Reasons and Options for Updating an Opponent Model in Persuasion Dialogues

Elizabeth Black[1] and Anthony Hunter[2]

[1] Department of Informatics, King's College London
elizabeth.black@kcl.ac.uk
[2] Department of Computer Science, University College London
anthony.hunter@cs.ucl.ac.uk

Abstract. Dialogical argumentation allows agents to interact by constructing and evaluating arguments through a dialogue. Numerous proposals have been made for protocols for dialogical argumentation, and recently there is interest in developing better strategies for agents to improve their own outcomes from the interaction by using an opponent model to guide their strategic choices. However, there is a lack of clear formal reasons for why or how such a model might be useful, or how it can be maintained. In this paper, we consider a simple type of persuasion dialogue, investigate options for using and updating an opponent model, and identify conditions under which such use of a model is beneficial.

1 Introduction

Argument dialogues are an established agreement technology; they provide a principled way of structuring rational interactions between participants (machine or human) who argue about the validity of certain claims in order to resolve their conflicting information, competing goals, incompatible intentions or opposing views of the world [1]. Such dialogues are typically defined by the *moves* that can be made and rules to determine which moves are permissible at any point in the dialogue. Much existing work in the field focusses on defining argument dialogues that allow achievement of a particular goal; for example, to persuade the other participant to accept some belief [2] or to agree on some action to achieve a shared goal [3]. However, successful achievement of a participant's dialogue goal normally depends on the *strategy* it employs to determine which of the permissible moves to make during the dialogue; the development of effective argument dialogue strategies is thus an important area of active research [4].

Recent work on argument dialogue strategy assumes the strategiser has some uncertain model of what its interlocutor knows, derived somehow from the strategiser's past interactions, which it uses to guide its choice of moves [5, 6]. However, there is a lack of formal investigation into how such a model can be maintained and under what circumstances it can be useful. Rienstra *et al.* propose a mechanism for updating an opponent model with the addition of arguments proposed or received by the opponent [6], Black *et al.*'s approach involves also removing

from the opponent model anything that is inconsistent with the observed opponent behaviour [5], while Hadjinikolis *et al.* consider how an agent can develop a model of the likelihood that an opponent will know a particular argument if it asserts some other argument [7, 8]; however, none of these works formally investigate the impact of the model update mechanism on the dialogue outcome.

We are interested in understanding the different options for updating an opponent model and investigating the circumstances under which such a model can be useful. We consider a simple type of persuasion dialogue with two participants, the *persuader* and the *responder*. The persuader (who has an uncertain *model* of the responder) aims to convince the responder to accept the topic of the dialogue by asserting beliefs, while the responder replies honestly to indicate whether it is currently convinced of the acceptability of the topic.

We investigate the performance of two model update mechanisms, based on those used by Rienstra *et al.* [6] and Black *et al.* [5]. In the first (which we refer to as *basic*), beliefs asserted by the persuader are added to its model of the responder, while the second mechanism (which we refer to as *smart*) also removes from the persuader's model of the responder anything that is inconsistent with the moves the responder makes (under the assumption that the responder is honest). We do not focus here on how the persuader determines which beliefs to assert and which order to assert them in; we assume the persuader has a mechanism for determining some total ordering over its beliefs (which we refer to as its *strategy* and corresponds to the order in which it will assert its beliefs) and instead focus on whether it uses its model of the responder to decide when to give up trying to persuade the responder. We consider the case where the persuader will not give up until it has exhausted all its beliefs (called an *exhaustive* persuader) and the case where the persuader will give up as soon as, given its model of the responder, it believes it is not possible to successfully persuade the responder no matter which beliefs it asserts (called an *economical* persuader).

We formally investigate the performance of our model update mechanisms by identifying the situations in which it is possible that, *when following the same strategy*, a persuader of one type will successfully persuade the responder, while a persuader of another type will unsuccessfully terminate the dialogue before it has achieved its goal. This paper thus contributes to our understanding of when it can be advantageous to use a particular model update mechanism.

2 Simple persuasion dialogues

In our simple persuasion dialogues (adapted from the simple persuasion dialogues of Black *et al.* [5]) the persuader aims to convince the responder to accept the topic of the dialogue by asserting beliefs. We make no prescription as to which semantics the participants use to reason about the acceptability of beliefs. We assume only a finite logical language $\mathcal{L}$ and some function for determining the set of *acceptable* claims given some knowledge base of $\mathcal{L}$.

Definition 1. *For a knowledge base* $\Phi \subseteq \mathcal{L}$*, the function* $\mathsf{Acceptable} : \wp(\mathcal{L}) \rightarrow \wp(\mathcal{L})$ *returns the set of* **acceptable claims** *of* Φ *such that:* $\mathsf{Acceptable}(\Phi) = \{\alpha \in \mathcal{L} \mid \alpha$ *is acceptable given* Φ *under the chosen acceptability semantics*$\}$.

There are many formalisms and associated acceptability semantics that may be used to instantiate Definition 1. For example: we could consider a standard propositional language $\mathcal{L}$ and specify that a claim α is acceptable given $\Phi \subseteq \mathcal{L}$ if and only if α can be classically entailed from Φ; it could be that $\mathcal{L}$ consists of atoms that represent abstract arguments and a claim α is acceptable given $\Phi \subseteq \mathcal{L}$ if and only if α is in the grounded extension of the argument framework constructed from Φ according to a particular defeat relation defined over $\mathcal{L}$ [9]; or $\Phi \subseteq \mathcal{L}$ may represent an ASPIC+ knowledge base and we may specify that α is acceptable given $\Phi \subseteq \mathcal{L}$ if and only if α is the claim of an admissible argument from the Dung-style argument framework constructed from Φ and a particular ASPIC+ argumentation system [10].

A simple persuasion dialogue has a *topic* (a wff of $\mathcal{L}$) and involves two *participants*, the *persuader* and the *responder*. Each participant has a *position* (a subset of $\mathcal{L}$) and the persuader has an uncertain *model* of the responder, which is a set consisting of those subsets of $\mathcal{L}$ that the persuader believes may be the responder's position. (Note that, unlike in [5], here we do not consider probabilities associated with the persuader's model.) We define a *dialogue scenario* by the participants, the participants' positions, the persuader's model of the responder and the topic. A dialogue scenario is *accurate* if the responder's position is a member of the persuader's model of the responder.

Definition 2. *A* **dialogue scenario** *is a tuple* $(Ag, \mathsf{P}_0, \Upsilon_0, \tau)$ *where:*

- $Ag = \{ag_P, ag_R\}$ *is the set of* **participants**, ag_P *is the* **persuader** *and* ag_R *is the* **responder**;
- $\mathsf{P}_0 : Ag \rightarrow \wp(\mathcal{L})$ *is a function that returns each participant's* **position**;
- $\Upsilon_0 \subseteq \wp(\mathcal{L})$ *is the persuader's* **model** *of the responder;*
- $\tau \in \mathcal{L}$ *is the* **topic** *of the dialogue.*

$(Ag, \mathsf{P}_0, \Upsilon_0, \tau)$ *is* **accurate** *iff* $\mathsf{P}_0(ag_R) \in \Upsilon_0$*. The set of all dialogue scenarios is denoted* $\mathcal{S}$*. The set of all accurate dialogue scenarios is denoted* $\mathcal{S}_{acc}$.

Example 1. Let $ds = (Ag, \mathsf{P}_0, \Upsilon_0, f)$ be a dialogue scenario. If $\mathsf{P}_0(ag_P) = \{a, b, c, d, e\}$, $\mathsf{P}_0(ag_R) = \{b\}$ and $\Upsilon_0 = \{\{a, b\}, \{a, c\}\}$, then ds is not accurate. If $\mathsf{P}_0(ag_P) = \{a, b, c, d, e\}$, $\mathsf{P}_0(ag_R) = \{b\}$ and $\Upsilon_0 = \{\{a, b\}, \{a, c\}, \{b\}\}$, then ds is accurate.

The set of *moves* used in simple persuasion dialogues is $\mathcal{M} = \{(\texttt{open}, \tau), (\texttt{assert}, \phi), (\texttt{pass}), (\texttt{close})\}$ where $\tau \in \mathcal{L}$ is the topic of the dialogue, $\phi \in \mathcal{L}$, and the function $\mathsf{Sender} : \mathcal{M} \rightarrow Ag$ returns the *sender* of a move. A *simple persuasion dialogue* is a sequence of *dialogue states*, where each state consists of a *move* being made, a function that returns each participant's *position* after that move has been made, and a set that represents the persuader's *model* of the responder after that move has been made. The participants take it in turn to make moves. The persuader always starts by *opening* with the dialogue topic, following which it can *assert* a wff of $\mathcal{L}$ or make a *close* move. The responder

can make a *close* or a *pass* move. The persuader cannot repeat assertions. The last move of the dialogue, and only the last move, is always a close move (and so either participant can chose to terminate the dialogue by making a close move); if this move is made by the responder, the dialogue is successful, otherwise it is unsuccessful. The *length* of a dialogue is equal to the number of dialogue states.

Definition 3. *A* **simple persuasion dialogue** *of dialogue scenario* $(Ag, \mathsf{P}_0, \Upsilon_0, \tau)$ *is a sequence of* **dialogue states** $[(m_1, \mathsf{P}_1, \Upsilon_1), \ldots, (m_t, \mathsf{P}_t, \Upsilon_t)]$ *where* $\forall\, s$ *such that* $1 \leq s \leq t$*:*

1. $\mathsf{P}_s : Ag \rightarrow \wp(\mathcal{L})$*;*
2. $\Upsilon_s \subseteq \wp(\mathcal{L})$*;*
3. $m_s \in \{(\texttt{open}, \tau), (\texttt{assert}, \phi), (\texttt{pass}), (\texttt{close})\}$ *where* $\phi \in \mathcal{L}$*;*
4. $m_s = (\texttt{open}, \tau)$ *iff* $s = 1$*;*
5. *if* s *is odd, then* $\mathsf{Sender}(m_s) = ag_P$ *and* $m_s \in \{(\texttt{open}, \tau), (\texttt{assert}, \phi), (\texttt{close})\}$*;*
6. *if* s *is even, then* $\mathsf{Sender}(m_s) = ag_R$ *and* $m_s \in \{(\texttt{pass}), (\texttt{close})\}$*;*
7. $m_s = (\texttt{close})$ *iff* $s = t$*;*
8. *if* $m_s = (\texttt{assert}, \phi)$*, then* $\forall\, i$ *such that* $1 \leq i < s$*,* $m_i \neq (\texttt{assert}, \phi)$*.*

Let $d = [(m_1, \mathsf{P}_1, \Upsilon_1), \ldots, (m_t, \mathsf{P}_t, \Upsilon_t)]$ *be a simple persuasion dialogue: the* **length** *of* d*, denoted* $\mathsf{Length}(d)$*, is* t*;* d *is* **successful** *iff* $\mathsf{Sender}(m_t) = ag_R$*.*

The previous definition defines the protocol that participants of a simple persuasion dialogue must abide by. We also make some assumptions about the behaviour of the dialogue participants, namely: the persuader's position does not change during the dialogue (so it is not engaged with any processes external to the dialogue); the persuader only asserts things that are part of its own position (so it is honest); the responder's position is updated only to include things asserted by the persuader (so the responder trusts the persuader and is not engaged with any processes external to the dialogue); and the responder's moves accurately reflect whether it has been successfully convinced of the topic (so it is honest). If these assumptions hold we say the dialogue is *regular*.

Definition 4. *A simple persuasion dialogue* $[(m_1, \mathsf{P}_1, \Upsilon_1), \ldots, (m_t, \mathsf{P}_t, \Upsilon_t)]$ *of dialogue scenario* $(Ag, \mathsf{P}_0, \Upsilon_0, \tau)$ *is a* **regular simple persuasion dialogue** *iff* $\forall\, s$ *such that* $1 \leq s \leq t$*:*

1. $\mathsf{P}_s(ag_P) = \mathsf{P}_0(ag_P)$*;*
2. *if* $m_s = (\texttt{assert}, \phi)$*, then* $\phi \in \mathsf{P}_0(ag_P)$*, and* $\mathsf{P}_s(ag_R) = \mathsf{P}_{s-1}(ag_R) \cup \{\phi\}$*;*
3. $\tau \in \mathsf{Acceptable}(\mathsf{P}_{s-1}(ag_R))$ *iff* $s = t$ *and* $\mathsf{Sender}(m_s) = ag_R$*.*

The set of all regular simple persuasion dialogues of a dialogue scenario ds *is denoted* $\mathsf{Dialogues}_{\mathsf{reg}}(ds)$*.*

The responder of a regular simple persuasion dialogue has no choice over the moves it can make; since we assume it to be honest, it terminates the dialogue with a close move if and only if it finds the topic to be acceptable, otherwise it makes a pass move. The persuader, however, can chose which beliefs to assert and whether to (unsuccessfully) terminate the dialogue; we consider conditions under which different types of persuader will terminate the dialogue in Section 3.

Example 2. Let $ds = (Ag, \mathsf{P}_0, \Upsilon_0, f)$ be a dialogue scenario with topic f such that $\mathsf{P}_0(ag_P) = \{a, b, c, d, e\}$, $\mathsf{P}_0(ag_R) = \{b\}$, $\Upsilon_0 = \{\{a, b\}, \{a, c\}\}$. The only sets of beliefs that determine the topic f to be acceptable (i.e., the only sets Φ such that $f \in \mathsf{Acceptable}(\Phi)$) are $\{b, c, d\}$ and $\{a, b, c\}$.
The following are each simple persuasion dialogues of ds. (In dialogues $d1$ and $d3$ the persuader's model of the responder is not updated, while in dialogues $d2$ and $d4$ the persuader's model is updated to include beliefs asserted by the persuader. We formally define model update methods in Section 3.)
$d1 = [((\mathtt{open}, f), \mathsf{P}_1, \Upsilon_1), ((\mathtt{pass}), \mathsf{P}_2, \Upsilon_2), ((\mathtt{close}), \mathsf{P}_3, \Upsilon_3)]$ where
- $\forall\, i$ such that $1 \leq i \leq 3$, $\mathsf{P}_i(ag_P) = \{a, b, c, d, e\}$,
- $\forall\, i$ such that $1 \leq i \leq 3$, $\mathsf{P}_i(ag_R) = \{b\}$,
- $\forall\, i$ such that $1 \leq i \leq 3$, $\Upsilon_i = \{\{a, b\}, \{a, c\}\}$.

$d1$ is a regular unsuccessful dialogue.
$d2 = [((\mathtt{open}, f), \mathsf{P}_1, \Upsilon_1), ((\mathtt{pass}), \mathsf{P}_2, \Upsilon_2), ((\mathtt{assert}, b), \mathsf{P}_3, \Upsilon_3), ((\mathtt{pass}), \mathsf{P}_4, \Upsilon_4),$
$((\mathtt{assert}, c), \mathsf{P}_5, \Upsilon_5), ((\mathtt{pass}), \mathsf{P}_6, \Upsilon_6), ((\mathtt{close}), \mathsf{P}_7, \Upsilon_7)]$ where
- $\forall\, i$ such that $1 \leq i \leq 7$, $\mathsf{P}_i(ag_P) = \{a, b, c, d, e\}$,
- $\forall\, i$ such that $1 \leq i \leq 4$, $\mathsf{P}_i(ag_R) = \{b\}$,
- $\forall\, i$ such that $5 \leq i \leq 7$, $\mathsf{P}_i(ag_R) = \{b, c\}$,
- $\Upsilon_1 = \Upsilon_2 = \{\{a, b\}, \{a, c\}\}$,
- $\Upsilon_3 = \Upsilon_4 = \{\{a, b\}, \{a, b, c\}\}$,
- $\forall\, i$ such that $5 \leq i \leq 7$, $\Upsilon_i = \{\{a, b, c\}\}$.

$d2$ is a regular unsuccessful dialogue.
$d3 = [((\mathtt{open}, f), \mathsf{P}_1, \Upsilon_1), ((\mathtt{pass}), \mathsf{P}_2, \Upsilon_2), ((\mathtt{assert}, b), \mathsf{P}_3, \Upsilon_3), ((\mathtt{pass}), \mathsf{P}_4, \Upsilon_4),$
$((\mathtt{assert}, c), \mathsf{P}_5, \Upsilon_5), ((\mathtt{pass}), \mathsf{P}_6, \Upsilon_6), ((\mathtt{assert}, a), \mathsf{P}_7, \Upsilon_7), ((\mathtt{close}), \mathsf{P}_8, \Upsilon_8)]$ where
- $\forall\, i$ such that $1 \leq i \leq 7$, $\mathsf{P}_i(ag_P) = \{a, b, c, d, e\}$,
- $\forall\, i$ such that $1 \leq i \leq 4$, $\mathsf{P}_i(ag_R) = \{b\}$,
- $\mathsf{P}_5(ag_R) = \mathsf{P}_6(ag_R) = \{b, c\}$,
- $\mathsf{P}_7(ag_R) = \mathsf{P}_8(ag_R) = \{a, b, c\}$,
- $\forall\, i$ such that $1 \leq i \leq 8$, $\Upsilon_i = \Upsilon_{i-1}$.

$d3$ is a regular successful dialogue.
$d4 = [((\mathtt{open}, f), \mathsf{P}_1, \Upsilon_1), ((\mathtt{pass}), \mathsf{P}_2, \Upsilon_2), ((\mathtt{assert}, b), \mathsf{P}_3, \Upsilon_3), ((\mathtt{close}), \mathsf{P}_4, \Upsilon_4)]$ where
- $\forall\, i$ such that $1 \leq i \leq 4$, $\mathsf{P}_i(ag_P) = \{a, b, c, d, e\}$,
- $\forall\, i$ such that $1 \leq i \leq 4$, $\mathsf{P}_i(ag_R) = \{b\}$,
- $\Upsilon_1 = \Upsilon_2 = \{\{a, b\}, \{a, c\}\}$,
- $\Upsilon_3 = \Upsilon_4 = \{\{a, b\}, \{a, b, c\}\}$.

$d4$ is not a regular dialogue, since at the point the responder successfully terminates the dialogue it does not find the topic to be acceptable.

We have now defined some assumptions about the behaviour of participants in a regular simple persuasion dialogue, however, we are yet to consider how the persuader updates or uses its model of the responder. In the following section we define different types of persuader according to the mechanism it uses to update its model of the responder and according to whether it will choose to unsuccessfully terminate the dialogue once, according to its (possibly incorrect) model, it believes it is impossible to convince the responder.

3 Updating and using an opponent model

We first consider how a persuader may use its model of the responder to determine when to give up trying to persuade the responder. An *economical* persuader only makes a close move (and will always make a close move) when, according to its model of the responder, it believes there is no sequence of assertions it can make that will lead to a successful dialogue; that is, for every set Ψ that it believes could possibly be the responder's position, there is no subset of its own position that it can assert (i.e., contains no beliefs already asserted) and that, when combined with Ψ, would determine the topic to be acceptable.

Definition 5. *Let* $d = [(m_1, \mathsf{P}_1, \Upsilon_1), \ldots, (m_t, \mathsf{P}_t, \Upsilon_t)] \in \mathsf{Dialogues}_{\mathsf{reg}}((Ag, \mathsf{P}_0, \Upsilon_0, \tau))$. *We say* d *has an* **economical persuader** *iff:*

1. *if* $\mathsf{Sender}(m_t) = ag_P$, *then* $\forall \Psi \in \Upsilon_{t-1}$, $\nexists \Phi \subseteq \mathsf{P}_{t-1}(ag_P)$ *such that:*
   (a) $\Phi \neq \emptyset$,
   (b) $\Phi \cap \{\phi \mid \exists\, s$ *such that* $1 \leq s < t$ *and* $m_s = (\mathsf{assert}, \phi)\} = \emptyset$,
   (c) $\tau \in \mathsf{Acceptable}(\Psi \cup \Phi)$;
2. $\forall s$ *such that* $1 \leq s < t$ *and* s *is odd,* $\exists \Psi \in \Upsilon_{s-1}$ *such that* $\exists \Phi \subseteq \mathsf{P}_{s-1}(ag_P)$ *such that:*
   (a) $\Phi \neq \emptyset$,
   (b) $\Phi \cap \{\phi \mid \exists\, i$ *such that* $1 \leq i < s$ *and* $m_i = (\mathsf{assert}, \phi)\} = \emptyset$,
   (c) $\tau \in \mathsf{Acceptable}(\Psi \cup \Phi)$;

Example 3. Of the three regular dialogues given in Example 2 ($d1$, $d2$ and $d3$) only $d2$ and $d3$ have an economical persuader.

We now define three types of persuader whose performance we will later explore. An *exhaustive persuader* does not maintain its model of the responder and will only terminate the dialogue once it has exhausted all beliefs it can assert (i.e., does not consider its model when deciding whether to terminate the dialogue). A *basic persuader* is an economical persuader that only updates its opponent model to reflect that the responder is aware of things the persuader has asserted. A persuader is *smart* if it is an economical persuader that updates its opponent model to reflect that the responder is aware of things the persuader has asserted and also removes from its model sets that are inconsistent with the responder's behaviour (assuming a regular dialogue). Thus, if the responder makes a pass move, a smart persuader removes from its model any sets that determine the topic to be acceptable (since it assumes the responder is honest).

Definition 6. *Let* $d = [(m_1, \mathsf{P}_1, \Upsilon_1), \ldots, (m_t, \mathsf{P}_t, \Upsilon_t)] \in \mathsf{Dialogues}_{\mathsf{reg}}((Ag, \mathsf{P}_0, \Upsilon_0, \tau))$.

d *has an* **exhaustive persuader** *iff: if* $\mathsf{Sender}(m_t) = ag_P$, *then* $\forall\, \phi \in \mathsf{P}_{t-1}(ag_P)$, $\exists\, s$ *such that* $1 \leq s < t$ *and* $m_s = (\mathsf{assert}, \phi)$ *and* $\forall\, s$ *such that* $1 \leq s \leq t$*:* $\Upsilon_s = \Upsilon_{s-1}$.

d *has a* **basic persuader** *iff* d *has an economical persuader and* $\forall\, s$ *such that* $1 \leq s \leq t$*: if* $m_s = (\mathtt{assert}, \phi)$, *then* $\Upsilon_s = \{\Psi \cup \{\phi\} \mid \Psi \in \Upsilon_{s-1}\}$*; otherwise* $\Upsilon_s = \Upsilon_{s-1}$.

d has a **smart persuader** *iff d has an economical persuader and* $\forall\, s$ *such that* $1 \leq s \leq t$*: if* $m_s = (\texttt{assert}, \phi)$*, then* $\Upsilon_s = \{\Psi \cup \{\phi\} \mid \Psi \in \Upsilon_{s-1}\}$*; if* $m_s = (\texttt{pass})$*, then* $\Upsilon_s = \{\Psi \in \Upsilon_{s-1} \mid \tau \notin \mathsf{Acceptable}(\Psi)\}$*; otherwise* $\Upsilon_s = \Upsilon_{s-1}$*.*

Example 4. Let $ds = (Ag, \mathsf{P}_0, \Upsilon_0, f)$ be the dialogue scenario given in Example 2 where $\mathsf{P}_0(ag_P) = \{a, b, c, d, e\}$, $\mathsf{P}_0(ag_R) = \{b\}$, $\Upsilon_0 = \{\{a, b\}, \{a, c\}\}$ and the only sets of beliefs that determine the topic of the dialogue to be acceptable (i.e., the only sets Φ such that $f \in \mathsf{Acceptable}(\Phi)$) are $\{b, c, d\}$ and $\{a, b, c\}$.
The persuader of the dialogue $d1$ given in Example 2 is neither exhaustive, basic nor smart. The persuader of the dialogue $d2$ given in Example 2 is basic. The persuader of the dialogue $d3$ given in Example 2 is exhaustive.

$d5 = [((\texttt{open}, f), \mathsf{P}_1, \Upsilon_1), ((\texttt{pass}), \mathsf{P}_2, \Upsilon_2), ((\texttt{assert}, b), \mathsf{P}_3, \Upsilon_3), ((\texttt{pass}), \mathsf{P}_4, \Upsilon_4), ((\texttt{assert}, c), \mathsf{P}_5, \Upsilon_5), ((\texttt{pass}), \mathsf{P}_6, \Upsilon_6), ((\texttt{close}), \mathsf{P}_7, \Upsilon_7)]$ where
- $\forall\, i$ such that $1 \leq i \leq 7$, $\mathsf{P}_i(ag_P) = \{a, b, c, d, e\}$,
- $\forall\, i$ such that $1 \leq i \leq 4$, $\mathsf{P}_i(ag_R) = \{b\}$,
- $\forall\, i$ such that $5 \leq i \leq 7$, $\mathsf{P}_i(ag_R) = \{b, c\}$,
- $\Upsilon_1 = \Upsilon_2 = \{\{a, b\}, \{a, c\}\}$,
- $\Upsilon_3 = \{\{a, b\}, \{a, b, c\}\}$,
- $\Upsilon_4 = \{\{a, b\}\}$,
- $\Upsilon_5 = \{\{a, b, c\}\}$,
- $\Upsilon_6 = \Upsilon_7 = \emptyset$.

$d5$ is a regular unsuccessful dialogue with a smart persuader.

$d6 = [((\texttt{open}, f), \mathsf{P}_1, \Upsilon_1), ((\texttt{pass}), \mathsf{P}_2, \Upsilon_2), ((\texttt{assert}, a), \mathsf{P}_3, \Upsilon_3), ((\texttt{pass}), \mathsf{P}_4, \Upsilon_4), ((\texttt{assert}, d), \mathsf{P}_5, \Upsilon_5), ((\texttt{pass}), \mathsf{P}_6, \Upsilon_6), ((\texttt{assert}, c), \mathsf{P}_7, \Upsilon_7), ((\texttt{pass}), \mathsf{P}_8, \Upsilon_8), ((\texttt{assert}, e), \mathsf{P}_9, \Upsilon_9), ((\texttt{pass}), \mathsf{P}_{10}, \Upsilon_{10}), ((\texttt{assert}, b), \mathsf{P}_{11}, \Upsilon_{11}), ((\texttt{pass}), \mathsf{P}_{12}, \Upsilon_{12}), ((\texttt{close}), \mathsf{P}_{13}, \Upsilon_{13})]$ where
- $\forall\, i$ such that $1 \leq i \leq 7$, $\mathsf{P}_i(ag_P) = \{a, b, c, d, e\}$,
- $\mathsf{P}_1(ag_R) = \mathsf{P}_2(ag_R) = \{b\}$,
- $\mathsf{P}_3(ag_R) = \mathsf{P}_4(ag_R) = \{a, b\}$,
- $\mathsf{P}_5(ag_R) = \mathsf{P}_6(ag_R) = \{a, b, d\}$,
- $\mathsf{P}_7(ag_R) = \mathsf{P}_8(ag_R) = \{a, b, c, d\}$,
- $\forall\, i$ such that $9 \leq i \leq 13$, $\mathsf{P}_i(ag_R) = \{a, b, c, d, e\}$,
- $\forall\, i$ such that $1 \leq i \leq 13$, $\Upsilon_i = \Upsilon i - 1$.

$d6$ is a regular unsuccessful dialogue with an exhaustive persuader.

It follows from our definitions that if a regular dialogue of an *accurate* scenario has a basic or smart persuader, the persuader's model will remain accurate throughout the dialogue (i.e., the responder's actual beliefs will always be a member of the persuader's model).

Lemma 1. *If* $ds \in \mathcal{S}_{acc}$ *and* $[(m_1, \mathsf{P}_1, \Upsilon_1), \ldots, (m_t, \mathsf{P}_t, \Upsilon_t)] \in \mathsf{Dialogues}_{\mathsf{reg}}(ds)$ *has a basic or smart persuader, then for all* i *such that* $1 \leq i \leq t$, $\mathsf{P}_i(ag_R) \in \Upsilon_i$*.*

In the following section, we define how a dialogue is generated from a particular dialogue scenario by a particular type of persuader.

4 Generating dialogues

We are interested in exploring the usefulness of our model update mechanisms when the persuader uses its model of the responder to decide when to unsuccessfully terminate the dialogue. We are not concerned here with the strategical choices the persuader makes to determine which beliefs to assert and which order to assert them in, but rather assume that the persuader has some mechanism for determining this. We define a *strategy* for a dialogue scenario as a sequence of beliefs that is some permutation of the persuader's position, corresponding to the order in which it will assert beliefs. Different dialogues may be generated from the same dialogue scenario by persuaders of different types following the same strategy, since an economical persuader will choose to terminate the dialogue once it thinks it is in a hopeless position according to its model of the responder, and a basic and a smart persuader's models may diverge.

Definition 7. *A* **strategy** *of a dialogue scenario* $(Ag, \mathsf{P}_0, \Upsilon_0, \tau) \in \mathcal{S}$ *is a sequence* $[\alpha_1, \ldots, \alpha_n]$ *such that* $\{\alpha_1, \ldots, \alpha_n\} = \mathsf{P}_0(ag_P)$ *and* $\forall\, i, i'$ *such that* $1 \leq i, i' \leq n$, $\alpha_i = \alpha_{i'}$ *iff* $i = i'$.

Example 5. Let $ds = (Ag, \mathsf{P}_0, \Upsilon_0, f)$ be the dialogue scenario given in Example 2 where $\mathsf{P}_0(ag_P) = \{a, b, c, d, e\}$, $\mathsf{P}_0(ag_R) = \{b\}$ and $\Upsilon_0 = \{\{a, b\}, \{a, c\}\}$.
Examples of strategies of ds are $[b, c, a, d, e]$ and $[b, c, a, e, d]$. However, $[b, c, e]$ is not a strategy of ds and $[b, c, a, c, d, e]$ is not a strategy of ds.

Whether the persuader makes a close move is determined by its initial position, the assertions it has already made and (in the case of an economical persuader) its model of the responder; whether the responder makes a close move is determined by its initial position and the assertions made by the persuader. Thus each possible strategy maps to exactly one dialogue for each type of persuader and a given dialogue scenario, where the assertions made during the dialogue correspond to a prefix of the strategy; we say this is the *dialogue of the persuader type generated by the strategy from the dialogue scenario.*

Definition 8. *Let* $ds \in \mathcal{S}$, $\mathrm{T} \in \{\mathsf{exh}, \mathsf{bas}, \mathsf{sm}\}$ *and* $st = [\alpha_1, \ldots, \alpha_n]$ *be a strategy of* ds. *The* **dialogue of type T generated by** st **from** ds, *denoted* $\mathsf{Dialogue}(ds, \mathrm{T}, st)$, *is* $d = [(m_1, \mathsf{P}_1, \Upsilon_1), \ldots, (m_t, \mathsf{P}_t, \Upsilon_t)]$ *($t \leq 2n + 2$) such that*

1. $d \in \mathsf{Dialogues}_{\mathsf{reg}}(ds)$,
2. $\forall\, i$ *such that* $1 < i < t$ *and* i *is odd,* $m_i = (\mathtt{assert}, \alpha_x)$ *where* $x = \frac{i-1}{2}$,
3. *if* $\mathrm{T} = \mathsf{exh}$, *then* d *has an exhaustive persuader,*
4. *if* $\mathrm{T} = \mathsf{bas}$, *then* d *has a basic persuader,*
5. *if* $\mathrm{T} = \mathsf{sm}$, *then* d *has a smart persuader.*

Example 6. Let $ds = (Ag, \mathsf{P}_0, \Upsilon_0, f)$ be the dialogue scenario given in Example 2 where $\mathsf{P}_0(ag_P) = \{a, b, c, d, e\}$, $\mathsf{P}_0(ag_R) = \{b\}$, $\Upsilon_0 = \{\{a, b\}, \{a, c\}\}$ and the only sets of beliefs that determine the topic f to be acceptable (i.e., the only sets Φ such that $f \in \mathsf{Acceptable}(\Phi)$) are $\{b, c, d\}$ and $\{a, b, c\}$.
Let $st1 = [b, c, a, d, e]$, $st2 = [b, c, a, e, d]$, $st3 = [a, d, c, e, b]$ be strategies of ds.

$\mathsf{Dialogue}(ds, \mathsf{bas}, st1) = \mathsf{Dialogue}(ds, \mathsf{bas}, st2) = d2$ (as given in Example 2).
$\mathsf{Dialogue}(ds, \mathsf{sm}, st1) = \mathsf{Dialogue}(ds, \mathsf{sm}, st2) = d5$ (as given in Example 4).
$\mathsf{Dialogue}(ds, \mathsf{exh}, st3) = d6$ (as given in Example 4).

It follows from Definition 6 that if a basic and a smart persuader each follow the same strategy, the smart persuader's model will be a subset of the basic persuader's model at corresponding points in the two dialogues produced.

Lemma 2. *If* $ds \in \mathcal{S}$ *and* st *is a strategy of* ds *such that* $\mathsf{Dialogue}(ds, \mathsf{bas}, st) = [(m_{b_1}, \mathsf{P}_{b_1}, \Upsilon_{b_1}), \ldots, (m_{b_n}, \mathsf{P}_{b_n}, \Upsilon_{b_n})]$ *and* $\mathsf{Dialogue}(ds, \mathsf{sm}, st) = [(m_{s_1}, \mathsf{P}_{s_1}, \Upsilon_{s_1}), \ldots, (m_{s_m}, \mathsf{P}_{s_m}, \Upsilon_{s_m})]$, *then* $\forall\, i$ *such that* $1 \leq i \leq m$, *if* $i \leq n$, *then* $\Upsilon_{s_i} \subseteq \Upsilon_{b_i}$.

It follows from our definitions and the previous lemma, that the smart dialogue generated by a particular strategy from a dialogue scenario is never longer than the basic dialogue generated with the same strategy, which is never longer than the exhaustive dialogue generated.

Proposition 1. *If* $ds \in \mathcal{S}$ *and* st *is a strategy of* ds, *then*
$\mathsf{Length}(\mathsf{Dialogue}(ds, \mathsf{exh}, st)) \geq \mathsf{Length}(\mathsf{Dialogue}(ds, \mathsf{bas}, st)) \geq \mathsf{Length}(\mathsf{Dialogue}(ds, \mathsf{sm}, st))$.

In the following section, we compare the performance of the different types of persuader we have defined (exhaustive, basic, smart). In particular, we identify the situations in which a persuader of one type can be successful while a persuader of another type may be unsuccessful.

5 Performance of model update mechanisms

We are interested in identifying the situations when a persuader of one type can have an advantage over a persuader of another type; i.e., when, following a particular strategy, a persuader of one type will successfully convince the responder, while a persuader of another type will not. We show that, for accurate scenarios, there is no difference in success of the different persuader types (Table 1). For accurate scenarios, the only difference in the dialogues produced by the different types of persuader with a particular strategy is that, if the dialogues produced are unsuccessful, a smart persuader may terminate the dialogue before a basic persuader, who may terminate before an exhaustive persuader.

We show that, for scenarios that are not accurate, it is possible for an exhaustive persuader to generate a successful dialogue, while a basic and a smart persuader each generate an unsuccessful dialogue with the same strategy. We also show that it is possible for an exhaustive and a basic persuader to each generate a successful dialogue, while a smart persuader generates an unsuccessful dialogue with the same strategy. (These results are summarised in Table 2.)

5.1 Performance of mechanisms for accurate scenarios

If we consider only accurate scenarios, if a persuader of a particular type generates a successful dialogue with a given strategy, then a persuader of either of the other types will generate the same dialogue with the same strategy. This follows from Lemma 1, the definitions of basic, smart and exhaustive persuaders (Def. 6) and the assumptions we have made about regular dialogues (Def. 4).

Lemma 3. *If* $ds \in \mathcal{S}_{acc}$ *and* st *is a strategy of* ds *such that* $\mathsf{Dialogue}(ds, \mathrm{T}, st)$ *is successful (where* $\mathrm{T} \in \{\mathsf{exh}, \mathsf{bas}, \mathsf{sm}\}$*) then* $\forall\, \mathrm{T}' \in \{\mathsf{exh}, \mathsf{bas}, \mathsf{sm}\}$, $\mathsf{Dialogue}(ds, \mathrm{T}, st) = \mathsf{Dialogue}(ds, \mathrm{T}', st)$.

If we again consider only accurate scenarios, but with a strategy that generates an unsuccessful dialogue for one persuader type, then the same strategy will also generate an unsuccessful dialogue for each of the other persuader types (in this case the dialogue generated by a smart persuader may be shorter than the dialogue generated by a basic persuader, which may be shorter than the dialogue generated by an exhaustive persuader, which must be of length $2n+3$ where n is the size of the persuader's position). Again, this follows from Lemma 1, the definitions of basic, smart and exhaustive persuaders (Def. 6) and the assumptions we have made about regular dialogues (Def. 4).

Lemma 4. *If* $ds \in \mathcal{S}_{acc}$ *and* st *is a strategy of* ds *such that* $\mathsf{Dialogue}(ds, \mathrm{T}, st)$ *is unsuccessful (where* $\mathrm{T} \in \{\mathsf{exh}, \mathsf{bas}, \mathsf{sm}\}$*) then* $\forall\, \mathrm{T}' \in \{\mathsf{exh}, \mathsf{bas}, \mathsf{sm}\}$, $\mathsf{Dialogue}(ds, \mathrm{T}', st)$ *is also unsuccessful.*

It is clear from the above results that there are no accurate dialogue scenarios for which there is any difference in success of the different agent types.

Proposition 2. $\nexists\, ds \in \mathcal{S}_{acc}$ *such that* st *is a strategy of* ds, $\mathsf{Dialogue}(ds, \mathrm{T}, st)$ *is successful,* $\mathsf{Dialogue}(ds, \mathrm{T}', st)$ *is unsuccessful,* $\mathrm{T}, \mathrm{T}' \in \{\mathsf{exh}, \mathsf{bas}, \mathsf{sm}\}$ *and* $\mathrm{T}' \neq \mathrm{T}$.

It is straightforward to construct examples to show that there are accurate dialogue scenarios in which, when following the same strategy, all agent types will be successful (similarly unsuccessful). This gives us the following propositions.

Proposition 3. $\exists\, ds \in \mathcal{S}_{acc}$ *such that* st *is a strategy of* ds *and* $\forall\, \mathrm{T} \in \{\mathsf{exh}, \mathsf{bas}, \mathsf{sm}\}$, $\mathsf{Dialogue}(ds, \mathrm{T}, st)$ *is successful.*

Proposition 4. $\exists\, ds \in \mathcal{S}_{acc}$ *such that* st *is a strategy of* ds *and* $\forall\, \mathrm{T} \in \{\mathsf{exh}, \mathsf{bas}, \mathsf{sm}\}$, $\mathsf{Dialogue}(ds, \mathrm{T}, st)$ *is unsuccessful.*

These results are summarised in Table 1.

5.2 Performance of mechanisms for scenarios that are not accurate

For any dialogue scenario (accurate or not), if the dialogue generated by an exhaustive persuader with a particular strategy is unsuccessful, then the dialogue generated by a basic persuader with the same strategy is unsuccessful;

Outcome by persuader type			Outcome combination possible for accurate dialogue scenarios?
Exhaustive	**Basic**	**Smart**	
Successful	Successful	Successful	Yes (Proposition 3)
Unsuccessful	Unsuccessful	Unsuccessful	Yes (Proposition 4)
Successful	Unsuccessful	Unsuccessful	No (Proposition 2)
Unsuccessful	Successful	Unsuccessful	No (Proposition 2)
Unsuccessful	Unsuccessful	Successful	No (Proposition 2)
Unsuccessful	Successful	Successful	No (Proposition 2)
Successful	Unsuccessful	Successful	No (Proposition 2)
Successful	Successful	Unsuccessful	No (Proposition 2)

Table 1. For an accurate dialogue scenario and a particular strategy, identifies possible combinations of outcomes by persuader type.

similarly, if the dialogue generated by a basic persuader with a particular strategy is unsuccessful, then the dialogue generated by a smart persuader with the same strategy is unsuccessful. This follows from Lemma 2, the definition of an exhaustive persuader (Def. 6) and the assumptions we have made about regular dialogues (Def. 4).

Lemma 5. *Let $ds \in \mathcal{S}$ and st be a strategy of ds.*
If $\mathsf{Dialogue}(ds, \mathsf{exh}, st)$ *is unsuccessful, then* $\mathsf{Dialogue}(ds, \mathsf{bas}, st)$ *is unsuccessful.*
If $\mathsf{Dialogue}(ds, \mathsf{bas}, st)$ *is unsuccessful, then* $\mathsf{Dialogue}(ds, \mathsf{sm}, st)$ *is unsuccessful.*

It follows straightforwardly from the above lemma that there are no dialogue scenarios for which (when following the same strategy) a smart persuader will be successful while either an exhaustive or a basic persuader will be unsuccessful, nor are there any dialogue scenario for which a basic persuader will be successful but an exhaustive persuader (with the same strategy) will be unsuccessful.

Proposition 5. *$\nexists\, ds \in \mathcal{S} \backslash \mathcal{S}_{acc}$ such that st is a strategy of ds,* $\mathsf{Dialogue}(ds, \mathsf{sm}, st)$ *is successful,* $\mathsf{Dialogue}(ds, \mathrm{T}, st)$ *is unsuccessful, and* $\mathrm{T} \in \{\mathsf{exh}, \mathsf{bas}\}$.

Proposition 6. *$\nexists\, ds \in \mathcal{S} \backslash \mathcal{S}_{acc}$ such that st is a strategy of ds,* $\mathsf{Dialogue}(ds, \mathsf{bas}, st)$ *is successful and* $\mathsf{Dialogue}(ds, \mathsf{exh}, st)$ *is unsuccessful.*

We now show by example that all other combinations of difference in outcome from the different persuader types are possible. First, we show that there exists a dialogue scenario that is not accurate in which all persuader types are successful.

Proposition 7. *$\exists\, ds \in \mathcal{S} \setminus \mathcal{S}_{acc}$ such that st is a strategy of ds, and $\forall\, \mathrm{T} \in \{\mathsf{exh}, \mathsf{bas}, \mathsf{sm}\}$,* $\mathsf{Dialogue}(ds, \mathrm{T}, st)$ *is successful.*

Proof. *Let $ds = (Ag, \mathsf{P}_0, \Upsilon_0, f)$ be a dialogue scenario with topic f such that $\mathsf{P}_0(ag_P) = \{a, b, c, d, e\}$, $\mathsf{P}_0(ag_R) = \{b\}$, $\Upsilon_0 = \{\{d\}\}$. The only sets of beliefs that determine the topic f to be acceptable (i.e., the only sets Φ such that $f \in \mathsf{Acceptable}(\Phi)$) are $\{a, b, c\}$ and $\{a, c, d, e\}$.*
Following the strategy $[a, c, e, d, b]$, each of the persuader types produces a successful dialogue where after it has asserted a and then c the responder will close the dialogue, indicating it has been persuaded.

We now show that there exists a dialogue scenario that is not accurate in which all persuader types are unsuccessful.

Proposition 8. *$\exists\, ds \in \mathcal{S} \setminus \mathcal{S}_{acc}$ such that st is a strategy of ds, and $\forall\, \mathrm{T} \in \{\mathsf{exh}, \mathsf{bas}, \mathsf{sm}\}$, $\mathsf{Dialogue}(ds, \mathrm{T}, st)$ is unsuccessful.*
Proof. *Let $ds = (Ag, \mathsf{P}_0, \Upsilon_0, f)$ be a dialogue scenario with topic f such that $\mathsf{P}_0(ag_P) = \{a, b, c, d, e\}$, $\mathsf{P}_0(ag_R) = \{b, d\}$, $\Upsilon_0 = \{\{a\}, \{d, e\}\}$. The only sets of beliefs that determine the topic f to be acceptable (i.e., the only sets Φ such that $f \in \mathsf{Acceptable}(\Phi)$) are $\{a, b, c\}$ and $\{a, c, d, e\}$.*
For all possible strategies of this dialogue scenario, each of the different persuader types will produce an unsuccessful dialogue (since there is no superset of the responder's initial position that determines the topic to be acceptable).

We now show the existence of a dialogue scenario that is not accurate in which, when following a particular strategy, an exhaustive persuader will be successful but both a basic and a smart persuader will be unsuccessful.

Proposition 9. *$\exists\, ds \in \mathcal{S} \setminus \mathcal{S}_{acc}$ such that st is a strategy of ds, $\mathsf{Dialogue}(ds, \mathsf{exh}, st)$ is successful, and $\forall\, \mathrm{T} \in \{\mathsf{bas}, \mathsf{sm}\}$, $\mathsf{Dialogue}(ds, \mathrm{T}, st)$ is unsuccessful.*
Proof. *Let $ds = (Ag, \mathsf{P}_0, \Upsilon_0, f)$ be a dialogue scenario with topic f such that $\mathsf{P}_0(ag_P) = \{a, b, c, d\}$, $\mathsf{P}_0(ag_R) = \{b, e\}$, $\Upsilon_0 = \{\{a, b\}, \{b, c\}\}$. The only sets of beliefs that determine the topic f to be acceptable (i.e., the only sets Φ such that $f \in \mathsf{Acceptable}(\Phi)$) are $\{a, c, d\}$ and $\{b, d, e\}$.*
No matter what strategy they are following, both a smart and a basic persuader will choose to terminate the dialogue unsuccessfully without asserting any beliefs, since according to their model of the responder they believe there is no way the responder can successfully be persuaded (as there is no superset of any element of its model that determines the topic to be acceptable). However, an exhaustive persuader with a strategy that chooses to assert d first will be successful.

Finally, we show the existence of a dialogue scenario that is not accurate in which, when following a particular strategy, both an exhaustive and a basic persuader will be successful but a smart persuader will be unsuccessful.

Proposition 10. *$\exists\, ds \in \mathcal{S} \setminus \mathcal{S}_{acc}$ such that st is a strategy of ds, $\forall\, \mathrm{T} \in \{\mathsf{exh}, \mathsf{bas}\}$ $\mathsf{Dialogue}(ds, \mathrm{T}, st)$ is successful, and $\mathsf{Dialogue}(ds, \mathsf{sm}, st)$ is unsuccessful.*
Proof. *Let $ds = (Ag, \mathsf{P}_0, \Upsilon_0, f)$ be a dialogue scenario with topic f such that $\mathsf{P}_0(ag_P) = \{a, b, c, d, e\}$, $\mathsf{P}_0(ag_R) = \{a, b\}$, $\Upsilon_0 = \{\{b, e\}, \{b, d\}, \{c\}\}$. The only sets of beliefs that determine the topic f to be acceptable (i.e., the only sets Φ such that $f \in \mathsf{Acceptable}(\Phi)$) are $\{a, b, d\}$, $\{a, b, e\}$, $\{b, c\}$ and $\{a, b, d, e\}$.*
Consider the strategy $st = [a, e, b, d, c]$.

$\mathsf{Dialogue}(ds, \mathsf{exh}, st) = [((\mathtt{open}, f), \mathsf{P}_1, \Upsilon_0), ((\mathtt{pass}), \mathsf{P}_2, \Upsilon_0), ((\mathtt{assert}, a), \mathsf{P}_3, \Upsilon_0), ((\mathtt{pass}), \mathsf{P}_4, \Upsilon_0), ((\mathtt{assert}, e), \mathsf{P}_5, \Upsilon_0), ((\mathtt{close}), \mathsf{P}_6, \Upsilon_0)]$ *where*

- *$\forall\, i$ such that $1 \leq i \leq 6$, $\mathsf{P}_i(ag_P) = \{a, b, c, d, e\}$,*
- *$\forall\, i$ such that $1 \leq i \leq 4$, $\mathsf{P}_i(ag_R) = \{a, b\}$,*
- *$\mathsf{P}_5(ag_R) = \mathsf{P}_6(ag_R) = \{a, b, e\}$.*

$\mathsf{Dialogue}(ds, \mathsf{bas}, st) = [((\mathtt{open}, f), \mathsf{P}_1, \Upsilon_1), ((\mathtt{pass}), \mathsf{P}_2, \Upsilon_2), ((\mathtt{assert}, a), \mathsf{P}_3, \Upsilon_3), ((\mathtt{pass}), \mathsf{P}_4, \Upsilon_4), ((\mathtt{assert}, e), \mathsf{P}_5, \Upsilon_5), ((\mathtt{close}), \mathsf{P}_6, \Upsilon_6)]$ *where*

- $\forall\, i$ *such that* $1 \leq i \leq 6$, $\mathsf{P}_i(ag_P) = \{a, b, c, d, e\}$,
- $\forall\, i$ *such that* $1 \leq i \leq 4$, $\mathsf{P}_i(ag_R) = \{a, b\}$,
- $\mathsf{P}_5(ag_R) = \mathsf{P}_6(ag_R) = \{a, b, e\}$,
- $\Upsilon_1 = \Upsilon_2 = \{\{b, e\}, \{b, d\}, \{c\}\}$,
- $\Upsilon_3 = \Upsilon_4 = \{\{a, b, e\}, \{a, b, d\}, \{a, c\}\}$,
- $\Upsilon_3 = \Upsilon_4 = \{\{a, b, e\}, \{a, b, d\}, \{a, c\}\}$,
- $\Upsilon_5 = \Upsilon_6 = \{\{a, b, e\}, \{a, b, d, e\}, \{a, c, e\}\}$.

$\mathsf{Dialogue}(ds, \mathsf{sm}, st) = [((\mathtt{open}, f), \mathsf{P}_1, \Upsilon_1), ((\mathtt{pass}), \mathsf{P}_2, \Upsilon_2), ((\mathtt{assert}, a), \mathsf{P}_3, \Upsilon_3), ((\mathtt{pass}), \mathsf{P}_4, \Upsilon_4), ((\mathtt{close}), \mathsf{P}_5, \Upsilon_5)]$ *where*

- $\forall\, i$ *such that* $1 \leq i \leq 6$, $\mathsf{P}_i(ag_P) = \{a, b, c, d, e\}$,
- $\forall\, i$ *such that* $1 \leq i \leq 5$, $\mathsf{P}_i(ag_R) = \{a, b\}$,
- $\Upsilon_1 = \Upsilon_2 = \{\{b, e\}, \{b, d\}, \{c\}\}$,
- $\Upsilon_3 = \{\{a, b, e\}, \{a, b, d\}, \{a, c\}\}$,
- $\Upsilon_4 = \Upsilon_5 = \{\{a, c\}\}$.

Thus we see that while the exhaustive and basic persuader types are each successful, the smart persuader incorrectly perceives there to be no chance of convincing the responder and terminates the dialogue unsuccessfully.

These results are summarised in Table 2. They demonstrate the potential disadvantage of behaving economically (that is, choosing to give up trying to persuade the responder as soon as, according to one's image of the responder, success seems impossible) in the case where the persuader's image of the responder may not be accurate. Furthermore, they show that a smart persuader may incorrectly perceive its position to be hopeless while a basic persuader may not. We now consider the conditions under which an exhaustive persuader is successful but an economical persuader (basic or smart) is not, following which we consider the conditions under which an exhaustive and a basic persuader are successful but a smart persuader is not.

For a dialogue scenario that is not accurate, it follows from our definitions that the dialogue produced by an exhaustive persuader with a particular strategy is successful but the dialogue produced by a basic persuader with the same strategy is unsuccessful if and only if: the arguments asserted by the basic persuader are a strict prefix of those asserted by the exhaustive persuader (condition 1 in Prop. 11); the topic of the dialogue is determined to be acceptable by the union of the responder's initial position with the arguments asserted by the exhaustive persuader (condition 2); there is no strict prefix of the arguments asserted by the exhaustive persuader that, when combined with the responder's initial position, determines the topic to be acceptable (condition 3); for every element of the persuader's initial model of the responder, there is no subset of the persuader's initial position that contains the arguments asserted by the basic persuader and that, when combined with the responder's initial position, determines the topic to be acceptable (condition 4); for every strict prefix of the arguments asserted by the basic persuader, there is some subset of the persuader's initial position that contains that prefix and some element of the persuader's initial model such

Outcome by persuader type			Outcome combination possible for dialogue scenarios that are not accurate?
Exhaustive	**Basic**	**Smart**	
Successful	Successful	Successful	Yes (Proposition 7)
Unsuccessful	Unsuccessful	Unsuccessful	Yes (Proposition 8)
Successful	Unsuccessful	Unsuccessful	Yes (Proposition 9)
Unsuccessful	Successful	Unsuccessful	No (Proposition 6)
Unsuccessful	Unsuccessful	Successful	No (Proposition 5)
Unsuccessful	Successful	Successful	No (Proposition 5/6)
Successful	Unsuccessful	Successful	No (Proposition 5)
Successful	Successful	Unsuccessful	Yes (Proposition 10)

Table 2. For a not-accurate dialogue scenario that is not accurate and a particular dialogue strategy, identifies possible combinations of outcomes by persuader type.

that the union of the two determines the topic to be acceptable (condition 5). Furthermore, it follows from these results that for every element Ψ of the persuader's initial model of the responder: if Ψ is a proper subset of the responder's initial position, then there is a belief that the responder is aware of and of which the persuader has no knowledge; if the responder's initial position is a proper subset of Ψ, then there is some belief in Ψ that is not in the responder's initial position and is not asserted by the exhaustive persuader; otherwise there is something in the responder's initial position that is not in Ψ and there is something in Ψ that is not in the responder's initial position, and either there is a belief in the responder's initial position that is not present in the persuader's initial position, or there is a belief in Ψ that is not in the responder's initial position and is not asserted by the exhaustive persuader (condition 6).

Proposition 11. *Let* $ds = (Ag, \mathsf{P}_0, \Upsilon_0, \tau) \in \mathcal{S} \setminus \mathcal{S}_{acc}$ *and* $st = [\alpha_1, \ldots, \alpha_n]$ *be a strategy of* ds.
$\mathsf{Dialogue}(ds, \mathsf{bas}, st)$ *is unsuccessful and* $\mathsf{Dialogue}(ds, \mathsf{exh}, st)$ *is successful where* $\mathsf{Dialogue}(ds, \mathsf{bas}, st) = [(\mathtt{open}, \tau), (\mathtt{pass}), (\mathtt{assert}, \alpha_1), (\mathtt{pass}), \ldots, (\mathtt{assert}, \alpha_j), (\mathtt{pass}), (\mathtt{close})]$ *and* $\mathsf{Dialogue}(ds, \mathsf{exh}, st) = [(\mathtt{open}, \tau), (\mathtt{pass}), (\mathtt{assert}, \alpha_1), (\mathtt{pass}), \ldots, (\mathtt{assert}, \alpha_k), (\mathtt{close})]$ *iff*

1. $j < k$,
2. $\tau \in \mathsf{Acceptable}(\{\alpha_1, \ldots, \alpha_k\} \cup \mathsf{P}_0(ag_R))$,
3. $\nexists\, i$ *such that* $1 \leq i < k$ *and* $\tau \in \mathsf{Acceptable}(\{\alpha_1, \ldots, \alpha_i\} \cup \mathsf{P}_0(ag_R))$,
4. $\forall \Psi \in \Upsilon_0, \nexists \Phi \subseteq \mathsf{P}_0(ag_P)$ *such that* $\{\alpha_1, \ldots, \alpha_j\} \subset \Phi$ *and* $\tau \in \mathsf{Acceptable}(\Psi \cup \Phi)$,
5. $\forall i$ *such that* $1 \leq i < j$, $\exists \Psi \in \Upsilon_0$ *such that* $\exists \Phi \subseteq \mathsf{P}_0(ag_P)$ *such that* $\{\alpha_1, \ldots, \alpha_i\} \subset \Phi$ *and* $\tau \in \mathsf{Acceptable}(\Psi \cup \Phi)$, *and*
6. $\forall \Psi \in \Upsilon_0$, *either*
   - $\Psi \subset \mathsf{P}_0(ag_R)$ *and* $\exists\, \phi \in \mathsf{P}_0(ag_R) \setminus \Psi$ *such that* $\phi \notin \mathsf{P}_0(ag_P)$,
   - $\mathsf{P}_0(ag_R) \subset \Psi$ *and* $\exists\, \phi$ *such that* $\phi \in \Psi \setminus \mathsf{P}_0(ag_R)$ *and* $\phi \notin \{\alpha_1, \ldots, \alpha_k\}$, *otherwise*
   - $\exists\, \phi \in \mathsf{P}_0(ag_R) \setminus \Psi$, $\exists\, \psi \in \Psi \setminus \mathsf{P}_0(ag_R)$, *and either*
     - $\mathsf{P}_0(ag_R) \setminus \mathsf{P}_0(ag_P) \neq \emptyset$, *or*
     - $\exists\, \phi$ *such that* $\phi \in \Psi \setminus \mathsf{P}_0(ag_R)$ *and* $\phi \notin \{\alpha_1, \ldots, \alpha_k\}$.

Proof. Left to right. Condition 1 follows directly from the definition of successful dialogues (Def. 3) and Prop. 1. Conditions 2-5 follow directly from the definitions of a basic persuader, an exhaustive persuader, an economical persuader and a regular dialogue (Defs. 4, 5 and 6.)
Since $ds \notin \mathcal{S}_{acc}$, it cannot be the case that $\mathsf{P}_0(ag_R) \in \Upsilon_0$ (Def. 2), thus $\forall \Psi \in \Upsilon_0$, either $\Psi \subset \mathsf{P}_0(ag_R)$, $\mathsf{P}_0(ag_R) \subset \Psi$, or $\exists\, \phi \in \mathsf{P}_0(ag_R) \setminus \Psi$ and $\exists\, \psi \in \Psi \setminus \mathsf{P}_0(ag_R)$. We now consider these three cases in turn.
Let $\Psi \in \Upsilon_0$ such that $\Psi \subset \mathsf{P}_0(ag_R)$. Since $\tau \in \mathsf{Acceptable}(\{\alpha_1, \ldots, \alpha_k\} \cup \mathsf{P}_0(ag_R))$ and $\Psi \subset \mathsf{P}_0(ag_R)$, $\tau \in \mathsf{Acceptable}(\Lambda \cup \Psi \cup \{\alpha_1, \ldots, \alpha_k\})$ where $\Lambda = \mathsf{P}_0(ag_R) \setminus \Psi$. From 4, $\nexists \Phi \subseteq \mathsf{P}_0(ag_P)$ such that $\{\alpha_1, \ldots, \alpha_j\} \subset \Phi$ and $\tau \in \mathsf{Acceptable}(\Psi \cup \Phi)$. Therefore, $\exists\, \phi \in \Lambda$ such that $\phi \notin \mathsf{P}_0(ag_P)$ and thus $\exists\, \phi \in \mathsf{P}_0(ag_R) \setminus \Psi$ such that $\phi \notin \mathsf{P}_0(ag_P)$.
Let $\Psi \in \Upsilon_0$ such that $\mathsf{P}_0(ag_R) \subset \Psi$. It follows from 4 that $\tau \notin \mathsf{Acceptable}(\Psi \cup \{\alpha_1, \ldots, \alpha_k\})$. Since $\tau \in \mathsf{Acceptable}(\{\alpha_1, \ldots, \alpha_k\} \cup \mathsf{P}_0(ag_R))$ and $\mathsf{P}_0(ag_R) \subset \Psi$, it must be the case that $\exists\, \phi$ such that $\phi \in \Psi \setminus \mathsf{P}_0(ag_R)$ and $\phi \notin \{\alpha_1, \ldots, \alpha_k\}$.
Let $\Psi \in \Upsilon_0$ such that $\exists\, \phi \in \mathsf{P}_0(ag_R) \setminus \Psi$ and $\exists\, \psi \in \Psi \setminus \mathsf{P}_0(ag_R)$. Assume $\mathsf{P}_0(ag_R) \setminus \mathsf{P}_0(ag_P) = \emptyset$ (i.e., $\mathsf{P}_0(ag_R) \subseteq \mathsf{P}_0(ag_P)$). Since $\mathsf{P}_0(ag_R) \subseteq \mathsf{P}_0(ag_P)$, it follows from 4 that $\tau \notin \mathsf{Acceptable}(\Psi \cup \{\alpha_1, \ldots, \alpha_k\} \cup \mathsf{P}_0(ag_R))$. Since $\tau \in \mathsf{Acceptable}(\{\alpha_1, \ldots, \alpha_k\} \cup \mathsf{P}_0(ag_R))$, it follows that $\exists\, \phi \in \Psi$ such that $\phi \notin (\{\alpha_1, \ldots, \alpha_n\} \cup \mathsf{P}_0(ag_R))$. Therefore either $\mathsf{P}_0(ag_R) \setminus \mathsf{P}_0(ag_P) \neq \emptyset$, or $\exists\, \phi$ such that $\phi \in \Psi \setminus \mathsf{P}_0(ag_R)$ and $\phi \notin \{\alpha_1, \ldots, \alpha_k\}$.
Right to left. Follows from conditions 2-5 and from the definitions of a basic persuader, an exhaustive persuader, an economical persuader and a regular dialogue (Defs. 4, 5 and 6.).

Also considering only non-accurate dialogue scenarios, it similarly follows from our definitions that the dialogue produced by a basic persuader with a particular strategy is successful but the dialogue produced by a smart persuader with the same strategy is unsuccessful if and only if: the arguments asserted by the smart persuader are a strict prefix of those asserted by the basic persuader (condition 1, Prop. 12); the topic of the dialogue is determined to be acceptable by the union of responder's initial position with the arguments asserted by the basic persuader (condition 2); there is no strict prefix of the arguments asserted by the basic persuader that, when combined with the responder's initial position, determines the topic to be acceptable (condition 3); for every element Ψ of the persuader's initial model of the responder, either Ψ determines the topic to be acceptable, or there is some strict prefix of the arguments asserted by the smart persuader that when combined with Ψ determines the topic to be acceptable, or there is no subset of the persuader's initial position that contains the arguments asserted by the smart persuader and when combined with Ψ determines the topic to be acceptable (condition 4); for every strict prefix $p1$ of the arguments asserted by the smart persuader, there is some element Ψ of the persuader's initial model such that there is no strict prefix $p2$ of $p1$ (including the empty prefix) that when combined with Ψ determines the topic to be acceptable and such that there exists some subset of the persuader's initial position that contains $p1$ and when combined with Ψ determines the topic to be acceptable.

Proposition 12. *Let* $ds = (Ag, \mathsf{P}_0, \Upsilon_0, \tau) \in \mathcal{S} \setminus \mathcal{S}_{acc}$ *and* $st = [\alpha_1, \ldots, \alpha_n]$ *be a strategy of* ds.
$\mathsf{Dialogue}(ds, \mathsf{sm}, st)$ *is unsuccessful and* $\mathsf{Dialogue}(ds, \mathsf{bas}, st)$ *is successful where* $\mathsf{Dialogue}(ds, \mathsf{sm}, st) = [(\mathtt{open}, \tau), (\mathtt{pass}), (\mathtt{assert}, \alpha_1), (\mathtt{pass}), \ldots, (\mathtt{assert}, \alpha_j), (\mathtt{pass}), (\mathtt{close})]$ *and* $\mathsf{Dialogue}(ds, \mathsf{bas}, st) = [(\mathtt{open}, \tau), (\mathtt{pass}), (\mathtt{assert}, \alpha_1), (\mathtt{pass}), \ldots, (\mathtt{assert}, \alpha_k), (\mathtt{close})]$ *iff*

1. $j < k$,
2. $\tau \in \mathsf{Acceptable}(\{\alpha_1, \ldots, \alpha_k\} \cup \mathsf{P}_0(ag_R))$,
3. $\forall\, i$ *such that* $1 \leq i < k$, $\exists \Psi \in \Upsilon_0$ *such that* $\exists \Phi \subseteq \mathsf{P}_0(ag_P)$ *such that* $\{\alpha_1, \ldots, \alpha_i\} \subset \Phi$ *and* $\tau \in \mathsf{Acceptable}(\Psi \cup \Phi)$.
4. $\forall \Psi \in \Upsilon_0$ *such that* $\tau \notin \mathsf{Acceptable}(\Psi)$ *and* $\nexists\, i$ *such that* $1 \leq i < j$ *and* $\tau \in \mathsf{Acceptable}(\Psi \cup \{\alpha_1, \ldots, \alpha_i\})$, $\nexists\, \Phi \subseteq \mathsf{P}_0(ag_P)$ *such that* $\{\alpha_1, \ldots, \alpha_j\} \subset \Phi$ *and* $\tau \in \mathsf{Acceptable}(\Psi \cup \Phi)$,
5. $\forall\, i$ *such that* $1 \leq i < j$, $\exists \Psi \in \Upsilon_0$ *such that* $\tau \notin \mathsf{Acceptable}(\Psi)$, $\nexists\, h$ *such that* $1 \leq h < i$ *and* $\tau \in \mathsf{Acceptable}(\Psi \cup \{\alpha_1, \ldots, \alpha_h\})$, *and* $\exists \Phi \subseteq \mathsf{P}_0(ag_P)$ *such that* $\{\alpha_1, \ldots, \alpha_i\} \subset \Phi$ *and* $\tau \in \mathsf{Acceptable}(\Psi \cup \Phi)$.

Proof. Follows from the definitions of successful dialogues, a basic persuader, a smart persuader, an economical persuader and a regular dialogue (Defs. 3, 4, 5 and 6) and Prop. 1.

Propositions 11 and 12 identify the necessary and sufficient conditions under which, while following the same strategy, a persuader of one type will successfully convince the responder, while a persuader of another type will not. These results help us to understand the situations in which the use of the different update mechanisms considered here can be disadvantageous.

6 Discussion

In this paper we have formally investigated the use of two approaches (basic and smart) for updating an uncertain opponent model in simple persuasion dialogues, where the persuader uses such a model to determine whether there is any chance of the dialogue leading to success, giving up and unsuccessfully terminating the dialogue as soon as it believes this not to be the case. We have shown that, if the persuader's initial model of the responder is accurate (i.e., represents the responder's actual position as being possible) there is no difference in the outcomes produced by the different persuader types with a particular strategy (where a strategy here predetermines the sequence of assertions to make), but a smart persuader may produce a shorter dialogue than a basic persuader, and a basic persuader may produce a shorter dialogue than an exhaustive persuader. In the case where the persuader's initial model of the responder does not represent the responder's actual initial position as a possibility, we have shown that it is possible for an exhaustive persuader to succeed in persuading the responder, while both a basic and a smart persuader following the same strategy will fail, and that it is possible for an exhaustive and basic persuader to be successful while a smart persuader following the same strategy will fail, and identified the conditions under which these cases occur.

These results help us to understand the situations under which it can be useful to apply the basic or the smart model update mechanism. If shorter dialogues are desirable and it is certain that the responder's actual initial position is captured as a possibility in the persuader's model, then a smart persuader will produce the best outcome, only producing an unsuccessful dialogue if neither a basic nor exhaustive persuader would succeed with the same strategy, but potentially terminating the dialogue at an earlier stage than the other types of persuader. If the persuader's model might not contain the responder's actual initial position as a possibility, then both a smart and a basic persuader risks incorrectly perceiving its position to be hopeless and unsuccessfully terminating the dialogue when in fact continuing with its strategy would lead to success.

Other works have investigated the use of a model of what is known to the opponent in order to generate a proponent's dialogue strategy. Rienstra *et al.* [6] apply a variation of the maxmin algorithm to an uncertain opponent model in order to determine the moves that produce the best expected outcome, while Black *et al.* [5] use automated planning techniques to generate a strategy with the highest chance of success given an uncertain opponent model. In contrast, here we do not consider here the generation of a dialogue strategy (we assume a sequence of assertions to make); however, our results can be beneficial in such settings, particularly in understanding the situations in which a possible strategy might be incorrectly classified as hopeless (using the results from Props. 11 and 12). In their work, Rienstra *et al.* [6] apply the basic model update mechanism and Black *et al.* [5] use the smart approach, however neither explicitly considers the effect the update mechanism has on the outcome of the dialogue. Hadjinikolis *et al.* [7, 8] propose a method an agent can use to augment an opponent model with extra information, based on previous dialogue experience, however they do not consider how this relates to dialogue outcome.

Hunter [11] also considers different mechanisms for updating an opponent model during a dialogue, where this opponent model represents the strength of belief the persuader believes its opponent has in different arguments (in the sense that it finds them convincing). In contrast, our opponent model represents the beliefs the persuader believes the responder is aware of and the responder's belief in the claims of arguments can be captured with the Acceptable function. Hunter considers how the accuracy of a user model can be improved through the use of moves that query the opponent's beliefs; it will be interesting to consider how an opponent model in our setting can be improved with such moves.

While the persuasion dialogue we consider here is simple, in that it is unidirectional and the responder's choice of moves is determined by its position and what has been asserted by the persuader, we believe that the results we present here provide useful foundations for exploring the behaviour of such model update functions in more complex persuasion situations, where each participant is asserting beliefs with the aim of persuading the other. The intuition underlying each of the basic and the smart update functions (to add to one's model beliefs that are asserted and to remove from one's model anything that is inconsistent with the opponent's behaviour) are also applicable in the symmetric persua-

sion setting, and we plan to adapt the results presented here to the symmetric persuasion setting in future work.

We also plan in future work to allow the assignment of probabilities to our uncertain opponent model and adapt our model update mechanisms to manage these, as is considered by Rienstra *et al.* [6] and Hunter [11]. The combination of probability with argumentation is a growing area of interest; e.g., recent work has proposed a framework for analysing the expected utility of probabilistic strategies for argument dialogues [12], while Oren *et al.* consider the use of a probabilistic audience model to determine convincing arguments to move in a monologue [13]. It will be interesting to investigate how the choice of update mechanism for a probabilistic opponent model impacts on dialogue outcome.

Acknowledgements. This work was partially supported by the the UK Engineering and Physical Sciences Research Council, grant ref. EP/M01892X/1.

References

1. Modgil, S., *et al.*: The added value of argumentation. In Ossowski, S., ed.: Agreement Technologies. Springer (2013) 357–403
2. Prakken, H.: Formal systems for persuasion dialogue. The Knowledge Engineering Review **21**(02) (2006) 163–188
3. Black, E., Atkinson, K.: Choosing persuasive arguments for action. In: 10th Int. Conf. on Autonomous Agents and Multiagent Systems. (2011) 905–912
4. Thimm, M.: Strategic argumentation in multi-agent systems. Künstliche Intelligenz, Special Issue on Multi-Agent Decision Making **28**(3) (2014) 159–168
5. Black, E., Coles, A.J., Bernardini, S.: Automated planning of simple persuasion dialogues. In: 15th Int. Workshop on Computational Logic in Multi-Agent Systems. (2014) 87–104
6. Rienstra, T., Thimm, M., Oren, N.: Opponent models with uncertainty for strategic argumentation. In: 23rd Int. Joint Conf. on Artificial Intelligence. (2013) 332–338
7. Hadjinikolis, C., Siantos, Y., Modgil, S., Black, E., McBurney, P.: Opponent modelling in persuasion dialogues. In: 23rd Int. Joint Conf. on Artificial Intelligence. (2013) 164–170
8. Hadjinikolis, C., Modgil, S., Black, E.: Building support-based opponent models in persuasion dialogues. In: Theory and Applications of Formal Argumentation: Third International Workshop, TAFA 2015, Buenos Aires, Argentina, July 25-26, 2015, Revised Selected papers., Springer LNAI 9524 (2016)
9. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence **77**(2) (1995) 321–357
10. Modgil, S., Prakken, H.: A general account of argumentation with preferences. Artificial Intelligence **195**(0) (2013) 361 – 397
11. Hunter, A.: Modelling the persuadee in asymmetric argumentation dialogues for persuasion. In: 24th Int. Joint Conf. on Artificial Intelligence. (2015) 3055–3061
12. Hunter, A.: Probabilistic strategies in dialogical argumentation. In: 8th Int. Conf. on Scalable Uncertainty Management. (2014) 190–202
13. Oren, N., Atkinson, K., Li, H.: Group persuasion through uncertain audience modelling. In: 4th Int. Conf. on Computational Models of Argument. (2012) 350–357