \in \mathcal{A}_i \wedge S \neq \emptyset] \}$ .

<sup>4</sup>Note that this is a non-standard use of the word “domain.” Usually the “domain” of the function  $\mathcal{A}_i : (\mathcal{T}(\mathcal{L}) \times \mathcal{L}) \rightarrow 2^W$  would be identified with the set  $\mathcal{T}(\mathcal{L}) \times \mathcal{L}$ . But our primary interest will be in those elements  $(t, \varphi)$  in  $\mathcal{T}(\mathcal{L}) \times \mathcal{L}$  for which  $\mathcal{A}_i(t, \varphi)$  is a non-empty set. So this is the reason for our non-standard definition of “domain.” Nonetheless, we can twist the standard definition so as to line-up with our non-standard definition if we view  $\mathcal{A}_i$  as a *partial function*  $\mathcal{T}(\mathcal{L}) \times \mathcal{L} \rightarrow 2^W$ , by which we mean that  $\mathcal{A}_i$  is a function between a subset of  $\mathcal{T}(\mathcal{L}) \times \mathcal{L}$  (in fact, the subset  $\text{dom}(\mathcal{A}_i)$ ) and  $2^W$ , as opposed to a function between the full set  $\mathcal{T}(\mathcal{L}) \times \mathcal{L}$  and  $2^W$ . Viewed in this way,  $\text{dom}(\mathcal{A}_i)$  is exactly what is meant when one refers to the standard definition of the “domain” of the *partial function*  $\mathcal{A}_i$ . So as to avoid any potential notational and explanatory baggage that may come along with defining  $\mathcal{A}_i$  as a partial function (as opposed to a proper function), we have settled on this non-standard definition of “domain.”

<sup>5</sup>Since each component of  $U$  is finite, then we can in principle write down some complicated symbolic expression that completely describes  $U$ .

the elimination  $\mathcal{A}(u)$ , though the agents' uncertainty as to which elimination is actually initiated is given by the structure of the underlying finite Kripke frame  $F$  and the way  $\mathcal{A}$  assigns eliminations within this structure. This allows us to describe not just the action of a single elimination but also the affect of a single elimination that sits within an arbitrarily complicated network of agent uncertainty over various possible eliminations.

After getting all the definitions straight, we will write the formula  $[U, u]\varphi$  to say that  $\varphi$  is true after the occurrence of the communicative update  $(U, u)$ . But before we can lay out the details of how this works, we need a bit more machinery.

**Definition 2.7** (Extension). Let  $F = (W, R)$  be a Kripke frame for an agent set  $A$ , and let  $\mathcal{L}$  be a language. To say that an evidence labeling  $\mathcal{A}$  (in  $\mathcal{L}$  on  $F$ ) *extends* (or *is an extension of*) an evidence labeling  $\mathcal{A}'$  (in  $\mathcal{L}$  on  $F$ ), written

$$\mathcal{A} \subseteq \mathcal{A}' ,$$

means that for each  $i \in A$  and each  $(t, \varphi) \in \mathcal{T}(\mathcal{L}) \times \mathcal{L}$ , we have  $\mathcal{A}_i(t, \varphi) \subseteq \mathcal{A}'_i(t, \varphi)$ . For a property  $\mathfrak{P}$  of evidence labelings in  $\mathcal{L}$  on  $F$ , to say that  $\mathcal{A}_*$  is the *smallest (evidence labeling in  $\mathcal{L}$  on  $F$ ) satisfying  $\mathfrak{P}$*  means that for each evidence labeling  $\mathcal{A}''$  in  $\mathcal{L}$  on  $F$  that satisfies  $\mathfrak{P}$ , we have that  $\mathcal{A}_* \subseteq \mathcal{A}''$ . Observation: if there exists a smallest evidence labeling in  $\mathcal{L}$  on  $F$  satisfying a property, then this is the unique smallest evidence labeling in  $\mathcal{L}$  on  $F$  satisfying that property.

From the perspective of semantics, having  $\mathcal{A} \subseteq \mathcal{A}'$  says that  $\mathcal{A}$  makes fewer relevance assertions true than does  $\mathcal{A}'$ . From the perspective of syntax, having  $\mathcal{A} \subseteq \mathcal{A}'$  says that  $\mathcal{A}$  eliminates fewer pieces of evidence than does  $\mathcal{A}'$ .

**Definition 2.8** ( $\mathcal{A}^{\mathcal{L}, F}$ ). Let  $F = (W, R)$  be a Kripke frame for an agent set  $A$ , and let  $\mathcal{L}$  be a language. The *full evidence labeling (in  $\mathcal{L}$  on  $F$ )*, written  $\mathcal{A}^{\mathcal{L}, F}$ , is defined by setting

$$\mathcal{A}_i^{\mathcal{L}, F}(t, \varphi) = W$$

for each  $i \in A$  and each  $(t, \varphi) \in \mathcal{T}(\mathcal{L}) \times \mathcal{L}$ . It is easy to see that for each evidence labeling  $\mathcal{A}$  in  $\mathcal{L}$  on  $F$ , we have that  $\mathcal{A} \subseteq \mathcal{A}^{\mathcal{L}, F}$ .

The full evidence labeling will have a special role within a few proofs but will otherwise not play a significant role.

**Definition 2.9** (Intersection). Let  $F = (W, R)$  be a Kripke frame for an agent set  $A$ , and let  $\mathcal{L}$  be a language. For a nonempty set  $\mathcal{S}$  of evidence labelings in  $\mathcal{L}$  on  $F$ , the *intersection of  $\mathcal{S}$* , written  $\bigcap \mathcal{S}$ , is the evidence labeling (in  $\mathcal{L}$  on  $F$ ) defined by setting

$$(\bigcap \mathcal{S})_i(t, \varphi) := \bigcap_{\mathcal{A} \in \mathcal{S}} \mathcal{A}_i(t, \varphi)$$

for each  $i \in A$  and each  $(t, \varphi) \in \mathcal{T}(\mathcal{L}) \times \mathcal{L}$ .

The operation of intersection will likewise play a role only within proofs.

## 2.3 The Update Language

Let us now fill in the details about the internal structure of the component  $U$  of a communicative update  $(U, u)$ .

**Definition 2.10.** Let  $A$  be an agent set and  $\mathcal{L}$  be a language. An *update frame* (in  $\mathcal{L}$  for  $A$ ) is a tuple

$$U = (W, R, f, \mathcal{A})$$

satisfying each of the following.

- $(W, R)$  is a Kripke frame for  $A$  that is finite and transitive.  
We call  $(W, R)$  the Kripke frame *underlying*  $U$ , and say that  $U$  is the update frame *based on*  $(W, R)$ . We write  $u \in U$  to mean that  $u \in W$ . For each  $u \in U$ , we say that  $u$  is a *world* (in  $U$ ).
- $f : W \rightarrow \mathcal{L}$  is a function mapping each world  $u \in W$  to a formula  $f(u) \in \mathcal{L}$ .
- $\mathcal{A}$  is a finite evidence labeling in  $\mathcal{L}$  on  $(W, R)$ . We call  $\mathcal{A}$  the *elimination basis* (in  $U$ ).

A *pointed update frame* (in  $\mathcal{L}$  for  $A$ ) is a pair  $(U, u)$  consisting of an update frame in  $\mathcal{L}$  for  $A$  and a world  $u \in U$ ; we call  $u$  the *point* of the pair  $(U, u)$ .

A few words on the definition of update frame. First, we require the underlying finite Kripke frame to be transitive so as to ensure that our communicative updates preserve a property we will impose on agent belief.<sup>6</sup>

Second, let us say something about the function  $f$ . This function is a formula-labeling function that assigns a formula to each world in the finite Kripke model  $F = (W, R)$ . We use this labeling function to describe the network of possible formulas that might be communicated to the agents by a communicative update based on  $U$ . This part of our work is taken out of the standard approach in Dynamic Epistemic Logic [6, 18, 19]. The basic idea: in addition to providing a description of the agents' uncertainty over a network of possible eliminations given by  $\mathcal{A}$ , the finite Kripke frame  $F$  also provides a description of the agents' uncertainty over a network of possible communications of a formula. For each world  $w \in F$ , the formula  $f(w)$  is communicated. So the pointed update frame  $(U, u)$  will specify a communicative update that initiates the communication of  $f(u)$  and the elimination  $\mathcal{A}(u)$ , though the agents' uncertainty as to which combination of communication and elimination is actually initiated is given by the structure of the finite Kripke frame  $F$  in conjunction with the way  $f$  assigns communications and  $\mathcal{A}$  assigns eliminations within this structure.

---

<sup>6</sup>To wit: we will impose the property of *positive introspection* on agent belief. This means that each agent will believe his own beliefs (in the sense that if agent  $i$  believes  $\varphi$ , then he believes that he believes  $\varphi$ ). This is perhaps a philosophically contentious property, but we have adopted it so as to ensure the smoothness in developing our first system for reasoning about belief dynamics and evidence elimination. There are various ways that this property can be abandoned [17], but we wish to avoid discussing variant systems and their provable properties in the present paper in the interest of remaining focused on the key issues at hand: belief dynamics and evidence elimination.

To simplify much of what is to come, we introduce the following notation for update frames.

**Notation 2.11.** Let  $A$  be an agent set, let  $\mathfrak{L}$  be a language, and let  $U = (W, R, f, \mathcal{A})$  be an update frame in  $\mathfrak{L}$  for  $A$ .

- We write  $U$  where one would expect a set as an abbreviation for  $W$ .  
Example: “ $(\Delta, v) \in (S \times U)$ ” is an abbreviation for “ $(\Delta, v) \in (S \times W)$ ”.
- We write  $U$  where one would expect a Kripke frame as an abbreviation for  $(W, R)$ .  
Example: we will later define a theory that takes a Kripke frame  $F$ , an evidence labeling  $\mathcal{A}$  on  $F$ , and a world  $w \in F$  and then allows us to write assertions such as  $(F, \mathcal{A}), w \vdash t \gg_i \varphi$ . Accordingly, “ $(U, \mathcal{A}), w \vdash t \gg_i \varphi$ ” will be an abbreviation for “ $((W, R), \mathcal{A}), w \vdash t \gg_i \varphi$ ”.
- For each  $i \in A$ , we write  $wU_i v$  as an abbreviation for  $wR_i v$ .
- We write  $U^l$  as an abbreviation for the function  $f$ .
- For each  $u \in U$ , we write  $U(u)$  as an abbreviation for the  $\mathfrak{L}$ -formula  $U^l(u)$ .
- We write  $U^e$  as an abbreviation for the elimination basis  $\mathcal{A}$  in  $U$ .
- For each  $i \in A$ , we write  $U_i(t, \psi)$  as an abbreviation for  $U_i^e(t, \psi)$ .

Adopting this notation allows us to refer to the components of an update frame  $U$  without having to explicitly name these components, thereby obviating the need for us to introduce extra symbols to name the components of an update frame that is under discussion.

Finally, we define our *update language*.

**Definition 2.12** ( $UL^A$ ). Let  $A$  be an agent set. The *update language (for  $A$ )*, written  $UL^A$ , consists of the *terms*  $t$  and the *formulas*  $\varphi$  formed by the following grammar.

$$\begin{aligned}
t & ::= c_k \mid x_k \mid t_1 \cdot_{\varphi} t_2 \mid t_1 + t_2 \mid !t \mid t^{U,u} \\
\varphi & ::= p_k \mid \perp \mid \top \mid \varphi_1 \star \varphi_2 \mid \neg\varphi \mid B_i\varphi \mid t \gg_i \varphi \mid t :_i \varphi \mid t \gg_i^{U,u} \varphi \mid [U, u]\varphi \\
& \quad k \in \mathbb{N}, \star \in \{\supset, \wedge, \vee, \equiv\}, i \in A
\end{aligned}$$

Crucial requirement: in the above grammar,  $(U, u)$  is a metavariable ranging over pointed update frames for  $A$  that satisfy each of the following items.

- For each  $v \in U$ , we have that  $U(v)$  is a formula that has been constructed using the above grammar.
- The language of the elimination basis  $U^e$  consists of formulas that have been constructed using the above grammar.<sup>7</sup>

---

<sup>7</sup>So for the elimination basis  $U^e : A \rightarrow (\mathcal{T}(\mathfrak{L}) \times \mathfrak{L} \rightarrow 2^W)$  in  $U$ , we have that each formula in  $\mathfrak{L}$  was constructed using the grammar in Definition 2.12.

A formula of the form  $t \gg_i^{U,u} \varphi$  is called an *elimination assertion* and is assigned the informal reading “[update]  $(U, u)$  eliminates  $t$  for (agent)  $i$ .” A formula of the form  $[U, u]\varphi$  is called an *update assertion* and is assigned the informal reading “after [update]  $(U, u)$ , [we have that]  $\varphi$  [is true].” The modals  $[U, u]$  in the language  $\text{UL}^A$  make up the set of *update modals* (for  $A$ ).

Here we have introduced two new rules of formula formation. The first rule allows us to build formulas of the form  $t \gg_i^{U,u} \varphi$ . These formulas allow the update language to “look inside” the internal structure of an update modal  $[U, u]$ , in that the formula  $t \gg_i^{U,u} \varphi$  tells us that the elimination  $U^e(u)$  will cause agent  $i$  to eliminate the evidence  $t$  for  $\varphi$ .

The second new rule of formula formation allows us to build formulas of the form  $[U, u]\varphi$ , which tell us that  $\varphi$  is true after the communicative update  $(U, u)$  occurs. Note that we also introduce a new rule of term formation, allowing us to build terms of the form  $t^{U,u}$ . As we described above, the term operations correspond to logical operations of evidence combination. While we described the intended meanings of most of these operations on page 9, we have not yet said the intended meaning for the operation that lets us form the term  $t^{U,u}$  from an existing term  $t$ . So let us say this now.

- If  $t$  is evidence for  $\varphi$ , then  $t^{U,u}$  is evidence for  $[U, u]\varphi$ . So if we think of  $t$  as very strong evidence that  $\varphi$  must always hold, then  $t^{U,u}$  explains that  $\varphi$  will accordingly hold even after an occurrence of the communicative update  $(U, u)$ .

Now that we have completed the definition of our update language, let us proceed toward defining its intended semantics.

### 3 Semantics

The intended semantics of the evidence language is called the *Fitting semantics*. This semantics, due to Fitting [9] and Mkrtychev [15], is obtained from Kripke’s semantics for modal logic [7, 13] by the addition of *evidence functions*, which regulate the truth of relevance assertions  $t \gg_i \varphi$  within models. We describe evidence functions in detail in the next subsection.

The semantics for the update language is an extension of the Fitting semantics that is based closely on the work of Baltag, Moss, and Solecki (BMS) [3, 4, 5]. We will extend the *BMS semantics* for update formulas  $[U, u]\varphi$  by providing an account of how the elimination basis  $U^e$  affects the evidence function in a way that eliminates relevant evidence, assuring us that certain relevance assertions  $t \gg_i \varphi$  will be false after the occurrence of the update described by  $(U, u)$ . In this way, an elimination basis will specify how it is that evidence is to be eliminated. We describe all of this in detail in a moment, once we have introduced the notions of *evidence function* and *elimination function*.

### 3.1 Evidence Functions

An *evidence function* is an evidence labeling on a Kripke frame that satisfies certain properties that together guarantee that the term-forming operations of the evidence language (Definition 2.3) have their intended meanings.

**Definition 3.1.** Let  $F = (W, R)$  be a Kripke frame for an agent set  $A$  and let  $S$  be a set. An  $S$ -*evidence function* (on  $F$ ) is an evidence labeling (in  $\text{UL}^A$  on  $F$ ) satisfying each of the following schematic properties.

- *Constant Specification  $S$ .* For each  $k \in \mathbb{N}$  and each  $(c_k :_i \varphi) \in (S \cap \text{UL}^A)$ , we have  $\mathcal{A}_i(c_k, \varphi) = W$ .

Note that the property of Constant Specification  $S$  requires us to provide a set  $S$  as a parameter. Due to the way we have defined this property, we may assume without loss of generality that the parameter set  $S$  is merely a set of  $\text{UL}^A$ -formulas.

- *Application.*  $\mathcal{A}_i(t, \varphi \supset \psi) \cap \mathcal{A}_i(s, \varphi) \subseteq \mathcal{A}_i(t \cdot_\varphi s, \psi)$ .
- *Sum.*  $\mathcal{A}_i(t, \varphi) \cup \mathcal{A}_i(s, \varphi) \subseteq \mathcal{A}_i(t + s, \varphi)$ .
- *Checker.*  $\mathcal{A}_i(t, \varphi) \subseteq \mathcal{A}_i(!t, t :_i \varphi)$ .
- *Monotonicity.*  $\Gamma R_i \Delta$  and  $\Gamma \in \mathcal{A}_i(t, \varphi)$  together imply that  $\Delta \in \mathcal{A}_i(t, \varphi)$ .
- *Update.*  $\mathcal{A}_i(t, \varphi) \subseteq \mathcal{A}_i(t^{U,u}, [U, u]\varphi)$ .

When the particular set  $S$  is unimportant or may be inferred from context, we may drop the prefix “ $S$ -” in referring to an  $S$ -evidence function.

We will eventually connect the truth of the relevance assertion  $t \gg_i \varphi$  at a world  $\Gamma$  with the statement that  $\Gamma \in \mathcal{A}_i(t, \varphi)$  holds for a given evidence function  $\mathcal{A}$ . So we see that evidence functions will be used to regulate the truth of relevance assertions.

The set  $S$  in an evidence function represents the set of “basic” assertions that can be accepted without justification. In a more nuanced presentation than that undertaken in the current paper, one could take  $S$  to be whatever set of  $\text{UL}^A$ -formulas that one wishes to take as “basic” for a given application at hand (including the possibility  $S = \emptyset$ ). But for present purposes, where our focus is on the issues of belief dynamics and evidence elimination, we will take  $S$  as the set of all axioms in our to-be-defined axiomatic theory. In this way, the basic assertions that we allow our agents to accept without justification are simply the to-be-defined axioms.

Let us now provide intuitive descriptions of the defining properties of an evidence function (Definition 3.1). The property of Constant Specification  $S$  says that each agent uses the constants as the pieces of evidence that are relevant to the basic assertions, the axioms. The property of Application ensures that the term-forming operation  $\cdot_\varphi$  is used to represent closure of relevant evidence under the application of Modus Ponens. The property of Sum ensures that the term-forming operation  $+$  is used to represent the monotonic combination

of relevant pieces of evidence, so that  $t+s$  is a piece of evidence relevant to anything to which either of  $t$  or  $s$  is a relevant piece of evidence. The property of Checker ensures that the term-forming operation  $!$  is used to represent an agent’s ability to internally verify relevant evidence assertions: if  $t$  is evidence relevant to  $\varphi$  for agent  $i$ , then  $!t$  is evidence relevant to the assertion that  $t :_i \varphi$  (“agent  $i$  has evidence  $t$  for  $\varphi$ ”). The property of Monotonicity says that an agent believes his own evidence: if  $t$  is a piece of evidence relevant to  $\varphi$  for agent  $i$ , then for each possible state of affairs envisioned by the agent,  $t$  is evidence relevant to  $\varphi$ . The property of Update says that an agent has a strong respect for his evidence: if  $t$  is evidence relevant to  $\varphi$ , then  $t^{U,u}$  is evidence relevant to the assertion that  $\varphi$  is true (even) after the update  $(U, u)$  occurs.

The following lemma guarantees that every evidence labeling may be extended in a minimal way so as to guarantee that the resulting evidence labeling is in fact an evidence function. This process produces what we call the evidence function *generated by* the original evidence labeling.

**Lemma 3.2** (Generation Lemma; [14]). Let  $F = (W, R)$  be a Kripke frame for an agent set  $A$ , let  $\mathcal{A}$  be an evidence labeling in  $\text{UL}^A$  on  $F$ , and let  $S$  be a set. There exists a smallest  $S$ -evidence function  $\mathcal{A}^{*,S}$  on  $F$  that extends  $\mathcal{A}$ . We call  $\mathcal{A}^{*,S}$  the  *$S$ -evidence function generated by  $\mathcal{A}$* .

*Proof.* Kuznets’ thesis [14] studied in detail the issue of whether there exists a smallest evidence functions satisfying certain properties, and the proof of the present lemma follows by a trivial adaptation of the tools and techniques presented in that work. Nonetheless, it is illustrative for us to write the argument here in full.

Let  $\mathcal{S}^*$  be the set of all evidence labelings  $\mathcal{A}'$  in  $\text{UL}^A$  on  $F$  such that  $\mathcal{A} \subseteq \mathcal{A}'$  and  $\mathcal{A}'$  is an  $S$ -evidence function on  $F$ . Observe that  $\mathcal{S}^*$  is nonempty because  $\mathcal{A}^{\text{UL}^A, F} \in \mathcal{S}^*$  and, further, it is clear that  $\mathcal{A}^{\text{UL}^A, F}$  satisfies Constant Specification  $S$ , Application, Sum, Checker, Monotonicity, and Update (see Definition 2.8). Now define the evidence labeling  $\mathcal{A}^{*,S}$  by setting  $\mathcal{A}^{*,S}(t, \varphi) := \bigcap \mathcal{S}^*$ . It is not difficult to see that  $\mathcal{A} \subseteq \mathcal{A}^{*,S}$  and that  $\mathcal{A}^{*,S}$  satisfies Constant Specification  $S$ , Application, Sum, Checker, Monotonicity, and Update. Further, for each  $\mathcal{A}' \in \mathcal{S}^*$ , it is clear from our use of an intersection in defining  $\mathcal{A}^{*,S}$  that we have  $\mathcal{A}^{*,S} \subseteq \mathcal{A}'$ . It follows that  $\mathcal{A}^{*,S}$  is the smallest  $S$ -evidence function on  $F$  that extends  $\mathcal{A}$ .  $\square$

Kuznets defined a class of axiomatic systems called *\*-calculi* that can be used to study how to generate evidence labelings that meet various criteria [14]. One of these \*-calculi allows us to provide an axiomatic characterization of the  $S$ -evidence functions generated by an evidence labeling. This \*-calculus will be useful to us in the remainder of this paper, so we now repackage this particular \*-calculus for our own purposes. We call the repackaged version the  $(F, \mathcal{A}, S)$ -calculus.

**Definition 3.3** ( $(F, \mathcal{A}, S)$ -calculus; [14]). Let  $F = (W, R)$  be a Kripke frame for an agent set  $A$ , let  $\mathcal{A}$  be an evidence labeling in  $\text{UL}^A$  on  $F$ , and let  $S$  be a set. The  $(F, \mathcal{A}, S)$ -calculus is the theory defined in Figure 1, where the expressions appearing on the left-hand side of the turnstile (“ $\vdash$ ”) are to be understood as metavariables ranging over worlds in  $F$  and the

AXIOM SCHEMES

- CS.  $w \vdash c_k \gg_i \varphi$  if  $(c_k :_i \varphi) \in (S \cap \text{UL}^A)$   
 EX.  $w \vdash t \gg_i \varphi$  if  $w \in \mathcal{A}_i(t, \varphi)$

RULES

$$\frac{w \vdash t \gg_i (\varphi \supset \psi) \quad w \vdash s \gg_i \varphi}{w \vdash (t \cdot_\varphi s) \gg_i \psi} \text{ (A)}$$

$$\frac{w \vdash t \gg_i \varphi}{w \vdash (t + s) \gg_i \varphi} \text{ (SL)} \quad \frac{w \vdash s \gg_i \varphi}{w \vdash (t + s) \gg_i \varphi} \text{ (SR)}$$

$$\frac{w \vdash t \gg_i \varphi}{w \vdash !t \gg_i (t :_i \varphi)} \text{ (C)}$$

$$\frac{w \vdash t \gg_i \varphi \quad w R_i v}{v \vdash t \gg_i \varphi} \text{ (M)}$$

$$\frac{w \vdash t \gg_i \varphi}{w \vdash t^{U,u} \gg_i [U, u] \varphi} \text{ (U)}$$

(Note: adapted from Kuznets' \*-calculi [14])

Figure 1: The  $(F, \mathcal{A}, S)$ -calculus (with  $F = (W, R)$ )

expressions appearing on the right-hand side of the turnstile are to be understood as schemes for formulas in the language  $\text{UL}^A$ . Notation: for a relevance assertion  $t \gg_i \chi$  in the language  $\text{UL}^A$  and a world  $w \in F$ , we write  $(F, \mathcal{A}, S), w \vdash t \gg_i \chi$  to mean that there is a derivation of  $w \vdash t \gg_i \chi$  in the  $(F, \mathcal{A}, S)$ -calculus, and we write  $(F, \mathcal{A}, S), w \not\vdash t \gg_i \chi$  to mean that there is not a derivation of  $w \vdash t \gg_i \chi$  in the  $(F, \mathcal{A}, S)$ -calculus. In using this notation, we may omit mention of “ $(F, \mathcal{A}, S)$ ” when doing so ought not cause confusion.

In the following lemma, we show that the  $(F, \mathcal{A}, S)$ -calculus describes the evidence function generated by the evidence labeling  $\mathcal{A}$  on  $F$ .

**Lemma 3.4** (Axiomatic Generation Lemma; [14]). Let  $F = (W, R)$  be a Kripke frame for an agent set  $A$ , let  $\mathcal{A}$  be an evidence labeling in  $\text{UL}^A$  on  $F$ , and let  $S$  be a set of  $\text{UL}^A$ -formulas. Define the evidence labeling  $\mathcal{A}'$  (in  $\text{UL}^A$  on  $F$ ) by setting

$$\mathcal{A}'_i(t, \varphi) := \{w \in F : (F, \mathcal{A}, S), w \vdash t \gg_i \varphi\}$$

for each  $i \in A$  and each  $(t, \varphi) \in \mathcal{T}(\text{UL}^A) \times \text{UL}^A$ . Then  $\mathcal{A}'$  is  $\mathcal{A}^{*,S}$ , the  $S$ -evidence function generated by  $\mathcal{A}$  (Lemma 3.2).

*Proof.* It was shown in the proof of Lemma 3.2 that if we let  $\mathcal{S}^*$  be the nonempty set of all evidence labelings  $\mathcal{A}''$  in  $\text{UL}^A$  on  $F$  such that  $\mathcal{A} \subseteq \mathcal{A}''$  and  $\mathcal{A}''$  is an  $S$ -evidence function on

$F$ , then  $\mathcal{A}^{*,S}$  satisfies the equality  $\mathcal{A}_i^{*,S}(t, \varphi) = \bigcap \mathcal{S}^*$ . Note that  $\mathcal{A}' \in \mathcal{S}^*$  follows because we have that  $\mathcal{A} \subseteq \mathcal{A}'$  by Axiom EX and, further, that  $\mathcal{A}'$  satisfies Constant Specification  $S$  by Axiom CS, Application by Rule A, Sum by Rules SL and SR, Checker by Rule C, Monotonicity by Rule M, and Update by Rule U. So it suffices for us to show that for each  $\mathcal{A}'' \in \mathcal{S}^*$ , we have that  $\mathcal{A}' \subseteq \mathcal{A}''$ . To do this, we make an arbitrary choice of an evidence labeling  $\mathcal{A}'' \in \mathcal{S}^*$  and of a world  $w \in F$  and then show by induction on the length of a derivation in the  $(F, \mathcal{A}, S)$ -calculus that  $w \vdash t \gg_i \varphi$  implies  $w \in \mathcal{A}_i''(t, \varphi)$ .

In the base case of this induction,  $w \vdash t \gg_i \varphi$  is an axiom. In case this axiom is Axiom CS (see Figure 1), then  $t = c_k$  with  $(c_k :_i \varphi) \in (S \cap \text{UL}^A)$  and so it follows that  $w \in \mathcal{A}_i''(t, \varphi)$  because  $\mathcal{A}''$  satisfies Constant Specification  $S$ . In case  $w \vdash t \gg_i \varphi$  is Axiom EX, then  $w \in \mathcal{A}_i''(t, \varphi)$  because  $\mathcal{A} \subseteq \mathcal{A}''$ . This completes the base cases of our induction, so let us proceed with the inductive cases by examining each of the rules of inference from Figure 1.

- Inductive Case: the last step of the derivation of  $w \vdash t \gg_i \varphi$  made use of Rule A.

It follows from the form of Rule A that  $t = s_1 \cdot_\psi s_2$  and that each of  $w \vdash s_1 \gg_i (\psi \supset \varphi)$  and  $w \vdash s_2 \gg_i \psi$  occurs earlier in the derivation. Applying the inductive hypothesis, we have that  $w \in \mathcal{A}_i''(s_1, \psi \supset \varphi)$  and  $w \in \mathcal{A}_i''(s_2, \psi)$ . But since  $\mathcal{A}''$  satisfies Application, it follows that  $w \in \mathcal{A}_i''(s_1 \cdot_\psi s_2, \varphi)$ . Since  $t = s_1 \cdot_\psi s_2$ , we conclude that  $w \in \mathcal{A}''(t, \varphi)$ .

- Inductive Case: the last step of the derivation of  $w \vdash t \gg_i \varphi$  made use of Rule SL.

It follows from the form of Rule SL that  $t = s_1 + s_2$  and that  $w \vdash s_1 \gg_i \varphi$  occurs earlier in the derivation. Applying the inductive hypothesis, we have that  $w \in \mathcal{A}_i''(s_1, \varphi)$ . But since  $\mathcal{A}''$  satisfies Sum, it follows that  $w \in \mathcal{A}_i''(s_1 + s_2, \varphi)$ . Since  $t = s_1 + s_2$ , we conclude that  $w \in \mathcal{A}''(t, \varphi)$ .

- Inductive Case: the last step of the derivation of  $w \vdash t \gg_i \varphi$  made use of Rule SR.

The argument is almost identical to the case for Rule SL.

- Inductive Case: the last step of the derivation of  $w \vdash t \gg_i \varphi$  made use of Rule C.

It follows from the form of Rule C that  $t = !s$ , that  $\varphi = s :_i \psi$ , and that  $w \vdash s \gg_i \psi$  occurs earlier in the derivation. Applying the inductive hypothesis, we have that  $w \in \mathcal{A}_i''(s, \psi)$ . But since  $\mathcal{A}''$  satisfies Checker, it follows that  $w \in \mathcal{A}_i''(!s, s :_i \psi)$ . Since  $t = !s$  and  $\varphi = s :_i \psi$ , we conclude that  $w \in \mathcal{A}''(t, \varphi)$ .

- Inductive Case: the last step of the derivation of  $w \vdash t \gg_i \varphi$  made use of Rule M.

It follows from the form of Rule M that  $v \vdash t \gg_i \varphi$  occurs earlier in the derivation for some world  $v \in F$  satisfying  $vR_i w$ . Applying the inductive hypothesis, we have that  $v \in \mathcal{A}_i''(t, \varphi)$ . But since  $\mathcal{A}''$  satisfies Monotonicity and  $vR_i w$ , it follows that  $w \in \mathcal{A}_i''(t, \varphi)$ .

- Inductive Case: the last step of the derivation of  $w \vdash t \gg_i \varphi$  made use of Rule U.

The argument is quite similar to the case for Rule C, except that  $t = s^{U,u}$ ,  $\varphi = [U, u]\psi$ , and we make use of the fact that  $\mathcal{A}''$  satisfies Update.

This completes the induction. Conclusion:  $\mathcal{A}' \subseteq \mathcal{A}''$ . Since  $\mathcal{A}' \in \mathcal{S}^*$  and we showed that  $\mathcal{A}' \subseteq \mathcal{A}''$  for an arbitrary  $\mathcal{A}'' \in \mathcal{S}^*$ , we have proven that  $\mathcal{A}' = \mathcal{A}^{*,S}$ .  $\square$

We have introduced evidence functions and shown how an evidence function may be axiomatically generated from a given evidence labeling. While this provides more than enough for us to define the Fitting semantics, our definition of the update semantics requires to define another kind of evidence labeling that we call an *elimination function*.

## 3.2 Elimination Functions

An *elimination function*, like an evidence function, is an evidence labeling satisfying certain properties. These properties describe the way in which the semantic operation of evidence elimination interacts with the term-forming operations.

**Definition 3.5.** Let  $F = (W, R)$  be a Kripke frame for an agent set  $A$ . An *elimination function* (on  $F$ ) is an evidence labeling (in  $\text{UL}^A$  on  $F$ ) satisfying each of the following schematic properties.

- *Eliminatory Constant Specification.* For each  $k \in \mathbb{N}$ , we have  $\mathcal{A}_i(c_k, \varphi) = \emptyset$ .
- *Eliminatory Application.*  $\mathcal{A}_i(t \cdot_{\varphi} s, \psi) = \mathcal{A}_i(t, \varphi \supset \psi) \cup \mathcal{A}_i(s, \varphi)$ .
- *Eliminatory Sum.*  $\mathcal{A}_i(t + s, \varphi) = \mathcal{A}_i(t, \varphi) \cap \mathcal{A}_i(s, \varphi)$ .
- *Eliminatory Checker.*  $\mathcal{A}_i(!t, t :_i \varphi) = \mathcal{A}_i(t, \varphi)$ .
- *Eliminatory Monotonicity.*  $\Gamma R_i \Delta$  and  $\Delta \in \mathcal{A}_i(t, \varphi)$  together imply that  $\Gamma \in \mathcal{A}_i(t, \varphi)$ .
- *Eliminatory Update.*  $\mathcal{A}_i(t^{U,u}, [U, u]\varphi) = \mathcal{A}_i(t, \varphi)$ .

Each of the properties found in Definition 3.5 is different from the similarly named property found in Definition 3.1. These differences are essential, so the reader is advised to take note of these differences now so as to avoid confusion later.

To understand the meaning of the above-stated properties of an elimination function, we think of the assertion  $\Gamma \in \mathcal{A}_i(t, \varphi)$ , where  $\mathcal{A}$  is a *elimination function*, as saying that our semantic elimination operation should make it the case that the relevance assertion  $t \gg_i \varphi$  is false. Thus to have  $\Gamma \in \mathcal{A}_i(t, \varphi)$  for an elimination function  $\mathcal{A}$  is to say that the evidence  $t$  for  $\varphi$  should be eliminated for agent  $i$ , so that, after the occurrence of the semantic update described by this elimination function, agent  $i$  will not consider  $t$  as evidence relevant to  $\varphi$ .

Under this reading, the property of Eliminatory Constant Specification says that no agent may eliminate a constant. The reason for this is that we do not want eliminations to undermine the use of constants as unanalyzed pieces of evidence relevant to basic assertions. Therefore for no world  $\Gamma$  should an agent eliminate evidence  $c_k$  relevant to  $\varphi$ .<sup>8</sup>

---

<sup>8</sup>In future work, we may loosen this restriction by allowing elimination of a constant  $c_k$  relevant to a formula  $\varphi$  as long as  $\varphi$  is not a member of the set  $S$  of “basic” assertions. But for the moment we take a much more simple approach and simply disallow all eliminations of a constant  $c_k$  relevant to a formula  $\varphi$ .

The property of Eliminary Application says that an agent eliminates  $t \cdot_{\varphi} s$  as evidence relevant to  $\psi$  if and only if she eliminates  $t$  as evidence relevant to  $\varphi \supset \psi$  or she eliminates  $s$  as evidence relevant to  $\varphi$ . The idea is that the construction of the piece  $t \cdot_{\varphi} s$  of evidence ostensibly relies upon the pieces of evidence  $t$  and  $s$ , so eliminating either of the latter two pieces of evidence ought to be equivalent to eliminating the first.

The property of Eliminary Sum says that an agent eliminates  $t + s$  as evidence relevant to  $\varphi$  if and only if she eliminates each of  $t$  and  $s$  as evidence relevant to  $\varphi$ . Like Eliminary Application, the piece of evidence  $t + s$  relevant to  $\varphi$  ostensibly depends on having one of  $t$  or  $s$  relevant to  $\varphi$ , so eliminating each of the latter two pieces of evidence ought to be equivalent to eliminating the first.

The property of Eliminary Checker says that agent  $i$  eliminates  $!t$  as evidence relevant to  $t :_i \varphi$  if and only if she eliminates  $t$  as evidence relevant to  $\varphi$ . As before, the evidence  $!t$  ostensibly relies upon the evidence  $t$ , and so eliminating one ought to be equivalent to eliminating the other.

The property of Eliminary Monotonicity says that an elimination must not be undone by an agent's belief in her own evidence; that is, in case agent  $i$  eliminates evidence  $t$  as relevant to  $\varphi$  at  $\Delta$  and  $\Gamma R_i \Delta$ , then she must also eliminate  $t$  as relevant to  $\varphi$  at  $\Gamma$ . This allows her to avoid situations such as the following: agent  $i$  is to eliminate  $t$  for  $\varphi$  at  $\Delta$  even though she maintains  $t$  for  $\varphi$  at  $\Gamma$  with  $\Gamma R_i \Delta$ , contrary to Monotonicity's (Definition 3.1) requirement that she believe her evidence  $t$  relevant to  $\varphi$  at  $\Gamma$  and so hold  $t$  relevant to  $\varphi$  at  $\Delta$  by the fact that  $\Gamma R_i \Delta$ . So as to avoid such a conflict, Eliminary Monotonicity says she must perform her eliminations in a way that they are not undermined by her belief in her own evidence.

The property of Eliminary Update says that an agent eliminates  $t^{U,u}$  as evidence relevant to  $[U, u]\varphi$  if and only if she eliminates  $t$  as evidence for  $\varphi$ . As before, the evidence  $t^{U,u}$  ostensibly relies upon the evidence  $t$ , and so eliminating one ought to be equivalent to eliminating the other.

We would like to generate an elimination function from an arbitrary evidence labeling similar to the way we generated an evidence function from an arbitrary evidence labeling using the Generation Lemma (Lemma 3.2). The only difficulty is that an arbitrary evidence labeling might violate one or more of the properties of elimination functions; as an example, an arbitrary evidence labeling might violate the property of Eliminary Constant Specification. But as it turns out, all such violations can be traced to the behavior of the evidence labeling on terms made up of constants. We will therefore adopt the convention that when we generate an elimination function from an arbitrary evidence labeling, we will ignore the action of the evidence labeling on everything but variables. This ignoring is captured in the following definition.

**Definition 3.6** ( $\mathcal{A}^{\sharp}$ ). Let  $F = (W, R)$  be a Kripke frame for an agent set  $A$  and let  $\mathcal{A}$  be an evidence labeling in  $\text{UL}^A$  on  $F$ . We define the evidence labeling  $\mathcal{A}^{\sharp}$  (in  $\text{UL}^A$  on  $F$ ) as follows.

$$\mathcal{A}_i^{\sharp}(t, \varphi) := \begin{cases} \emptyset & \text{if } t \neq x_k \text{ for any } k \in \mathbb{N}, \\ \mathcal{A}_i(x_k, \varphi) & \text{if } t = x_k \text{ for some } k \in \mathbb{N}. \end{cases}$$

AXIOM SCHEMES

V.  $w \vdash x_k \gg_i \varphi$  if  $w \in \mathcal{A}_i(x_k, \varphi)$

RULES

$$\frac{w \vdash t \gg_i (\varphi \supset \psi)}{w \vdash (t \cdot_{\varphi} s) \gg_i \psi} \text{ (EAL)} \quad \frac{w \vdash s \gg_i \varphi}{w \vdash (t \cdot_{\varphi} s) \gg_i \psi} \text{ (EAR)}$$

$$\frac{w \vdash t \gg_i \varphi \quad w \vdash s \gg_i \varphi}{w \vdash (t + s) \gg_i \varphi} \text{ (ES)}$$

$$\frac{w \vdash t \gg_i \varphi}{w \vdash !t \gg_i (t :_i \varphi)} \text{ (EC)}$$

$$\frac{w \vdash t \gg_i \varphi \quad v R_i w}{v \vdash t \gg_i \varphi} \text{ (EM)}$$

$$\frac{w \vdash t \gg_i \varphi}{w \vdash t^{U,u} \gg_i [U, u] \varphi} \text{ (EU)}$$

Figure 2: The  $(F, \mathcal{A})$ -calculus (with  $F = (W, R)$ )

Using this definition, we may then specify how it is that an elimination function is generated from an arbitrary evidence labeling. The proof of this lemma is almost identical to the proof of the Generation Lemma (Lemma 3.2).

**Lemma 3.7** (Eliminatory Generation Lemma). Let  $F = (W, R)$  be a Kripke frame for an agent set  $A$  and let  $\mathcal{A}$  be an evidence labeling in  $\text{UL}^A$  on  $F$ . There exists a smallest elimination function  $\mathcal{A}^-$  on  $F$  that extends  $\mathcal{A}^\sharp$ . We call  $\mathcal{A}^-$  the *elimination function generated by  $\mathcal{A}$* .

And, as in the case of axiomatically generated evidence functions, we show how it is that elimination functions can be axiomatically generated using the following axiomatic system.

**Definition 3.8** ( $(F, \mathcal{A})$ -calculus). Let  $F = (W, R)$  be a Kripke frame for an agent set  $A$  and let  $\mathcal{A}$  be an evidence labeling in  $\text{UL}^A$  on  $F$ . The  $(F, \mathcal{A})$ -calculus is the theory defined in Figure 2. We adopt similar notation for the  $(F, \mathcal{A})$ -calculus as we did in Definition 3.3 for the  $(F, \mathcal{A}, S)$ -calculus.

It is sometimes helpful to point out the obvious: we can easily distinguish between the  $(F, \mathcal{A}, S)$ -calculus of Definition 3.3 and the  $(F, \mathcal{A})$ -calculus of Definition 3.8 because the name of the former begins with the tuple  $(F, \mathcal{A}, S)$ , a tuple with three elements, whereas the name of the latter begins with the tuple  $(F, \mathcal{A})$ , a tuple with just two elements.

The proof of the following lemma is quite similar to the proof of the Axiomatic Generation Lemma (Lemma 3.4).

**Lemma 3.9** (Eliminatory Axiomatic Generation Lemma). Let  $F = (W, R)$  be a Kripke frame for an agent set  $A$  and let  $\mathcal{A}$  be an evidence labeling in  $\text{UL}^A$  on  $F$ . Define the evidence labeling  $\mathcal{A}'$  (in  $\text{UL}^A$  on  $F$ ) by setting

$$\mathcal{A}'_i(t, \varphi) := \{w \in F : (F, \mathcal{A}), w \vdash t \gg_i \varphi\}$$

for each  $i \in A$  and each  $(t, \varphi) \in \mathcal{T}(\text{UL}^A) \times \text{UL}^A$ . Then  $\mathcal{A}'$  is  $\mathcal{A}^-$ , the elimination function generated by  $\mathcal{A}$  (Lemma 3.7).

Finally, as an elimination function says what pieces of evidence that the agents are to eliminate, it is useful to understand what pieces of evidence that an elimination function says the agents are *not* to eliminate. For this we define a calculus that is complementary to that of the  $(F, \mathcal{A})$ -calculus, in a sense described in Lemma 3.11.

**Definition 3.10** ( $(F, \mathcal{A})^c$ -calculus). Let  $F = (W, R)$  be a Kripke frame for an agent set  $A$  and let  $\mathcal{A}$  be an evidence labeling in  $\text{UL}^A$  on  $F$ . The  $(F, \mathcal{A})^c$ -calculus is the theory defined in Figure 3. We adopt similar notation for the  $(F, \mathcal{A})^c$ -calculus as we did in Definition 3.3 for the  $(F, \mathcal{A}, S)$ -calculus.

**Lemma 3.11** (Complementary Axiomatic Generation Lemma). Let  $F = (W, R)$  be a Kripke frame for an agent set  $A$  and let  $\mathcal{A}$  be an evidence labeling in  $\text{UL}^A$  on  $F$ . Define the evidence labeling  $\mathcal{A}^c$  (in  $\text{UL}^A$  on  $F$ ) by setting

$$\mathcal{A}^c_i(t, \varphi) := \{w \in F : (F, \mathcal{A})^c, w \vdash t \gg_i \varphi\}$$

for each  $i \in A$  and each  $(t, \varphi) \in \mathcal{T}(\text{UL}^A) \times \text{UL}^A$ . Then we have the following schematic statement:  $w \in \mathcal{A}^c_i(t, \varphi)$  if and only if  $w \notin \mathcal{A}^-_i(t, \varphi)$ . Said informally:  $\mathcal{A}^c$  is the *complement* of the elimination function  $\mathcal{A}^-$  generated by  $\mathcal{A}$  (Lemma 3.7). By Lemma 3.9, saying that  $\mathcal{A}^c$  and  $\mathcal{A}^-$  are complementary (in the sense just described) is equivalent to the following schematic biconditional about provability in the  $(F, \mathcal{A})^c$ -calculus and the  $(F, \mathcal{A})$ -calculus:

$$(F, \mathcal{A})^c, w \vdash t \gg_i \varphi \text{ if and only if } (F, \mathcal{A}), w \not\vdash t \gg_i \varphi .$$

*Proof.* By Lemma 3.9, it suffices for us to prove the above schematic biconditional about complementary provability. We will prove each direction in turn. But let us first make the following auxiliary definition: the *depth*  $d(t)$  of a term  $t \in \text{UL}^A$  is given by the inductive definition in Figure 4 (on Page 25). Let us now observe that for each rule of the  $(F, \mathcal{A})$ -calculus (Figure 2), if  $t$  is the term occurring immediately to the right of the turnstile (“ $\vdash$ ”) in the hypothesis of the rule and  $t'$  is any term occurring immediately to the right of the turnstile in a conclusion of the rule, then we have that  $d(t') \geq d(t)$ . In fact, we have  $d(t') = d(t)$  for the rule EM, and we have  $d(t') > d(t)$  for all non-EM rules. Referring to the term appearing immediately to the right of a turnstile as the *evidencing term*, we see that the rules of the  $(F, \mathcal{A})$ -calculus do not decrease the depth of evidencing terms. In fact, the rule

### AXIOM SCHEMES

CS<sup>c</sup>.  $w \vdash c_k \gg_i \varphi$

V<sup>c</sup>.  $w \vdash x_k \gg_i \varphi$  if  $wR_i^*v$  implies  $v \notin \mathcal{A}_i(x_k, \varphi)$

### RULES

$$\frac{w \vdash t \gg_i (\varphi \supset \psi) \quad w \vdash s \gg_i \varphi}{w \vdash (t \cdot_\varphi s) \gg_i \psi} \text{ (A}^c\text{)}$$

$$\frac{w \vdash t \gg_i \varphi}{w \vdash (t + s) \gg_i \varphi} \text{ (SL}^c\text{)} \quad \frac{w \vdash s \gg_i \varphi}{w \vdash (t + s) \gg_i \varphi} \text{ (SR}^c\text{)}$$

$$\frac{w \vdash t \gg_i \varphi}{w \vdash !t \gg_i (t :_i \varphi)} \text{ (C}^c\text{)}$$

$$\frac{w \vdash t \gg_i \varphi \quad wR_iv}{v \vdash t \gg_i \varphi} \text{ (M}^c\text{)}$$

$$\frac{w \vdash t \gg_i \varphi}{w \vdash t^{U,u} \gg_i [U, u]\varphi} \text{ (U}^c\text{)}$$

Note: In Axiom V<sup>c</sup>, the symbol  $R_i^*$  denotes the *reflexive-transitive closure* of  $R_i$ . That is,  $wR_i^*v$  means that either  $w = v$  or else there is a finite nonempty sequence  $\{u_j\}_{j=0}^n$  of worlds in  $F$  such that  $u_0 = w$ ,  $u_n = v$ , and  $u_j R_i u_{j+1}$  for each  $j = 0, 1, 2, \dots, n-1$ .

Figure 3: The  $(F, \mathcal{A})^c$ -calculus (with  $F = (W, R)$ )

$$\begin{aligned}
d(q) &:= 1, \text{ for each } q \in \{p_k, \perp, \top\} \\
d(\varphi \star \psi) &:= 1 + \max\{d(\varphi), d(\psi)\}, \text{ for each } \star \in \{\supset, \wedge, \vee, \equiv\} \\
d(\neg\varphi) &:= 1 + d(\varphi) \\
d(t :_i \varphi) &:= 4 + \max\{d(t), d(\varphi)\} \\
d(B_i \varphi) &:= 1 + d(\varphi) \\
d(t \gg_i \varphi) &:= 2 + \max\{d(t), d(\varphi)\} \\
d(t \gg_i^{U,u} \varphi) &:= 1 + \max\{d(t), d(\varphi)\} \\
d([U, u] \varphi) &:= (4 + d(U))^{4+d(\varphi)} \cdot d(\varphi) \\
d(U) &:= |U| + \max_{u \in U} d(U(u)) \\
d(c_k) &:= 1 \\
d(x_k) &:= 1 \\
d(t \cdot_\varphi s) &:= 1 + \max\{d(t), d(s), d(\varphi)\} \\
d(t + s) &:= 1 + \max\{d(t), d(s)\} \\
d(!t) &:= 1 + d(t) \\
d(t^{U,u}) &:= 1 + d(t)
\end{aligned}$$

- Notes:
- $|U|$  denotes the number of worlds in  $U$ .
  - This definition is adapted from [19].

Figure 4: Definition of a function  $d : \text{UL}^A \rightarrow \mathbb{N}$ .

EM maintains the depth of evidencing terms and each of the non-EM rules strictly increases the depth of evidencing terms. Similarly, we observe that the rules of the  $(F, \mathcal{A})^c$ -calculus (Figure 3) do not decrease the depth of evidencing terms. In fact, the rule  $M^c$  maintains the depth of evidencing terms and each of the non- $M^c$  rules strictly increases the depth of evidencing terms. With these observations, let us proceed with our proof.

We first argue by induction on the length of derivation in the  $(F, \mathcal{A})^c$ -calculus (Figure 3) that if  $w \vdash t \gg_i \varphi$  appears in a line of a derivation in the  $(F, \mathcal{A})^c$ -calculus, then  $(F, \mathcal{A}), v \not\vdash t \gg_i \varphi$  for each  $v \in F$  satisfying  $wR_i^*v$ . (Recall that  $R_i^*$  is the reflexive-transitive closure of  $R_i$ ; see Figure 3 for details.)

- Base Case:  $w \vdash c_k \gg_i \varphi$  is an instance of Axiom  $CS^c$  appearing in a line of a derivation in the  $(F, \mathcal{A})^c$ -calculus (Figure 3).

A line in a derivation of the  $(F, \mathcal{A})$ -calculus (Figure 2) is either an instance of Axiom V or else follows from earlier lines by a rule. However, Axiom V does not have an evidencing term of the form  $c_k$ , each non-EM rule of the  $(F, \mathcal{A})$ -calculus strictly increases the depth of evidencing terms, and the Rule EM leaves its evidencing term unchanged. It follows that no line of a derivation in the  $(F, \mathcal{A})$ -calculus can have the form of Axiom  $CS^c$ . Thus  $(F, \mathcal{A}), v \not\vdash c_k \gg_i \varphi$  for each  $v \in F$  satisfying  $wR_i^*v$ .

- Base Case:  $w \vdash x_k \gg_i \varphi$  is an instance of Axiom  $V^c$  appearing in a line of a derivation in the  $(F, \mathcal{A})^c$ -calculus (Figure 3).

A line in a derivation of the  $(F, \mathcal{A})$ -calculus (Figure 2) is either an instance of Axiom V or else follows from earlier lines by a rule. Further, each non-EM rule strictly increases the depth of evidencing terms. So it follows that  $v \vdash x_k \gg_i \varphi$  appears in a line  $L$  of a derivation in the  $(F, \mathcal{A})$ -calculus for some  $v \in F$  satisfying  $wR_i^*v$  if and only if  $L$  either is an instance of Axiom V or else follows from an earlier line by Rule EM. We consider each of these cases in turn.

If  $L$  is an instance of Axiom V, then we have  $v \in \mathcal{A}_i(x_k, \varphi)$  by the statement of Axiom V. But our assumption that  $w \vdash x_k \gg_i \varphi$  is an instance of Axiom  $V^c$  (Figure 3) implies  $v \notin \mathcal{A}_i(x_k, \varphi)$  by the statement of Axiom  $V^c$  and the fact that  $wR_i^*v$ . Thus we have  $v \in \mathcal{A}_i(x_k, \varphi)$  and  $v \notin \mathcal{A}_i(x_k, \varphi)$ , a contradiction. So  $L$  cannot be an instance of Axiom V.

If  $L$  follows from an earlier line  $L_1$  by Rule EM, then  $L_1$  must itself either be an instance of Axiom V or else follow from an even earlier line  $L_2$  by Rule EM. If the latter situation obtains, then  $L_2$  likewise is either an instance of Axiom V or else follows from an even earlier line  $L_3$  by Rule EM. In this way, we have a finite sequence of lines  $L_0, L_1, L_2, L_3, \dots, L_n$  such that  $L_0 := L$ , that  $n \geq 1$ , that  $L_j$  follows from  $L_{j+1}$  by Rule EM for each  $L_j$  and  $L_{j+1}$  within the range of the sequence, and that  $L_n$  is an instance of Axiom V. It also follows from a hypothesis of Rule EM that whenever  $L_j$  follows from  $L_{j+1}$  by Rule EM, we have worlds  $u_j \in F$  and  $u_{j+1} \in F$  such that  $u_j R_i u_{j+1}$ , with  $u_0 = v$ . Since  $u_n$  is an instance of Axiom V, it follows that  $u_n \in \mathcal{A}_i(x_k, \varphi)$  (Figure 2). And yet we also have  $wR_i^*v$ ,  $v = u_0$ , and  $u_0 R_i^* u_n$ , and thus that  $wR_i^* u_n$ . But the

latter implies that  $u_n \notin \mathcal{A}_i(x_k, \varphi)$  by our assumption that  $w \vdash x_k \gg_i \varphi$  is an instance of Axiom V<sup>c</sup> (Figure 3). Hence  $u_n \in \mathcal{A}_i(x_k, \varphi)$  and  $u_n \notin \mathcal{A}_i(x_k, \varphi)$ , a contradiction. So  $L$  cannot have followed from an earlier line by Rule EM.

We have argued that  $v \vdash x_k \gg_i \varphi$  appears in a line  $L$  of a derivation in the  $(F, \mathcal{A})$ -calculus for some  $v \in F$  satisfying  $wR_i^*v$  if and only if  $L$  either is an instance of Axiom V or else follows from an earlier line by Rule EM, and we showed that each of the latter two possibilities leads to a contradiction under our assumption that  $w \vdash x_k \gg_i \varphi$  appears in a line of a derivation in the  $(F, \mathcal{A})^c$ -calculus. Conclusion:  $(F, \mathcal{A}), v \not\vdash x_k \gg_i \varphi$  for each  $v \in F$  satisfying  $wR_i^*v$ .

- Inductive Case:  $w \vdash (t \cdot_{\varphi} s) \gg_i \psi$  appears in the line of a derivation of the  $(F, \mathcal{A})^c$ -calculus (Figure 3) by an application of Rule A<sup>c</sup>.

It follows from the statement of Rule A<sup>c</sup> that  $w \vdash t \gg_i (\varphi \supset \psi)$  and  $w \vdash s \gg_i \varphi$  each appear earlier in the derivation. Applying the induction hypothesis, we have that  $(F, \mathcal{A}), v_1 \not\vdash t \gg_i (\varphi \supset \psi)$  for each  $v_1 \in F$  satisfying  $wR_i^*v_1$  and that  $(F, \mathcal{A}), v_2 \not\vdash s \gg_i \varphi$  for each  $v_2 \in F$  satisfying  $wR_i^*v_2$ . But the only way for us to derive  $v \vdash (t \cdot_{\varphi} s) \gg_i \psi$  for some  $v \in F$  satisfying  $wR_i^*v$  within the  $(F, \mathcal{A})$ -calculus (Figure 2) is to use Rule EAL or EAR to conclude that  $v' \vdash (t \cdot_{\varphi} s) \gg_i \psi$  for some  $v' \in F$  satisfying  $vR_i^*v'$ . But  $wR_i^*v$  and  $vR_i^*v'$  together imply that  $wR_i^*v'$ , and so neither the hypothesis  $v' \vdash t \gg_i (\varphi \supset \psi)$  of EAL nor the hypothesis  $v' \vdash s \gg_i \varphi$  of EAR is provable in the  $(F, \mathcal{A})$ -calculus by what we have just argued. It follows that  $(F, \mathcal{A}), v \not\vdash (t \cdot_{\varphi} s) \gg_i \psi$  for each  $v \in F$  satisfying  $wR_i^*v$ .

- Inductive Case for Rules SL<sup>c</sup>, SR<sup>c</sup>, C<sup>c</sup> and U<sup>c</sup>: analogous to the case for Rule A<sup>c</sup>.
- Inductive Case:  $v \vdash t \gg_i \varphi$  appears in a line of a derivation of the  $(F, \mathcal{A})^c$ -calculus (Figure 3) by an application of Rule M<sup>c</sup>.

It follows from the statement of rule M<sup>c</sup> that for some  $w \in F$  satisfying  $wR_iv$ , we have that  $w \vdash t \gg_i \varphi$  appears earlier in the derivation. By the induction hypothesis, we have that  $(F, \mathcal{A}), u \not\vdash t \gg_i \varphi$  for each  $u \in F$  satisfying  $wR_i^*u$ . Since  $wR_iv$ , we have that  $vR_i^*u'$  implies  $wR_i^*u'$ . So it follows by what we have shown that  $(F, \mathcal{A}), u' \not\vdash t \gg_i \varphi$  for each  $u' \in F$  satisfying  $vR_i^*u'$ .

This completes inductive proof that if  $w \vdash t \gg_i \varphi$  appears in a line of a derivation in the  $(F, \mathcal{A})^c$ -calculus, then  $(F, \mathcal{A}), v \not\vdash t \gg_i \varphi$  for each  $v \in F$  satisfying  $wR_i^*v$ . Conclusion:  $(F, \mathcal{A})^c, w \vdash t \gg_i \varphi$  implies  $(F, \mathcal{A}), w \not\vdash t \gg_i \varphi$ .

We now argue by induction on the length of derivation in the  $(F, \mathcal{A})$ -calculus (Figure 2) that if  $v \vdash t \gg_i \varphi$  appears in a line of a derivation in the  $(F, \mathcal{A})$ -calculus, then we have that  $(F, \mathcal{A})^c, v \not\vdash t \gg_i \varphi$  for each  $v \in F$  satisfying  $vR_i^*w$ .

- Base Case:  $w \vdash x_k \gg_i \varphi$  is an instance of Axiom V appearing in a line of a derivation in the  $(F, \mathcal{A})$ -calculus (Figure 2).

A line of a derivation of the  $(F, \mathcal{A})^c$ -calculus (Figure 3) is either an instance of Axiom CS<sup>c</sup>, an instance of Axiom V<sup>c</sup>, or else follows from earlier lines by a rule. Further,

the evidencing term of Axiom  $\text{CS}^c$  does not have the form  $x_k$  and each non- $\text{M}^c$  rule strictly increases the depth of evidencing terms. So it follows that  $v \vdash x_k \gg_i \varphi$  appears in a line  $L_0$  of a derivation in the  $(F, \mathcal{A})^c$ -calculus for some  $v \in F$  satisfying  $vR_i^*w$  if and only if we have a finite sequence  $L_0, L_1, L_2, \dots, L_n$  of lines such that  $n \geq 0$ ,  $L_j$  follows from  $L_{j+1}$  by Rule  $\text{M}^c$  for each  $L_j$  and  $L_{j+1}$  within the range of the sequence, and  $L_n$  is an instance of Axiom  $\text{V}^c$ . It also follows from a hypothesis of Rule  $\text{M}^c$  that whenever  $L_j$  follows from  $L_{j+1}$  by Rule  $\text{M}^c$ , we have worlds  $u_j \in F$  and  $u_{j+1} \in F$  such that  $u_{j+1}R_iu_j$ , with  $u_0 = v$ . But then  $u_nR_i^*u_0$ ,  $u_0 = v$ , and  $vR_i^*w$ . It follows that  $u_nR_i^*w$  and hence that  $w \notin \mathcal{A}_i(x_k, \varphi)$  by the fact that  $L_n$  is an instance of Axiom  $\text{V}^c$ . But by our assumption that  $w \vdash x_k \gg_i \varphi$  is an instance of Axiom  $\text{V}$  appearing in a line of a derivation in the  $(F, \mathcal{A})$ -calculus (Figure 2), it follows from the statement of Axiom  $\text{V}$  that  $w \in \mathcal{A}_i(x_k, \varphi)$ . Thus  $w \notin \mathcal{A}_i(x_k, \varphi)$  and  $w \in \mathcal{A}_i(x_k, \varphi)$ , a contradiction. Conclusion:  $(F, \mathcal{A})^c, v \not\vdash x_k \gg_i \varphi$  for each  $v \in F$  satisfying  $vR_i^*w$ .

- Inductive Case:  $w \vdash (t + s) \gg_i \varphi$  appears in a line of a derivation of the  $(F, \mathcal{A})$ -calculus (Figure 2) by an application of Rule  $\text{ES}$ .

It follows from the statement of Rule  $\text{ES}$  that  $w \vdash t \gg_i \varphi$  and  $w \vdash s \gg_i \varphi$  each appear earlier in the derivation. Applying the induction hypothesis, we have that  $(F, \mathcal{A})^c, v_1 \not\vdash t \gg_i \varphi$  for each  $v_1 \in F$  satisfying  $v_1R_i^*w$  and that  $(F, \mathcal{A})^c, v_2 \not\vdash s \gg_i \varphi$  for each  $v_2 \in F$  satisfying  $v_2R_i^*w$ . But the only way for us to derive  $v \vdash (t + s) \gg_i \varphi$  in the  $(F, \mathcal{A})^c$ -calculus (Figure 3) for some  $v \in F$  satisfying  $vR_i^*w$  is to use Rule  $\text{SL}^c$  or  $\text{SR}^c$  to conclude that  $v' \vdash (t + s) \gg_i \varphi$  for some  $v' \in F$  satisfying  $v'R_i^*v$ . But  $v'R_i^*v$  and  $vR_i^*w$  together imply that  $v'R_i^*w$ , and so neither the hypothesis  $v' \vdash t \gg_i \varphi$  of  $\text{SL}^c$  nor the hypothesis  $v' \vdash s \gg_i \varphi$  of  $\text{SR}^c$  is provable in the  $(F, \mathcal{A})^c$ -calculus by what we have just argued. It follows that  $(F, \mathcal{A})^c, v \not\vdash (t + s) \gg_i \varphi$  for each  $v \in F$  satisfying  $vR_i^*w$ .

- Inductive Case for Rules  $\text{EAL}$ ,  $\text{EAR}$ ,  $\text{EC}$ , and  $\text{EU}$ : analogous to the case for Rule  $\text{ES}$ .
- Inductive Case:  $v \vdash t \gg_i \varphi$  appears in a line of a derivation of the  $(F, \mathcal{A})$ -calculus (Figure 2) by an application of Rule  $\text{EM}$ .

It follows from the statement of rule  $\text{EM}$  that for some  $w \in F$  satisfying  $vR_iw$ , we have that  $w \vdash t \gg_i \varphi$  appears earlier in the derivation. By the induction hypothesis, we have that  $(F, \mathcal{A})^c, u \not\vdash t \gg_i \varphi$  for each  $u \in F$  satisfying  $uR_i^*w$ . Since  $vR_iw$ , we have that  $u'R_i^*v$  implies  $u'R_i^*w$ . So it follows by what we have shown that  $(F, \mathcal{A})^c, u' \not\vdash t \gg_i \varphi$  for each  $u' \in F$  satisfying  $u'R_i^*v$ .

This completes inductive proof that if  $w \vdash t \gg_i \varphi$  appears in a line of a derivation in the  $(F, \mathcal{A})$ -calculus, then  $(F, \mathcal{A})^c, v \not\vdash t \gg_i \varphi$  for each  $v \in F$  satisfying  $vR_i^*w$ . Conclusion:  $(F, \mathcal{A}), w \vdash t \gg_i \varphi$  implies  $(F, \mathcal{A})^c, w \not\vdash t \gg_i \varphi$ . Since we also proved the converse of this conclusion, the overall proof is complete.  $\square$

This completes our study of elimination functions. We are now in a position to define the semantics of the update language.

### 3.3 Models and Truth

The notion of truth for the update language is given by induction on the construction of  $\text{UL}^A$ -formulas. To specify the base case of truth for propositional letters, we will use a *valuation*.

**Definition 3.12.** Let  $A$  be an agent set and  $F = (W, R)$  be a Kripke frame for  $A$ . A *valuation (on  $F$ )* is a function

$$V : \{p_k : k \in \mathbb{N}\} \rightarrow 2^W$$

mapping each propositional letter  $p_k$  to a set  $V(p_k)$  of worlds in  $F$ .

To construct models for the update language, we first add an evidence function to a Kripke frame to obtain a *Fitting frame* [9].

**Definition 3.13.** Let  $A$  be an agent set and  $S$  be a set. An  *$S$ -Fitting frame (for  $A$ )* is a triple  $F = (W, R, \mathcal{A})$  satisfying each of the following.

- $(W, R)$  is a transitive Kripke frame for  $A$ .  
 $(W, R)$  is called the Kripke frame *underlying  $F$* , and  $F$  is called the  *$S$ -Fitting frame based on  $(W, R)$* .
- $\mathcal{A}$  is an  $S$ -evidence function on  $(W, R)$ .

We write  $\Gamma \in F$  to mean that  $\Gamma \in W$ ; for each  $\Gamma \in F$ , we say that  $\Gamma$  is a *world (in  $F$ )*. When the particular set  $S$  is unimportant or may be inferred from context, we may drop the prefix “ $S$ ” in referring to an  $S$ -Fitting frame.

To a Fitting frame we add a valuation, which gives us a *Fitting model* [9].

**Definition 3.14.** Let  $A$  be an agent set and  $S$  be a set. An  *$S$ -Fitting model (for  $A$ )* is a pair  $(F, V)$  consisting of an  $S$ -Fitting frame  $F$  for  $A$  and a valuation  $V$  on the Kripke frame underlying  $F$ . We say that  $F$  is the  *$S$ -Fitting frame underlying  $M$*  and that  $M$  is the  *$S$ -Fitting model based on  $F$* ; we also say that the Kripke frame  $F'$  underlying  $F$  is the Kripke frame *underlying  $M$*  and that  $M$  is the  *$S$ -Fitting model based on  $F'$* . If  $M = (F, V)$  is an  $S$ -Fitting model, we write  $\Gamma \in M$  to mean that  $\Gamma \in F$ ; for each  $\Gamma \in M$ , we say that  $\Gamma$  is a *world (in  $M$ )*. A *pointed  $S$ -Fitting model (for  $A$ )* is a pair  $(M, \Gamma)$  consisting of an  $S$ -Fitting model  $M$  and a world  $\Gamma \in M$ ; we call  $\Gamma$  the *point* of the pair  $(M, \Gamma)$ . For convenience of exposition, we may conflate the  $S$ -Fitting model  $((W, R, \mathcal{A}), V)$  with the tuple  $(W, R, \mathcal{A}, V)$ ; likewise, we may conflate the pointed  $S$ -Fitting model  $((W, R, \mathcal{A}), V, \Gamma)$  with the tuple  $(W, R, \mathcal{A}, V, \Gamma)$ . When the particular set  $S$  is unimportant or may be inferred from context, we may drop the prefix “ $S$ ” in referring to an  $S$ -Fitting model or to a pointed  $S$ -Fitting model.

Our notion of truth for  $\text{UL}^A$ -formulas is defined with respect to a given pointed Fitting model  $(M, \Gamma)$  by an induction on formula construction.

**Definition 3.15** (Truth). Let  $A$  be an agent set and  $S$  be a set. For a  $\text{UL}^A$ -formula  $\varphi$  and a pointed  $S$ -Fitting model  $(M, \Gamma)$  for  $A$ , we write  $M, \Gamma \models \varphi$  to mean that  $\varphi$  is *true at*  $(M, \Gamma)$ , and we write  $M, \Gamma \not\models \varphi$  to mean that  $\varphi$  is not true at  $(M, \Gamma)$ . Truth of a  $\text{UL}^A$ -formula at a pointed  $S$ -Fitting model  $(M, \Gamma) = ((W, R, \mathcal{A}, V), \Gamma)$  for  $A$  is defined by the following induction on  $\text{UL}^A$ -formula construction.

- $M, \Gamma \models p_k$  means that  $\Gamma \in V(p_k)$ .
- $M, \Gamma \not\models \perp$  and  $M, \Gamma \models \top$ .
- $M, \Gamma \models \varphi_1 \star \varphi_2$  means that  $M, \Gamma \models \varphi_1$  star  $M, \Gamma \models \varphi_2$  for  $\star \in \{\supset, \wedge, \vee, \equiv\}$ .<sup>9</sup>
- $M, \Gamma \models \neg\varphi$  means that  $M, \Gamma \not\models \varphi$ .
- $M, \Gamma \models B_i\varphi$  means that  $M, \Delta \models \varphi$  for each  $\Delta \in M$  such that  $\Gamma R_i \Delta$ .
- $M, \Gamma \models t \gg_i \varphi$  means that  $\Gamma \in \mathcal{A}_i(t, \varphi)$ .
- $M, \Gamma \models t :_i \varphi$  means that  $\Gamma \in \mathcal{A}_i(t, \varphi)$  and  $M, \Delta \models \varphi$  for each  $\Delta \in M$  such that  $\Gamma R_i \Delta$ .
- $M, \Gamma \models t \gg_i^{U, u} \varphi$  means that  $(U, U^e), u \vdash t \gg_i \varphi$ .  
Here we use the  $(U, U^e)$ -calculus (Figure 2).
- $M, \Gamma \models [U, u]\varphi$  means that either  $M, \Gamma \not\models U(u)$  or else that both  $M, \Gamma \models U(u)$  and  $M[U], (\Gamma, u) \models \varphi$ , where the components of the tuple

$$M[U] := (W[U], R[U], \mathcal{A}[U], V[U])$$

are given as follows.

- $W[U] := \{(\Delta, v) \in (W \times U) : M, \Delta \models U(v)\}$ .
- $R[U]_i := \{((\Delta, v), (\Delta', v')) \in (W[U] \times W[U]) : (\Delta R_i \Delta') \wedge (v U_i v')\}$ .
- $\mathcal{A}[U]_i(t, \psi) := \{(\Delta, v) \in W[U] : \Delta \in \mathcal{A}_i(t, \psi) \wedge v \notin U_i^-(t, \psi)\}$ .<sup>10</sup>
- $V[U](p_k) := \{(\Delta, v) \in W[U] : \Delta \in V(p_k)\}$ .

The construction of the model  $M[U]$  may be understood in the following way: a world  $v \in U$  represents the combined affect of two events.

<sup>9</sup>The word “star” is to be replaced by the English reading for the binary logical connective  $\star$ ; in particular,  $\supset$  is read “implies”,  $\wedge$  is read “and”,  $\vee$  is read “or”, and  $\equiv$  is read “if and only if”. Note that the connectives  $\supset$  and  $\equiv$  are to be understood as being defined in the appropriate way in terms of the *material conditional*, which is given by saying that “ $\varphi$  implies  $\psi$ ” is true exactly when  $\varphi$  is false or  $\psi$  is true.

<sup>10</sup> $U_i^-(t, \psi)$  denotes the result of  $(U^e)^-$ , the elimination function generated by  $U^e$ , on the tuple  $(i, (t, \psi))$ ; see Lemma 3.7 for details. Also, recall that the  $(U, U^e)^c$ -calculus is an axiomatic theory for proving that  $v \notin U_i^-(t, \psi)$ ; see Figure 3 and Lemma 3.11 for details.

1. The communication of the formula  $U(v)$ .

Communication of the formula  $U(v)$  amounts to eliminating all worlds  $\Delta \in M$  in which  $U(v)$  is false. This elimination causes the agents to jointly eliminate from consideration all non- $U(v)$  worlds in  $M$ , which provides a sense in which the formula  $U(v)$  is communicated.

2. The execution of the evidence elimination described by  $U^e(v)$ .

Here we identify the expression  $U^e(v)$  with the following semantic operation of evidence elimination: for each relevance assertion  $t \gg_i \psi$  such that  $(U, U^e)^c, v \vdash t \gg_i \psi$ ,<sup>11</sup> we have that  $t \gg_i \psi$  is made false, thereby eliminating  $t$  as evidence relevant to  $\psi$  for agent  $i$ . It is in this sense that we use the expression  $U^e(v)$  to describe a multi-agent elimination of evidence.

In this way, a world  $v \in U$  represents an event that combines the communication of  $U(v)$  with the elimination given by  $U^e(v)$ . But the agents are uncertain as to which joint communication-elimination event  $v \in U$  actually occurs, since if the event  $u \in U$  occurs, agent  $i$  thinks it possible that each of the communication-elimination events  $v \in U$  satisfying  $uU_i v$  may be the one that is occurring. So we see that the operation  $(M, \Gamma) \mapsto (M[U], (\Gamma, u))$  represents the occurrence of the communication-elimination event given by  $u \in U$ , though the agents have uncertainty as to which communication-elimination event in fact occurs. It is in this way that the structure of the update frame  $U$  represents subtle mixtures of privacy and deceit with respect to multi-agent communication and evidence elimination.

We use the elimination assertion  $t \gg_i^{U,u} \varphi$  to express the fact that the elimination basis  $U^e(u)$  will cause agent  $i$  to bring about the elimination of evidence  $t$  as relevant to  $\varphi$ , so the operation  $(M, \Gamma) \mapsto (M[U], (\Gamma, u))$  will bring about a situation in which we have that  $t \gg_i \varphi$  is false. As described in the semantics above, we use the elimination assertion  $t \gg_i^{U,u} \varphi$  merely as a shorthand for the derivability of  $u \vdash t \gg_i \varphi$  in the  $(U, U^e)$ -calculus (Definition 3.8).

The following lemma shows that whenever we are able to form the structure  $M[U]$  from a Fitting model  $M$ , we have that  $M[U]$  is itself a Fitting model. This ensures that our inductive definition of truth (Definition 3.15), an induction on  $\text{UL}^A$ -formula construction that provides a relationship between pointed Fitting models and  $\text{UL}^A$ -formulas, is a well-formed definition by induction.

**Lemma 3.16** (Update Correctness). Let  $A$  be an agent set,  $S$  be a set,

$$(M, \Gamma) = ((W, R, \mathcal{A}, V), \Gamma)$$

be a pointed  $S$ -Fitting model for  $A$ , and  $(U, u)$  be a pointed update frame in  $\text{UL}^A$  for  $A$ . If  $M, \Gamma \models U(u)$ , then  $M[U]$  is an  $S$ -Fitting model for  $A$ .

*Proof.* We are to show that the tuple  $M[U]$  from Definition 3.15 satisfies each of the defining properties of an  $S$ -Fitting model for  $A$  (Definition 3.14). It suffices to verify three items:

---

<sup>11</sup>This derivation uses the  $(U, U^e)^c$ -calculus (Figure 3).

that  $F[U] := (W[U], R[U])$  is a transitive Kripke frame for  $A$ , that  $\mathcal{A}[U]$  is an  $S$ -evidence function on  $F[U]$ , and that  $V[U]$  is a valuation on  $F[U]$ . We shall only focus our attention on the second item because the other two are easily verified.<sup>12</sup>

So let us argue that  $\mathcal{A}[U]$  is an  $S$ -evidence function on  $F[U]$ . By Definition 3.1, we must show that  $\mathcal{A}[U]$  satisfies Constant Specification  $S$ , Application, Sum, Checker, Monotonicity, and Update. We will check each item in turn, making frequent use of the Complementary Axiomatic Generation Lemma (Lemma 3.11) without explicit mention.

- *Constant Specification  $S$ .* For each  $k \in \mathbb{N}$  and each  $(c_k :_i \varphi) \in (S \cap \text{UL}^A)$ , we have  $\mathcal{A}[U]_i(c_k, \varphi) = W[U]$ .

Suppose that  $(\Delta, v) \in W[U]$ . We have that  $\Delta \in \mathcal{A}_i(c_k, \varphi)$  by the fact that  $\mathcal{A}$  is an  $S$ -evidence function and thus satisfies Constant Specification  $S$ . Further, we have that  $(U, U^e)^c, v \vdash c_k \gg_i \varphi$  by Axiom  $\text{CS}^c$  of the  $(U, U^e)^c$ -calculus (Figure 3). Thus  $v \notin U_i^-(c_k, \varphi)$ . But  $\Delta \in \mathcal{A}_i(c_k, \varphi)$  and  $v \notin U_i^-(c_k, \varphi)$  together imply that  $(\Delta, v) \in \mathcal{A}[U]_i(c_k, \varphi)$  by the definition of  $\mathcal{A}[U]$ . Since we chose  $(\Delta, v) \in W[U]$  arbitrarily, we have shown that  $\mathcal{A}[U]$  satisfies Constant Specification  $S$ .

- *Application.*  $\mathcal{A}[U]_i(t, \varphi \supset \psi) \cap \mathcal{A}[U]_i(s, \varphi) \subseteq \mathcal{A}[U]_i(t \cdot_\varphi s, \psi)$ .

Suppose that  $(\Delta, v) \in (\mathcal{A}[U]_i(t, \varphi \supset \psi) \cap \mathcal{A}[U]_i(s, \varphi))$ . It follows by the definition of  $\mathcal{A}[U]$  that  $\Delta \in (\mathcal{A}_i(t, \varphi \supset \psi) \cap \mathcal{A}_i(s, \varphi))$ ,  $v \notin U_i^-(t, \varphi \supset \psi)$ , and  $v \notin U_i^-(s, \varphi)$ . The first of these items implies  $\Delta \in \mathcal{A}_i(t \cdot_\varphi s)$  by the fact that  $\mathcal{A}$  is an  $S$ -evidence function and thus satisfies Application. The remaining two items imply that  $(U, U^e)^c, v \vdash t \gg_i (\varphi \supset \psi)$  and  $(U, U^e)^c, v \vdash s \gg_i \varphi$ . Applying Rule  $\text{A}^c$  of the  $(U, U^e)^c$ -calculus (Figure 3), it follows that  $(U, U^e)^c, v \vdash (t \cdot_\varphi s) \gg_i \psi$  and thus that  $v \notin U_i^-(t \cdot_\varphi s, \psi)$ . But  $\Delta \in \mathcal{A}_i(t \cdot_\varphi s)$  and  $v \notin U_i^-(t \cdot_\varphi s, \psi)$  together imply that  $(\Delta, v) \in \mathcal{A}[U]_i(t \cdot_\varphi s, \psi)$  by the definition of  $\mathcal{A}[U]$ . Since we chose  $(\Delta, v) \in (\mathcal{A}[U]_i(t, \varphi \supset \psi) \cap \mathcal{A}[U]_i(s, \varphi))$  arbitrarily, we have shown that  $\mathcal{A}[U]$  satisfies Application.

- *Sum.*  $\mathcal{A}[U]_i(t, \varphi) \cup \mathcal{A}[U]_i(s, \varphi) \subseteq \mathcal{A}[U]_i(t + s, \varphi)$ .

Suppose that  $(\Delta, v) \in \mathcal{A}[U]_i(t, \varphi)$ . It follows by the definition of  $\mathcal{A}[U]$  that  $\Delta \in \mathcal{A}_i(t, \varphi)$  and  $v \notin U_i^-(t, \varphi)$ . The former implies that  $\Delta \in \mathcal{A}_i(t + s, \varphi)$  because  $\mathcal{A}$  is an  $S$ -evidence function and thus satisfies Sum.  $v \notin U_i^-(t, \varphi)$  implies that  $(U, U^e)^c, v \vdash t \gg_i \varphi$  and thus that  $(U, U^e)^c, v \vdash (t + s) \gg_i \varphi$  by Rule  $\text{SL}^c$  of the  $(U, U^e)^c$ -calculus (Figure 3). The latter is equivalent to  $v \notin U_i^-(t + s, \varphi)$ . But  $\Delta \in \mathcal{A}_i(t + s, \varphi)$  and  $v \notin U_i^-(t + s, \varphi)$  together imply that  $(\Delta, v) \in \mathcal{A}[U]_i(t + s, \varphi)$  by the definition of  $\mathcal{A}[U]$ . Thus we have shown that  $(\Delta, v) \in \mathcal{A}[U]_i(t, \varphi)$  implies  $(\Delta, v) \in \mathcal{A}[U]_i(t + s, \varphi)$ .

By a similar argument (using Rule  $\text{SR}^c$  in place of Rule  $\text{SL}^c$ ), we see that  $(\Delta, v) \in \mathcal{A}[U]_i(s, \varphi)$  implies  $(\Delta, v) \in \mathcal{A}[U]_i(t + s, \varphi)$ . The result follows.

---

<sup>12</sup>For transitivity: the Kripke frame underlying  $M$  is transitive by the fact that  $M$  is an  $S$ -Fitting model (Definitions 3.13 and 3.14) and the Kripke frame underlying  $U$  is transitive by the fact that  $U$  is an update frame (Definition 2.10). That  $F[U]$  is transitive then follows from the definition of  $R[U]$  (Definition 3.15).

- *Checker.*  $\mathcal{A}[U]_i(t, \varphi) \subseteq \mathcal{A}[U]_i(!t, t :_i \varphi)$ .

Suppose that  $(\Delta, v) \in \mathcal{A}[U]_i(t, \varphi)$ , which implies that  $\Delta \in \mathcal{A}_i(t, \varphi)$  and  $v \notin U_i^-(t, \varphi)$  by the definition of  $\mathcal{A}[U]$ . The former implies that  $\Delta \in \mathcal{A}_i(!t, t :_i \varphi)$  by the fact that  $\mathcal{A}$  is an  $S$ -evidence function and thus satisfies Checker. Further,  $v \notin U_i^-(t, \varphi)$  implies  $(U, U^e)^c, v \vdash t \gg_i \varphi$ . By Rule C<sup>c</sup> of the  $(U, U^e)^c$ -calculus (Figure 3), it follows that  $(U, U^e)^c, v \vdash !t \gg_i (t :_i \varphi)$ , which is equivalent to  $v \notin U_i^-(!t, t :_i \varphi)$ . But  $\Delta \in \mathcal{A}_i(!t, t :_i \varphi)$  and  $v \notin U_i^-(!t, t :_i \varphi)$  together imply that  $(\Delta, v) \in \mathcal{A}[U]_i(!t, t :_i \varphi)$  by the definition of  $\mathcal{A}[U]$ . Thus we have shown that  $(\Delta, v) \in \mathcal{A}[U]_i(t, \varphi)$  implies  $(\Delta, v) \in \mathcal{A}[U]_i(!t, t :_i \varphi)$ . The result follows.

- *Monotonicity.*  $(\Delta, v)R_i(\Delta', v')$  and  $(\Delta, v) \in \mathcal{A}[U]_i(t, \varphi)$  together imply that  $(\Delta', v') \in \mathcal{A}[U]_i(t, \varphi)$ .

Suppose that  $(\Delta, v) \in \mathcal{A}[U]_i(t, \varphi)$  and  $(\Delta, v)R[U]_i(\Delta', v')$ , which implies that  $\Delta \in \mathcal{A}_i(t, \varphi)$  and  $v \notin U_i^-(t, \varphi)$  by the definition of  $\mathcal{A}[U]$  and that  $\Delta R_i \Delta'$  and  $v U_i v'$  by the definition of  $R[U]$ . Since  $\mathcal{A}$  is an  $S$ -evidence function, it also satisfies Monotonicity; therefore  $\Delta \in \mathcal{A}_i(t, \varphi)$  and  $\Delta R_i \Delta'$  together imply that  $\Delta' \in \mathcal{A}_i(t, \varphi)$ . Since  $v \notin U_i^-(t, \varphi)$  and  $v U_i v'$ , we have that  $(U, U^e)^c, v \vdash t \gg_i \varphi$  and  $v U_i v'$ ; therefore  $(U, U^e)^c, v' \vdash t \gg_i \varphi$  by Rule M<sup>c</sup> of the  $(U, U^e)^c$ -calculus (Figure 3). Further,  $(U, U^e)^c, v' \vdash t \gg_i \varphi$  is equivalent to  $v' \notin U_i^-(t, \varphi)$ . But  $\Delta' \in \mathcal{A}_i(t, \varphi)$  and  $v' \notin U_i^-(t, \varphi)$  together imply that  $(\Delta', v') \in \mathcal{A}[U]_i(t, \varphi)$  by the definition of  $\mathcal{A}[U]$ . The result follows.

- *Update.*  $\mathcal{A}[U]_i(t, \varphi) \subseteq \mathcal{A}[U]_i(t^{U,u}, [U, u]\varphi)$ .

Similar to the argument for Checker, except that we use the Update property in place of the Checker property and the rule U<sup>c</sup> in place of the rule C<sup>c</sup>.  $\square$

Finally, we introduce the standard notions of validity in terms of our notion of truth (from Definition 3.15).

**Definition 3.17** (Validity). Let  $A$  be an agent set,  $S$  be a set,  $\varphi$  be a  $\text{UL}^A$ -formula,  $M$  be an  $S$ -Fitting model for  $A$ , and  $\mathcal{I}$  be a set of  $S$ -Fitting models for  $A$ .

- To say that  $\varphi$  is *valid in*  $M$ , written  $M \models \varphi$ , means that  $M, \Gamma \models \varphi$  for each world  $\Gamma \in M$ .
- To say that  $\varphi$  is *valid for*  $\mathcal{I}$ , written  $\mathcal{I} \models \varphi$ , means that for each  $S$ -Fitting model  $M \in \mathcal{I}$ , we have that  $M \models \varphi$ .
- To say that  $\varphi$  is  *$S$ -valid*, written  $S \models \varphi$ , means that  $\varphi$  is valid for the set of all  $S$ -Fitting models for  $A$ .

As before, when it is unimportant or otherwise ought not cause confusion, we may omit the prefix “ $S$ -” when we refer to the concept of  $S$ -validity.

It will be our task in the next section to study an axiomatic characterization of the  $S$ -valid  $\text{UL}^A$ -formulas for a to-be-defined set  $S$  of “basic” assertions.

## 4 Axiomatics

In order to state our axiomatics and to prove its correctness (soundness and completeness), we introduce a notion of composition for update frames, which allows us to combine the communications and eliminations of two update frames  $U$  and  $U'$  into a single update frame  $U \circ U'$ . Our notion of composition is an adaptation of the standard composition operation in Dynamic Epistemic Logic [19].

**Definition 4.1** (Composition of Update Frames). Let  $A$  be an agent set and let  $U = (W, R, f, \mathcal{A})$  and  $U' = (W', R', f', \mathcal{A}')$  be update frames in  $\text{UL}^A$  for  $A$ . The *composition of  $U$  and  $U'$* , written  $U \circ U'$ , is the update frame  $(W^\circ, R^\circ, f^\circ, \mathcal{A}^\circ)$  in  $\text{UL}^A$  for  $A$  whose components are defined as follows.

- $W^\circ := W \times W'$ .
- For each  $i \in A$ , set  $R_i^\circ := \{((v, v'), (w, w')) \in W^\circ : vR_i w \text{ and } v'R'_i w'\}$ .
- For each  $(v, v') \in W^\circ$ , set  $f^\circ(v, v') := \neg[U, v]\neg U'(v')$ .
- For each  $i \in A$  and each  $(t, \varphi) \in \mathcal{T}(\text{UL}^A) \times \text{UL}^A$ , set

$$\mathcal{A}_i^\circ(t, \varphi) := \{(v, v') \in W^\circ : v \in \mathcal{A}_i(t, \varphi) \text{ or } v' \in \mathcal{A}'_i(t, \varphi)\} .$$

The composition  $U \circ U'$  has been designed so that the operation  $M \mapsto M[U \circ U']$  has the same effect with respect to eliminations as does the composite operation

$$M \mapsto M[U] \mapsto M[U][U'] .$$

The sense of “same effect with respect to eliminations” is given by the following lemma.

**Lemma 4.2** (Composition Lemma). Let  $\mathcal{A}$  be an agent set and let  $U$  and  $U'$  be update frames in  $\text{UL}^A$  for  $A$ . The following two statements are equivalent.

- Both  $(U, U^e)^c, w \vdash t \gg_i \varphi$  and  $(U', U'^e)^c, w' \vdash t \gg_i \varphi$ .
- $(U \circ U', (U \circ U')^e)^c, (w, w') \vdash t \gg_i \varphi$ .

See Figure 3 for the axiomatics corresponding to the calculi named in the statements above.

*Proof.* So as to reduce complexity of the many symbolic statements we will write in proving this lemma, let us adopt the following abbreviations for the remainder of the proof.

- $w \vdash t \gg_i \varphi$  abbreviates  $(U, U^e)^c, w \vdash t \gg_i \varphi$ .
- $w' \vdash t \gg_i \varphi$  abbreviates  $(U', U'^e)^c, w' \vdash t \gg_i \varphi$ .
- $(w, w') \vdash^\circ t \gg_i \varphi$  abbreviates  $(U \circ U', (U \circ U')^e)^c, (w, w') \vdash t \gg_i \varphi$ .

So we now prove that  $(w, w') \vdash^\circ t \gg_i \varphi$  if and only if  $w \vdash t \gg_i \varphi$  and  $w' \vdash' t \gg_i \varphi$  by induction on the construction of  $t \in \mathcal{T}(\text{UL}^A)$ .

- Base Case:  $t = c_k$ .

By Axiom CS<sup>c</sup>, we have that  $(w, w') \vdash^\circ c_k \gg_i \varphi$ , that  $w \vdash c_k \gg_i \varphi$ , and that  $w' \vdash' c_k \gg_i \varphi$ .

- Base Case:  $t = x_k$ .

Assume  $(w, w') \vdash^\circ x_k \gg_i \varphi$ . By inspection of the axiomatics, it follows that this assumption is equivalent to the statement that  $(w, w')(U \circ U')_i^*(v, v')$  implies  $(v, v') \notin (U \circ U')_i(x_k, \varphi)$ . By the definition of  $U \circ U'$  (Definition 4.1), this is equivalent to the statement that  $wU_i^*v$  and  $w'U_i'^*v'$  together imply  $v \notin U_i(x_k, \varphi)$  and  $v' \notin U_i'(x_k, \varphi)$ . But by our inspection of the axiomatics, the latter is equivalent to the statement that  $w \vdash x_k \gg_i \varphi$  and  $w' \vdash' x_k \gg_i \varphi$ .

- Induction Case:  $t = r \cdot_\psi s$ .

Assume  $(w, w') \vdash^\circ (r \cdot_\psi s) \gg_i \varphi$ . By inspection of the axiomatics, it follows that this assumption is equivalent to the statement that  $(v, v') \vdash^\circ r \gg_i (\psi \supset \varphi)$  and  $(v, v') \vdash^\circ s \gg_i \psi$  for some  $(v, v') \in (U \circ U)$  satisfying  $(v, v')(U \circ U')_i^*(w, w')$ . By the induction hypothesis and the definition of  $U \circ U'$  (Definition 4.1), this is equivalent to our having both that  $v \vdash r \gg_i (\psi \supset \varphi)$  and  $v \vdash s \gg_i \psi$  with  $vU_i^*w$  and also that  $v' \vdash' r \gg_i (\psi \supset \varphi)$  and  $v' \vdash' s \gg_i \psi$  with  $v'U_i'^*w'$ . By inspection of the axiomatics, it follows that the latter is equivalent to the statement that  $w \vdash (r \cdot_\psi s) \gg_i \varphi$  and  $w' \vdash' (r \cdot_\psi s) \gg_i \varphi$ .

- The Induction Cases for  $t = r + s$ ,  $t = !s$ , and  $t = s^{U'', u''}$  are handled similarly.  $\square$

The Composition Lemma (Lemma 4.2) and the Complementary Axiomatic Generation Lemma (Lemma 3.11) together tell us that  $(U \circ U')^e(u, u')$  does *not* eliminate  $t \gg_i \varphi$  if and only if neither  $U^e(u)$  nor  $U'^e(u')$  eliminates  $t \gg_i \varphi$ . Therefore, we have that  $(U \circ U')^e(u, u')$  eliminates  $t \gg_i \varphi$  if and only if one or more of  $U^e(u)$  or  $U'^e(u')$  eliminations  $t \gg_i \varphi$ . Thus we see that agent  $i$ 's evidence  $t$  relevant for  $\varphi$  will be eliminated in the operation

$$(M, \Gamma) \mapsto (M[U \circ U'], (\Gamma, (u, u')))$$

if and only if her evidence  $t$  relevant for  $\varphi$  is eliminated in at least one of the two steps in the composite operation

$$(M, \Gamma) \mapsto (M[U], (\Gamma, u)) \mapsto (M[U][U'], ((\Gamma, u), u')) .$$

It is in this sense that the operation  $M \mapsto M[U \circ U']$  has the same effect with respect to eliminations as does the composite operation  $M \mapsto M[U][U']$ .

The following simple theory, called  $\text{AX}^A$ , is used to formulate the axioms of our eventual target theory,  $\text{JLcd}^A$ . The reason for this separation of theories is that we want to ensure that whenever  $\varphi$  is an axiom, we have that  $c_k \vdash_i \varphi$  is also an axiom for each  $k \in \mathbb{N}$  and  $i \in A$ . In this way, the constants will serve as the agents' evidence for their explicit common belief in the axioms.

**Definition 4.3.** Let  $A$  be an agent set. The theory  $\text{AX}^A$  is defined in Figure 5. For a formula  $\varphi \in \text{UL}^A$ , we write  $\text{AX}^A \vdash \varphi$  to mean that  $\varphi$  is derivable in  $\text{AX}^A$ . The negation of the statement  $\text{AX}^A \vdash \varphi$  is written  $\text{AX}^A \not\vdash \varphi$ .

Finally, we define our theory of multi-agent Justification Logic with communication and evidence elimination,  $\text{JLce}^A$ .

**Definition 4.4.** The theory of multi-agent Justification Logic with communication and evidence elimination, written  $\text{JLce}^A$ , is defined in Figure 6. For a formula  $\varphi \in \text{UL}^A$ , we write  $\text{JLce}^A \vdash \varphi$  to mean that  $\varphi$  is derivable in  $\text{JLce}^A$ . The negation of the statement  $\text{JLce}^A \vdash \varphi$  is written  $\text{JLce}^A \not\vdash \varphi$ . Similarly, we will write  $\psi_1, \psi_2, \dots, \psi_n \vdash_{\text{JLce}^A} \varphi$  to mean that  $\varphi$  is derivable in  $\text{JLce}^A$  using the  $\text{UL}^A$ -formulas  $\psi_1, \dots, \psi_n$  as assumptions, where  $n \in \mathbb{N}$ . The negation of this statement of derivation from assumptions is written by replacing the turnstile (“ $\vdash$ ”) by a negated turnstile (“ $\not\vdash$ ”).

The following lemma, whose proof is a dull exercise in  $\text{JLce}^A$ -axiomatics,<sup>13</sup> presents a few important theorems of  $\text{JLce}^A$ . This lemma shows that evidence (and not just relevant evidence) is closed under the term-forming operations  $\cdot$ ,  $+$ , and  $!$ .

**Lemma 4.5.** Let  $A$  be an agent set. Then we have each of the following.

- $\text{JLce}^A \vdash (t :_i (\varphi \supset \psi)) \supset ((s :_i \varphi) \supset (t \cdot \varphi s) :_i \psi)$ .
- $\text{JLce}^A \vdash ((t :_i \varphi) \vee (s :_i \varphi)) \supset (t + s) :_i \varphi$ .
- $\text{JLce}^A \vdash (t :_i \varphi) \supset !t :_i (t :_i \varphi)$ .

One of the most important results in Justification Logic is Artemov’s Internalization Theorem [2]. Informally, this theorem says that agents can reason within the theory itself.

**Theorem 4.6** (Artemov’s Internalization Theorem; [2]). Let  $A$  be an agent set. For each  $i \in A$  and each  $\text{JLce}^A$ -theorem  $\chi$ , there is a variable-free term  $s$  such that  $s :_i \chi$  is also a  $\text{JLce}^A$ -theorem.

*Proof.* Throughout the remainder of the proof, whenever we write a turnstile (“ $\vdash$ ”), we mean the turnstile with the theory  $\text{JLce}^A$  written to the left. We prove by induction on the length of the  $\text{JLce}^A$ -derivation of  $\chi$  that there is a variable-free term  $s \in \text{UL}^A$  such that  $\vdash s :_i \chi$ .

- Base Case:  $\text{AX}^A \vdash \chi$ .  
 $\text{AX}^A \vdash \chi$  implies  $\text{AX}^A \vdash c_0 :_i \chi$  by Rule CN of the theory  $\text{AX}^A$ . We therefore have that  $\vdash c_0 :_i \chi$ , so take  $s := c_0$ . Observe that  $c_0$  is variable-free.

---

<sup>13</sup>See Figures 3.4, 3.5, and 3.8 (respectively) in [17] for full proofs, though note that all occurrences of  $\square$  should be replaced by  $B_i$ .

### CLASSICAL BELIEF AND EVIDENCE

- CL. Axiom schemes for classical propositional logic
- BK.  $B_i(\varphi \supset \psi) \supset (B_i\varphi \supset B_i\psi)$
- B4.  $B_i\varphi \supset B_i(B_i\varphi)$
- EA.  $(t \gg_i(\varphi \supset \psi)) \supset ((s \gg_i \varphi) \supset (t \cdot_\varphi s) \gg_i \psi)$
- ES.  $((t \gg_i \varphi) \vee (s \gg_i \varphi)) \supset ((t + s) \gg_i \varphi)$
- EC.  $(t \gg_i \varphi) \supset (!t \gg_i (t :_i \varphi))$
- EM.  $(t \gg_i \varphi) \supset B_i(t \gg_i \varphi)$
- EU.  $(t \gg_i \varphi) \supset (t^{U,u} \gg_i [U, u]\varphi)$
- C.  $(t :_i \varphi) \supset (t \gg_i \varphi)$   
 $(t :_i \varphi) \supset B_i\varphi$   
 $(t \gg_i \varphi) \supset (B_i\varphi \supset (t :_i \varphi))$

### ELIMINATION AND UPDATE

- E.  $(t \gg_i^{U,u} \varphi) \equiv \begin{cases} \top & \text{if } (U, U^e), u \vdash t \gg_i \varphi \\ \perp & \text{otherwise} \end{cases}$
- UA.  $[U, u]q \equiv U(u) \supset q$ , for  $q \in \{p_k, \perp, \top\}$
- U $\star$ .  $[U, u](\varphi \star \psi) \equiv [U, u]\varphi \star [U, u]\psi$ , for  $\star \in \{\supset, \wedge, \vee, \equiv\}$
- UN.  $[U, u]\neg\varphi \equiv U(u) \supset \neg[U, u]\varphi$
- UB.  $[U, u]B_i\varphi \equiv U(u) \supset \bigwedge_{uU_i v} B_i[U, v]\varphi$
- UE.  $[U, u](t \gg_i \varphi) \equiv U(u) \supset ((t \gg_i \varphi) \wedge \neg(t \gg_i^{U,u} \varphi))$
- UE.  $[U, u](t \gg_i^{U',u'} \varphi) \equiv U(u) \supset (t \gg_i^{U',u'} \varphi)$
- UU.  $[U, u][U', u']\varphi \equiv [U \circ U', (u, u')]\varphi$

### RULES

$$\frac{k \in \mathbb{N} \quad i \in A \quad \vdash \varphi}{\vdash c_k :_i \varphi} \text{ (CN)}$$

- Notes:
- Scheme C says that  $(t :_i \varphi) \equiv (t \gg_i \varphi) \wedge B_i\varphi$ .
  - Scheme E uses the  $(U, U^e)$ -calculus (see Definition 3.8 and Figure 2).

Figure 5: The theory  $\text{AX}^A$

RULES

$$\frac{\text{AX}^A \vdash \varphi}{\vdash \varphi} \text{ (AX)} \quad \frac{\vdash \varphi \supset \psi \quad \vdash \varphi}{\vdash \psi} \text{ (MP)}$$

$$\frac{\vdash \varphi}{\vdash B_i \varphi} \text{ (BN)} \quad \frac{\vdash \varphi}{\vdash [U, u] \varphi} \text{ (UN)}$$

Figure 6: The theory  $\text{JLce}^A$

- Induction Case:  $\vdash \varphi_1 \supset \varphi_2$  and  $\vdash \varphi_1$ ; further,  $\chi = \varphi_2$  follows from  $\varphi_1 \supset \varphi_2$  and  $\varphi_1$  by Rule MP.

Applying the induction hypothesis, we have that  $\vdash s_1 :_i (\varphi_1 \supset \varphi_2)$  and  $\vdash s_2 :_i \varphi_1$  for variable-free terms  $s_1, s_2 \in \mathcal{T}(\text{UL}^A)$ . But then it follows that  $\vdash (s_1 \cdot_{\varphi_1} s_2) :_i \varphi_2$  by the derivation in Figure 7. So take  $s := s_1 \cdot_{\varphi_1} s_2$ . Since  $s_1$  and  $s_2$  are variable-free,  $s$  is itself variable-free.

- Induction Case:  $\vdash \varphi$  and  $\chi = B_i \varphi$  follows from  $\varphi$  by Rule BN.

Applying the induction hypothesis, we have that  $\vdash r :_i \varphi$  for a variable-free term  $r \in \mathcal{T}(\text{UL}^A)$ . But then it follows that  $\vdash (c_0 \cdot_{r :_i \varphi} !r) :_i (B_i \varphi)$  by the derivation in Figure 8. So take  $s := c_0 \cdot_{r :_i \varphi} !r$ . Since  $r$  is variable-free, it follows that  $s$  is also variable-free.

- Induction Case:  $\vdash \varphi$  and  $\chi = [U, u] \varphi$  follows from  $\varphi$  by Rule UN.

Applying the induction hypothesis, we have that  $\vdash r :_i \varphi$  for a variable-free term  $r \in \mathcal{T}(\text{UL}^A)$ . In addition, we also have that  $\vdash \varphi$  by our inductive assumption. But then it follows that  $\vdash r^{U, u} :_i [U, u] \varphi$  by the derivation in Figure 9. So take  $s := r^{U, u}$ . Since  $r$  is variable-free, it follows that  $s$  is variable-free.  $\square$

Our Soundness Theorem says that if we take the set  $S$  of  $\text{AX}^A$ -theorems as the “basic” assertions that do not require detailed justification, then each  $\text{JLce}^A$ -theorem is  $S$ -valid.

**Theorem 4.7** (Soundness). Let  $A$  be an agent set. Define  $S$  by setting

$$S := \{\varphi \in \text{UL}^A : \text{AX}^A \vdash \varphi\} .$$

Then  $\text{JLce}^A \vdash \chi$  implies  $S \models \chi$ .

*Proof.* We first show by induction on the length of a derivation  $\text{AX}^A \vdash \chi$  that  $\chi$  is  $S$ -valid. In the base case of this induction,  $\chi$  is an  $\text{AX}^A$ -axiom. So let us examine each  $\text{AX}^A$ -axiom  $\chi$  in turn, arguing that  $\chi$  is  $S$ -valid.

1.	$s_1 :_i (\varphi_1 \supset \varphi_2)$	hypothesis
2.	$s_2 :_i \varphi_1$	hypothesis
3.	$(s_1 :_i (\varphi_1 \supset \varphi_2)) \supset (s_1 \gg_i (\varphi_1 \supset \varphi_2))$	AX
4.	$s_1 \gg_i (\varphi_1 \supset \varphi_2)$	MP: 1, 3
5.	$(s_2 :_i \varphi_1) \supset (s_2 \gg_i \varphi_1)$	AX
6.	$s_2 \gg_i \varphi_1$	MP: 2, 5
7.	$(s_1 \gg_i (\varphi_1 \supset \varphi_2)) \supset ((s_2 \gg_i \varphi_1) \supset (s_1 \cdot_{\varphi_1} s_2) \gg_i \varphi_2)$	AX
8.	$(s_1 \cdot_{\varphi_1} s_2) \gg_i \varphi_2$	MP: 4, 6, 7
9.	$(s_1 :_i (\varphi_1 \supset \varphi_2)) \supset B_i(\varphi_1 \supset \varphi_2)$	AX
10.	$B_i(\varphi_1 \supset \varphi_2)$	MP: 1, 9
11.	$(s_2 :_i \varphi_1) \supset B_i \varphi_1$	AX
12.	$B_i \varphi_1$	MP: 2, 11
13.	$B_i(\varphi_1 \supset \varphi_2) \supset (B_i \varphi_1 \supset B_i \varphi_2)$	AX
14.	$B_i \varphi_2$	MP: 10, 12, 13
15.	$((s_1 \cdot_{\varphi_1} s_2) \gg_i \varphi_2) \supset (B_i \varphi_2 \supset (s_1 \cdot_{\varphi_1} s_2) :_i \varphi_2)$	AX
16.	$(s_1 \cdot_{\varphi_1} s_2) :_i \varphi_2$	MP: 8, 14, 15

Figure 7: Proof that  $s_1 :_i (\varphi_1 \supset \varphi_2), s_2 :_i \varphi_1 \vdash_{\text{JLce}^A} (s_1 \cdot_{\varphi_1} s_2) :_i \varphi_2$ .

**CL.**  $S \models \chi$  if  $\chi$  is an axiom of classical propositional logic.

$\chi$  is  $S$ -valid by the standard truth-table arguments for classical propositional logic.

**BK.**  $S \models B_i(\varphi \supset \psi) \supset (B_i \varphi \supset B_i \psi)$ .

BK is  $S$ -valid by the standard modal logic argument for validity of BK on Kripke models [7].

**B4.**  $S \models B_i \varphi \supset B_i(B_i \varphi)$ .

The Kripke frame underlying each  $S$ -Fitting model is transitive (Definitions 3.13 and 3.14), so it follows that B4 is  $S$ -valid by the standard argument for validity of B4 on transitive Kripke frames [7].

**EA.**  $S \models (t \gg_i (\varphi \supset \psi)) \supset ((s \gg_i \varphi) \supset (t \cdot_{\varphi} s) \gg_i \psi)$ .

Let  $(M, \Gamma) = ((W, R, \mathcal{A}, V), \Gamma)$  be a pointed  $S$ -Fitting model.  $\mathcal{A}$  satisfies Application (Definitions 3.1 and 3.13), which means that

$$\mathcal{A}_i(t, \varphi \supset \psi) \cap \mathcal{A}_i(s, \varphi) \subseteq \mathcal{A}_i(t \cdot_{\varphi} s, \psi) .$$

Applying this to the definition of truth (Definition 3.15), it follows that EA is true at  $(M, \Gamma)$ . Since  $(M, \Gamma)$  is an arbitrary pointed  $S$ -Fitting model, we have shown that EA is  $S$ -valid.

1.	$r :_i \varphi$	hypothesis
2.	$(r :_i \varphi) \supset (r \gg_i \varphi)$	AX
3.	$r \gg_i \varphi$	MP: 1, 2
4.	$(r \gg_i \varphi) \supset (!r \gg_i (r :_i \varphi))$	AX
5.	$!r \gg_i (r :_i \varphi)$	MP: 3, 4
6.	$c_0 :_i ((r :_i \varphi) \supset B_i \varphi)$	AX
7.	$(c_0 :_i ((r :_i \varphi) \supset B_i \varphi)) \supset (c_0 \gg_i ((r :_i \varphi) \supset B_i \varphi))$	AX
8.	$c_0 \gg_i ((r :_i \varphi) \supset B_i \varphi)$	MP: 6, 7
9.	$(c_0 \gg_i ((r :_i \varphi) \supset B_i \varphi)) \supset$ $((!r \gg_i (r :_i \varphi)) \supset (c_0 \cdot_{r :_i \varphi} !r) \gg_i (B_i \varphi))$	AX
10.	$(c_0 \cdot_{r :_i \varphi} !r) \gg_i (B_i \varphi)$	MP: 5, 8, 9
11.	$(r :_i \varphi) \supset B_i \varphi$	AX
12.	$B_i \varphi$	MP: 1, 11
13.	$B_i \varphi \supset B_i(B_i \varphi)$	AX
14.	$B_i(B_i \varphi)$	MP: 12, 13
16.	$((c_0 \cdot_{r :_i \varphi} !r) \gg_i (B_i \varphi)) \supset$ $(B_i(B_i \varphi) \supset (c_0 \cdot_{r :_i \varphi} !r) :_i (B_i \varphi))$	AX
17.	$(c_0 \cdot_{r :_i \varphi} !r) :_i (B_i \varphi)$	MP: 10, 14, 16

Figure 8: Proof that  $r :_i \varphi \vdash_{\text{JLce}^A} (c_0 \cdot_{r :_i \varphi} !r) :_i (B_i \varphi)$ .

1.	$\varphi$	hypothesis
2.	$r :_i \varphi$	hypothesis
3.	$(r :_i \varphi) \supset (r \gg_i \varphi)$	AX
4.	$r \gg_i \varphi$	MP: 2, 3
5.	$(r \gg_i \varphi) \supset (r^{U,u} \gg_i [U, u] \varphi)$	AX
6.	$r^{U,u} \gg_i [U, u] \varphi$	MP: 4, 5
7.	$[U, u] \varphi$	UN: 1
8.	$B_i [U, u] \varphi$	BN: 7
9.	$(r^{U,u} \gg_i [U, u] \varphi) \supset (B_i [U, u] \varphi \supset (r^{U,u} :_i [U, u] \varphi))$	AX
10.	$r^{U,u} :_i [U, u] \varphi$	MP: 6, 8, 9

Figure 9: Proof that  $\varphi, r :_i \varphi \vdash_{\text{JLce}^A} r^{U,u} :_i [U, u] \varphi$ .

**ES.**  $S \models ((t \gg_i \varphi) \vee (s \gg_i \varphi)) \supset (t + s) \gg_i \varphi$ .

$S$ -evidence functions satisfy Sum, so argue as for EA.

**EC.**  $S \models (t \gg_i \varphi) \supset (!t \gg_i (t :_i \varphi))$ .

$S$ -evidence functions satisfy Checker, so argue as for EA.

**EM.**  $S \models (t \gg_i \varphi) \supset B_i(t \gg_i \varphi)$ .

Let  $(M, \Gamma) = ((W, R, \mathcal{A}, V), \Gamma)$  be a pointed  $S$ -Fitting model.  $\mathcal{A}$  satisfies Monotonicity (Definitions 3.1 and 3.13), which means that for each  $\Delta \in M$ , if  $\Gamma R_i \Delta$  and  $\Gamma \in \mathcal{A}_i(t, \varphi)$ , then  $\Delta \in \mathcal{A}_i(t, \varphi)$ . Applying the definition of truth (Definition 3.15), the latter is equivalent to the statement that  $M, \Gamma \models t \gg_i \varphi$  implies  $M, \Gamma \models B_i(t \gg_i \varphi)$ . But then EM is true at  $(M, \Gamma)$ . Since  $(M, \Gamma)$  is an arbitrary pointed  $S$ -Fitting model, we have shown that EM is  $S$ -valid.

**EU.**  $S \models (t \gg_i \varphi) \supset (t^{U,u} \gg_i [U, u]\varphi)$ .

$S$ -evidence functions satisfy Update, so argue as for EA.

**C.**  $S \models (t :_i \varphi) \equiv (t \gg_i \varphi) \wedge B_i \varphi$ .

C is  $S$ -valid by the definition of truth (Definition 3.15).

**E.**  $S \models (t \gg_i^{U,u} \varphi) \equiv \begin{cases} \top & \text{if } (U, U^e), u \vdash t \gg_i \varphi \\ \perp & \text{otherwise} \end{cases}$

E is  $S$ -valid by the definition of truth (Definition 3.15).

**UA.**  $S \models [U, u]q \equiv (U(u) \supset q)$  for  $q \in \{p_k, \perp, \top\}$ .

This argument is standard in Dynamic Epistemic Logic [19], but it will be instructive for us to work through the details.

Let  $(M, \Gamma) = ((W, R, \mathcal{A}, V), \Gamma)$  be a pointed  $S$ -Fitting model. Now in case  $q = \perp$  or  $q = \top$ , then it follows almost immediately from the definition of truth (Definition 3.15) that UA is true at  $(M, \Gamma)$ . So let us check the case where  $q = p_k$ . We may assume that  $M, \Gamma \models U(u)$ , for otherwise UA is certainly true at  $(M, \Gamma)$ . Applying the definition of truth, we then observe that  $M[U], (\Gamma, u) \models p_k$  means  $\Gamma \in V[U](p_k)$ , and the latter is equivalent to  $\Gamma \in V(p_k)$ , which is what it means to write  $M, \Gamma \models p_k$ . Since we assumed that  $M, \Gamma \models U(u)$ , we have shown that  $M[U], (\Gamma, u) \models p_k$  is equivalent to  $M, \Gamma \models U(u) \supset p_k$ , from which it follows by the definition of truth that UA is true at  $(M, \Gamma)$ . Since  $(M, \Gamma)$  is an arbitrary pointed  $S$ -Fitting model, we have shown that UA is  $S$ -valid.

**U $\star$ .**  $S \models [U, u](\varphi \star \psi) \equiv ([U, u]\varphi \star [U, u]\psi)$  for  $\star \in \{\supset, \wedge, \vee, \equiv\}$ .

This argument is also standard in Dynamic Epistemic Logic [19].

Let  $(M, \Gamma) = ((W, R, \mathcal{A}, V), \Gamma)$  be a pointed  $S$ -Fitting model. We may assume that  $M, \Gamma \models U(u)$ , for otherwise it follows easily by the definition of truth (Definition 3.15)

that  $U\star$  is true at  $(M, \Gamma)$ . Applying the definition of truth, we then observe that  $M[U], (\Gamma, u) \models \varphi \star \psi$  is equivalent to the statement that  $M[U], (\Gamma, u) \models \varphi$  star  $M[U], (\Gamma, u) \models \psi$ , where the word “star” is to be replaced by the English reading for the binary logical connective  $\star$ .<sup>14</sup> Since  $M, \Gamma \models U(u)$ , the latter implication is equivalent to  $M, \Gamma \models [U, u]\varphi \star [U, u]\psi$ . So we see that  $U\star$  is true at  $(M, \Gamma)$ . Since  $(M, \Gamma)$  is an arbitrary pointed  $S$ -Fitting model, we have shown that  $U\star$  is  $S$ -valid.

**UN.**  $S \models [U, u]\neg\varphi \equiv (U(u) \supset \neg[U, u]\varphi)$ .

This argument is also standard in Dynamic Epistemic Logic [19].

Let  $(M, \Gamma) = ((W, R, \mathcal{A}, V), \Gamma)$  be a pointed  $S$ -Fitting model. We may assume that  $M, \Gamma \models U(u)$ , for otherwise it follows easily by the definition of truth (Definition 3.15) that UN is true at  $(M, \Gamma)$ . Proceeding with the assumption  $M, \Gamma \models U(u)$ , it follows from the definition of truth that  $M[U], (\Gamma, u) \models \neg\varphi$ , which is itself equivalent to  $M[U], (\Gamma, u) \not\models \varphi$  by the definition of truth. Since we assumed that  $M, \Gamma \models U(u)$ , the latter is itself equivalent to  $M, \Gamma \not\models [U, u]\varphi$  by the definition of truth. But  $M, \Gamma \not\models [U, u]\varphi$  is what it means to have  $M, \Gamma \models \neg[U, u]\varphi$ . Again by our assumption that  $M, \Gamma \models U(u)$ , the latter is equivalent to  $M, \Gamma \models U(u) \supset \neg[U, u]\varphi$ . So we see that UN is true at  $(M, \Gamma)$ . Since  $(M, \Gamma)$  is an arbitrary pointed  $S$ -Fitting model, we have shown that UN is  $S$ -valid.

**UB.**  $S \models [U, u]B_i\varphi \equiv (U(u) \supset \bigwedge_{uU_iv} B_i[U, v]\varphi)$ .

This argument is also standard in Dynamic Epistemic Logic [19].

Let  $(M, \Gamma) = ((W, R, \mathcal{A}, V), \Gamma)$  be a pointed  $S$ -Fitting model. We may assume that  $M, \Gamma \models U(u)$ , for otherwise it follows easily by the definition of truth (Definition 3.15) that UB is true at  $(M, \Gamma)$ . Proceeding with the assumption  $M, \Gamma \models U(u)$ , it follows from the definition of truth that  $M[U], (\Gamma, u) \models B_i\varphi$  is equivalent to having  $M[U], (\Delta, v) \models \varphi$  for each  $(\Delta, v) \in M$  satisfying  $(\Gamma, u)R[U]_i(\Delta, v)$ . The latter is itself equivalent to the statement that  $M[U], (\Delta, v) \models \varphi$  for each  $\Delta \in M$  and each  $v \in U$  such that  $\Gamma R_i\Delta$ ,  $uU_iv$ , and  $M, \Delta \models U(v)$ . This is equivalent to saying that for each  $v \in U$  satisfying  $uU_iv$  and each  $\Delta \in M$  satisfying  $\Gamma R_i\Delta$ , if we have that  $M, \Delta \models U(v)$ , then  $M[U], (\Delta, v) \models \varphi$ . But this is equivalent to saying that for each  $v \in U$  satisfying  $uU_iv$  and each  $\Delta \in M$  satisfying  $\Gamma R_i\Delta$ , we have  $M, \Delta \models [U, u]\varphi$ . But the latter is itself equivalent to  $M, \Gamma \models \bigwedge_{uU_iv} B_i[U, v]\varphi$ . Under our assumption  $M, \Gamma \models U(u)$ , our statement in the previous sentence is equivalent to  $M, \Gamma \models U(u) \supset \bigwedge_{uU_iv} B_i[U, v]\varphi$ . So we see that UB is true at  $(M, \Gamma)$ . Since  $(M, \Gamma)$  is an arbitrary pointed  $S$ -Fitting model, we have shown that UB is  $S$ -valid.

**UE.**  $S \models [U, u](t \gg_i \varphi) \equiv (U(u) \supset ((t \gg_i \varphi) \wedge \neg(t \gg_i^{U, u} \varphi)))$ .

Let  $(M, \Gamma) = ((W, R, \mathcal{A}, V), \Gamma)$  be a pointed  $S$ -Fitting model. We may assume that  $M, \Gamma \models U(u)$ , for otherwise it follows easily by the definition of truth (Definition 3.15)

---

<sup>14</sup> $\supset$  is read “implies”,  $\wedge$  is read “and”,  $\vee$  is read “or”, and  $\equiv$  is read “if and only if”.

that UE is true at  $(M, \Gamma)$ . Proceeding with the assumption  $M, \Gamma \models U(u)$ , it follows from the definition of truth that  $M[U], (\Gamma, u) \models t \gg_i \varphi$ . But the latter is equivalent to  $(\Gamma, u) \in \mathcal{A}[U]_i(t, \varphi)$ , which is itself equivalent to the statement that  $\Gamma \in \mathcal{A}_i(t, \varphi)$  and  $u \notin U_i^-(t, \varphi)$ . By Lemma 3.9, the latter statement is equivalent to  $\Gamma \in \mathcal{A}_i(t, \varphi)$  and  $(U, U^e), u \not\vdash t \gg_i \varphi$ . Applying the definition of truth, this is equivalent to the statement that  $M, \Gamma \models t \gg_i \varphi$  and  $M, \Gamma \models \neg(t \gg_i^{U, u} \varphi)$ . Since we assumed that  $M, \Gamma \models U(u)$ , the latter statement is equivalent to  $M, \Gamma \models U(u) \supset ((t \gg_i \varphi) \wedge \neg(t \gg_i^{U, u} \varphi))$ . Since  $(M, \Gamma)$  is an arbitrary pointed  $S$ -Fitting model, we have shown that UE is  $S$ -valid.

**UE.**  $S \models [U, u](t \gg_i^{U', u'} \varphi) \equiv (U(u) \supset (t \gg_i^{U', u'} \varphi))$ .

Let  $(M, \Gamma) = ((W, R, \mathcal{A}, V), \Gamma)$  be a pointed  $S$ -Fitting model. We may assume that  $M, \Gamma \models U(u)$ , for otherwise it follows easily by the definition of truth (Definition 3.15) that UE is true at  $(M, \Gamma)$ . Proceeding with the assumption  $M, \Gamma \models U(u)$ , it follows from the definition of truth that  $M[U], (\Gamma, u) \models t \gg_i^{U', u'} \varphi$ . But the latter is equivalent to  $(U', U^e), u' \vdash t \gg_i \varphi$  by the definition of truth. Applying the definition of truth again,  $(U', U^e), u' \vdash t \gg_i \varphi$  is what it means to have  $M, \Gamma \models t \gg_i^{U', u'} \varphi$ . Since we assumed  $M, \Gamma \models U(u)$ , the latter is equivalent to  $M, \Gamma \models U(u) \supset (t \gg_i^{U', u'} \varphi)$ . Since  $(M, \Gamma)$  is an arbitrary  $S$ -Fitting model, we have shown that UE is  $S$ -valid.

**UU.**  $S \models [U, u][U', u']\varphi \equiv [U \circ U', (u, u')]\varphi$ .

Many of the components of this argument are standard in Dynamic Epistemic Logic [19], though our argument here must take into account the extra features that come into play due to the structure of our update frames. So let us proceed with the argument in full.

Let  $(M, \Gamma) = ((W, R, \mathcal{A}, V), \Gamma)$  be a pointed  $S$ -Fitting model. We may assume that  $M, \Gamma \models \neg[U, u]\neg U'(u')$ , for otherwise it follows easily by the definition of truth (Definition 3.15) that UU is true at  $(M, \Gamma)$ . Proceeding with the assumption  $M, \Gamma \models \neg[U, u]\neg U'(u')$ , for us to prove that UU is true at  $(M, \Gamma)$ , it suffices for us to show that the  $S$ -Fitting models  $M[U][U']$  and  $M[U \circ U']$  are isomorphic, by which we mean that there is a bijection

$$f : M[U][U'] \rightarrow M[U \circ U']$$

mapping each world  $\Omega \in M[U][U']$  to a world  $f(\Omega) \in M[U \circ U']$  such that each of the following three properties is satisfied.

- $\Omega_1 R[U][U']_i \Omega_2$  if and only if  $f(\Omega_1) R[U \circ U']_i f(\Omega_2)$  for each  $i \in A$ .
- $\Omega \in \mathcal{A}[U][U']_i(t, \psi)$  if and only if  $f(\Omega) \in \mathcal{A}[U \circ U']_i(t, \psi)$  for each  $i \in \mathcal{A}$  and each  $(t, \psi) \in \mathcal{T}(\mathbf{UL}^A) \times \mathbf{UL}^A$ .
- $\Omega \in V[U][U'](p_k)$  if and only if  $f(\Omega) \in V[U \circ U'](p_k)$  for each  $k \in \mathbb{N}$ .

Proceeding, define  $f$  by setting  $f((\Delta, v), v') := (\Delta, (v, v'))$  for each  $((\Delta, v), v') \in M[U][U']$ . Observe that  $f$  is a well-defined bijection: we have  $((\Delta, v), v') \in M[U][U']$  if and only if  $M, \Delta \models \neg[U, v]\neg U'(v')$ , and the latter holds if and only if we have

$(\Delta, (v, v')) \in M[U \circ U']$  by the definition of  $(U \circ U')(v, v')$  as the formula  $\neg[U, v]\neg U'(v')$  (Definition 4.1). So let us now argue that  $f$  satisfies each of the three properties above.

For the first property, we have  $((\Delta, v), v')R[U][U']_i((\Omega, w), w')$  if and only if both  $(\Delta, v)R[U]_i(\Omega, w)$  and  $v'U'_i w'$  if and only if  $\Delta R_i \Omega$ ,  $vU_i w$ , and  $v'U'_i w'$ . But the latter is equivalent to the statement that  $\Delta R_i \Omega$  and  $(v, v')(U \circ U')_i(w, w')$  by the definition of  $U \circ U'_i$  (Definition 4.1). But this statement is itself equivalent to the statement that  $(\Delta, (v, v'))R[U \circ U']_i(\Omega, (w, w'))$  by the definition of truth.

For the second property, suppose  $((\Delta, v), v') \in \mathcal{A}[U][U']_i(t, \psi)$ . This means that  $\Delta \in \mathcal{A}_i(t, \psi)$ ,  $v \notin U_i^-(t, \psi)$ , and  $v' \notin U'_i(t, \psi)$  by the definition of truth. By Lemma 3.11, the latter is equivalent to  $\Delta \in \mathcal{A}_i(t, \psi)$ ,  $(U, U^e)^c, v \vdash t \gg_i \psi$ , and  $(U', U'^e)^c, v' \vdash t \gg_i \psi$ . By the Composition Lemma (Lemma 4.2), the latter is itself equivalent to the statement that both  $\Delta \in \mathcal{A}_i(t, \psi)$  and

$$(U \circ U', (U \circ U')^e)^c, (v, v') \vdash t \gg_i \psi .$$

By Lemma 3.11, this is itself equivalent to  $\Delta \in \mathcal{A}_i(t, \psi)$  and  $(v, v') \notin (U \circ U')_i^-(t, \psi)$ . By the definition of truth, this is equivalent to  $(\Delta, (v, v')) \in \mathcal{A}[U \circ U']_i(t, \psi)$ .

For the third and final property, we have  $((\Delta, v), v') \in V[U][U'](p_k)$  if and only if  $(\Delta, v) \in V[U](p_k)$  if and only if  $\Delta \in V(p_k)$  if and only if  $(\Delta, (v, v')) \in V[U \circ U']$ .

So we have shown that  $f$  is indeed an isomorphism and thus that  $M[U][U']$  and  $M[U \circ U']$  are isomorphic. It follows that UU is true at  $(M, \Gamma)$ . Since  $(M, \Gamma)$  is an arbitrary  $S$ -Fitting model, we have shown that UU is  $S$ -valid.

We have completed the base case of an inductive argument showing that every  $\text{AX}^A$ -theorem is  $S$ -valid. For the induction case of this argument, we assume that we have derived  $c_k :_i \varphi$  using Rule CN of the theory  $\text{AX}^A$ . Applying the induction hypothesis, we have that  $\varphi$  is  $S$ -valid. Letting  $(M, \Gamma) = ((W, R, \mathcal{A}, V), \Gamma)$  be a pointed  $S$ -Fitting model, it follows from the  $S$ -validity of  $\varphi$  that we have  $M, \Delta \models \varphi$  for each  $\Delta \in M$  satisfying  $\Gamma R_i \Delta$ . Further, since  $(c_k :_i \varphi) \in (S \cap \text{UL}^A)$ , it follows that  $\Gamma \in \mathcal{A}_i(c_k, \varphi)$  by the fact that  $\mathcal{A}$  is an  $S$ -evidence function and thus satisfies Constant Specification  $S$ . But then we have that  $\Gamma \in \mathcal{A}_i(c_k, \varphi)$  and that  $M, \Delta \models \varphi$  for each  $\Delta \in M$  satisfying  $\Gamma R_i \Delta$ . Applying the definition of truth, it follows that  $M, \Gamma \models c_k :_i \varphi$ . Since  $(M, \Gamma)$  was an arbitrary  $S$ -Fitting model, we have proved that  $c_k :_i \varphi$  is  $S$ -valid. Conclusion: every  $\text{AX}^A$ -theorem is  $S$ -valid. This conclusion allows us to prove the statement of the present theorem; so let us now argue by induction on the length of  $\text{JLce}^A$ -derivation that each  $\text{JLce}^A$ -theorem is  $S$ -valid.

- Base Case: Rule AX was used to derive  $\varphi$ .

It follows that  $\varphi$  is an  $\text{AX}^A$ -theorem, and we already argued that every  $\text{AX}^A$ -theorem is  $S$ -valid.

- Induction Case: Rule MP was used to derive  $\psi$ .

The standard argument for validity of the rule of Modus Ponens in Kripke models shows that MP is  $S$ -valid [7].

- Induction Case: Rule BN was used to derive  $B_i\varphi$ .

The standard argument for validity of the rule of modal necessitation in Kripke models shows that BN is  $S$ -valid [7].

- Induction Case: Rule UN was used to derive  $[U, u]\varphi$ .

By the induction hypothesis, we have that  $\varphi$  is  $S$ -valid. Now let  $(M, \Gamma)$  be an arbitrary pointed  $S$ -Fitting model. If we have that  $M, \Gamma \not\models U(u)$ , then it follows that  $M, \Gamma \models [U, u]\varphi$  by the definition of truth. If we have that  $M, \Gamma \models U(u)$ , then it follows from Lemma 3.16 that  $(M[U], (\Gamma, u))$  is a pointed  $S$ -Fitting model and thus that  $M[U], (\Gamma, u) \models \varphi$  by the  $S$ -validity of  $\varphi$ . Applying the definition of truth, we see that  $M, \Gamma \models [U, u]\varphi$  if we have that  $M, \Gamma \models U(u)$ . Thus we see that no matter whether  $M, \Gamma \models U(u)$ , it follows that  $M, \Gamma \models [U, u]\varphi$ . Since we chose the pointed  $S$ -Fitting model  $(M, \Gamma)$  arbitrarily, we have shown that  $[U, u]\varphi$  is  $S$ -valid.

Conclusion: each  $\text{JLce}^A$ -theorem is  $S$ -valid. □

Our Completeness Theorem says that if we take the set  $S$  of  $\text{AX}^A$ -theorems as the “basic” assertions that do not require detailed justification, then each  $S$ -valid  $\text{UL}^A$ -formula is a  $\text{JLce}^A$ -theorem.

**Theorem 4.8** (Completeness). Let  $A$  be an agent set. Define the set  $S$  by setting

$$S := \{\varphi \in \text{UL}^A : \text{AX}^A \vdash \varphi\} .$$

Then  $S \models \chi$  implies  $\text{JLce}^A \vdash \chi$ .

*Proof.* Our proof proceeds by a canonical model argument, so let us first make some preliminary definitions. The *conjunction* of a finite set of  $\text{UL}^A$ -formulas is just the conjunction whose conjuncts are the members of that finite set. To say that a  $\text{UL}^A$ -formula  $\varphi$  *implies*  $\perp$  means that  $\text{JLce}^A \vdash \varphi \supset \perp$ . To say that a set of  $\text{UL}^A$ -formulas is *consistent* means that no conjunction of a finite subset implies  $\perp$ . To say that a set of  $\text{UL}^A$ -formulas is *inconsistent* means that the set is not consistent. To say that a set of  $\text{UL}^A$ -formulas is *maximal consistent* means that the set is consistent and adding any  $\text{UL}^A$ -formula not already present in the set will result in an inconsistent set. By a Lindenbaum Argument, any consistent set of  $\text{UL}^A$ -formulas may be extended to a maximal consistent set. The *canonical model* is the structure  $M = (((W, R), \mathcal{A}), V)$  whose components are defined as follows.

- $W$  is the set of all maximal consistent sets of  $\text{UL}^A$ -formulas.
- $R : A \rightarrow 2^{W \times W}$  is defined by setting  $\Gamma R_i \Delta$  whenever  $\{\varphi \in \text{UL}^A : B_i\varphi \in \Gamma\} \subseteq \Delta$ .
- $\mathcal{A} : A \rightarrow (\mathcal{T}(\text{UL}^A) \times \text{UL}^A \rightarrow 2^W)$  is defined by  $\mathcal{A}_i(t, \varphi) := \{\Gamma \in W : (t \gg_i \varphi) \in \Gamma\}$ .
- $V : \{p_k : k \in \mathbb{N}\} \rightarrow 2^W$  is defined by  $V(p_k) := \{\Gamma \in W : p_k \in \Gamma\}$ .

Let us now prove that the canonical model is in fact an  $S$ -Fitting model. This requires us to prove that  $M$  satisfies each of the defining properties of an  $S$ -Fitting model for  $A$  (Definition 3.14). It suffices to verify three items: that  $F := (W, R)$  is a transitive Kripke frame for  $A$ , that  $\mathcal{A}$  is an  $S$ -evidence function on  $F$ , and that  $V$  is a valuation on  $F$ . The third of these items is obvious, so we will only focus our attention on the first two.

- $F := (W, R)$  is a transitive Kripke frame for  $A$ .

It follows from the soundness of  $\mathbf{JLce}^A$  with respect to  $S$ -Fitting models (Theorem 4.7) that  $W$  is nonempty. Further, since  $\mathbf{JLce}^A$  proves the scheme  $B_i\varphi \supset B_i(B_i\varphi)$ , it follows that  $R_i$  is transitive for each  $i \in A$  using the standard modal argument showing that provability of this scheme brings about a transitive canonical model [7]. Conclusion:  $F$  is a transitive Kripke frame for  $A$ .

- $\mathcal{A}$  is an  $S$ -evidence function on  $F$ .

By Definition 3.1, we must show that  $\mathcal{A}$  satisfies each of Constant Specification  $S$ , Application, Sum, Checker, Monotonicity, and Update.

- *Constant Specification  $S$ .* For each  $k \in \mathbb{N}$  and each  $(c_k :_i \varphi) \in (S \cap \mathbf{UL}^A)$ , we have  $\mathcal{A}_i(c_k, \varphi) = W$ .

Observe that  $\mathbf{JLce}^A \vdash c_k :_i \varphi$  by Rule AX. Applying Axiom C and propositional logic, it follows that  $\mathbf{JLce}^A \vdash c_k \gg_i \varphi$ . But then we have for each  $\Gamma \in W$  that  $(c_k \gg_i \varphi) \in \Gamma$  by the maximal consistency of  $\Gamma$ . Applying the definition of  $\mathcal{A}$ , the result follows.

- *Application.*  $\mathcal{A}_i(t, \varphi \supset \psi) \cap \mathcal{A}_i(s, \varphi) \subseteq \mathcal{A}_i(t \cdot_\varphi s, \psi)$ .

If  $\Gamma \in (\mathcal{A}_i(t, \varphi \supset \psi) \cap \mathcal{A}_i(s, \varphi))$ , then we have that  $(t \gg_i (\varphi \supset \psi)) \in \Gamma$  and  $(s \gg_i \varphi) \in \Gamma$  by the definition of  $\mathcal{A}$ . But then  $((t \cdot_\varphi s) \gg_i \psi) \in \Gamma$  by Axiom EA and the maximal consistency of  $\Gamma$ . Applying the definition of  $\mathcal{A}$ , the result follows.

- *Sum, Checker, and Update* are shown much as was Application, except that we make use of Axioms ES, EC, and EU (respectively).

- *Monotonicity.*  $\Gamma R_i \Delta$  and  $\Gamma \in \mathcal{A}_i(t, \varphi)$  together imply that  $\Delta \in \mathcal{A}_i(t, \varphi)$ .

Suppose  $\Gamma R_i \Delta$  and  $\Gamma \in \mathcal{A}_i(t, \varphi)$ . The latter means that  $(t \gg_i \varphi) \in \Gamma$  by the definition of  $\mathcal{A}$ . But then  $B_i(t \gg_i \varphi) \in \Gamma$  by Axiom EM and the maximal consistency of  $\Gamma$ . Since  $\Gamma R_i \Delta$  and  $B_i(t \gg_i \varphi) \in \Gamma$ , it follows by the definition of  $R_i$  that  $(t \gg_i \varphi) \in \Delta$  and thus that  $\Delta \in \mathcal{A}_i(t, \varphi)$  by the definition of  $\mathcal{A}$ .

Thus we have shown that  $\mathcal{A}$  is an  $S$ -evidence function on  $F$ .

We conclude that  $M$  is in fact an  $S$ -Fitting model. The key property of  $M$  that we now wish to prove is known as the *Truth Lemma*: for each  $\varphi \in \mathbf{UL}^A$  and each  $\Gamma \in W$ , we have that  $\varphi \in \Gamma$  if and only if  $M, \Gamma \models \varphi$ . But before we prove the Truth Lemma, let us recall the function  $d : \mathbf{UL}^A \rightarrow \mathbb{N}$  that we defined in Figure 4 (on Page 25). Following the lead in [19], we state and prove a number of properties of  $d$  that will come up in the proof of the Truth Lemma. These properties and their proofs are as follows.

$$\begin{aligned}
& d(U(u) \supset q) \\
&= 1 + \max\{d(U(u)), 1\} \\
&= 1 + d(U(u)) \\
&< (4 + d(U))^5 \\
&= d([U, u]q)
\end{aligned}$$

Figure 10: Proof that  $d(U(u) \supset q) < d([U, u]q)$  for  $q \in \{p_k, \perp, \top\}$ .

$$\begin{aligned}
& d([U, u]\varphi \supset [U, u]\psi) \\
&= 1 + \max\{(4 + d(U))^{4+d(\varphi)} \cdot d(\varphi), (4 + d(U))^{4+d(\psi)} \cdot d(\psi)\} \\
&= 1 + (4 + d(U))^{4+\max\{d(\varphi), d(\psi)\}} \cdot \max\{d(\varphi), d(\psi)\} \\
&< (4 + d(U))^{5+\max\{d(\varphi), d(\psi)\}} \cdot (1 + \max\{d(\varphi), d(\psi)\}) \\
&= d([U, u](\varphi \supset \psi))
\end{aligned}$$

Figure 11: Proof that  $d([U, u]\varphi \star [U, u]\psi) < d([U, u](\varphi \star \psi))$  for each  $\star \in \{\supset, \wedge, \vee, \equiv\}$ .

- If  $\psi$  is a proper subformula of  $\varphi$ , then  $d(\psi) < d(\varphi)$ .

This follows by an inspection of Figure 4 (on Page 25).

- For the  $\text{AX}^A$ -axioms UA,  $U\star$ , UN, UB, UE, UE, and UU, each of which is of the form  $[U, u]\varphi \equiv \psi$ , we have that  $d(\psi) < d([U, u]\varphi)$ .

The results for UA,  $U\star$ , UN, UB, UE, UE, and UU are proved, respectively, in Fig-

$$\begin{aligned}
& d(U(u) \supset \neg[U, u]\varphi) \\
&= 1 + \max\{d(U(u)), 1 + (4 + d(U))^{4+d(\varphi)} \cdot d(\varphi)\} \\
&= 2 + (4 + d(U))^{4+d(\varphi)} \cdot d(\varphi) \\
&< (4 + d(U))^{5+d(\varphi)} \cdot (1 + d(\varphi)) \\
&= d([U, u]\neg\varphi)
\end{aligned}$$

Figure 12: Proof that  $d(U(u) \supset \neg[U, u]\varphi) < d([U, u]\neg\varphi)$ .

$$\begin{aligned}
& d(U(u) \supset \bigwedge_{wU_i v} B_i[U, v]\varphi) \\
\leq & 1 + \max\{d(U(u)), |U| + 1 + (4 + d(U))^{4+d(\varphi)} \cdot d(\varphi)\} \\
= & 2 + |U| + (4 + d(U))^{4+d(\varphi)} \cdot d(\varphi) \\
< & (4 + d(U))^{5+d(\varphi)} + (4 + d(U))^{5+d(\varphi)} \cdot d(\varphi) \\
= & (4 + d(U))^{5+d(\varphi)} \cdot (1 + d(\varphi)) \\
= & d([U, u]B_i\varphi)
\end{aligned}$$

Figure 13: Proof that  $d(U(u) \supset \bigwedge_{wU_i v} B_i[U, v]\varphi) < d([U, u]B_i\varphi)$ .

$$\begin{aligned}
& d(U(u) \supset ((t \gg_i \varphi) \wedge \neg(t \gg_i^{U, u} \varphi))) \\
= & 1 + \max\{d(U(u)), 1 + \max\{d(t \gg_i \varphi), 1 + d(t \gg_i^{U, u} \varphi)\}\} \\
= & 1 + \max\{d(U(u)), 1 + d(t \gg_i \varphi)\} \\
< & 2 + d(U) + d(t \gg_i \varphi) \\
< & (4 + d(U))^{4+d(t \gg_i \varphi)} \cdot d(t \gg_i \varphi) \\
= & d([U, u](t \gg_i \varphi))
\end{aligned}$$

Figure 14: Proof that  $d(U(u) \supset ((t \gg_i \varphi) \wedge \neg(t \gg_i^{U, u} \varphi))) < d([U, u](t \gg_i \varphi))$ .

$$\begin{aligned}
& d(U(u) \supset (t \gg_i^{U',u'} \varphi)) \\
&= 1 + \max\{d(U(u)), d(t \gg_i^{U',u'} \varphi)\} \\
&< 1 + d(U) + d(t \gg_i^{U',u'} \varphi) \\
&< (4 + d(U))^{4+d(t \gg_i^{U',u'} \varphi)} \cdot d(t \gg_i^{U',u'} \varphi) \\
&= d([U, u](t \gg_i^{U',u'} \varphi))
\end{aligned}$$

Figure 15: Proof that  $d(U(u) \supset (t \gg_i^{U',u'} \varphi)) < d([U, u](t \gg_i^{U',u'} \varphi))$ .

ures 10, 11, 12, 13, 15, 15, and 16.

- $d((t \gg_i \varphi) \wedge B_i \varphi) < d(t :_i \varphi)$ .  
This is proved in Figure 17.
- $d([U, u]((t \gg_i \varphi) \wedge B_i \varphi)) < d([U, u]t :_i \varphi)$ .  
This is proved in Figure 18.

We now use the function  $d$  to define the  $d$ -ordering of  $\text{UL}^A$ -formulas:  $\varphi$  comes before  $\psi$  in the  $d$ -ordering if and only if  $d(\varphi) < d(\psi)$ . Let us now prove the Truth Lemma by induction on the  $d$ -ordering of  $\text{UL}^A$ -formulas (Definition 2.12).

- Base Cases: for each  $q \in \{p_k, \perp, \top\}$ , we have that  $q \in \Gamma$  if and only if  $M, \Gamma \models q$ .  
If  $q = p_k$ , then the result follows immediately from our definition of  $V$ . Otherwise, the result follows by the maximal consistency of  $\Gamma$  and the definition of truth.
- Induction Case:  $(\varphi \star \psi) \in \Gamma$  if and only if  $M, \Gamma \models \varphi \star \psi$  for each  $\star \in \{\supset, \wedge, \vee, \equiv\}$ .  
Since  $d(\varphi) < d(\varphi \star \psi)$  and  $d(\psi) < d(\varphi \star \psi)$ , the induction hypothesis applies to  $\varphi$  and to  $\psi$ . The result follows easily.
- Induction Case:  $\neg\varphi \in \Gamma$  if and only if  $M, \Gamma \models \neg\varphi$ .  
Since  $d(\varphi) < d(\neg\varphi)$ , the induction hypothesis applies to  $\varphi$ . The result follows easily.
- Induction Case:  $B_i \varphi \in \Gamma$  if and only if  $M, \Gamma \models B_i \varphi$ .  
Suppose  $B_i \varphi \in \Gamma$  and  $\Gamma R_i \Delta$ . It follows that  $\varphi \in \Delta$  by the definition of  $R_i$ . Since  $d(\varphi) < d(B_i \varphi)$ , we may apply the induction hypothesis to conclude that  $M, \Delta \models \varphi$ . But then we have shown that  $\Gamma R_i \Delta$  implies  $M, \Delta \models \varphi$ , which is the meaning of  $M, \Gamma \models B_i \varphi$ .

$$\begin{aligned}
& d([U \circ U', (u, u')] \varphi) \\
= & (4 + d(U \circ U'))^{4+d(\varphi)} \cdot d(\varphi) \\
= & (4 + |U| \cdot |U'| + \max_{(v, v') \in (U \circ U')} \{d(\neg[U, v] \neg U'(v'))\})^{4+d(\varphi)} \cdot d(\varphi) \\
= & (5 + |U| \cdot |U'| + \max_{v' \in U'} \{(4 + d(U))^{5+d(U'(v'))} \cdot (1 + d(U'(v')))\})^{4+d(\varphi)} \cdot d(\varphi) \\
\leq & (5 + |U| \cdot |U'| + (4 + d(U))^{4+d(U')} \cdot d(U'))^{4+d(\varphi)} \cdot d(\varphi) \\
< & (2 \cdot (4 + d(U))^{4+d(U')} \cdot d(U'))^{4+d(\varphi)} \cdot d(\varphi) \\
< & ((4 + d(U))^{5+d(U')} \cdot d(U'))^{4+d(\varphi)} \cdot d(\varphi) \\
< & (4 + d(U))^{(5+d(U')) \cdot (4+d(\varphi))} \cdot (4 + d(U'))^{4+d(\varphi)} \cdot d(\varphi) \\
= & (4 + d(U))^{20+5d(\varphi)+4d(U')+d(U')d(\varphi)} \cdot (4 + d(U'))^{4+d(\varphi)} \cdot d(\varphi) \\
< & (4 + d(U))^{4+(4+d(U'))^5 \cdot d(\varphi)} \cdot (4 + d(U'))^{4+d(\varphi)} \cdot d(\varphi) \\
< & (4 + d(U))^{4+(4+d(U'))^{4+d(\varphi)} \cdot d(\varphi)} \cdot (4 + d(U'))^{4+d(\varphi)} \cdot d(\varphi) \\
= & d([U, u][U', u'] \varphi)
\end{aligned}$$

Figure 16: Proof that  $d([U \circ U', (u, u')] \varphi) < d([U, u][U', u'] \varphi)$ .

$$\begin{aligned}
& d((t \gg_i \varphi) \wedge B_i \varphi) \\
= & 1 + \max\{2 + \max\{d(t), d(\varphi)\}, 1 + d(\varphi)\} \\
= & 3 + \max\{d(t), d(\varphi)\} \\
< & 4 + \max\{d(t), d(\varphi)\} \\
= & d(t :_i \varphi)
\end{aligned}$$

Figure 17: Proof that  $d((t \gg_i \varphi) \wedge B_i \varphi) < d(t :_i \varphi)$ .

$$\begin{aligned}
& d([U, u]((t \gg_i \varphi) \wedge B_i \varphi)) \\
= & (4 + d(U))^{4+d((t \gg_i \varphi) \wedge B_i \varphi)} \cdot d((t \gg_i \varphi) \wedge B_i \varphi) \\
< & (4 + d(U))^{4+d(t :_i \varphi)} \cdot d(t :_i \varphi) \\
= & d([U, u]t :_i \varphi)
\end{aligned}$$

Figure 18: Proof that  $d([U, u]((t \gg_i \varphi) \wedge B_i \varphi)) < d([U, u]t :_i \varphi)$ .

Conversely, suppose  $B_i\varphi \notin \Gamma$ . Let  $\Delta' := \{\neg\varphi\} \cup \{\psi \in \text{UL}^A : B_i\psi \in \Gamma\}$ . We claim that  $\Delta'$  is consistent. Proof by contradiction: if  $\Delta'$  is inconsistent, then it follows from the consistency of  $\Gamma$  that would have  $\text{JLce}^A \vdash (\bigwedge_{j=0}^n \psi_j) \supset \varphi$  for some  $n \in \mathbb{N}$  such that  $B_i\psi_j \in \Gamma$  for each  $j = 0, 1, 2, \dots, n$ . Applying modal reasoning, we would then have  $\text{JLce}^A \vdash (\bigwedge_{j=0}^n B_i\psi_j) \supset B_i\varphi$  and thus that  $B_i\varphi \in \Gamma$  by the maximal consistency of  $\Gamma$ . But  $B_i\varphi \in \Gamma$  contradicts our original assumption that  $B_i\varphi \notin \Gamma$ . Therefore  $\Delta'$  is in fact consistent and so may be extended to a maximal consistent set  $\Delta \in M$ . Since  $\Delta' \subseteq \Delta$ , it follows by the definition of  $R_i$  that  $\Gamma R_i \Delta$ . Further, we have that  $\neg\varphi \in \Delta$  and thus that  $\varphi \notin \Delta$  by the consistency of  $\Delta$ . Since  $d(\varphi) < d(B_i\varphi)$ , we may apply the induction hypothesis to conclude that  $M, \Delta \not\models \varphi$ . But since  $\Gamma R_i \Delta$ , it follows that  $M, \Gamma \not\models B_i\varphi$ .

- Induction Case:  $(t \gg_i \varphi) \in \Gamma$  if and only if  $M, \Gamma \models t \gg_i \varphi$ .

By the definition of  $\mathcal{A}$ , we have that  $(t \gg_i \varphi) \in \Gamma$  if and only if  $\Gamma \in \mathcal{A}_i(t, \varphi)$ . The result then follows by the definition of truth.

- Induction Case:  $(t :_i \varphi) \in \Gamma$  if and only if  $M, \Gamma \models t :_i \varphi$ .

We have that  $\text{JLce}^A \vdash (t :_i \varphi) \equiv (t \gg_i \varphi) \wedge B_i\varphi$ . Thus it follows by the soundness of  $\text{JLce}^A$  over  $S$ -Fitting models (Theorem 4.7) that  $S \models (t :_i \varphi) \equiv (t \gg_i \varphi) \wedge B_i\varphi$ . Further, we have that  $d((t \gg_i \varphi) \wedge B_i\varphi) < d(t :_i \varphi)$  (Figure 17), and so the induction hypothesis applies. The result follows easily.

- Induction Case:  $(t \gg_i^{U,u} \varphi) \in \Gamma$  if and only if  $M, \Gamma \models t \gg_i^{U,u} \varphi$ .

We have that  $(t \gg_i^{U,u} \varphi) \in \Gamma$  if and only if  $\text{JLce}^A \vdash (t \gg_i^{U,u} \varphi) \equiv \top$  and  $(U, U^e), u \vdash t \gg_i \varphi$  by the consistency of  $\Gamma$  and the statement of Axiom E of  $\text{AX}^A$ . But  $(U, U^e), u \vdash t \gg_i \varphi$  is equivalent to  $M, \Gamma \models t \gg_i^{U,u} \varphi$  by the definition of truth.

- Induction Case:  $[U, u]\varphi \in \Gamma$  if and only if  $M, \Gamma \models [U, u]\varphi$ .

We perform a sub-induction on the  $d$ -ordering of  $\text{UL}^A$ -formulas. Most of these sub-cases are standard in Dynamic Epistemic Logic [19].

- Base Sub-Case:  $[U, u]q \in \Gamma$  if and only if  $M, \Gamma \models [U, u]q$  for each  $q \in \{p_k, \perp, \top\}$ .  
We have that  $\text{JLce}^A \vdash [U, u]q \equiv (U(u) \supset q)$ . Thus it follows by the soundness of  $\text{JLce}^A$  over  $S$ -Fitting models (Theorem 4.7) that  $S \models [U, u]q \equiv (U(u) \supset q)$ . Proceeding, we see that  $[U, u]q \in \Gamma$  is equivalent to  $(U(u) \supset q) \in \Gamma$  by the maximal consistency of  $\Gamma$ . But  $d(U(u) \supset q) < d([U, u]q)$  (Figure 10) and so it follows from the (outer) induction hypothesis that  $(U(u) \supset q) \in \Gamma$  is equivalent to  $M, \Gamma \models U(u) \supset q$ . But the latter is equivalent to  $M, \Gamma \models [U, u]q$ .
- Induction Sub-Case:  $[U, u](\varphi \star \psi) \in \Gamma$  if and only if  $M, \Gamma \models [U, u](\varphi \star \psi)$  for each  $\star \in \{\supset, \wedge, \vee, \equiv\}$ .  
We have that  $\text{JLce}^A \vdash [U, u](\varphi \star \psi) \equiv ([U, u]\varphi \star [U, u]\psi)$ . Thus it follows by the soundness of  $\text{JLce}^A$  over  $S$ -Fitting models (Theorem 4.7) that  $S \models [U, u](\varphi \star \psi) \equiv ([U, u]\varphi \star [U, u]\psi)$ . Proceeding, we see that  $[U, u](\varphi \star \psi) \in \Gamma$  is equivalent to

$([U, u]\varphi \star [U, u]\psi) \in \Gamma$  by the maximal consistency of  $\Gamma$ . But  $d([U, u]\varphi \star [U, u]\psi) < d([U, u](\varphi \star \psi))$  (Figure 11) and so it follows from the sub-induction hypothesis that  $([U, u]\varphi \star [U, u]\psi) \in \Gamma$  is equivalent to  $M, \Gamma \models [U, u]\varphi \star [U, u]\psi$ . But the latter is equivalent to  $M, \Gamma \models [U, u](\varphi \star \psi)$ .

- Induction Sub-Case:  $[U, u]\neg\varphi \in \Gamma$  if and only if  $M, \Gamma \models [U, u]\neg\varphi$ .  
Similar the sub-case for formulas of the form  $[U, u](\varphi \star \psi)$ , though we make use of the fact that UN is both a  $\text{JLce}^A$ -theorem and  $S$ -valid, and we also make use of the fact that  $d(U(u) \supset \neg[U, u]\varphi) < d([U, u]\neg\varphi)$  (Figure 12), which allows us to apply the sub-induction hypothesis.
- Induction Sub-Case:  $[U, u]B_i\varphi \in \Gamma$  if and only if  $M, \Gamma \models [U, u]B_i\varphi$ .  
Similar the sub-case for formulas of the form  $[U, u](\varphi \star \psi)$ , though we make use of the fact that UB is both a  $\text{JLce}^A$ -theorem and  $S$ -valid, and we also make use of the fact that  $d(U(u) \supset \bigwedge_{wU_i, v} B_i[U, v]\varphi) < d([U, u]B_i\varphi)$  (Figure 13), which allows us to apply the sub-induction hypothesis.
- Induction Sub-Case:  $[U, u](t \gg_i \varphi) \in \Gamma$  if and only if  $M, \Gamma \models [U, u](t \gg_i \varphi)$ .  
Similar the sub-case for formulas of the form  $[U, u](\varphi \star \psi)$ , though we make use of the fact that UE is a  $\text{JLce}^A$ -theorem and is  $S$ -valid. We also make use of the fact that  $d(U(u) \supset ((t \gg_i \varphi) \wedge \neg(t \gg_i^{U, u} \varphi))) < d([U, u](t \gg_i \varphi))$  (Figure 15), which allows us to apply the sub-induction hypothesis.
- Induction Sub-Case:  $[U, u](t :_i \varphi) \in \Gamma$  if and only if  $M, \Gamma \models [U, u](t :_i \varphi)$ .  
Similar the sub-case for formulas of the form  $[U, u](\varphi \star \psi)$ , though we make use of the fact that  $[U, u](t :_i \varphi) \equiv [U, u]((t \gg_i \varphi) \wedge B_i\varphi)$  is a  $\text{JLce}^A$ -theorem and is  $S$ -valid. We also make use of the fact that  $d([U, u]((t \gg_i \varphi) \wedge B_i\varphi)) < d([U, u](t :_i \varphi))$  (Figure 18), which allows us to apply the sub-induction hypothesis.
- Induction Sub-Case:  $[U, u](t \gg_i^{U', u'} \varphi) \in \Gamma$  if and only if  $M, \Gamma \models [U, u](t \gg_i^{U', u'} \varphi)$ .  
Similar the sub-case for formulas of the form  $[U, u](\varphi \star \psi)$ , though we make use of the fact that UE is a  $\text{JLce}^A$ -theorem and is  $S$ -valid. We also make use of the fact that  $d(U(u) \supset ((t \gg_i \varphi) \wedge \neg(t \gg_i^{U, u} \varphi))) < d([U, u](t \gg_i \varphi))$  (Figure 15), which allows us to apply the sub-induction hypothesis.
- Induction Sub-Case:  $[U, u][U', u']\varphi \in \Gamma$  if and only if  $M, \Gamma \models [U, u][U', u']\varphi$ .  
Similar the sub-case for formulas of the form  $[U, u](\varphi \star \psi)$ , though we make use of the fact that UU is a  $\text{JLce}^A$ -theorem and is  $S$ -valid. We also make use of the fact that  $d([U \circ U', (u, u')]\varphi) < d([U, u][U', u']\varphi)$  (Figure 16), which allows us to apply the sub-induction hypothesis.

This completes the proof of the Truth Lemma. The completeness argument is then easy: suppose  $\text{JLce}^A \not\models \chi$ . It follows that  $\{\neg\chi\}$  is consistent and so may be extended to a maximal consistent set  $\Gamma \in M$ . Since  $\neg\chi \in \Gamma$ , it follows by the Truth Lemma that  $M, \Gamma \models \neg\chi$  and thus that  $M, \Gamma \not\models \chi$  by the definition of truth. Since  $M$  is an  $S$ -Fitting model, we we have shown that  $\text{JLce}^A \not\models \chi$  implies  $S \not\models \chi$ . It follows that  $S \models \chi$  implies  $\text{JLce}^A \vdash \chi$ , completing the proof.  $\square$

So we see that if we take the set  $S$  of  $\text{AX}^A$ -theorems as the “basic” assertions that do not require detailed justification, then the  $S$ -valid  $\text{UL}^A$ -formulas are exactly the  $\text{JLce}^A$ -provable  $\text{UL}^A$ -formulas.

## 5 Formalized Example

We now wish to formalize our email example from the beginning of the paper in the theory  $\text{JLce}^A$ . In this example, the set of agents in which we are interested is

$$A := \{B, C\} ,$$

where  $B$  represents Bryan and  $C$  represents Charlie.<sup>15</sup> The first four email messages from Anne and Charlie present us with the following initial assumptions.

1.  $x_1 :_B (J \supset M)$

In words:  $x_1$  is Bryan’s evidence that “if Anne gets the job, then Anne moves to Columbus.”

2.  $x_2 :_B (x_1 :_C (J \supset M))$

In words:  $x_2$  is Bryan’s evidence that  $x_1$  is Charlie’s evidence that “if Anne gets the job, then Anne moves to Columbus.”

3.  $x_3 :_B J$

In words:  $x_3$  is Bryan’s evidence that Anne gets the job.

4.  $x_4 :_B (x_3 :_C J)$

In words:  $x_4$  is Bryan’s evidence that  $x_3$  is Charlie’s evidence that Anne gets the job.

Let the symbol  $X$  denote the conjunction whose conjuncts consist of assumptions 1, 2, 3, and 4. Further, wherever it makes sense in the remaining text of this section, use of the turnstile (“ $\vdash$ ”) without mention of a theory is to be understood as use of the turnstile with the theory  $\text{JLce}^A$  appearing to the left.

It follows from Lemma 4.5 (on the closure of evidence under the term-forming operations) along with assumptions 1 and 3 that

$$\vdash X \supset (x_1 \cdot_J x_3) :_B M ;$$

that is, “ $x_1 \cdot_J x_3$  is Bryan’s evidence that Anne is moving to Columbus.” Indeed, Bryan has evidence  $x_1$  that Anne will move to Columbus if she gets the job, and he has evidence  $x_3$  that she got the job. Combining these pieces of evidence, he obtains the evidence  $x_1 \cdot_J x_3$

---

<sup>15</sup>Anne’s evidence and beliefs do not play a significant role in the email example.

that Anne will move to Columbus. Written symbolically, Bryan applied his evidence  $x_1$  and  $x_3$  according to the  $\text{JLce}^A$ -theorem (Lemma 4.5)

$$(x_1 :_B (J \supset M)) \supset ((x_3 :_B J) \supset (x_1 \cdot_J x_3) :_B M) \quad (4)$$

to conclude that  $(x_1 \cdot_J x_3) :_B M$ . But a similar theorem also applies with respect to Charlie's reasoning. That is, we have that

$$\vdash (x_1 :_C (J \supset M)) \supset ((x_3 :_C J) \supset (x_1 \cdot_J x_3) :_C M) . \quad (5)$$

Applying the Internalization Theorem (Theorem 4.6) and reasoning in  $\text{JLce}^A$ , it follows that

$$\vdash t :_B ((x_1 :_C (J \supset M)) \supset ((x_3 :_C J) \supset (x_1 \cdot_J x_3) :_C M)) \quad (6)$$

for a variable-free term  $t \in \mathcal{T}(\text{UL}^A)$ . The term  $t$  represents Bryan's evidence that Charlie is able to reason according to the principle (5); that is,  $t$  is Bryan's explanation of the way in which Charlie can combine evidence  $x_1$  (about Anne's conditional moving plans) with evidence  $x_3$  (about the truth of the antecedent of the conditional) to produce the evidence  $x_1 \cdot_J x_3$  that Anne is moving.

But Bryan in fact has evidence  $x_2$  that Charlie has evidence  $x_1$  (about Anne's conditional moving plans), and Bryan also has evidence  $x_4$  that Charlie has evidence  $x_3$  (about the truth of the antecedent of the conditional). So Bryan can then combine his evidence  $t$  (about how Charlie can reason about evidence) with his evidence  $x_2$  and  $x_4$  (that Charlie has the requisite evidence mentioned in the explanation  $t$ ) to conclude that Charlie himself has evidence that Anne will move. That is, we have the following  $\text{JLce}^A$ -derivation.

1.  $\vdash X \supset t :_B ((x_1 :_C (J \supset M)) \supset ((x_3 :_C J) \supset (x_1 \cdot_J x_3) :_C M))$  (6)
2.  $\vdash X \supset x_2 :_B (x_1 :_C (J \supset M))$  Assumption
3.  $\vdash X \supset x_4 :_B (x_3 :_C J)$  Assumption
4.  $\vdash X \supset (t \cdot_{x_1 :_C (J \supset M)} x_2) :_B ((x_3 :_C J) \supset (x_1 \cdot_J x_3) :_C M)$  Lemma 4.5: 1, 2
5.  $\vdash X \supset ((t \cdot_{x_1 :_C (J \supset M)} x_2) \cdot_{x_2 :_C J} x_4) :_B ((x_1 \cdot_J x_3) :_C M)$  Lemma 4.5: 3, 4

For convenience, let us write  $s$  as an abbreviation for the term

$$(t \cdot_{x_1 :_C (J \supset M)} x_2) \cdot_{x_2 :_C J} x_4 .$$

So we have shown that  $s$  is Bryan's evidence that Charlie has the evidence  $x_1 \cdot_J x_3$  that Anne is moving to Columbus. Bryan's evidence  $s$  was built by taking his understanding  $t$  of how Charlie can reason about evidence and then combining this with the evidence  $x_2$  and  $x_4$  that Charlie has access to the evidence mentioned in  $t$ . Of course, this makes sense: after Bryan reads Charlie's email replies  $x_2$  and  $x_4$  to Anne's messages about her conditional move to Columbus and about how the condition for her move ended up being satisfied, then Bryan can see that Charlie can conclude that Anne will move to Columbus.

Now let us examine how Anne's last email affects this situation. In this email, Anne tells Bryan privately that her job offer was rescinded. This says that, sadly, Anne will not get the

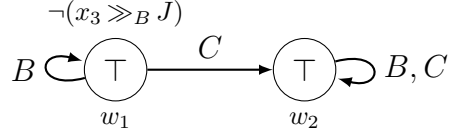


Figure 19: Diagrammatic representation of the private elimination of Bryan’s evidence  $x_3$  relevant to  $J$ .

job. The effect of this final email message is to eliminate Bryan’s evidence  $x_3$  that Anne got the job. But since the final email was sent privately to Bryan—a private message in which Anne says that she has not yet told Charlie that the job offer was rescinded—the final email does not affect Bryan’s view of Charlie’s has evidence  $x_3$  that Anne got the job. After all, as far as Bryan knows, Charlie is not yet aware that the job offer has been rescinded and so Bryan thinks that Charlie still believes Anne is moving to Columbus. Thus we say that Anne’s last message is a *private elimination* of Bryan’s evidence  $x_3$  (that Anne gets the job), since this elimination only affects Bryan’s evidence about Anne’s job (and not his evidence about Charlie’s evidence).

It will now be our task to define an update frame  $U$  that represents the private elimination of Bryan’s evidence  $x_3$ . We define the components of this update frame  $U = (W, R, f, \mathcal{A})$  in the following way.

- $W = \{w_1, w_2\}$ .
- $R_B := \{(w_1, w_1), (w_2, w_2)\}$  and  $R_C := \{(w_1, w_2), (w_2, w_2)\}$ .
- $f(w_1) := \top$  and  $f(w_2) := \top$ .
- $\mathcal{A}_B(r, \varphi) := \begin{cases} \{w_1\} & \text{if } (r, \varphi) = (x_3, J), \\ \emptyset & \text{otherwise;} \end{cases}$  and  $\mathcal{A}_C(r, \varphi) := \emptyset$ .

See Figure 19 for a diagrammatic representation of the update frame  $U$ .

We may understand the structure of  $U$  in the following way. The world  $w_1$  consists of the communication  $\top$  along with the elimination of Bryan’s evidence  $x_3$  relevant to  $J$ , while the world  $w_2$  consists of the communication  $\top$  and a null elimination. Now for a formula  $\varphi$ , we have identified the *communication of  $\varphi$*  with the operation that has the agents jointly eliminate all  $\neg\varphi$  possibilities. But since  $\top$  is the propositional constant for truth, the communication of  $\top$  does not eliminate any possibilities whatsoever (since  $\neg\top$  is always false), so we conclude that the communication of  $\top$  is simply a null communication.

The elimination of Bryan’s evidence  $x_3$  relevant to  $J$  brings about a situation in which  $x_3 \gg_B J$  is false; that is, this elimination makes it the case so that Bryan no longer has evidence  $x_3$  relevant to  $J$ . So if we execute the communication and elimination given by world  $w_1$ , then Bryan’s evidence  $x_3$  relevant to  $J$  will be eliminated. And if we execute the communication and elimination given by world  $w_2$ —a world that has a null communication and a null elimination—then nothing happens.

Now let us see how the structure of the update frame  $U$  affects Bryan and Charlie's belief of which communication and elimination is occurring. In case the communication and elimination of world  $w_1$  occurs, Bryan knows it occurs (since this is the only world he considers possible), and yet Charlie, who only thinks it possible that the communication and elimination of world  $w_2$  occurs, mistakenly thinks that nothing (the null communication and the null elimination) occurs. It is in this sense that the update  $U$  brings about the *private elimination* of Bryan's evidence  $x_3$  relevant to  $J$ .

We now examine how our axiomatics ensures that the update  $U$  indeed functions according to the intuitive description given in the previous paragraph. First, we use the  $(U, \mathcal{A})$ -calculus (Figure 2) to prove some properties about the elimination brought about at world  $w_1$ .

1.  $(U, \mathcal{A}), w_1 \vdash x_3 \gg_B J$  V
2.  $(U, \mathcal{A}), w_1 \vdash (x_1 \cdot_J x_3) \gg_B M$  EAR: 1

Applying Axiom E (Figure 5), we have that

$$\vdash x_3 \gg_B^{U, w_1} J \quad \text{and} \quad \vdash (x_1 \cdot_J x_3) \gg_B^{U, w_1} M . \quad (7)$$

So the update  $(U, w_1)$  not only eliminates Bryan's evidence  $x_3$  that Anne gets the job but also his evidence  $x_1 \cdot_J x_3$  that Anne moves to Columbus. This makes sense: Bryan's evidence  $x_1 \cdot_J x_3$  is predicated on the validity of the evidence  $x_3$  that Anne got the job, and yet it turns out that Anne did not get the job.

Now let us see how the elimination assertions in (7) cause the update  $(U, w_1)$  to eliminate Bryan's evidence  $x_3$  and his evidence  $x_1 \cdot_J x_3$ . We observe that an instance of Axiom UE (Figure 5) says that

$$[U, w_1](x_3 \gg_B J) \equiv \top \supset ((x_3 \gg_B J) \wedge \neg(x_3 \gg_B^{U, w_1} J))$$

and that another instance of Axiom UE says that

$$[U, w_1]((x_1 \cdot_J x_3) \gg_B M) \equiv \top \supset (((x_1 \cdot_J x_3) \gg_B M) \wedge \neg((x_1 \cdot_J x_3) \gg_B M)) .$$

But then (7) implies that the right-hand side of each of the above two instances of Axiom UE is equivalent to  $\perp$  in the theory  $\mathbf{JLce}^A$ . So we see that

$$\vdash \neg[U, w_1](x_3 \gg_B J) \quad \text{and} \quad \vdash \neg[U, w_1]((x_1 \cdot_J x_3) \gg_B M) .$$

Since  $U(w_1) = \top$ , it follows from Axiom UN (Figure 5) that  $\vdash \neg[U, w_1]\varphi \equiv [U, w_1]\neg\varphi$  for each  $\varphi \in \mathbf{UL}^A$ . So what we have shown is that

$$\vdash [U, w_1]\neg(x_3 \gg_B J) \quad \text{and} \quad \vdash [U, w_1]\neg((x_1 \cdot_J x_3) \gg_B M) .$$

In words: “after update  $(U, w_1)$ , Bryan does not have evidence  $x_3$  relevant (to the statement that) Anne gets the job” and “after update  $(U, w_1)$ , Bryan does not have evidence  $x_1 \cdot_J x_3$  relevant to (the statement that) Anne moves to Columbus.” Applying Axiom C (Figure 5) along with  $\mathbf{JLce}^A$ -reasoning, this implies that

$$\vdash [U, w_1]\neg(x_3 :_B J) \quad \text{and} \quad \vdash [U, w_1]\neg((x_1 \cdot_J x_3) :_B M) .$$

That is, “after update  $(U, w_1)$ , Bryan does not have evidence  $x_1$  that Anne gets the job” and “after update  $(U, w_1)$ , Bryan does not have evidence  $x_1 \cdot_J x_3$  that Anne moves to Columbus.”

So we have seen that the update  $(U, w_1)$  indeed eliminates Bryan’s evidence about Anne’s getting the job along with his evidence about her moving to Columbus. But let us now see that Bryan is aware that this update is a private elimination of his evidence; that is, we shall show that  $(U, w_1)$  does not affect Bryan’s evidence about Charlie’s reasoning. We will do this in two stages: we first show that the update  $(U, w_1)$  preserves Bryan’s belief that Charlie has evidence that Anne is moving, and then we show that this update also preserves Bryan’s evidence relevant to the assertion that Charlie has evidence that Anne is moving. We will then combine these two facts to obtain the conclusion that Bryan still has evidence that Charlie has evidence that Anne is moving to Columbus.

So let us first show that after the update  $(U, w_1)$ , Bryan still believes that Charlie has evidence  $x_1 \cdot_J x_3$  that Anne is moving; that is, we wish to show that

$$\vdash X \supset [U, w_1]B_B((x_1 \cdot_J x_3) :_C M) . \quad (8)$$

Using Axiom UB (Figure 5), the formula to the right of the turnstile (“ $\vdash$ ”) is equivalent in  $\mathbf{JLce}^A$  to  $X \supset B_B[U, w_1]((x_1 \cdot_J x_3) :_C M)$  because  $U(w_1) = \top$ . But by further reasoning in  $\mathbf{JLce}^A$ , the latter formula is equivalent to

$$X \supset ( B_B[U, w_1]((x_1 \cdot_J x_3) \gg_C M) \wedge B_B[U, w_1]B_C M ) ,$$

which is itself equivalent to

$$X \supset ( B_B((x_1 \cdot_J x_3) \gg_C M) \wedge B_B\neg((x_1 \cdot_J x_3) \gg_C^{U, w_1} M) \wedge B_B B_C M ) \quad (9)$$

because  $U(w_1) = \top$ . Now the first and third conjuncts in the conclusion of (9) are implied by  $X$  in the theory  $\mathbf{JLce}^A$  because  $\vdash X \supset s :_B ((x_1 \cdot_J x_3) :_C M)$ . As for the second conjunct in the conclusion of (9), consider the following  $(U, \mathcal{A})^c$ -derivation (Figure 3).

1.  $w_1 \vdash x_1 \gg_C (J \supset M) \quad \text{V}^c$
2.  $w_1 \vdash x_3 \gg_C J \quad \text{V}^c$
3.  $w_1 \vdash (x_1 \cdot_J x_3) \gg_C M \quad \text{A}^c: 1, 2$

Thus it follows from the Complementary Axiomatic Generation Lemma (Lemma 3.11) that  $\vdash \neg((x_1 \cdot_J x_3) \gg_C^{U, w_1} M)$  and thus that

$$\vdash B_B\neg((x_1 \cdot_J x_3) \gg_C^{U, w_1} M)$$

by Rule BN. So we have shown that each of the conjuncts in the conclusion of (9) is implied by  $X$  in the theory  $\mathbf{JLce}^A$ , from which it follows that (8) indeed holds; that is, “after update  $(U, w_1)$ , Bryan (still) believes that Charlie has evidence  $x_1 \cdot_J x_3$  that Anne is moving to Columbus.”

Now let us argue that after the update  $(U, w_1)$ , Bryan still has evidence relevant to the assertion that Charlie has evidence that Anne is moving; that is, we show that

$$\vdash X \supset [U, w_1](s \gg_B (x_1 \cdot_J x_3) :_C M) . \quad (10)$$

Applying Axiom UE (Figure 5), the formula to the right of the turnstile (“ $\vdash$ ”) is equivalent in  $\text{JLce}^A$  to

$$X \supset \left( (s \gg_B (x_1 \cdot_J x_3) :_C M) \wedge \neg (s \gg_B^{U, w_1} (x_1 \cdot_J x_3) :_C M) \right) . \quad (11)$$

The left conjunct in the conclusion of (11) is implied by  $X$  in the theory  $\text{JLce}^A$  because  $X \vdash s :_B (x_1 \cdot_J x_3) :_C M$ . To see that the right conjunct of (11) is implied by  $X$  in the theory  $\text{JLce}^A$ , consider the following  $(U, \mathcal{A})^c$ -derivation (Figure 3).

1.  $w_1 \vdash t \gg_B ((x_1 :_C (J \supset M)) \supset ((x_3 :_C J) \supset (x_1 \cdot_J x_3) :_C M))$   $t$  is variable-free
2.  $w_1 \vdash x_2 \gg_B (x_1 :_C (J \supset M))$   $\text{V}^c$
3.  $w_1 \vdash x_4 \gg_B (x_3 :_C J)$   $\text{V}^c$
4.  $w_1 \vdash (t \cdot_{x_1 :_C (J \supset M)} x_2) \gg_B ((x_3 :_C J) \supset (x_1 \cdot_J x_3) :_C M)$   $\text{A}^c: 1, 2$
5.  $w_1 \vdash ((t \cdot_{x_1 :_C (J \supset M)} x_2) \cdot_{x_2 :_C J} x_4) \gg_B ((x_1 \cdot_J x_3) :_C M)$   $\text{A}^c: 3, 4$

And so it follows from the Complementary Axiomatic Generation Lemma (Lemma 3.11) that  $\vdash \neg (s \gg_B^{U, w_1} (x_1 \cdot_J x_3) :_C M)$  and thus that the right conjunct of (11) is implied by  $X$  in the theory  $\text{JLce}^A$ . Thus (11) holds, from which it follows that (10) holds. That is, we have shown that “after update  $(U, w_1)$ , Bryan (still) has evidence  $s$  relevant to (the assertion) that Charlie has evidence  $x_1 \cdot_J x_3$  that Anne is moving to Columbus.” But we already showed that (8): after update  $(U, w_1)$ , Bryan still believes that Charlie has evidence  $x_1 \cdot_J x_3$  that Anne is moving to Columbus. Taken together, we have shown that

$$\vdash X \supset \left( [U, w_1] B_B((x_1 \cdot_J x_3) :_C M) \wedge [U, w_1] (s \gg_B (x_1 \cdot_J x_3) :_C M) \right) ,$$

which is equivalent in  $\text{JLce}^A$  to the statement

$$\vdash X \supset [U, w_1] (s :_B (x_1 \cdot_J x_3) :_C M) . \quad (12)$$

That is, “after update  $(U, w_1)$ , Bryan (still) has evidence  $s$  that Charlie has evidence  $x_1 \cdot_J x_3$  that Anne is moving to Columbus.”

We have shown that the update  $(U, w_1)$  alters Bryan’s evidence about Anne’s move:

$$\vdash X \supset ((x_1 \cdot_J x_3) :_B M) \quad \text{and} \quad \vdash X \supset [U, w_1] \neg ((x_1 \cdot_J x_3) :_B M) .$$

But we also showed that the update  $(U, w_1)$  does not alter Bryan’s evidence about Charlie’s evidence about Anne’s move:

$$\vdash X \supset s :_B ((x_1 \cdot_J x_3) :_C M) \quad \text{and} \quad \vdash X \supset [U, w_1] s :_B ((x_1 \cdot_J x_3) :_C M) .$$

This makes sense: the update  $(U, w_1)$  was brought about by the email Anne sent privately to Bryan. So while Bryan is to eliminate his own evidence, he can see that Charlie’s evidence ought to be left intact.

## 6 Conclusion

We have presented  $JLce^A$ , a multi-agent theory of Justification Logic with communication and evidence elimination. Using an email example, we showed how it is that this logic can be used to reason about everyday eliminations of evidence. In future work, we plan to extend our notion of evidence elimination to allow for *evidence introduction*, an operation that would make true certain relevance assertions  $t \gg_i \varphi$ . Such a joint system of communication with evidence introduction and elimination would provide a fuller account of the dynamics of evidence and communication in systems of interacting rational agents.

## References

- [1] Sergei Artemov. Justification logic. Technical Report TR-2007019, CUNY Graduate Center Ph.D. Program in Computer Science, 2007.
- [2] Sergei N. Artemov. Explicit provability and constructive semantics. *The Bulletin of Symbolic Logic*, 7(1):1–36, 2001.
- [3] Alexandru Baltag and Lawrence S. Moss. Logics for epistemic programs. *Synthese*, 139(2):165–224, 2004.
- [4] Alexandru Baltag, Lawrence S. Moss, and Sławomir Solecki. The logic of common knowledge, public announcements, and private suspicions. In Itzhak Gilboa, editor, *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK VII)*, pages 43–56, Evanston, IL, USA, 1998.
- [5] Alexandru Baltag, Lawrence S. Moss, and Sławomir Solecki. Logics for epistemic actions: completeness, decidability, expressivity. Manuscript, 2005.
- [6] Alexandru Baltag, Hans P. van Ditmarsch, and Lawrence S. Moss. Epistemic logic and information update. In Pieter Adriaans and Johan van Benthem, editors, *Handbook on the Philosophy of Information*, pages 369–463. Elsevier, 2008.
- [7] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*. Cambridge University Press, 2001.
- [8] Melvin Fitting. Reasoning with justifications. Forthcoming in *Studia Logica*.
- [9] Melvin Fitting. The Logic of Proofs, semantically. *Annals of Pure and Applied Logic*, 132(1):1–25, 2005.
- [10] Jelle Gerbrandy. *Bisimulations on Planet Kripke*. PhD thesis, University of Amsterdam, 1999.
- [11] Jelle Gerbrandy and Willem Groeneveld. Reasoning about information change. *Journal of Logic, Language, and Information*, 6:147–169, 1997.
- [12] Kurt Gödel. Vortrag bei Zilsel/Lecture at Zilsel’s (\*1938a). In Solomon Feferman, John W. Dawson, Jr., Warren Goldfarb, Charles Parsons, and Robert M. Solovay, editors, *Unpublished*

- essays and lectures*, volume III of *Kurt Gödel Collected Works*, pages 86–113. Oxford University Press, 1995.
- [13] Saul A. Kripke. A completeness theorem in modal logic. *The Journal of Symbolic Logic*, 24(1):1–14, 1959.
  - [14] Roman Kuznets. *Complexity Issues in Justification Logic*. PhD thesis, CUNY Ph.D. Program in Computer Science, 2008.
  - [15] Alexey Mkrtychev. Models for the Logic of Proofs. In Sergei Adian and Anil Nerode, editors, *Logical Foundations of Computer Science*, volume 1234 of *Lecture Notes in Computer Science*, pages 266–275. Springer, 1997.
  - [16] Jan A. Plaza. Logics of public communications. In Zbigniew W. Ras, editor, *Proceedings of the Fourth International Symposium on Methodologies for Intelligent Systems (ISMIS 1989)*. North-Holland, 1989.
  - [17] Bryan Renne. *Dynamic Epistemic Logic with Justification*. PhD thesis, The City University of New York, 2008.
  - [18] Bryan Renne. A survey of Dynamic Epistemic Logic. Manuscript, July 2008.
  - [19] Hans van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. *Dynamic Epistemic Logic*. Springer, 2007.
  - [20] Tatiana Yavorskaya. Multi-agent explicit knowledge. In D. Grigoriev, J. Harrison, and E.A. Hirsch, editors, *Computer Science Russia 2006*, volume 3967 of *Lecture Notes in Computer Science*, pages 369–380. Springer-Verlag, 2006.