EFFICIENCY IN PERSUASION DIALOGUES

Katie Atkinson[1], Priscilla Bench-Capon[2] and Trevor Bench-Capon[1]

[1]*Department of Computer Science, University of Liverpool, Liverpool, U.K.*
[2]*The Open University, Milton Keynes, U.K.*

Keywords: Dialogue, Persuasion, Argumentation.

Abstract: Inquiry, Persuasion and Deliberation Dialogues are all designed to transfer information between agents so that their beliefs and opinions may be revised in the light of the new information, and all make use of a similar set of speech acts. These dialogues also have significant differences. We define success conditions for some different dialogue types in this family and note the pragmatic implications of the speech acts they employ. Focusing on persuasion we consider how successful persuasion dialogues can be conducted efficiently, in terms of minimising the expected transfer of information. We observe that a strategy for efficient persuasion can be developed by considering the pragmatic implications. We present results showing that our strategy is an optimal strategy in a range of representative persuasion scenarios.

1 INTRODUCTION

The influential dialogue typology of Walton and Krabbe (Walton and Krabbe, 1995) has been taken as a basis for developing dialogue protocols for use in agent systems by several authors. For example, see (Black and Hunter, 2009) for inquiry dialogues; (Rahwan et al., 2003) for negotiation; (Prakken, 2006) for persuasion; and (McBurney et al., 2007) for deliberation. It is, Walton and Krabbe argue, important that the agents recognise the type of dialogue they are participating in, since otherwise misunderstandings arise, fallacies become possible and the conversation may break down. None the less, the distinctions have rarely been made precise, and confusion is increased because inquiry, persuasion and deliberation all make use of a similar set of speech acts. In this paper we will give a precise characterisation of the distinctive features of persuasion dialogues that show how these features distinguish persuasion dialogues from deliberation and inquiry dialogues. Our analysis will draw attention to the pragmatic meaning of utterances used in persuasion dialogues, which give such dialogues their particular character. Following this analysis, we will focus further on persuasion, with a view to obtaining a clear specification of, and an optimal strategy for, persuasion. Our contribution is thus the analysis yielding the distinguishing features of persuasion dialogues, as well as the optimal strategy we present for a common class of persuasion dialogues.

One common situation in which both persuasion and deliberation occur is when an agent must choose between several options, each of which has several features which can serve as reasons for and against the choice, to which different agents will ascribe different degrees of importance. When buying a car one person will be most interested in safety, another in speed, another in comfort. For cameras, some will require the best quality, whereas others will value compactness or a low price more highly. In such situations an agent will often need to deal with a series of salespeople, each trying to overturn the agent's currently preferred option, or to consider recommendations from other agents. A very common example in AI conference papers is choosing a restaurant for lunch or an evening dinner. Typically the individual agents will have only incomplete and often ill-founded or outdated beliefs about the local eateries, and so they will pool their knowledge to get a more complete picture of the situation before deciding. Thus one agent may solicit recommendations, and another agent may attempt to persuade that agent of its own favourite venue. We will use this as an illustration throughout this paper.

In (Walton and Krabbe, 1995), dialogue types are characterised by an initial situation, a collective goal, and individual goals, all stated informally. For inquiry dialogues the initial situation is that both agents are not certain of some proposition p. Both individual goals and the collective goal are the same: to determine whether or not p. In persuasion one agent will

23

argue that p, or that some action ϕ should be done[1], in order that the other agent will come to agreement. The collective goal is to resolve whether p is true or ϕ should be done. With regard to individual goals, persuasion is asymmetric: the persuader wishes to convince the persuadee, whereas the persuadee wishes to explore the possibility that its current opinion should be revised in the light of information known to the persuader: the persuadee is interested in what is true, whether it be p or $\neg p$. A different case of persuasion is what Walton terms a *dispute*. In this case the persuadee also wishes to convince the other agent that its own original position is correct, so that its individual goal is now that the other should believe $\neg p$ or that ϕ should not be done: we will not consider disputes further in this paper. Deliberation is generally held to concern actions: initially both agents are unsure whether or not to ϕ, and individually *and collectively* they wish to come to agreement as to whether or not to ϕ. In the next section we will explore the distinctions further, with a view to precisely characterising persuasion dialogues in particular.

2 DISTINGUISHING THE DIALOGUE TYPES

Instead of distinguishing between actions and propositions, we believe that the correct distinction is related to *directions of fit*, a distinction made by Searle (Searle, 2003). Searle distinguishes theoretical reasoning, reasoning about what is the case, from practical reasoning, reasoning about what it is desired to be the case, and what should be done to realise those desires. In the first case it is necessary to fit one's beliefs to the world, whereas in the second the idea is to make the world fit one's desires, in so far as one has the capacity to do so. In these terms, inquiry represents an attempt to better fit the beliefs of the agents to the world, and deliberation how best to make the world fit the collective desires of the agents. Persuasion can be about either. Note, however, that when we have two (or more) participating agents, we have two (or more), probably different, sets of desires to consider. In deliberation no set of desires should be given pre-eminence, but rather the group as a whole needs to come to an agreement on what desires they will adopt collectively. In contrast, as discussed in (Bench-Capon, 2002), in persuasion it is the desires

of the *persuadee* that matter: a persuadee is fully entitled to use its own preferences to assess any proposition or proposal, without any need to consider what the persuader desires. The construction of a set of collective desires introduces an additional order of complication, and puts deliberation beyond the scope of this paper. Therefore in what follows we will focus exclusively on persuasion.

2.1 Definitions

An example, including example use of our notation, is given in Section 3. The reader might find it helpful to refer to this in conjunction with the following definitions for concrete illustrations of their use.

The knowledge bases of agents can be partitioned into *factual elements*, on which agents should agree[2], used when the direction of fit is from world to beliefs, and *preference elements*, which represent their own individual desires, tastes and aspirations, and are used when the direction of fit is from desires to the world. Thus, the preference elements represent the way the agent evaluates possible futures to determine what it wishes to bring about, and how it evaluates objects and situations for value judgements such as *best car* and *acceptable restaurant*.

Definition 1. *Let AG denote a set of agents, each of which, $Ag \in AG$, has a knowledge base KB^{Ag}. KB^{Ag} is partitioned into factual elements denoted by KB_F^{Ag} and preference information denoted by KB_P^{Ag}. KB_F^{Ag} comprises facts, strict rules and defeasible rules. KB_P^{Ag} comprises rules to determine the utility for Ag of certain items based on their attributes, and the weights used by these rules. These preference elements are defined below.*

Agents expand their KB_F by taking information from one another, but KB_P remains fixed throughout the dialogue. Whereas, because it is intended to fit the world, KB_F is objective, KB_P represents the personal preferences of an individual agent, and is entirely local to the agent concerned. We will use f for factual propositions, and p_{Ag} (to be read as "p is the case for Ag") for propositions based on preferences. We will not represent actions separately, so that p_{Ag} may represent either propositions such as *Roux Brothers is an*

---

[1]There has been some disagreement as to whether persuasion can be over actions. Walton in (Walton, 1998) seems to suggest not. None the less it is clear that we are, in ordinary language, fully prepared to speak of persuading someone to do something.

[2]Of course, this does assume that there is a set of propositions which are objectively true. In practice there may be room for dispute. None the less we will assume agreement in *judgements* (even for matters such as whether a restaurant is near, or whether it is of good quality), reserving subjectivity for differences in *taste*, based on preferences and choice.

acceptable restaurant for Ag or propositions such as *it is desirable for Ag_1 that Ag_2 ϕ*.

Definition 2. *Let PROP be a set of atomic propositions and let $f \in PROP$ be a* factual *proposition and $p_{Ag} \in PROP$ be a proposition based on preferences. Let $KB_F^{Ag} \rightsquigarrow f$ denote that a factual proposition can be defeasibly shown, using an appropriate reasoning engine[3], without using preference information. KB_P^{Ag} records the preferences of Ag using clauses with preference based propositions as heads and bodies comprising factual propositions. Now $KB_F^{Ag} \cup KB_P^{Ag} \rightsquigarrow p_{Ag}$ denotes that the preferences of Ag are required for p_{Ag} to be shown.*

This reflects that a value judgement such as *Rolls Royce make the best cars*, cannot be considered true *simpliciter*, but is true (or false) relative to an agent, and determined using that agent's individual preferences. Next we may need to distinguish between the knowledge base of an agent at the start of the dialogue, and the knowledge base of that agent at the end of the dialogue[4].

Definition 3. *Let the knowledge base of an agent Ag at the start of a dialogue be denoted by KB^{Ag_0}, and its knowledge base after n steps of dialogue be denoted by KB^{Ag_n}.*

We now introduce some further definitions needed for the protocol and strategy. Firstly, agents need to be able to discuss the options under consideration within the dialogue and the attributes of these options. For example, in a dialogue about where to go on holiday, the options might cover different countries such as France and Spain with relevant attributes being the food and the weather. In the restaurant example the various local restaurants will be the options and various features such as distance, quality, type of cuisine, ambience and the like are attributes of restaurants that agents may use as reasons to choose between them. Furthermore, agents can give individual weightings to these attributes to reflect their different tastes and aspirations.

Definition 4. *Let O be a set of options that agents can propose during the course of a dialogue. O has an* associated set of attributes A_O. An agent associates each $a_j \in A_O$ with a weight $w_j \geq 0$ to form a set of weights W. Each set W is individual to an agent.*

Next we need to be able to determine the truth value of attributes of options, e.g. stating that the option Spain does indeed have the attribute of good weather, or that the Mogul Palace serves Indian food. We also need to be able to determine the weight that an agent assigns to an attribute of an option. The following two functions enable the above.

Definition 5. *Let $\tau : O \times A_O \rightarrow \{0, 1\}$ be a truth function that returns the truth value given by $\tau(O_i, a_j) = \tau_{ij}$ for option $O_i \in O$ and attribute $a_j \in A_O$.*

Definition 6. *Let $w : AG \times A_O \rightarrow \mathbb{N} \cup \{0\}$ be a weight function that returns the weight $w(Ag, a_j) = w_{Ag}(a_j)$ of an attribute a_j for agent $Ag \in A_O$. Where the agent Ag is clear from the context, we use w_j.*

Next we introduce notation to enable us to refer to sets of attributes of options as determined by their truth status (as will be required in section 4). Thus, attributes of options will fall within one the following disjoint sets: *verified true*, *verified false*, *unverified true* and *unverified false*, where verified options have been the subject of an inquiry dialogue (and so have been agreed by the agents) and unverified have not (and so could change in the light of information that could be elicited by an inquiry dialogue).

Definition 7. *$\tau_{ij} = 1$ if option O_i has attribute a_j. If this attribute for this option has been the subject of an inquiry dialogue, τ_{ij} has been verified. Attributes of options for which $\tau_{ij} = 1$ has been verified form a set VT_{ag}; those for which $\tau_{ij} = 0$ has been verified form a set VF_{ag}, for both agents. For a particular agent, if for an unverified attribute $KB_F^{Ag} \rightsquigarrow \tau_{ij} = 1$, the attribute is unverified true and so an element of UT_{ag}: otherwise the attribute is unverified false and so is an element of UF_{ag}. $VT_{ag} \cup UT_{ag} \cup VF_{ag} \cup UF_{ag} = A_O$.*

Finally, we are able to define the *utility* of an option for an agent, based on attributes of the option that are true.

Definition 8. *Let $A \subseteq A_O$ be a set of attributes true of O_i. Then the utility of O_i for an Agent Ag with respect to these attributes, $U_i(A)$, is $\sum_{a_j \in A} w_{Ag}(a_j)$. Now the current utility is $U_i(VT_i \cup UT_i)$. This we will sometimes abbreviate to U_i when there is no ambiguity. The maximum utility is $U_i(VT_i \cup UT_i \cup UF_i)$ and the minimum utility is $U_i(VT_i)$.*

---

[3](Black and Hunter, 2009) use Defeasible Logic (García and Simari, 2004).

[4]Here only KB_F^{Ag} changes during a dialogue. It is of course possible that an agent may be persuaded to change its preferences, but we will not consider this higher order persuasion in this paper. This is not a straightforward matter of simply adding to the existing KB_P, as is the case with facts.

2.2 Inquiry and Persuasion Dialogues

We can now characterise the distinction between inquiry and persuasion. Suppose we have two agents, Wilma and Bert, so that $AG = \{W, B\}$: following the conventions of chess, W will initiate the dialogues. We may say that for an inquiry dialogue between Wilma and Bert concerning a proposition f (inquiry dialogues concern *only* factual propositions), the following two conditions should hold:

I1 Initially either Wilma does not believe that f: $KB_F^{W_0} \not\leadsto f$ or Bert does not believe that f : $KB_F^{B_0} \not\leadsto f$ or both.

I2 At the end of the dialogue both believe that f if and only if f is a consequence of the joint knowledge bases: $(KB_F^{W_1} \leadsto f) \land (KB_F^{B_1} \leadsto f) \leftrightarrow KB_F^{W_0} \cup KB_F^{B_0} \leadsto f$.

An inquiry dialogue will always result in agreement, since, if agreement does not arise before, the dialogue will terminate only after Bert and Wilma have exchanged all their knowledge, so that, if the dialogue has taken n steps, $KB_F^{W_n} = KB_F^{B_n} = KB_F^{W_0} \cup KB_F^{B_0}$. Some have argued that neither should believe that f at the outset, but we wish to allow Wilma to start an inquiry dialogue to confirm her beliefs, since, given the defeasible reasoning mechanism we are using, it is always possible that Bert may supply information resulting in Wilma revising her beliefs.

The second condition is plausible (since we are assuming that factual propositions are objectively assessed by the agents) and is the condition used to show soundness and completeness of the inquiry dialogues in Black and Hunter (Black and Hunter, 2009). Since that paper shows soundness and completeness for their inquiry dialogues, we shall suppose that where Wilma and Bert wish to establish the truth of some factual proposition, they will use a dialogue as described there.

In contrast, persuasion can concern matters with both directions of fit, although probably evaluative propositions are more usually the topic than factual ones. This is because if the persuader is correct (and honest), for factual topics an inquiry dialogue will serve to achieve the goals of a persuasion dialogue. It may be, however, that the persuader has some interest in establishing the proposition, and this interest will persist even if it becomes aware of information suggesting that the proposition is in fact false. If one of the agents is a salesperson, for example, this will be what it will do. Since persuasion may have this adversarial element we distinguish *open persuasion*, where the persuader is expected to tell the truth, the whole truth and nothing but the truth, and *partial[5] persuasion* where the persuader must tell only the truth and nothing but the truth, but need not tell the whole truth if that would damage its cause. Open persuasion about a fact is simply an inquiry dialogue where the persuader initially believes the proposition under consideration, so that these dialogues can also be characterised by I1 and I2 above. We will henceforward confine ourselves to persuasion dialogues that concern matters of preference, so that there will subjective elements dependent of tastes and preferences as well as fact.

Open Persuasion

OP1: Wilma believes that Bert does not believe that p_B: $KB^{B_0} \not\leadsto p_B$

OP2: Both wish that at the end of the dialogue Bert believes that p_B if and only if p_B is a consequence of their joint knowledge bases and Bert's preferences: $KB^{B_t} \leadsto p_B \leftrightarrow KB_F^{W_0} \cup KB^{B_0} \leadsto p_B$, where the dialogue terminates in t steps.

Note that at the end of an open dispute, Bert and Wilma may differ as to whether p or $\neg p$, since their beliefs reflect their own individual preferences applied to the shared information. Again the dialogue will always terminate because either the agent is persuaded, or all factual knowledge has been exchanged.

For partial persuasion we include the desire to, as it were, win the dialogue, irrespective of the truth of the matter. Although partial persuasion can be conducted regarding a matter of fact, we will consider here only persuasion relating to matters involving elements of preference. The initial goal remains OP1, and Bert's final goal remains OP2, but Wilma has a different goal, PP3. PP4 represents the condition under which Wilma can legitimately satisfy her goal. Again the dialogue terminates in t steps.

Partial Persuasion Using Preferences

PP3: Wilma's goal is that Bert should believe that p: $KB^{B_t} \leadsto p_B$

PP4: Wilma can succeed if she has information to enable Bert to believe p: $KB^{B_t} \leadsto p_B \leftrightarrow (\exists S) \subseteq KB_F^{W_0}$ such that $S \cup KB^{B_0} \leadsto p_B$.

Note that open persuasion begins with a conflict, but has the same goals for both participants, whereas in partial persuasion they have different views as to what constitutes a successful termination. Both agents can realise their individual goals if p_B does follow from their collective knowledge and Bert's preferences. If not, strategic considerations may affect the

---

[5]*partial* as in *biased, not impartial*, rather than as in incomplete.

outcome: it may be that there is a subset of Wilma's KB which could be used to persuade Bert, but that she reveals too much, so that Bert is unpersuaded. Here Wilma could have succeeded according to PP4, but in fact Bert is rightly unpersuaded, satisfying OP2.

Having made these distinctions we will now focus on a particular type of persuasion dialogue, open persuasion involving preferences. This is the simplest kind of dialogue after inquiry dialogues, in that only one side is doing the persuasion, and the participants have common goals. Our particular scenario is the common situation where an agent is seeking a recommendation, or comparing options, concerning things such as restaurants, cars, digital cameras, insurance, or any other product where competing options have some out of a large number of features which vary in importance for different agents. The persuading agent will want what is best for the other agent, and will have no particular interest in having its recommendation accepted. The particular topic we will consider is *what is the best restaurant for Bert*?

3 EXAMPLE

For our example we will spell out choice of restaurant in detail. Let us suppose that Wilma and Bert are standing outside Burger World. Bert can see that Burger World is close by and appears cheap, but as a stranger to the town knows nothing about any other restaurant. Wilma, on the other hand, as a native to the city, has complete information. So, initially, Bert finds Burger World acceptable but Wilma will attempt to persuade him of the merits of the Thai Palace. Moreover suppose Bert wants a good quality restaurant, that is cheap, close by and licenced. Bert weights these attributes 6,2,1 and 2 respectively. All other features of restaurants, such as whether they have music, and the type of cuisine, are matters on which Bert is indifferent, and so have weights of 0, and need not be considered. We can summarise the situation:

Set of Agents AG = {Wilma, Bert}. Wilma is the persuader and Bert is the persuadee.

Set of Options O = {BurgerWorld, ThaiPalace} = {O_1 , O_2}

Set of Attributes A_O = {goodQuality, cheap, close, licenced} = {a_1, a_2, a_3, a_4}

Sets of weights. $W_{Wilma}(a_i)$ = {6,4,0,2}; $W_{Bert}(a_i)$ = {6,2,1,2}; for i = 1,2,3,4.

Truth values τ_{ij}. Burger world (j = 1) {0,1,1,0} (cheap and close). Thai Palace (j = 2) {1,1,0,1}: (good quality, cheap and licenced).

Each agent, for each option, partitions A_O into four subsets, depending on its own knowledge base. Attributes that are not known and have not yet been the subject of an inquiry are *unverified*, while those that have been the subject of an inquiry are *verified*, and the agents are in agreement as to them.[6] So at the start of the dialogue,

Bert has UT_1 = {$cheap, close$}, UF_1 = {$good, licenced$}, VF_1 = VT_1 = $\emptyset$ for Burger World and

UF_2 = {$good, cheap, close, licenced$}, UT_2 = $VF_2 = VT_2 = \emptyset$ for the Thai Palace.

As the dialogue progresses, inquiries regarding attributes are made, and these attributes will move from unverified to verified. Bert's utility calculations for Burger World at the start of the dialogue are shown in Figure 1.

W1 I would go to the Thai palace. *Wilma starts the persuasion dialogue by making a recommendation.*

B1 Burger World is right here and it looks cheap. *Bert indicates two criteria which he values and which he believes are satisfied by Burger World*

W2 The Thai Palace is also cheap, but it is a walk away. *Wilma supplies information about the criterion satisfied by the Thai Palace. Bert increases the current and minimum utilities of the Thai Palace to 3.*

B2 Is Burger World good? *Bert seeks information about another valued criterion.*

W3 No. But the Thai Palace is. *Wilma indicates a point in favour of the Thai Palace. Bert must now adjust his utilities: while the current utility of Burger World remains 3, the maximum falls to 5. But the minimum utility of the Thai Palace is now 6, and so cannot be bettered by Burger World.*

B3 OK. *Bert now has sufficient information about both restaurants for the criteria he values: he does not ask about licensing because that can no longer change the order for him.*

At the end of the dialogue:

Bert has UT_1 = $\emptyset, UF_1$ = {$licenced$}, VF_1 = {$goodQuality$}, VT_1 = {$cheap, close$} for Burger World and

UT_2 = $\emptyset, UF_2$ = {$licenced$}, VF_2 = {$close$}, VT_2 = {$cheap, goodQuality$} for the Thai Palace.

The utility calculations for Bert and Burger World at the end of the dialogue are shown in Figure 1.

---

[6]If an attribute cannot be shown true, it is considered to be false since the defeasible reasoner uses negation as failure.

Calculation of Utilities for Bert for Burger World at the start of the dialogue:

Current Utility $= U_1(UT_1 \cup VT_1)) = U_1(\{cheap, close\}) = w_B(a_2) + w_B(a_3) = 2 + 1 = 3$.
Minimum Utility $= U_1(VT_1)) = 0$
Maximum Utility $= U_1(UT_1 \cup VT_1 \cup UF_1)) = U_1(\{goodQuality, cheap, close, licenced\}) = 6 + 2 + 1 + 2 = 11$.

Calculation of Utilities for Bert for Burger World at the end of the dialogue:

Current Utility $= U_1(UT_1 \cup VT_1)) = U_1(\{cheap, close\}) = w_B(a_2) + w_B(a_3) = 2 + 1 = 3$.
Minimum Utility $= U_1(VT_1)) = 3$
Maximum Utility $= U_1(UT_1 \cup VT_1 \cup UF_1)) = U_1(\{cheap, close, licenced\}) = 2 + 1 + 2 = 5$.

Figure 1: Utility Calculations for Bert.

This is a fairly efficient dialogue: restaurants have many attributes and so Wilma could have told Bert many things he did not know, whereas the dialogue is able to conclude after Bert has received just six items of information. How is this possible? It is because Wilma is able to infer things about Bert's criteria and current knowledge beyond what Bert explicitly states and asks, and so can recognise what will be relevant to Bert's opinion. This is Grice's notion of *conversational implicature*. In (Grice, 1975), Grice advanced four maxims intended to capture the pragmatics of utterances in cooperative dialogues. The maxims express the cooperation principle:

> Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged,

which arguably must be observed if misunderstandings and conversational breakdown are to be avoided. The four maxims relate to Quality (statements must be believed on adequate evidence), Quantity (contributions should be as informative as the situation requires, and no more), Relevance (which may be dependent on the dialogue type and its state) and Manner (contributions should be clear and unambiguous). Of these maxims, relevance will be of particular concern to us here.

Thus Bert begins in B1 by stating that Burger World satisfies two criteria. Burger World is not chosen at random from the options of which Bert is aware. It is Bert's currently preferred option, the one which the Thai Palace must overcome. Nor are the criteria just any criteria, but criteria which Bert believes Burger World does, and the Thai Palace does not, satisfy. Thus in Wilma's reply she can improve her case by stating that one of the criteria is satisfied by the Thai Palace. Note here, that once a criterion is put into play, it is considered for both options, thus verifying any existing beliefs Bert may have about the options. In B2 Bert asks a question. This again is not chosen at random but concerns a criterion which, if satisfied would put Burger World beyond reach. Wilma can, however, truthfully say that it does not satisfy this criterion, but that the Thai Palace does. Bert now draws the dialogue to a close since he has sufficient knowledge of the criteria relating to the two options relevant to him, according to his preferences. He does not ask about licencing because that can make no difference to his choice, even if Burger World is licenced and the Thai Palace is not. Note that, in this dialogue features of the Thai Palace that make it attractive to Wilma are not even mentioned, while she supplies information about a feature she is indifferent to: Bert is the sole arbiter of what makes the restaurant good (in complete contrast to deliberation). Bert could not ask about another criterion without misleading Wilma by *conversationally implying* that he valued a fourth criterion enough to overturn his current view, which would prolong the dialogue to no useful purpose. Note also the asymmetry in the use of the criteria of the two participants. Wilma's criteria may determine her recommendation, but play no part in the dialogue unless they are shown to be valued by Bert.

The dialogue is not, however, as efficient as it might be. Given his weights, if Bert discovered that either restaurant was, and the other was not, of good quality, he could stop immediately, since this criterion carries more weight than the other three combined. Thus Bert's best initial question would have been B2, with Wilma's reply in W3 enough to resolve the discussion. We could extend our notion of conversational implicature to suggest that mentioning a criterion means not only that it matters to the persuadee but additionally that there are no more important criteria not yet mentioned. This is what we will do in the algorithm developed in the next section. As we will see in section 5, this algorithm, based on conversational implicature, is optimal in terms of minimising the expected exchange of information.

Note also that Bert's strategy is good only with regard to open persuasion. Had this been a partial persuasion situation, Bert's question in B2 could only have received the response that the Thai Palace was good. Since Wilma is not obliged to tell the whole truth, even if she had believed that the Burger World

was good, she would have remained silent on that point, and indeed on all positive features of Burger World. The question is answered only if the answer does not improve Bert's assessment of that restaurant. In such a case there is no point in mentioning Bert's preferred option, except to check that what he believes is satisfied is indeed satisfied. In a partial persuasion where Bert trusts his current knowledge, therefore, Bert will ask only whether the Thai Palace is good. If Bert does not know some relevant facts about Burger World, he should attribute some kind of expected utility to them. In this case discovery that the Thai Palace was of good quality might not immediately displace Burger World, and the question as to price would still be important.

Beneath the surface of the dialogue, the pragmatic meaning of the utterances in the context of persuasion is that Bert is asking a sequence of questions about whether attributes that he values enough to be able to change his current preference are satisfied by the proposed option and (in open persuasion) by his currently preferred option. Wilma's role is to answer these questions. We suggest also that Bert should ask about criteria in order of their importance to him. We now develop an algorithm based on this principle and show that this represents an optimal strategy in section 5. Thus following Grice's maxims, and taking the pragmatics of the dialogue type seriously, has computational benefits.

4 PROTOCOL AND STRATEGY

The persuasion dialogue, (W1-B3 in the example), once it has been initiated by the persuader, is a series of statements and requests for information on the part of the persuadee. Either these will be a statement of the form $\tau_{1j} = 1$, or a question of the form whether $\tau_{1j} = 1$. In either case, in the terms of (Black and Hunter, 2009), the persuadee wishes to open inquiry dialogues with τ_{1j} and τ_{2j} as topics (in open persuasion) or with τ_{2j} as the only topic (in partial persuasion).

This can be accomplished using a series of dialogues designed for inquiry such as that of (Black and Hunter, 2009). Some difference in strategy within the inquiry dialogue may be required for partial persuasion, but the inquiry dialogue of (Black and Hunter, 2009), which has been shown there to be sound and compete, offers a means of ensuring that Bert will gain possession of the answer dictated by the joint knowledge of the two agents, which is the best information available. We will discuss the nature of the inquiry dialogues no further. When should the per-

suasion dialogue terminate? If O_1 is Bert's currently best option, and Wilma is trying to persuade him of O_2, Bert should continue to seek information until the minimum utility of O_1 is greater than the maximum utility of O_2, in which case Wilma's proposal can be rejected, or until minimum utility of O_2 exceeds the maximum utility of O_1, in which case her proposal can be accepted. This state will be approached by initiating inquiry dialogues to verify as yet unverified features of the options. When $UT_i = UF_i = \emptyset$, the dialogue must terminate, since the option with the highest current utility cannot be displaced. Often, however, it will terminate with only a subset of the attributes verified: in the example above it may terminate if good quality is true of one but not the other: if true of both or neither it could terminate for Wilma if cheap is true of one but not the other, and she will consider whether they are licenced only if the options are the same in terms of quality and price.

Definition 9. *Termination condition. Let $\mathcal{T}_1(X)$ be the condition that the algorithm has terminated with O_1 preferred when the status of all the attributes associated with the weights in $X \subseteq W$ are known to the persuadee.*

$\mathcal{T}_1(X)$ *holds if and only if*
$$\sum_{w_j \in X} \tau_{1j} w_j - \sum_{w_j \in X} \tau_{2j} w_j > \sum_{w_j \in W \setminus X} w_j.$$
Similarly for $\mathcal{T}_2(X)$.

The agent will attempt to resolve the dialogue in as few steps as possible, and so should choose the topics so as to minimise the expected number of steps. From our observations about what is conversationally implied in persuasion dialogues, we conjecture that an optimal strategy, S for an agent would be to inquire about attributes in descending order of their weights, and so employing a kind of greedy algorithm[7].

We now present the protocol and embedded strategy S for our persuasion dialogue. We require that agents can *open* and *close* the relevant dialogues, *propose* options and *inquire* about the status of attributes, for which they may use inquiry dialogues as defined in (Black and Hunter, 2009).

A persuasion dialogue follows the following protocol:

[0]: Wilma opens by proposing an option, O_2. If

---

[7]Note that this is a different problem from that considered in (Dunne and McBurney, 2003). That paper concerned the *actual* number of *locutions* in persuasion dialogues according to their rather different protocol, whereas we are concerned with the information exchanged, i.e. the *expected* number of inquiry *dialogues*: in our approach the total number of locutions will therefore depend on the protocol employed for inquiry.

Bert's currently preferred option is O_1 a persuasion dialogue will commence; otherwise he agrees immediately. Initially $A = X = \emptyset$; $W =$ the set of weights of all attributes about which Bert may inquire.

[1]: Bert opens inquiry dialogues with topics τ_{1j}, τ_{2j} for some attribute a_j with which Bert associates a positive weight w_j; A becomes $A \cup \{a_j\}$; X becomes $X \cup \{w_j\}$; increment the utilities U_i : $i \in \{1,2\}$ by $\tau_{ij}w_j$. This may change which option is currently preferred.

[2]: If $\mathcal{T}_1(X)$ holds terminate with O_1 preferred, else if $\mathcal{T}_2(X)$ hold terminate with O_2 preferred else if $X = W$, return the currently preferred option; else go to [1].

The inquiry about a single attribute in [1] and [2] will be termed a *step* in the remainder of this paper. Moves are subscripted with W or B depending on whether the move is made by the agent acting as the persuader (W) or persuadee (B). There are two options O_1 and O_2. Thus the dialogue begins with W proposing O_2. Now B either agrees, or states its preferred option and inquires about some attribute. Since only B's weights matter, w_j will refer to the weight given to a_j by B.

open dialogue$_W$
propose$_W$(O_2)
if *agree$_B$(O_2))*
    $\mathcal{T}_2(\emptyset)$, *end dialogue$_B$*
else
    $X = \emptyset; A = \emptyset$
    propose$_B$(O_1), such that $O_2 \neq O_1$
    open persuasion dialogue$_B$
    for all $a_j \in A_O$ **do**
        sort A_O into ordered descending list such that
        $w_j \geq w_{j+1}$
    end for
    $j=1$
    while $\neg\mathcal{T}_1(X)$ and $\neg\mathcal{T}_2(X)$ **do**
        inquire$_B$(τ_{1j}, τ_{2j})
        A becomes $A \cup \{a_j\}$;
        X becomes $X \cup \{w_j\}$;
        increment the utilities $U_i : i \in \{1,2\}$ by $\tau_{ij}w_j$.
        if $\tau_{1j} = 1$ **then** $a_j \in VT_1$ **else** $a_j \in VF_1$
        end if
        if $\tau_{2j} = 1$ **then** $a_j \in VT_2$ **else** $a_j \in VF_2$
        end if
        j++
    end while
    if $\mathcal{T}_1(X)$ **then** *disagree$_B$(O_2)* **else** *agree$_B$(O_2)*
    end if
end if
end dialogue$_W$

The algorithm given above is for open persuasion; for partial persuasion the inquiries will concern τ_{2j} only. As noted above, each pass through the while loop is referred to as a *step* in the remainder of the paper. Each step establishes the truth values for one attribute.

5 RESULTS

In this section we describe a series of results that show our strategy $\mathcal{S}$ to be an optimal strategy for a number of representative situations. The first of these is where the persuadee has no initial opinions as to the facts: the attributes of the options will be discovered from the dialogue, and are all considered equally likely at the outset. Theorem 1 shows that $\mathcal{S}$ is optimal for this scenario[8].

Theorem 1. Suppose we have two options O_1 and O_2 of equal prior utility, m attributes, which are equally likely to be true or false for each option, and m positive weights assigned to the attributes. Then an optimal strategy, in the sense that the algorithm terminates in the expected fewest number of steps, is to inquire about the attributes in descending order of weight (strategy $\mathcal{S}$). First we prove two preliminary lemmas.

Lemma 1. Given a set A of n attributes, the probability that the algorithm will terminate in not more than n steps is independent of the order in which we inquire about these attributes. We term this probability $P(X)$ where $X = \{w_{Ag}(a_j) : a_j \in A\}$.

Since we are concerned with the opinions of only one agent (the persuadee), we shall use the simplified notation w_j for $w_{Ag}(a_j)$.

Proof of Lemma 1. Let $X = \{w_1, ...w_n\} \subseteq W$ be the set of weights associated with the attributes examined. If $\mathcal{T}_1(X)$ holds, the algorithm terminates in at most n steps, by Definition 9.

Conversely if O_1 is determined in $n_1 \leq n$ steps after considering $X_1 \subseteq X$ then $\mathcal{T}_1(X_1)$ holds.

Subtracting $\sum\limits_{w_j \in X \setminus X_1} w_j$ from both sides of $\mathcal{T}_1(X_1)$ gives

[Inequality 1]. $\sum\limits_{w_j \in X_1} \tau_{1j}w_j - \sum\limits_{w_j \in X_1} \tau_{2j}w_j - \sum\limits_{w_j \in X \setminus X_1} w_j$

$> \sum\limits_{w_j \in W \setminus X} w_j,$

since $(W \setminus X_1) \setminus (X \setminus X_1) = W \setminus X$.

---

[8]We would like to thank Michael Bench-Capon for his insights regarding the proof strategy used in this section.

We aim to deduce that $\mathcal{T}_1(X)$ holds.

LHS of $\mathcal{T}_1(X) = \sum_{w_j \in X} \tau_{1j} w_j - \sum_{w_j \in X} \tau_{2j} w_j$

$\geq \sum_{w_j \in X_1} \tau_{1j} w_j - \left(\sum_{w_j \in X_1} \tau_{2j} w_j + \sum_{w_j \in X \setminus X_1} \tau_{2j} w_j \right)$

$\geq \sum_{w_j \in X_1} \tau_{1j} w_j - \sum_{w_j \in X_1} \tau_{2j} w_j - \sum_{w_j \in X \setminus X_1} w_j$ since

$\tau_{2j} \leq 1$

$> \sum_{w_j \in W \setminus X} w_j$ by Inequality 1

$=$ RHS of $\mathcal{T}_1(X)$, so $\mathcal{T}_1(X)$ holds.

Hence the algorithm returns O_1 in no more than n steps (by discovering the status of some or all elements of A) if and only if $\mathcal{T}_1(X)$ holds. Similarly for O_2. Hence the probability $P(X)$ is independent of the order in which the elements of X are considered. $\square$

Lemma 2. Let X, Y be n-element subsets of W which differ only in one element: $X = (X \cap Y) \cup \{w_x\}$ and $Y = (X \cap Y) \cup \{w_y\}$ with $w_x > w_y$. Then $P(X) \geq P(Y)$ where $P(X), P(Y)$ are as defined in Lemma 1.

Proof of Lemma 2. Let $X = \{w_1, ... w_{n-1}, w_x\}$, $Y = \{w_1, ... w_{n-1}, w_y\}$ where $w_x > w_y$. By Lemma 1, $P(X)$ and $P(Y)$ are well defined. We write TT, TF, FT, FF for the 4 possible values of $\langle \tau_{1j}, \tau_{2j} \rangle$. For each set X, Y, there are 4^n possible such assignments of truth values, all equally likely by our hypothesis. For example, when $n = 3$, one assignment is $\langle TT, FT, TT \rangle$, indicating the first and third attributes are true for both O_1 and O_2, but the second is true only for O_2. If the algorithm terminates for r out of the 4^n assignments, the probability of termination in at most n steps is $r/4^n$.

Suppose $\mathcal{T}_1(Y)$ holds for a particular assignment: we will show that it follows that $\mathcal{T}_1(X)$ holds for that assignment.

$\sum_{w_j \in X} \tau_{1j} w_j - \sum_{w_j \in X} \tau_{2j} w_j$

$= \sum_{w_j \in Y} (\tau_{1j} - \tau_{2j}) w_j + (\tau_{1j} - \tau_{2j})(w_x - w_y)$

$\geq \sum_{w_j \in Y} (\tau_{1j} - \tau_{2j}) w_j - (w_x - w_y)$

since $\tau_{1j} - \tau_{2j} \in \{-1, 0, 1\}$ and $w_x > w_y$.

$> \left(\sum_{w_j \in W \setminus Y} w_j \right) - (w_x + w_y)$ by $\mathcal{T}_1(Y)$.

$= \sum_{w_j \in W \setminus X} w_j$ so that $\mathcal{T}_1(X)$ holds.

Hence the number r_X of assignments for which the algorithm terminates in not more than n steps is at least r_Y.

$P(X) = r_X / 4^n$ and $P(Y) = r_Y / 4^n$ so that $P(X) \geq P(Y)$ as required. $\square$

Proof of Theorem 1. Suppose there exists a strategy $\mathcal{R}$ better than $\mathcal{S}$. Then there exists $n \in \mathbb{N}$ such that $\mathcal{R}$

is more likely to terminate in at most n steps than $\mathcal{S}$.

Let $X_\mathcal{S} = \{w_1, w_2, ..., w_n\}$, be the set of the largest n weights and let $X_\mathcal{R}$ be the set of n weights used by $\mathcal{R}$.

We construct a chain of n-element sets $X_\mathcal{S}$, $X_1, X_2, ... X_k, X_\mathcal{R}$ so that each set X and its successor Y satisfy the conditions of Lemma 2.

We retain elements of $X \cap Y$ and replace others in turn, substituting the largest in Y for the largest in X and continuing in descending order. For example, if $X = \{9,8,6,5\}$ and $Y = \{8,4,3,1\}$, the first intermediate sets are $\{4,8,6,5\}$, $\{4,8,3,5\}$.

Since $X_\mathcal{S}$ contains the largest weights, $w_x > w_y$ is satisfied for each pair in the chain. By Lemma 2, for each pair $\langle X, Y \rangle$, $P(X) \geq P(Y)$. Hence $P(X_\mathcal{S}) \geq P(X_\mathcal{R})$, a contradiction. So no such better strategy exists, and we conclude that $\mathcal{S}$ is optimal. $\square$

Suppose we relax the assumption that the persuadee knows nothing about the options in the initial situation, and instead starts with a set of initial (perhaps unverified) beliefs that lead to a preference for one of the options, but where there still remain attributes whose value for the options is unknown. This was the case for Bert in the Burger World example above. Corollary 1 demonstrates that $\mathcal{S}$ is an optimal strategy in this case also.

Corollary 1. If the utilities of O_1 and O_2 are initially unequal, $\mathcal{S}$ remains an optimal strategy.

Proof of Corollary 1. Let the current utilities be U_1, U_2 with $U_1 > U_2$ and $W = \{w_1 .., w_m\}$ be the set of weights.

Suppose, for contradiction, that a better strategy $\mathcal{R}$ than $\mathcal{S}$ exists, expected to terminate in $n(\mathcal{R})$ steps with $n(\mathcal{R}) < n(\mathcal{S})$. Let L be a number greater than any weight in W.

Suppose there were additional attributes b_1, b_2 with weights $L + U_1, L + U_2$. Now initiate a dialogue with weights $W \cup \{L + U_1, L + U_2\}$ and the initial utilities zero for both options.

Apply the following strategy.

Step 1: Inquire about b_1.

Step 2: Inquire about b_2.

After Step 2: if Step 1 assigns TF and Step 2 assigns FT, (difference between utilities is $U_1 - U_2$), continue as for $\mathcal{R}$, otherwise continue in descending order.

For this strategy the expected number of steps is $2 + 1/4 n(\mathcal{R}) + 3/4 n(\mathcal{S})$. The expected number of steps for the descending order strategy is $2 + n(\mathcal{S})$ which is greater, contradicting Theorem 1. So no such strategy $\mathcal{R}$ exists. $\square$

Next consider partial persuasion, where nothing improving the current estimated utility of the pre-

ferred option can be learned, since the persuader will choose to remain silent. Now all inquiry will take place concerning the option advocated by the persuader. That S remains an optimal strategy in the case is established by Theorem 2.

Theorem 2. If τ_{1j} is known in advance for all attributes a_j, S is an optimal strategy for inquiring about the τ_{2j}.

Sketch of Proof of Theorem 2. In this case U_1 is constant throughout. After examining attributes with weights in $X \subseteq W$, $U_2 = \sum_{w_j \in X} \tau_{2j} w_j$.

The algorithm terminates when **either** $U_2 > U_1$ **or** $U_1 - U_2 > \sum_{w_j \in W \setminus X} w_j = \sum_{w_j \in W} w_j - \sum_{w_j \in X} w_j$.

U_2 increases monotonically.

The expected value of U_2 is $\frac{1}{2} \sum_{w_j \in X} w_j$ so the algorithm is expected to terminate when

either $\frac{1}{2} \sum_{w_j \in X} w_j > U_1$

or $\frac{1}{2} \sum_{w_j \in X} w_j > \sum_{w_j \in W} w_j - U_1$.

The RHS of each inequality is constant, so the best strategy is to maximise the LHS at each step, that is to choose the largest remaining weight. But this is strategy S as in Theorem 1. $\square$

Finally we consider the case where the persuadee, agent B, has initial beliefs about attributes relating to both options, but has varying degrees of confidence in these beliefs, and wishes to confirm them during the dialogue. Theorem 3 shows S to be an optimal strategy on the assumption that the currently preferred option is at least as likely as the alternative proposed by the persuader to satisfy the criteria valued by the persuadee. This will be the case where the persuadee is quite sure that its preferred option satisfies some desirable attributes, but has only tenuous beliefs about the other option.

Theorem 3. S remains an optimal strategy if the probabilities p_{ij} are not equal, provided that $p_{1j} \geq p_{2j}$ for each attribute j, where $p_{ij} = P(\tau_{ij} = 1)$.

Sketch of Proof of Theorem 3. After examining attributes with weights in $X \subseteq W$, (expected value of U_i) $= \sum_{w_j \in X} p_{ij} w_j$. So the algorithm is expected to terminate when

$$\sum_{w_j \in X} (p_{1j} - p_{2j}) w_j > \sum_{w_j \in W} w_j - \sum_{w_j \in X} w_j.$$

Arguing as before, an optimal strategy is, at each step, to choose from the remaining weights so as to maximise $w_j(p_{1j} - p_{2j} + 1)$. $\square$

6 CONCLUDING REMARKS

In this paper we have considered the distinctive features of persuasion dialogues. We have made these precise, for the very common form of persuasion dialogue where one agent is trying to convince another that its currently preferred option is not as good as some other possibility known to the persuading agent. We have also drawn attention to the pragmatic meanings of utterances in these persuasion dialogues, as revealed by considering what the utterances conversationally imply in this context. We have used these pragmatic meanings to develop a protocol and strategy for agent persuasion dialogues, and have shown the strategy to be an optimal strategy in a range of representative scenarios for these persuasion dialogues.

In future work we intend to consider dialogues with three or more participants, and dialogues which attempt to change the preferences of the participants.

REFERENCES

Bench-Capon, T. (2002). Agreeing to differ: Modelling persuasive dialogue between parties without a consensus about values. *Informal Logic*, 22(32).

Black, E. and Hunter, A. (2009). An inquiry dialogue system. *Autonomous Agents and Multi-Agent Systems*, 19(2):173–209.

Dunne, P. E. and McBurney, P. (2003). Concepts of optimal utterance in dialogue: Selection and complexity. In *Workshop on Agent Communication Languages, Springer LNCS 2922*, pages 310–328.

García, A. J. and Simari, G. R. (2004). Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming*, 4(1-2):95–138.

Grice, H. (1975). Logic and conversation. In Cole, P. and Morgan, J., editors, *Syntax and semantics, vol 3*, pages 43–58. Academic Press, New York.

McBurney, P., Hitchcock, D., and Parsons, S. (2007). The eightfold way of deliberation dialogue. *International Journal of Intelligent Systems*, 22(1):95–132.

Prakken, H. (2006). Formal systems for persuasion dialogue. *Knowledge Engineering Review*, 21(2):163–188.

Rahwan, I., Ramchurn, S. D., Jennings, N. R., McBurney, P., Parsons, S., and Sonenberg, L. (2003). Argumentation-based negotiation. *Knowledge Engineering Review*, 18(4):343–375.

Searle, J. R. (2003). *Rationality in Action*. The MIT Press, Cambridge.

Walton, D. (1998). *The New Dialectic*. University of Toronto Press, Toronto.

Walton, D. and Krabbe, E. (1995). *Commitment in dialogue: Basic concepts of interpersonal reasoning*. State University of New York Press, Albany NY.