Comfort or Safety? Gathering and Using the Concerns of a Participant for Better Persuasion

Emmanuel Hadoux and Anthony Hunter
Department of Computer Science,
University College London,
London, UK

February 1, 2019

Abstract

Persuasion is an important and yet complex aspect of human intelligence. When undertaken through dialogue, the deployment of good arguments, and therefore counterarguments, clearly has a significant effect on the ability to be successful in persuasion. A key dimension for determining whether an argument is good is the impact that it has on the concerns of the intended audience of the argument (*e.g.*, the other participant(s) in the dialogue). In this paper, we investigate how we can acquire and represent concerns of a participant, and her preferences over them, and we show how this can be used for selecting good moves in a persuasion dialogue. We provide results from empirical studies showing that: (1) we can gather preferences over types of concern; (2) there is a common understanding of what is meant by concerns; (3) participants tend to make moves according to their preferences; and (4) the persuader can use these preferences to improve the persuasiveness of a dialogue.

1 Introduction

Persuasion is an activity that involves one party trying to induce another party to believe or disbelieve something or to do (or not do) something. It is an important and complex human ability. Obviously, it is essential in commerce and politics. But, it is equally important in many aspects of daily life. Consider for example, a child asking a parent for a raise in pocket money, a doctor trying to get a patient to enter a smoking cessation programme, a charity volunteer trying to raise funds for a poverty stricken area, or a government advisor trying to get people to avoid revealing personal details online that might be exploited by fraudsters.

Arguments are a crucial part of persuasion. They may be explicit, such as in a political debate, or they may be implicit, such as in an advert. In a

dialogue involving persuasion, counterarguments also need to be taken into account. Participants may take turns in the dialogue with each of them presenting arguments, some of which may be counterarguments to previously presented arguments. So the aim of the persuader is to persuade the persuadee through this exchange of arguments. Since some arguments may be more effective than others in such a dialogue, it is valuable for the persuader to have an understanding of the persuadee and of what might work better with her.

In this paper, we consider how arguments that relate better to the priorities and concerns of a persuadee can be more effective in a persuasion dialogue. Consider for example a doctor in a university health clinic who is trying to persuade a university student to take up regular exercise, and suppose the student says that she does not want to take up a sport because she finds sports boring. The doctor then needs to find a counterargument to the student's argument. Suppose the doctor has two options:

- Option 1: Doing sport will not only help your physical health, but it will help you study better.
- Option 2: Doing sport will get you in shape, and also help you make new friends.

The argument for Option 1 concerns physical health and getting a good degree, whereas the argument for Option 2 concerns physical health and social life. Now suppose the doctor has learnt through the conversation that the student does not prioritize physical health at all, ranks social life somewhat highly, and ranks getting a good degree very highly. In this case, the doctor will regard the argument in Option 1 as being a better counterargument to present to the student, since it appears to have a better chance of convincing the student.

Whilst this is a simple and intuitive idea, there is a lack of a general framework for using concerns in making strategic choices of move in the way suggested by the above example. The notion of a concern seems to be similar to the notion of a value. Often a value is considered as a moral or ethical principle that is promoted by an argument, though the notion of a value promoted by an argument can be more diverse and used to capture general goals of an agent [Atk06]. Values have been used in a version of abstract argumentation called value-based argumentation frameworks (VAFs) [BC03, BCAC05, BCA09], and furthermore, they have been considered in models of dialogical argumentation for persuasion [Ben02]. In VAFs, values are used for ignoring attacks by counterarguments where the value of the attacking argument is lower ranked than the attacked one. Thus, the role of values in VAFs is different to the role of concerns as suggested in the example above where the doctor chooses the argument of greater concern to the patient.

In this paper, we investigate how taking into account the concerns of a participant can be used in making strategic choices of move, and show how this can be incorporated in argument-based software for persuasion. We focus on the use of persuasion to change the belief in some goal argument (i.e., an

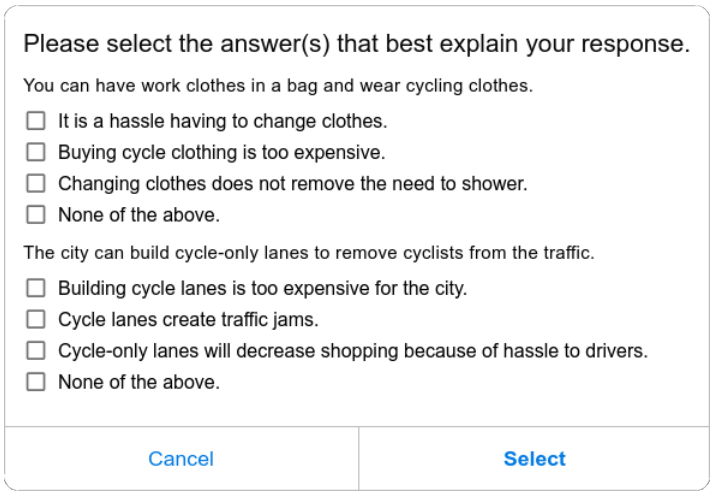


Figure 1: Interface for an asymmetric dialogue move for asking the user's counterarguments. In this example, the user is asked about two arguments with four possible counterarguments per argument.

argument that we want the persuadee to believe). A goal argument might reflect an "intention" to do something (as illustrated in the examples above) but this is not mandatory. For example, it could be that we want the persuadee to believe a particular explanation for a past event.

1.1 Background on automated persuasion systems

We assume that an automated persuasion system (APS) is a software application running on a desktop or mobile device. It aims to use convincing arguments in a dialogue in order to persuade the persuadee (the *user* of the system) to accept some persuasion goal (for example, to believe a specific argument) [Hun16a]. The dialogue may involve moves such as queries, claims, and importantly, posits of arguments and counterarguments, that are presented according to some protocol. The protocol specifies which moves are permitted or obligatory at each dialogue step. The dialogue may involve stages where the system finds out more about the persuadee. For instance, the system can ask questions about her beliefs. The system also needs to handle objections or doubts (represented by counterarguments) with the aim of convincing the user to accept the persuasion goal (*i.e.*, the argument that encapsulates the reason for a change of behaviour).

The dialogue may be asymmetric since the kinds of moves that the APS can present may be different to the moves that the persuadee may make. For instance, the persuadee might be restricted to only making arguments by selecting them from a menu in order to obviate the need for natural language processing of arguments being entered (as illustrated in Figure 1). In the extreme, it may

be that only the APS can make moves.

Whether an argument is convincing or not depends on the context of the dialogue and on the characteristics of the persuadee. An APS may maintain a model of the persuadee. This may be used to predict what arguments and counterarguments the persuadee knows about and/or believes. This can be harnessed by the strategy of the APS in order to choose good moves to make in the dialogue.

In our previous work on developing APSs, we have primarily focussed on beliefs in arguments as being the key aspect of a user model for making good choices of move in a dialogue. To represent and reason with beliefs in arguments, we can use the epistemic approach to probabilistic argumentation [Thi12, Hun13, BGV14, HT16, PHT17, HPT18, HPP18] which has been supported by experiments with participants [PH18]. In applying this approach to modelling a persuadee's beliefs in arguments, we have developed methods for: (1) updating beliefs during a dialogue [Hun15, Hun16b, HP17b]; (2) efficiently representing and reasoning with the probabilistic user model [HH16]; (3) modelling uncertainty in the modelling of persuadee beliefs [Hun16c, HH18]; (4) harnessing decision rules for optimizing the choice of argument based on the user model [HH17, HHC18]; (5) crowdsourcing the acquisition of user models [HP17a]; and (6) modelling a domain in a way that supports the use of the epistemic approach [CHH$^+$18]. So these developments offer a well-understood theoretical and computationally viable framework for taking belief into account in APSs for applications such as behaviour change.

1.2 Concerns in automated persuasion systems

In our previous work, there is no consideration of how the concerns of the persuadee could be taken into account in an APS. Therefore, our ultimate goal in this paper is to understand the notion of concerns in order to harness them for making better choices of move in asymmetric dialogues in APSs. To this end, we have undertaken empirical studies with participants showing that:

1. there is a common understanding amongst participants as to what is meant by concerns associated with arguments;

2. useful preferences over types of concern can be gathered from participants;

3. participants tend to select counterarguments according to their preferences over types of concern associated with the counterarguments;

4. preferences over types of concern can be deployed by an automated persuasion system in order to improve the persuasiveness of a dialogue.

In addition, the methods that we employed in the empirical studies can be used as pipeline of methods for acquiring and deploying types of concern, and preferences between them, in automated persuasion systems.

Note, in this paper, we focus exclusively on the role of concerns in persuasion, and so we do not consider role of beliefs in persuasion here. We leave the combination of beliefs and concerns to future work.

We proceed as follows. In Section 2, we consider how we can capture concerns in argumentation dialogues, then in Section 3, we present our empirical studies for investigating how concerns can be acquired and used in argumentation for persuasion. We situate our work with respect to related literature in Section 4, and finally discuss our work in Section 5.

2 Argumentation with concerns

In order to investigate the use of concerns in persuasion, we will draw on abstract argumentation for representing arguments and counterarguments arising in persuasion dialogues, and then we will consider how concerns arise in argumentation.

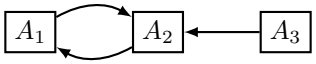
2.1 Abstract argumentation

In abstract argumentation, as proposed by Dung [Dun95], each argument is treated as an atom, and so no internal structure of the argument needs to be identified. This can be represented by an argument graph as follows.

Definition 1 *An* **argument graph** *is a pair $G = (\mathcal{A}, \mathcal{R})$ where $\mathcal{A}$ is a set and $\mathcal{R}$ is a binary relation over $\mathcal{A}$ (in symbols, $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$). Let $\mathsf{Nodes}(G)$ be the set of nodes in G (i.e. $\mathsf{Nodes}(G) = \mathcal{A}$) and let $\mathsf{Arcs}(G)$ be the set of arcs in G (i.e. $\mathsf{Arcs}(G) = \mathcal{R}$).*

So an argument graph is a directed graph. Each element $A \in \mathcal{A}$ is called an **argument** and $(A_i, A_j) \in \mathcal{R}$ means that A_i **attacks** A_j (accordingly, A_i is said to be an **attacker** of A_j). So A_i is a **counterargument** for A_j when $(A_i, A_j) \in \mathcal{R}$ holds.

Example 1 *Consider arguments A_1 = "Giving up smoking would be good for my health", A_2 = "When I try to give up smoking, I put on weight, and that is bad for my health", and A_3 = "I could try take up a new sport to get fit and avoid putting on weight". Here, we assume that A_1 and A_2 attack each other because they rebut each other (i.e. the claims of A_1 and A_2 contradict each other), and we assume that A_3 attacks A_2 because it provides an undercut to A_2. Hence, we get the following abstract argument graph:*



Given an argument graph, a natural question to ask is which arguments are acceptable (*i.e.*, which arguments are winning). Based on dialectical notions, Dung made some important proposals for acceptable subsets of arguments where each subset is conflict-free (*i.e.*, no argument in the subset attacks another

argument in the subset) and defensible (*i.e.*, for each argument in the subset that is attacked, there is an argument in the subset that attacks that attacking argument). Numerous proposals have been made that investigate variants and developments of Dung's proposal (see [BGGv18] for a comprehensive review).

In our work, we will use the definition of an argument graph. However, as we explain in the next subsection (*i.e.*, Section 2.2), we do not assume that the agents will use dialectical semantics. The reason for not using dialectical semantics is that we are not concerned with determining acceptable arguments according to normative principles but rather we wish to model how persuasion may occur in practice. We are aiming for a predictive model, and therefore do not want to impose conditions for when an agent should be persuaded (in a normative sense), but rather have a model that reflects aspects of how an agent is likely to behave.

2.2 Argumentation in persuasion dialogues

We assume that a **dialogue** is a sequence of moves $D = [m_1, \ldots, m_k]$. Equivalently, we use D as a function with an index position i to return the move at that index (*i.e.*, $D(i) = m_i$). In general, in formal models of argument, there is a wide variety of types of move that we could consider in our dialogues including positing of an argument, making a claim, conceding a claim, providing premises for an argument, etc.

In this paper, we will just consider two types of move. For both types, we assume that the arguments appearing in the move come from an argument graph that is known by the system.

- A **posit** move is a set of arguments selected by the system.
- A **menu** move is a set of arguments selected by the user from a set of arguments provided by the system.

We introduce the menu move as a way for the user to give her input into the discussion. We assume that we are unable to accept input from the user that would involve free text presentation of arguments since natural language processing technology is not yet able to adequately understand such input. Using menu moves leads to an *asymmetric* dialogue which means that the moves available to the system are different to those available to the user. Nonetheless, if the argument graph used by the system is sufficiently comprehensive, the user may be able to present her views faithfully (*i.e.*, that the menu always includes the user's arguments).

A feature of the posit and menu moves is that they involve a set of arguments at each step. To illustrate a real-world situation where a discussion that can proceed by participants exchanging sets of arguments, consider an email discussion between two colleagues. Here, one participant might start with a persuasion goal, and the second participant might provide some counterarguments to that persuasion goal. The first participant might then reply with a counterargument to each of the counterarguments, and so on.

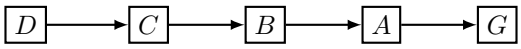


Figure 2: Example of an argument graph.

A **protocol** specifies what are allowed moves that can be made at each step of a dialogue. There are many possible protocols that we could define even with a limited range of moves. Here we just give one definition for a protocol as an example, which we will harness in our empirical studies. In this protocol, each posit is for a persuasion goal (*i.e.*, an argument that the persuader wants the persuadee to accept), or for an argument that directly or indirectly defends the persuasion goal, whereas each argument in a menu move is for an argument that directly or indirectly attacks the persuasion goal.

Note, by indirect attack, we mean that an argument is on a path to the persuasion goal with an odd number of arcs where that number is greater than 1, such as the path from C to G in Figure 2. By direct defence, we mean that if G is the persuasion goal, and A attacks G, and B attacks A, then B directly defends G (such as for B in Figure 2). Finally, by indirect defence, we mean that an argument is on a path to the persuasion goal with an even number of arcs where that number is greater than 2 (e.g., if D attacks C, and C attacks B, and B attacks A, and A attacks G, then A indirectly defends G, as in Figure 2).

To formalize our protocol, we require the following subsidiary functions. The first function gives the set of options which are the counterarguments of an argument that have not been used at a previous step in the dialogue. This function is used to ensure that the dialogue only allows for an argument to be presented at most once in the dialogue. The second function gives the arguments that appear in a menu.

Definition 2 *The set of* **options** *of argument A at step i in dialogue D is defined as follows, where for all $j < i$, $D(j) \subseteq \mathsf{Nodes}(G)$.*

$$\mathsf{Options}_i^D(A) = \{B \mid (B, A) \in \mathsf{Nodes}(G) \text{ and there is no } j < i \text{ s.t. } B \in D(j)\}$$

Definition 3 *The* **menu** *at step i in dialogue D is defined as follows, where for all $j < i$, $D(j) \subseteq \mathsf{Nodes}(G)$.*

$$\mathsf{Menu}^D(i) = \bigcup_{A \in D(i-1)} \mathsf{Options}_i^D(A)$$

Before we give the definition of the asymmetric posit dialogue protocol, we explain the conditions in the definition as follows.

1. For each step i, the move $D(i)$ is the posit of a set of arguments.
2. For the first step, the first move is by the system and it is a singleton set containing a persuasion goal (*i.e.*, an argument that the persuader wants the persuadee to accept).

3. The participants take turns, so if $D(i)$ is a move by the system, then $D(i+1)$ is a move by the user, $D(i+2)$ is another move by the system, and so on.

4. For each step i such that $2 < i \leq k$, if the step is for a system move, the system chooses a set of arguments from $\mathsf{Nodes}(G)$ satisfying the following conditions:

   (a) Each argument $B \in D(i)$ is a counterargument to an argument at $D(i-1)$.

   (b) For each argument A in $D(i-1)$, if there is a counterargument to A in $\mathsf{Nodes}(G)$ that has not appeared in previous step of the dialogue, then $D(i)$ contains a counterargument to A. For example, if B and C are the counterarguments to A, then at least one of them will be in $D(i)$.

5. For each step i such that $2 \leq i \leq k$, if the step is for a user move, the user is given a menu of arguments to choose from. This menu is the set containing the arguments B such that B is a counterargument to A for each $A \in D(i-1)$ and B has not been presented in a previous move. The user either selects any number of the counterarguments in the menu, or selects the "null argument" which means the user has no counterarguments or the user does have counterarguments but none in the menu are appropriate. So the move $D(i)$ is either the set of counterarguments selected from the menu by the user or the singleton set containing the null argument.

6. A dialogue ends when one of the following conditions holds:

   (a) The last move of the dialogue is $D(k) = \{A_1, \ldots, A_n\}$, and there is no argument B such that $(B, A) \in \mathsf{Arcs}(G)$ where $A \in \{A_1, \ldots, A_n\}$.

   (b) The last move of the dialogue is a user move, and the user selects the null argument (*i.e.*, the option that the user has no counterarguments or the user does have counterarguments, but none in the menu are appropriate).

We capture the above informal conditions in the following definition for the asymmetric posit dialogue protocol.

Definition 4 *An **asymmetric posit dialogue protocol** for an argument graph G is a dialogue D satisfying the following conditions, where the domain of D is $\{1, \ldots, k\}$ and Q is the assignment of a participant to each step:*

1. *For each step $i \in \{1, \ldots, k\}$, $D(i) \subseteq \mathsf{Nodes}(G)$.*

2. *For step $i = 1$, $D(i) = \{A\}$ where A is the persuasion goal.*

3. *For each step $i \in \{1, \ldots, k\}$, if i is odd, then $Q(i) = \mathsf{System}$, else $Q(i) = \mathsf{User}$.*

4. *For each step i such that $2 < i \leq k$, if $Q(i) = \mathsf{System}$,*
   (a) *for each argument $B \in D(i)$, there is an $A \in D(i-1)$ s.t. $(B, A) \in \mathsf{Arcs}(G)$;*
   (b) *and for each argument $A \in D(i-1)$, if $\mathsf{Options}_i^D(A) \neq \emptyset$, then $\mathsf{Options}_i^D(A) \cap D(i) \neq \emptyset$.*
5. *For each step i such that $2 \leq i \leq k$, if $Q(i) = \mathsf{User}$, then $D(i) \subseteq \mathsf{Menu}^D(i)$ or $D(i) = \{$"null argument"$\}$*
6. *For the final step k, one of the following conditions hold, and for all steps $j < k$, neither of the conditions holds.*
   (a) *For each argument $A \in D(k)$, $\mathsf{Options}_i^D(A) = \emptyset$.*
   (b) *$D(k) = \{$"null argument"$\}$.*

There are three main advantages of this protocol: (1) it supports a form of asymmetric dialogue where the user can only select arguments that are presented to her via the menu; (2) it allows for multiple arguments to be attacked at each step of the dialogue; and (3) it does not force all arguments to be addressed by counterarguments. For the latter point, the system can fail to counter some, though not all, of the user's arguments at a step without the dialogue terminating at that step. This is because we are not assuming any particular dialectical semantics (such as proposed by Dung [Dun95]). Rather, we are just concerned with what are allowed exchanges between the system and user.

Example 2 *For the argument graph in Figure 3, the following is a dialogue according to the asymmetric posit protocol. Note, that the system did not select a counterargument to A_4 and was unable to posit a counterargument to A_{11}.*

- *System move:* $D(1) = \{A_1\}$
- *User move:* $D(2) = \{A_2, A_3, A_4\}$
- *System move:* $D(3) = \{A_5, A_7\}$
- *User move:* $D(4) = \{A_8, A_9, A_{11}, A_{12}\}$
- *System move:* $D(5) = \{A_{14}, A_{18}\}$

Example 3 *For the argument graph in Appendix B, the following is a dialogue according to the asymmetric posit protocol where* G2 *is "You should cycle to work",* bikedangerous *is "Cycling in the city is dangerous",* rain *is "You are at risk of rain and your clothes can get wet",* bright *is "Wearing bright cycling clothes makes cycling much safer",* clothes *is "You can have work clothes in a bag and wear cycling clothes",* hassle *is "It is a hassle having to change clothes", and* habit *is "Changing clothes is just a habit that you do quickly at the start and end of the working day". Note, that the user did not select a counterargument to* bright *at step 3.*

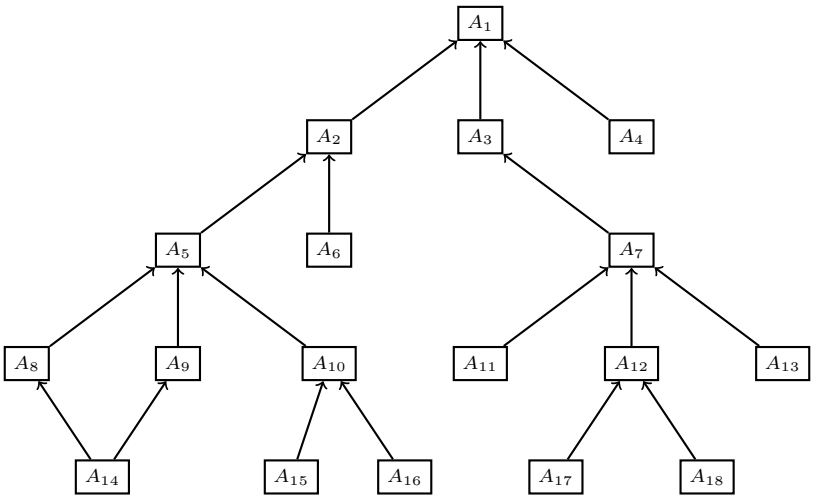


Figure 3: An argument graph for Example 2.

- *System move:* $D(1) = \{\text{G2}\}$
- *User move:* $D(2) = \{\text{bikedangerous}, \text{rain}\}$
- *System move:* $D(3) = \{\text{bright}, \text{clothes}\}$
- *User move:* $D(4) = \{\text{hassle}\}$
- *System move:* $D(5) = \{\text{habit}\}$

So the above protocol captures the behaviour where the agents exchange arguments from the argument graph until one of the following conditions holds: (1) the system has no further arguments to present to defend the persuasion goal; (2) the user has no further arguments to present to directly or indirectly attack the persuader's persuasion goal; or (3) the user has a list of counterarguments to choose from, but the user chooses the option that none of the counterarguments are appropriate (*i.e.*, the null argument).

At the end of the dialogue, the system (*i.e.*, the persuader) hopes to have persuaded the user (*i.e.*, the persuadee) to accept the persuasion goal (*i.e.*, the first argument in the dialogue). We do not use dialectical semantics to determine whether the persuasion goal is a winning argument (for example, if it is in a grounded or preferred extension of the subgraph of G induced by the arguments that appear in the dialogue). Rather, we ask the user at the end of the dialogue whether she believes the persuasion goal and to what degree.

This protocol is different to the dialogue protocols for abstract argumentation that are used for determining whether specific arguments are in the grounded extension [Pra05] or preferred extension [CP12]. It is also different to the dialogue protocols for arguments that are generated from logical knowledgebases (*e.g.*, [BH09, FT11]). These protocols are concerned with determining

the winning arguments in a dialogue and that this determination is sound and complete with respect to determining the winning arguments from the original knowledgebase. It is also noteworthy that many proposals for protocols for dialogical argumentation involved depth-first search (e.g., [Ben02]). This may have computational advantages but does not appear so natural to participants.

Note, the discussion of concerns that we consider in the next section does not rely on any specific protocol. We have used the asymmetric posit dialogue protocol in the empirical studies later in this paper (as we explain in Section 3.2). However, we believe that the findings we make from the empirical study are generalizable to a wide variety of dialogue protocols.

2.3 Modelling types of concern

A concern is something that is important to an agent. It may be something that she wants to maintain (for example, a student may wish to remain healthy during the exam period), or it may be something that she wants to bring about (for example, a student may wish to do well in the exams). Often arguments can be seen as either raising a concern or addressing a concern as we illustrate in the following example.

Example 4 *Consider the domain "Cycling in the city". Depending on the participants, the first argument could be addressing the concern of health whereas the second could be raising the concern of health.*

- *(A_1) Cycling will improve your health.*
- *(A_2) Cycling will cause you to breathe in a lot of exhaust pollution.*

When considering a set of agents, there may be a number of similar concerns, and it may be appropriate to group these into types of concern. For example, we could choose to define the type "Fitness" to cover a variety of arguments from those that flag problems associated with lack of exercise through to those that suggest solutions for people training for marathons. The actual types of concern we might consider, and the scope and granularity of them, depend on the application. However, we assume that they are atomic, and that ideally, the set is sufficient to be able to type all the possible arguments that we might want to consider using in the dialogue.

The types of concern may reflect possible motivations, agenda, or plans that the persuadee has in the domain (i.e., subject area) of the persuasion dialogue. They may also reflect worries or issues she might have in the domain. When an argument is labelled with a type of concern, it is meant to denote that the argument has an impact on that concern, irrespective of whether that impact may be positive or negative.

Example 5 *Returning to Example 4, consider the concern type "Health" and the domain "Cycling in the city". Both A_1 and A_2 are arguments with the type "Health". If we assume that the concern "Health" means that the agent wants to*

maintain or increase her health, then A_1 has a positive impact on the concern, and A_2 has a negative impact.

In this paper, we do not differentiate between positive and negative impacts. We just assume that when we label an argument with a type of concern, then the argument has an impact on that concern. Nevertheless, being aware that the impact may be positive or negative may help in identifying them in practice. We leave formalization of impacts on concerns for future work.

However, we do require information about preferences over types of concern. Given a set of types of concern for a given domain, an agent may have a preference ordering over those concerns that reflects the relative importance of the concerns to the agent. For instance, in the "Cycling in the city" domain, we may have the types "Health", "Safety", and "Personal Economy", and we may have an agent that regards "Health" as the most important concern for her, "Safety" as the next most important concern for her, and "Personal Economy" as the least important concern for her. Another agent may have a different ordering over these types of concern.

There are some potentially important choices for how we can represent the preferences. For instance, we can consider preferences as pairwise choices or we can assume a partial or even linear ordering. We are agnostic as to how preferences are represented, and assume an appropriate choice can be made for an application. There are also numerous techniques for acquiring preferences from participants (see *e.g.*, [CP04], for a survey). Again, we assume that an appropriate technique can be harnessed.

We will use the preferences over types of concern to make good choices of move. Since agents may differ in their preferences, we need to find out their preferences during a dialogue. We can do this by querying the user. In practice, we do not want to ask the user too many questions as it is likely to increase the risk of the user terminating the dialogue prematurely. Furthermore, it is normally not necessary to know about all the preferences of the user. To address this, we can acquire comprehensive data on the preferences of a set of participants, and then use this data to train a classifier to predict preferences for other participants based on a subset of questions. We investigate this further in the empirical study (Section 3).

2.4 Selecting moves based on concerns

In a dialogue, the system needs to choose which arguments to present. For instance, in the asymmetric posit protocol, the system chooses arguments from amongst those that are counterarguments to the arguments selected by the user. There are various criteria that can be used for selecting arguments. Here, we consider how the system can select arguments based on the concerns that the user has.

For this, we assume that we have a set of arguments, and that each argument is labelled with the type(s) of concern it has an impact on. Furthermore, the system has a model of the user in the form of a preference relation over the

Algorithm 1: Compare types

```
Function Comp(args)
    Initialize an array count to 0 for each argument in the set
    foreach a_1, a_2 ∈ args do
        if a_1 = a_2 then Continue
        else
            count_1 ← 0, count_2 ← 0
            foreach (t_1, t_2) ∈ types of (a_1, a_2) do
8               if t_1 ≻ t_2 then
                    count_1 ← count_1 + 1
10              else if t_2 ≻ t_1 then
                    count_2 ← count_2 + 1
            if count_1 > count_2 then
                count[a_1] ← count[a_1] + 1
            else if count_2 > count_1 then
                count[a_2] ← count[a_2] + 1
            else
17              (t_1, t_2) ← first type of (a_1, a_2)
18              if t_1 ≻ t_2 then count[a_1] ← count[a_1] + 1
19              else if t_2 ≻ t_1 then count[a_2] ← count[a_2] + 1
    Rank the arguments according to the count
    return top n
```

types of concern. We do not assume any structure for the preference relation. In particular, we do not assume it is transitive.

For each user argument A that the system wishes to attack with a counterargument, the set of attackers (the set of candidates) is identified (*i.e.*, the set of arguments B such that $(B, A) \in \mathsf{Arcs}(G)$). From this set of candidates, Algorithm 1 is called to select the most preferred one amongst this set. In other words, the argument returned will be the most preferred attacker of A according to the preference over concerns.

The algorithm considers each pair of arguments A and A' in the set of candidates, and determines for how many types of concern A is preferred to A', and for how many types of concern A' is preferred to A. In other words, if A is associated with concern t and A' is associated with t', and t is preferred to t', then A is preferred to A' according to one type of concern (as specified at lines 8 and 10). Once all the types of concern are considered, if A is preferred to A' according to more types of concern, then overall A is preferred to A'. Any ties between arguments are broken by comparing the top chosen types for the arguments (lines 17 to 19). It is possible that top chosen types are equal for the arguments (though, this did not arise in our experiments). To address this possibility, it is straightforward to extend the algorithm to consider subsequently

ranked types.

We can use the algorithm above to select moves in a dialogue, such as in part 4 of the asymmetric posit dialogue (Definition 4). We will harness this in the empirical study of persuasion dialogues in the next section.

3 Empirical studies

In this section, we will describe the aims, methods, results, and conclusions, for a sequence of empirical studies. All these studies have been approved by the UCL Research Ethics Committee.

3.1 Aims

The purpose of this sequence of studies is to investigate the nature of concerns of a participant in a dialogue. In particular, we want to consider the following questions.

- Question 1: Do participants tend to agree on the types of concern that should be associated with an argument?
- Question 2: Can participants give meaningful preferences over the types of concern? In other words, do participants express non-random preferences over concerns?
- Question 3: Are participants playing by their preferences over concerns? In other words, when a participant makes a choice of argument to present from a set of possible candidates, does she choose the argument with the highest ranking concern?
- Question 4: Can we use a participant's preferences over concerns to choose more preferred arguments so that the dialogue is more persuasive?

In order to use preferences to choose good moves in a persuasion dialogue, we would expect to have positive answers to questions 1 to 3. So Question 4 depends on the answers to the questions before it.

3.2 Methods

Building upon the discussion of types of concern given in the previous section, we now present the methods we require for our empirical study. These are split into the following steps which we explain in more details in the specified subsections.

1. Creating the argument graph (Section 3.2.1).
2. Crowdsourcing the associations between arguments and types of concern (Section 3.2.2).

3. Gathering the pairwise preferences the participants have on the types of concern (Section 3.2.3).

4. Checking whether participants are playing by their explicit preferences over concerns (Section 3.2.4).

5. Creating a decision tree for each pair of types that can allow us to predict a participant's preferences over types of concern (Section 3.2.5).

6. Building an automated persuasion system having a discussion with a human and reasoning about the preferences of this participant (Section 3.2.6). The decision trees developed in step 5 will be harnessed in the automated persuasion system to allow it to present fewer questions to the participants about their preferences.

We use these steps as the methods for this empirical study, but they can also be harnessed as a general framework for acquiring and deploying concerns in persuasion systems. They offer a pipeline of methods for the development of an automatic persuasion system, from the creation of the argument graph through to using the preferences on the types in an automatic dialogue. This pipeline includes the gathering of the type associations with the arguments and the learning of the preferences. So although we present our framework via its application to the "Cycling in the city" domain, it is general and can be applied to other domains.

Note that for the studies with participants, we recruited separate groups of participants for each of step 2 (i.e., the study described in Section 3.2.2), steps 3/4 (i.e., the study described in Sections 3.2.3 and 3.2.4), and step 6 (i.e., the study described in Section 3.2.6). The reason we recruited three disjoint groups of participants is that we wanted to remove any possibility of bias that might arise from participating in multiple studies.

3.2.1 Creation of the argument graph

This first step consists of listing the arguments associated with the issue at hand. We start by specifying a goal argument in the dialogue. This argument is what the persuader wants the persuadee to believe by the end of the dialogue. In the empirical study described in Section 3.2.6, the automated persuasion system starts with this argument in every dialogue.

In this paper, we undertook a web search on the pros and cons of city cycling. We consulted diverse sources including online forums and documentation on the topic (for instance, governmental websites promoting cycling). From this search, we manually identified a number of arguments, and attacks between them, concerning aspects of cycling in the city. Recall that we only deal with the attack relation in this work, and so we did not consider other kinds of relation such as support. Also, we did not attempt to distinguish between different kinds of attack (such as undercutting or undermining). Some arguments were edited to enable us to have reasonable depth (so that the dialogues were of a

reasonable length) and breadth (so that alternative dialogues were possible) to the argument graph.

We wrote a short statement specifying each argument. This resulted in 51 arguments about aspects of cycling in the city. The arguments are enthymemes (i.e., some premises/claims are implicit) as this offers more natural exchanges in the dialogues. The full list of the arguments can be found in Appendix A, and the argument graph can be found in Appendix B.

We flagged two of the 51 arguments as being potential persuasion goals: "You should cycle to work." and "The government should invest in supporting initiatives to increase cycling.". We used the former as the principal goal and the latter as a fall-back goal if the persuader gets into a situation where the principal goal cannot be defended, and so it can open new possible branches of dialogues.

As an alternative to manual creation of the argument graph, this step can be crowdsourced by asking for arguments and counterarguments from participants. This can be done incrementally as follows: create an argument graph composed of the goal argument and several obvious counterarguments; ask the participants for counterarguments to the arguments in the graph; create a new graph composed of the previous arguments and the new counterarguments; repeat steps 2, 3 and 4 until no additional arguments are added to the graph. Potentially, this alternative will provide better coverage of the arguments and counterarguments for a domain, but it may involve more time and effort to run the interactive and incremental process, and it may also involve filtering out duplicate arguments and rejecting poor quality arguments.

3.2.2 Association of types of concern to arguments

Once all the arguments have been defined, they need to be typed. The types of concern that can be associated with the arguments are topic dependent. In this work, we manually defined a set of 8 allowed types of concern and crowdsourced their associations with the arguments. These types are: "Time", "Fitness", "Health", "Environment", "Personal Economy", "City Economy", "Safety" and "Comfort". The difficulty of this step lies in presenting the participants with a set of types of concern that is sufficiently broad to cover all possibilities, but not too specific in order to not uselessly spread the votes and have too many types associated with a given argument.

The creation of the set of types of concern can alternatively be left to the participants as a pre-step. However, it bears the risk of having too many types (or even unique ones, *i.e.*, associated with a single argument) because each participant can potentially come with her categorization and her own wording. This would then create the need to filter and/or group the suggestions, and then undertake the testing prescribed in the rest of this subsection.

For this step, we used 20 participants from the Prolific[1] crowdsourcing website. At the recruitment stage, the participants were informed that the study

[1] `https://prolific.ac`

was about cycling in the city. We prescreened participants according to the following criteria: fluency in English; current country of residence was United Kingdom; and age was between 18 and 100. For all the studies in this paper, only the participants who passed the prescreening were able to participate in the study. The full survey description can be found in Appendix D.

For each argument described in Appendix A, we asked the participants to choose the type of concern they think is the most appropriate from the list presented at the beginning of this section. We also used two attention checks asking them explicitly to select "Environment" and "Personal Economy". This was to ensure that participants were concentrating on the instructions for each question.

3.2.3 Pairwise preferences on the types of concern

After the set of types of concern had been created, the next step was to determine the preferences that the users of our system could have on these types. Preference elicitation and preference aggregation are research domains by themselves and it would take more than a paper to fully investigate them all in our context. For this reason, in this work, we decided to use pairwise preferences because of the simplicity of their elicitation. The drawback is that participants can have cycles in their preferences, unlike when a linear complete ordering is forced by the elicitation process. However, in our case, this was not problematical as we later used the pairs of preferences themselves instead of a preorder created from the pairwise elicitation procedure.

For each set of counterarguments for every argument, we asked for the preferences in all the possible pairs of types of concern associated with the counterarguments in the set. So for each pair, we asked the participants to state if they preferred the first type, the second type, if they were equal or if they were incomparable.

For this step, we used 50 participants from the Prolific crowdsourcing website. At the recruitment stage, the participants were informed that the study was about cycling in the city. The prescreening was as follows: first language was English; current country of residence was United Kingdom; age was between 18 and 100. We used two attention checks explicitly asking the participants to choose the left (resp. right) type in two pairs.

3.2.4 Are participants playing by their explicit preferences?

In order to answer Question 3 (*i.e.*, are participants playing by their preferences over concerns?) from Section 3.1, we investigated whether the following principle holds: when a participant makes a choice of argument to present from a set of possible candidates, she chooses the argument from the set of candidates with the highest ranking concern.

To investigate whether this principle holds, we augmented the previous step of the study (given in Section 3.2.3), by asking the participants to choose the counterargument they "think is the most persuasive" for each of a series of 6

arguments. Using the short names for arguments given in Appendix A, the arguments were: (1) the main goal; (2) lanes; (3) savestime; (4) clothes; (5) expensivelanes; and (6) moretime. For each of these 6 arguments, all counter-arguments in the argument graph were given, and the participant had to select the counterargument that they think is most persuasive.

3.2.5 Creation of the decision trees

A potential problem with using preferences in an APS is the need to ask the user about their preferences. This may be an issue if the user is asked too many questions, and as a result, disengages. A potential solution is to train a classification system for predicting the preferences that a given user may have. For this, we created decision trees using information that we had obtained about participants.

In addition to asking for the preferences of the participants over the types of concern (as explained in the previous subsection), we asked them to take a personality test, and to provide some demographic information (such as age, sex, etc.), and domain dependent information such as situational information (living in a city, in the countryside, etc.). Note, after collecting this data, we decided to not use the domain dependent information as we were able to get good results without it. We used the Ten-Item Personality Inventory (TIPI) [GSRS03] to assess values on 5 features of personality based on the OCEAN model [MC87]. It consists of the "Openness to experience", the "Conscientiousness", the "Extroversion", the "Agreeableness" and the "Neuroticism" (the emotional instability).

Using all the data, we learnt a decision tree for each pair of types asked during the previous step using the Scikit-learn [2] Python library. The purpose was to be able to determine the preferred type (or the equality or incomparability) given these pieces of information.

As a first stage, we ran a meta-learning process in order to determine the best combination of tree depth and minimum number of samples at each leaf for each pair of types. The meta-learning process is the repeated application of the learning algorithm for different choices of these parameters (*i.e.*, tree depth and minimum number of samples at each leaf) until the best combination of parameters is found. The criterion to minimize is the difference between the prediction and the actual preferred type. In our case, traditional metrics to assess the performance of a classifier (*i.e.*, F1-score) could not directly be used. Indeed, if, for instance, a participant answers that two types are equal to her in terms of preference, if the classifier returns one of the two types instead of the "equal" class, we should still count it as a match and not a discrepancy. For this reason, we defined our own difference value as follows.

Definition 5 Type difference. *Let t_1 and t_2 be two types. The preference on a pair of types (t_1, t_2) can be: t_1 preferred to t_2 (denoted by $\{t_1\}$), t_2 preferred to t_1 (denoted by $\{t_2\}$), both are equal (denoted by $\{t_1, t_2\}$), or they cannot be*

[2] `http://scikit-learn.org`

Algorithm 2: Difference

```
Function Diff(t_a, t_b)
    if t_a is {t_1, t_2} then
        if t_b is {} then return 1
        else return 0
    else if t_a is {} then
        if t_b is {t_1, t_2} then return 1
        else return 0
    else if t_a ≠ t_b then
        return 1
    else return 0
```

compared (denoted by $\{\}$*). Also, let* t_a *be the prediction concerning the preference over* t_1 *and* t_2 *(i.e., one of* $\{t_1\}$*,* $\{t_2\}$*,* $\{t_1, t_2\}$ *and* $\{\}$*), and let* t_b *be the actual preferred type preference over* t_1 *and* t_2 *(i.e., one of* $\{t_1\}$*,* $\{t_2\}$*,* $\{t_1, t_2\}$ *and* $\{\}$*). We define Algorithm 2 to calculate the difference value to minimize. This algorithm measures a difference between the predicted preference and the actual preference by taking into account that if* t_1 *and* t_2 *are equal, one or the other is still a valid prediction.*

We used cross-validation in the meta-learning to determine the best combination of tree depth and minimum number of datapoints at each leaf. Once the best parameters were found for each pair of types, we then ran the actual learning part using these parameters with all the datapoints concerning the personality and demographic information. We thus obtained one decision tree for each pair of types that was used by the automated persuasion system in the final study.

3.2.6 Persuasion dialogue

In order to investigate Question 4 (*i.e.*, can we use a participant's preference over concerns to choose more preferred arguments so that the dialogue is more persuasive?) from Section 3.1, we designed two automated persuasion systems. The first system was the baseline system. It chose an argument at random from amongst the arguments that attack one of the arguments presented by the persuadee at the previous step. The second system was called the preference-based system, and it chose the next argument to play according to the model of the preferences of the user over the types of concern. So from amongst the arguments that attack one of the arguments presented by the persuadee at the previous step, it chose the argument that was associated with the most preferred type of concern.

Note that for the belief, as described in Appendix F, we asked for a value between -5 and 5 with a 0.01 step. We chose to do so for two reasons: first, it allows for a finer grained value without having too many decimal figures, and

second it may be easier for the participants to consider the scale if it is wide centred over 0 (unlike other scales such as [0,1] with 0.5 as the middle point). The participant entered this data via a slider bar and so this allowed them to not be excessively concerned by the precision of the value entered.

The decision trees are the last part needed by the preference-based system to be able to fully function. Once up and running, the system needs to respond to the arguments from the participant in order to maximize the chances of having her persuaded at the end of the dialogue.

In this study, we used 99 participants recruited from the Prolific crowdsourcing website: 50 for our method and 49 for the baseline (one participation for the baseline was not saved by the system). At the recruitment stage, the participants were informed that the study was about cycling in the city. We presented each participant with a chatbot composed of a front-end we coded in Javascript and a back-end in Python using the Flask web server library[3]. The Python code presented the arguments being posited by the system, constructed the menu of counterarguments that the participant can select from, and collected the selection made by the participant. The arguments used for this are given in Appendix B.

The prescreening was as follows: first language was English; current country of residence was United Kingdom; age was between 18 and 100; commute/travel to work was by any means except walking and cycling; no long-term health condition/disability; employment status was full-time or part-time; no full-time remote working; working hours were regular (*e.g.*, 9-5). We chose these constraints for several reasons. Firstly, we wanted to target people who are not already cycling to work (nor walking as it would invalidate all money saving arguments) but who did not have a condition that would prevent them from cycling. Secondly, we wanted people having a job that was not full-time remote in order for them to have a mandatory commute. Finally, we had the condition of regular working hours as we assumed cycling for late shifts might be more daunting if not dangerous. The full survey description and demographic statistics can be found in Appendix F.

3.3 Results

In this section, we present the results concerning the experiments specified in the methods section (i.e. Section 3.2). We analyse the results of each step and give insights into the data obtained[4]. Note that, as explained in Section 3.2.1, we manually created the argument graph which is given in Appendix B.

3.3.1 Association of types of concern to arguments

We start by considering results concerning Question 1 (Do participants tend to agree on the types of concern that should be associated with an argument?) from

[3]The code is available at `https://github.com/ComputationalPersuasion/Surveyer`.
[4]The data is at `http://www0.cs.ucl.ac.uk/staff/a.hunter/papers/concernsdata.zip`.

Section 3.1. Of the 20 participants for this step, we removed 2 participants from the pool as they failed one or both attention checks, thus obtaining 18 answers.

Table 4 in Appendix C shows the results regarding the associations. Five of the arguments have a single concern associated with them, and of the remaining arguments, most had the majority of people selecting the same concern.

The types in bold in Table 4 are the types kept for the following steps of the experiments. To select them, we decided to keep all the types having strictly more than half of the votes of the most voted type. For instance, if the most voted type gathered 8 votes over 18, all types with strictly more than 4 votes were kept in the association.

The κ value for the reliability of agreement using Fleiss' Kappa [Fle71] is 0.53, denoting a moderate agreement. This value measures the amount of agreement amongst the participants that is above what can be achieved by choosing randomly. Therefore, we consider this set of types sufficiently expressive and accurate. Furthermore, this experiment allows us to get a positive answer to Question 1. Note, this κ value is calculated as the exact match of type of concern. An alternative would be to calculate whether participants selected a specific type or not, and this could yield a higher κ value.

3.3.2 Pairwise preferences on types of concern

Next, we consider results concerning Question 2 (*i.e.*, Can participants give meaningful preferences over the types of concern?) from Section 3.1. Of the 50 participants we recruited, we removed one participant who failed the attention check and so we ended up with 49 responses. The full survey description and demographic statistics can be found in Appendix E. We analysed the results in two ways as follows.

First, we considered whether the choices made by the participants were different from random choices. Recall, that for each pair of types of concern t and t', the participant could choose one of four options, namely t is strictly preferred to t, t' is strictly preferred to t, t and t' are equally preferred, and t and t' are incomparable. So for each pair, we obtain the number of participants who preferred each of the four options. If the participants were making random choices, it would be reflected by a uniform distribution over the choices. However, for each pair, the distribution we obtained is significantly different from the uniform distribution. Out of these pairs, the maximum p value was less than 0.007 using a Chi-square test. This means that the participants were making genuine choices.

Second, we considered whether there was a structural pattern to the preferences expressed by each participant. Out of the 49 participants, 11 had a linear partial order (*i.e.*, where all types are either pairwise comparable or incomparable and there is no directed cycle). A further 11 had directed cycles but one or more equal or incomparable types dominated all the others. Therefore, we could clearly identify the most preferred concerns for 22 participants. However, 27 had either too many cycles or too many incomparable or equal types to be able to clearly designate a dominating type or subset of types. Table 1 shows

Health	Safety	Comfort	P. Eco.	Env.	Time	Fitness
54.5%	36%	9%	9%	9%	4.5%	4.5%

Table 1: Partition of dominating types (% is of the proportion of the 22 participants).

the partition of the dominating types amongst the 22 participants where they can be defined. Interestingly and unsurprisingly the "City Economy" is never a dominating type. Note that cycles in the preferences from a participant do not necessarily indicate a lack of coherence. It might reflect an error in reporting by the participant, or it may indicate that the participant is using multiple criteria to compare the concerns.

Since the participants are not making random choices, and they have given their pairwise preferences over all the types of concern, we have a positive answer to Question 2.

3.3.3 Are participants playing by their explicit preferences?

We now turn to Question 3 (*i.e.*, Are participants playing by their preferences over concerns?) from Section 3.1. Six arguments were considered. For each of them, multiple counterarguments were given. The participant selected the counterargument that they thought was most persuasive. We then calculated the average agreement ratio between their answers and their preferences.

The six arguments and their counterarguments were a subset of those considered in Section 3.2.2 and so for the analysis, the types of concerns were known for each of them.

We use $t \succeq t'$ to denote that t is more or equally preferred to t'. The agreement ratio is calculated as follows where $\mathbb{1}_{t \succeq t'}$ is the indicator function (*i.e.*, if $t \succeq t'$ holds, then it returns 1, otherwise it returns 0).

Definition 6 *Let A be the counterargument chosen by the participant, $C = \{A_1, \ldots, A_p\} \setminus \{A\}$ be the set of all the other possible counterarguments, T be the set of all the types of A and T' be the set of all the types of the arguments in C. We define the* ***agreement ratio*** *for one participant as:*

$$\frac{1}{|T| \times |T'|} \sum_{t \in T} \sum_{t' \in T'} \mathbb{1}_{t \succeq t'}.$$

We then average on all the participants. A value of 1 means that 100% of the participants played by their preferences and that all the types of the chosen argument are preferred to all the types of non-chosen arguments.

The results are presented in Table 2. As some chosen counterarguments have "City Economy" as one of their types, and given that this type is never preferred, a ratio of 1 cannot be reached. Hence, a value of 0.71 is a very high agreement ratio when using our calculation method.

1	2	3	4	5	6	average
0.69	0.57	0.78	0.76	0.72	0.73	0.71

Table 2: Agreement ratios between explicit and implicit preferences for specific arguments. Using the short names for arguments given in Appendix A, the arguments are: (1) the main goal; (2) lanes; (3) savestime; (4) clothes; (5) expensivelanes; and (6) moretime.

Note that the result for the second argument ("The city can build cycle-only lanes.") is marginally below the others because one counterargument that the participants could choose ("They create traffic jams.") was seen in the argument/types association step as having 3 completely different types: "Environment", "Time" and "City Economy". However, we assume that no participant could see all three of these types of concern. Therefore, the type(s) seen by the participant concerning this argument might be preferred for her, even though the other(s) is/are not, thus artificially decreasing the ratio. The only other argument with 3 types is "Even if it does take more time, it is a more relaxing way to travel if you avoid busy roads." but with "Time", "Comfort" and "Health" where the last two can be related (and sometimes exchanged).

Since the agreement ratios given in Table 2 are quite high, the participants do appear to be playing by their preferences, and hence we have a positive answer to Question 3.

3.3.4 Creation of the decision trees

We applied the methodology explained in Section 3.2.5 in order to learn the parameters and subsequently the decision trees. In our case, the depth is from 1 to 5 and the minimal number of samples at each leaf is between 1 and 13.

Given our argument graph, 22 pairs of types (and thus decision trees) out of 28 possible pairs occur. Note, we have 28 possible pairs because there are 8 types of concern, and hence, $(8 \times 7)/2$ different pairs. In 16 of them, the preferred type is deterministic, *i.e.*, it is the same independently of the input data. Note that this does not mean that every single participant chose the same type but rather that too few of them chose a different one so they could not be put into a different leaf of the tree without risking overfitting. An additional decision tree could be reduced to a single deterministic decision for simplification at the cost of a very small possible classification error. We thus obtained 5 decision trees with a maximum depth of 5, and 17 deterministic choices. Figure 4 shows the example of the decision learnt for the Time/Comfort pair of types where the categories for the occupation are given in Appendix E Question 3 and "E" stands for "Extraversion" in the OCEAN model.

In order to study the usefulness of this method, we compared it with a dummy classifier (as named by the Scikit-learn library). We tested three different types of dummy classifiers: "uniform", "most frequent" and "stratified".

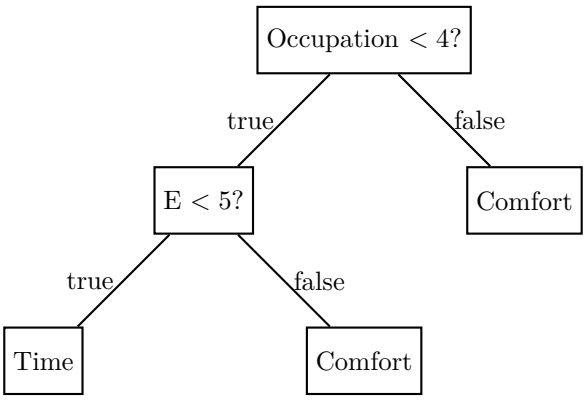


Figure 4: Example of a decision tree for the Time/Comfort pair where the categories for the occupation are given in Appendix E Question 3 and "E" stands for "Extraversion" in the OCEAN model.

The difference lies in how they predict the classes for the testing data given the learning data classes. Given a learning data class distribution (for instance, two classes c_1 and c_2 accounting for respectively 70% and 30% of the data):

- the "uniform" dummy predicts c_1 and c_2 with an equal probability when asked to classify the testing data;
- the "most frequent" dummy predicts c_1 for all testing data as it is the most represented in the learning data;
- the "stratified" dummy predicts c_1 (resp. c_2) with a probability of 0.7 (resp. 0.3), thus respecting the learning class distribution.

In our case the "stratified" dummy was more efficient than the other two. We thus used it as our comparison baseline.

Our decision tree method yielded an average difference (as of Algorithm 2) of 0.18 (the lower the better) on a 10-fold cross validation for each pair of types. The stratified dummy yielded an average result of 0.37, with a difference statistically significant with $p < 2.7e^{-10}$ meaning that the input information (personality and demographics) can be used and is indeed useful.

3.3.5 Persuasion dialogue

Finally, we consider the results concerning Question 4 (*i.e.*, can we use a participant's preference over concerns to choose more preferred arguments so that the dialogue is more persuasive?) from Section 3.1. For this, we compared the baseline system and the preference-based system that used the preference relation over concerns.

Figures 5 and 6 show the belief before and after the participants engaged in the dialogue with respectively the baseline chatbot and the preference-based

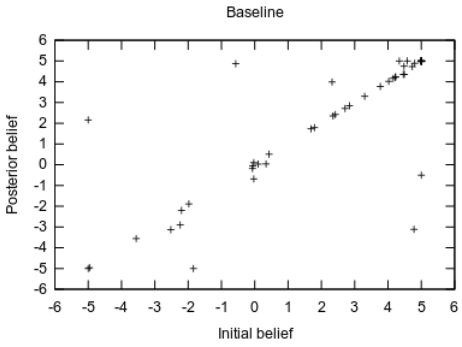


Figure 5: Initial and posterior belief for the baseline

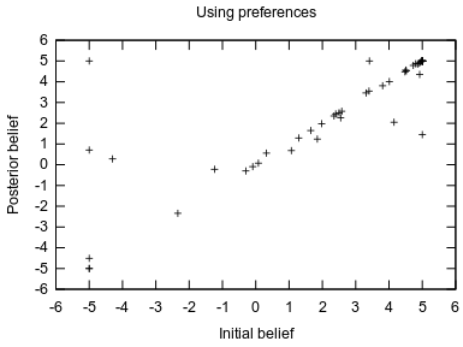


Figure 6: Initial and posterior belief for our method

one. As we can see, the majority of participants do not change their belief. However, although most people also already believe our goal argument, it is easy to see that more people change from negative belief to positive belief and fewer from positive to negative when using our preference-based method instead of the baseline method.

All the results are summarized in Table 3. The participants and the systems exchange 4.2 messages (by both persuader and persuadee) on average during the dialogues. To test the significance of the results, we put the participants into three groups: those who experienced positive changes in the belief; those who experienced negative changes; and those who experienced no change. We used a Chi-square test with the results of our preference-based method as the expected values and the results of the baseline as the observed ones. Note that, as the number of participants were different by one between our method and the baseline method, we repeated it three times with an increment in all the three different groups, one at a time. The difference between our method and the baseline is statistically significant with a maximum $p < 0.035$ (and minimum $p < 0.01$). We can therefore say that we have a positive answer to the Question 4 from Section 3.1 (*i.e.*, can we use a participant's preferences over concerns to choose more preferred arguments so that the dialogue is more persuasive?).

3.4 Conclusions

We now return to the aims of the empirical study that we presented in Section 3.1 and consider how the results have answered the questions raised.

- **Question 1: Do participants tend to agree on the types of concern that should be associated with an argument?** Our results show that given a sufficiently expressive yet not too precise set of types, participants tend to agree amongst themselves on the associations (Section 3.3.1). It means that the notion of types of concern is understood by

	Preference-based system	Baseline system
Positive change	26%	22%
Negative change	12%	24%
Avg. change	0.33	-0.06
Avg. change w/o 0	0.88	-0.14
Negative to positive	6%	6%
Positive to negative	0%	4%
Comments	46%	20%

Table 3: Results for the automated persuasion systems. We report our results using the following seven criteria: the ratio of participants with positive (resp. negative) change in the belief; the average change on all the participants (avg. change); the average change on the participants who did change (avg. change w/o 0); the ratio of participants going from a negative belief (resp. positive) to a positive belief (resp. negative); and the ratio of participants who took the time to provide further arguments (that can somewhat be related with the engagement).

the participants and that it can be used as an efficient input for machine learning models, similarly to what we did with decision trees (Section 3.3.4).

- **Question 2: Can participants give meaningful preferences over the types of concern?** Our results show that participants are not randomly expressing preferences over types of concern but rather presenting meaningful information about the relative importance of the concerns. However, the participants do not seem to have a linear complete ordering on the types (Section 3.3.2). Therefore, without a more complex preference elicitation process (that might, itself, turn out to be ineffective), we need to use pairwise preferences.

- **Question 3: Are participants playing by their preferences over concerns?** Our results show that people tend to argue following their preferences (Section 3.3.3). Conveniently, it means that what is acquired during the elicitation process can be used without complex preprocessing, yields good performances, and is easily interpretable.

- **Question 4: Can we use a participant's preferences over concerns to choose more preferred arguments so that the dialogue is more persuasive?** Our results show that preferences over concerns can indeed be used to improve the persuasiveness of a dialogue when compared with randomly generated dialogues Section 3.3.5.

So the empirical studies have provided a positive answer to each of the questions. However, further experiments are required to make these claims with

more confidence. For instance, different domains, and different experimental setups should be investigated in order to get a deeper understanding of the role of concerns in persuasion.

As well as answering the above questions, our empirical study demonstrates that our methods can be applied to the implementation of an automated persuasion system in the "Cycling in the city" domain, and that they appear transferable to other domains.

4 Literature review

Since the orginal proposals for formalizing dialogical argumentation [Ham71, Mac79], most approaches focus on protocols (*e.g.*, [ME98, AMP00a, AMP00b, DDKV00, HMP01, MP02a, MP02b, MvEPA03, Pra05, Pra06, CP12]) with strategies being under-developed. See [Thi14] for a review of strategies in multi-agent argumentation.

Some strategies have focussed on correctness with respect to argumentation undertaken directly with the knowledgebase, in other words, whether the argument graph constructed from the knowlegdebase yields the same acceptable arguments as those from the dialogue (*e.g.*, [BH09, FT11]). Strategies in argumentation have been analysed using game theory (*e.g.*, [RL08, RLT09], [FT12]), but these are more concerned with issues of mechanism design, rather than persuasion.

In [BCB14], a planning system is used by the persuader to optimize choice of arguments based on belief in premises, and in [BCH17], an automated planning approach is used for persuasion that accounts for the uncertainty of the proponent's model of the opponent by finding strategies that have a certain probability of guaranteed success no matter which arguments the opponent chooses to assert. Alternatively, heuristic techniques can be used to search the space of possible dialogues [MBL16]. Persuasion strategies can also be based on convincing participants according to what arguments they accept given their view of the structure of an argument graph [MBL$^+$18]. As well as trying to maximize the chances that a dialogue is won according to some dialectical criterion, a strategy can aim to minimize the number of moves made [ABB12].

There are some proposals for strategies using probability theory to, for instance, select a move based on what an agent believes the other is aware of [RTO13], or, to approximately predict the argument an opponent might put forward based on data about the moves made by the opponent in previous dialogues [HSM$^+$13]. Using the constellations approach to probabilistic argumentation, a decision-theoretic lottery can be constructed for each possible move [HT16]. Other works represent the problem as a probabilistic finite state machine with a restricted protocol [Hun14], and generalize it to POMDPs when there is uncertainty on the internal state of the opponent [HBM$^+$15].

In value-based argumentation, the values of the audience are taken into account. Often a value is considered as moral or ethical principle that is promoted by an argument. Though as considered by Atkinson and Bench-Capon [Atk06],

the notion of a value promoted by an argument can be more diverse and used to capture general goals of an agent. A value-based argumentation framework (VAF) extends an abstract argumentation framework by assigning a value to each argument, and for each type of audience, a preference relation over values. This preference relation which can then be used to give a preference ordering over arguments [BC03, BCAC05, Atk06, BCA09, AW13]. The preference ordering is used to ignore an attack relationship when the attackee is more preferred than the attacker, for that member of the audience. This means the extensions obtained can vary according to who the audience is. VAFs have been used in a dialogical setting to make strategic choics of move [Ben02]. Whilst, the notion of a concern seems to be similar to the notion of a value, the role of concerns is different to that of values. In our approach, we do not use concerns to change the structure of an argument graph but rather use concerns directly to select arguments to present to a user.

Whilst there may be some overlap in the notion of a value assignment and of a concern assignment, there do appear to be differences between them. In particular, values appear to capture ethical or moral dimensions that are promoted by an argument whereas we view concerns as being issues raised or addressed by an argument. In addition, the way they are used is different. For VAFs, it is about trying to determine what is acceptable to an audience using a static presentation of an argument graph, whereas for our approach, it is about trying to determine what are the best moves to make in dynamic setting during a dialogue to get the best outcome.

More recently, an alternative notion of values has been proposed for labelling arguments that have been obtained by crowdsourcing. In this alternative notion, a value is a category of motivation that is important in the life of the agent (*e.g.*, family, comfort, wealth, etc.), and a value assignment to an argument is a category of motivation for an agent if she were to posit this argument [CHHP18]. It was shown with participants that different people tend to apply the same (or similar) values to the same argument.

The notion of interests as arising in negotiation is also related to concerns. In psychological studies of negotiation, it has been shown that it is advantageous for a participant to determine which goals of the other participants are fixed and which are flexible [FU81]. In [RPSD09], this idea was developed into an argument-based approach to negotiation where meta-information about each agent's underlying goals can help improve the negotiation process. Argumentation has been used in another approach to co-operative problem solving where intentions are exchanged between agents as part of dialogue involving both persuasion and negotiation [DDKV00]. Even though the notions of interests and intentions are used in a different way to the way we use the notion of concerns in this paper, it would be worthwhile investigating the relationship between these concepts in future work.

The empirical approach taken in this paper is part of a trend in the field of computational argumentation for studies with participants. This includes studies that evaluate the accuracy of dialectical semantics of abstract argumentation for predicting behaviour of participants in evaluating arguments

[RMB+10, CTO14], studies comparing a confrontational approach to argumentation with argumentation based on appeal to friends, appeal to group, or appeal to fun [VSM+13, VSMO16], studies of appropriateness of probabilistic argumentation for modelling aspects of human argumentation [PH18], studies to investigate physiological responses of argumentation [VCJ+17], studies using reinforcement learning for persuasion [HL07], and studies of the use of predictive models of an opponent in argumentation to make strategic choices of move by the proponent [RK16]. There have also been studies in psycholinguistics to investigate the effect of argumentation style on persuasiveness [LAWW17].

Finally, there are a number of studies that indicate the potential for dialogical argumentation in behaviour change applications including dialogue games for health promotion [Gra98, CGJ99, GCJ00, Gra03], embodied conversational agents for encouraging exercise [NME07], and tailored assistive living systems for encouraging exercise [GNL16].

5 Discussion

In this paper, we proposed a pipeline of methods for acquiring and using types of concern, and preferences over them, to select moves in a dialogue. In addition, we have presented empirical studies to validate the pipeline including showing that people do agree on the concerns that can be associated with arguments, that people do have preferences over types of concern, and that people do tend to play by their preferences. We have also demonstrated that we can train classifiers to predict the preferences over types of concern for individuals. Finally, we have shown that we can use preferences over concerns to improve the persuasiveness of a dialogue.

The aim of our dialogues in this paper is to raise belief in goal arguments. A goal argument may reflect an intention but this is not mandatory. We focus on beliefs in arguments because belief is an important aspect of the persuasiveness of an argument (see for example [HP17a]). Furthermore, beliefs can be measured more easily than intentions in crowdsourced surveys.

Preferences over types of concern are an important aspect of persuadee modelling that we have investigated in this paper. We did not assume that the preferences are transitive, and we found that preferences given by participants are often not entirely coherent. This may reflect that participants may find it difficult to remember at each point in a survey what answers they have given previously. It may also reflect that participants change their minds about their preferences as the survey progresses. So we may regard some of the preferences given by participants as noise. Despite this, our investigations in this paper indicate that it is beneficial to harness this preference information for user modelling.

Another topic for future work is the specification of the protocol. Many protocols for dialogical argumentation involve a depth-first approach (e.g., [Ben02]). So when one agent presents an argument, the other agent may provides a counterargument, and then the first agent may provide a counter-counterargument.

In this way, a depth-first search of the argument graph is undertaken. With the aim of having more natural dialogues, we used a breadth-first approach. So when a user selects arguments from the menu, the system then may attack more than one of the arguments selected. For the argument graph we used in our study, this appeared to work well. However, for larger argument graphs, a breadth-first approach could also be unnatural. This then raises the questions of how to specify a protocol that interleaves depth-first and breadth-first approaches, and of how to undertake studies with participants to evaluate such protocols.

In future work, we plan to undertake empirical studies with the same research questions but a different domain. The aim would be to see whether we can get similar results in a new domain. In the final part of the empirical study reported in this paper, we had the majority of participants having a starting position of agreement with the stance that we were arguing for. In future studies, we will seek topics for discussion where fewer of the participants start by agreeing with the stance that we are arguing for. This may enable us to demonstrate a bigger effect. We also intend to combine the approach reported here of using concerns of the participant with the beliefs of the participant (as we reported in [HH17]) and undertake empirical studies to see if we can further improve the degree of persuasion.

Finally, a better understanding of concerns, and the potential impacts that arguments may have on them, may facilitate a better understanding of the nature of persuasion dialogues. Consider two agents in a dialogue. If one of them presents an argument A_1 with type T, and the other presents a counterargument A_2 to A_1 that also has type T, then normally, we would expect that A_2 has the opposite impact to that of A_1. So if A_1 has a positive impact on T, then A_2 has a negative impact on T, and if A_1 has a negative impact on T, then A_2 has a positive impact on T.

Example 6 *Consider the concern type "Personal Economy" and the domain "Cycling in the city". The following are arguments of type "Personal Economy". Assuming this type means that the agent has the concern that she wants to economize on personal expenditure, then the argument A3 has a positive impact on the concern "Personal Economy" and the counterargument A4 has a negative impact on the concern "Personal Economy".*

- *(A3) Cycling saves money on public transport.*
- *(A4) Bikes are expensive.*

We will investigate the patterns arising for types of concern, including whether there are advantages or disadvantages to maintaining or switching the type of concern during a dialogue in the future.

Acknowledgements

This research was funded by EPSRC Project EP/N008294/1 "Framework for Computational Persuasion". The authors are grateful to Lisa Chalaguine and Sylwia Polberg for valuable feedback on earlier versions of this paper. The authors are also grateful to the anonymous reviewers for numerous suggestions for improvements to the paper.

References

[ABB12] K. Atkinson, P. Bench-Capon, and T. Bench-Capon. Value-based argumentation for democratic decision support. In *Proceedings of the 4th International Conference on Agents and Artificial Intelligence. (ICAART'12)*, pages 23–32. Scitepress, 2012.

[AMP00a] L. Amgoud, N. Maudet, and S. Parsons. Arguments, dialogue and negotiation. In *Proceedings of the 14th European Conference on Artifcial Intelligence (ECAI'00)*, pages 338–342. IOS Press, 2000.

[AMP00b] L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In *Proceedings of the 4th International Conference on Multi-Agent Systems (ICMAS'00)*, pages 31–38. IEEE Computer Society, 2000.

[Atk06] K. Atkinson. Value-based argumentation for democratic decision support. In *Proceedings of the 1st International Conference on Computational Models of Argument (COMMA'06)*, pages 47–58. IOS Press, 2006.

[AW13] K. Atkinson and A. Wyner. The value of values in computational argumentation. In *From Knowledge Representation to Argumentation in AI, Law and Policy Making: A Festschrift in Honour of Trevor Bench-Capon on the Occasion of His 60th Birthday*, pages 39–62. College Publications, 2013.

[BC03] T. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.

[BCA09] T. Bench-Capon and K. Atkinson. Abstract argumentation and values. In *Argumentation in Artificial Intelligence*, pages 45–64. Springer, 2009.

[BCAC05] T. Bench-Capon, K. Atkinson, and A. Chorley. Persuasion and value in legal argument. *Journal of Logic and Computation*, 15(6):1075–1097, 2005.

[BCB14] E. Black, A. Coles, and S. Bernardini. Automated planning of simple persuasion dialogues. In *Proceedings of the International Workshop on Computational Logic in Multi-agent Systems (CLIMA'14)*, volume 8624 of *LNCS*, pages 87–104. Springer, 2014.

[BCH17] E. Black, A. Coles, and C. Hampson. Planning for persuasion. In *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'17)*, pages 933–942. International Foundation for Autonomous Agents and Multiagent Systems, 2017.

[Ben02] T. Bench-Capon. Agreeeing to differ: Modelling persuasive dialogue between parties with different values. *Informal Logic*, 22:231–246, 2002.

[BGGv18] P. Baroni, D. Gabbay, M. Giacomin, and L. van der Torre, editors. *Handbook Of Formal Argumentation*. College Publications, 2018.

[BGV14] P. Baroni, M. Giacomin, and P. Vicig. On rationality conditions for epistemic probabilities in abstract argumentation. In *Proceedings of the 5th International Conference on Computational Models of Argument (COMMA'14)*, pages 121–132. IOS Press, 2014.

[BH09] E. Black and A. Hunter. An inquiry dialogue system. *Autonomous Agents and Multi-Agent Systems*, 19(2):173–209, 2009.

[CGJ99] A. Cawsey, F. Grasso, and R. Jones. A conversational model for health promotion on the world wide web. In *Proceedings of the Joint European Conference on AI in Medicine and Medical Decision Making*, volume 1620 of *LNAI*, pages 379–388. Springer, 1999.

[CHH+18] L. Chalaguine, E. Hadoux, F. Hamilton, A. Hayward, A. Hunter, S. Polberg, and H. Potts. Domain modelling in computational persuasion for behaviour change in healthcare. *ArXiv*, 2018. arXiv:1802.10054 [cs.AI].

[CHHP18] L. Chalaguine, F. Hamilton, A. Hunter, and H. Potts. Argument harvesting using chatbots. In *Proceedings of the 7th International Conference on Computational Models of Argument*, pages 149–160. IOS Press, 2018.

[CP04] L. Chen and P. Pu. Survey of preference elicitation methods. Technical Report IC/2004/67, EPFL, 2004.

[CP12] M. Caminada and M. Podlaszewski. Grounded semantics as persuasion dialogue. In *Proceedings of the 4th International Conference on Computational Models of Argument (COMMA'12)*, pages 478–485. IOS Press, 2012.

[CTO14] F. Cerutti, N. Tintarev, and N. Oren. Formal arguments, preferences, and natural language interfaces to humans: an empirical evaluation. In *Proceedings of the 21st European Conference on Artificial Intelligence (ECAI'14)*, pages 1033–1034. IOS Press, 2014.

[DDKV00] F. Dignum, B. Dunin-Keplicz, and R. Verbrugge. Dialogue in team formation. In *Issues in Agent Communication*, volume 1916 of *LNCS*, pages 264–280. Springer, 2000.

[Dun95] P. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n-person games. *Artificial Intelligence*, 77:321–357, 1995.

[Fle71] J. Fleiss. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378, 1971.

[FT11] X. Fan and F. Toni. Assumption-based argumentation dialogues. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI'11)*, pages 198–203. AAAI Press, 2011.

[FT12] X. Fan and F. Toni. Mechanism design for argumentation-based persuasion. In *Proceedings of the 4th International Conference on Computational Models of Argument (COMMA'12)*, pages 322–333. IOS Press, 2012.

[FU81] R. Fisher and W. Ury. *Getting to Yes: Negotiating Agreement Without Giving In*. Penguin, 1981.

[GCJ00] F. Grasso, A. Cawsey, and R. Jones. Dialectical argumentation to solve conflicts in advice giving: a case study in the promotion of healthy nutrition. *International Journal of Human-Computer Studies*, 53(6):1077–1115, 2000.

[GNL16] E. Guerrero, J. Nieves, and H. Lindgren. An activity-centric argumentation framework for assistive technology aimed at improving health. *Argument and Computation*, 7:5–33, 2016.

[Gra98] F. Grasso. Exciting avocados and dull pears - combining behavioural and argumentative theory for producing effective advice. In *In Proceedings of the 20th Annual Meeting of the Cognitive Science Society*, pages 436–441. Lawrence Erlbaum Associates, 1998.

[Gra03] F. Grasso. Rhetorical coding of health promotion dialogues. In *Proceedings of the 9th Conference on Artificial Intelligence in Medicine (AIME'03)*, volume 2780 of *LNCS*, pages 179–188, 2003.

[GSRS03] S. Gosling, D. Samuel, P. Rentfrow, and W. Swann. A very brief measure of the big-five personality domains. *Journal of Research in Personality*, 37(6):504–528, 2003.

[Ham71] C. Hamblin. Mathematical models of dialogue. *Theoria*, 37:567–583, 1971.

[HBM+15] E. Hadoux, A. Beynier, N. Maudet, P. Weng, and A. Hunter. Optimization of probabilistic argumentation with Markov decision models. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI,15)*, pages 2004–2010. AAAI Press, 2015.

[HH16] E. Hadoux and A. Hunter. Computationally viable handling of beliefs in arguments for persuasion. In *Proceedings of the 28th International Conference on Tools with Artificial Intelligence (ICTAI'16)*, pages 319–326. IEEE Press, 2016.

[HH17] E. Hadoux and A. Hunter. Strategic sequences of arguments for persuasion using decision trees. In *Proceeding of the 31st AAAI Conference on Artificial Intelligence (AAAI'17)*, pages 1128–1134. AAAI Press, 2017.

[HH18] E. Hadoux and A. Hunter. Learning and updating user models for subpopulations in persuasive argumentation using beta distributions. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'18)*, pages 1141–1149. International Foundation for Autonomous Agents and Multiagent System, 2018.

[HHC18] E. Hadoux, A. Hunter, and J.-B. Corrégé. Strategic dialogical argumentation using multi-criteria decision making with application to epistemic and emotional aspects of arguments. In *Proceedings of the 10th Internataional Symposium on Foundations of Information and Knowledge Systems (Foiks'18)*, volume 10833 of *LNCS*, pages 207–224. Springer, 2018.

[HL07] S. Huang and F. Lin. The design and evaluation of an intelligent sales agent for online persuasion and negotiation. *Electronic Commerce Research and Applications*, 6:285–296, 2007.

[HMP01] D. Hitchcock, P. McBurney, and S. Parsons. A framework for deliberation dialogues. In *Proceedings of the 4th Biennial Conference of the Ontario Society for the Study of Argumentation (OSSA'01)*. The Ontario Society for the Study of Argumentation, 2001.

[HP17a] A. Hunter and S. Polberg. Empirical methods for modelling persuadees in dialogical argumentation. In *Proceedings of the 29th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'17)*. IEEE Computer Society Press, 2017.

[HP17b] A. Hunter and N. Potyka. Updating probabilistic epistemic states in persuasion dialogues. In *Proceedings of the European Conference on Symbolic and Quantitative Approaches to Reasoning with*

Uncertainty (ECSQARU'17), volume 10369 of *LNCS*, pages 46–56. Springer, 2017.

[HPP18] A. Hunter, S. Polberg, and S. Potyka. Proceedings of the 16th international conference on principles of knowledge representation and reasoning, (KR'18). pages 138–147. AAAI Press, 2018.

[HPT18] A. Hunter, S. Polberg, and M. Thimm. Epistemic graphs for representing and reasoning with positive and negative influences of arguments. *ArXiv*, 2018. arXiv:1802.07489 [cs.AI].

[HSM+13] C. Hadjinikolis, Y. Siantos, S. Modgil, E. Black, and P. McBurney. Opponent modelling in persuasion dialogues. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI'15)*, pages 164–170. AAAI Press, 2013.

[HT16] A. Hunter and M. Thimm. Optimization of dialectical outcomes in dialogical argumentation. *International Journal of Approximate Reasoning,*, 78:73–102, 2016.

[Hun13] A. Hunter. A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning*, 54(1):47–81, 2013.

[Hun14] A. Hunter. Probabilistic strategies in dialogical argumentation. In *Proceedings of the 8th International Conference on Scalable Uncertainty Management (SUM'14)*, volume 8720 of *LNCS*, pages 190–202. Springer, 2014.

[Hun15] A. Hunter. Modelling the persuadee in asymmetric argumentation dialogues for persuasion. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI'15)*, pages 3055–3061. AAAI Press, 2015.

[Hun16a] A. Hunter. Computational persuasion with applications in behaviour change. In *Proceedings of 6th International Conference on Computational Models of Argument (COMMA'16)*, pages 5–18. IOS Press, 2016.

[Hun16b] A. Hunter. Persuasion dialogues via restricted interfaces using probabilistic argumentation. In *Proceedings of the 10th International Conference in Scalable Uncertainty Management (SUM'16)*, volume 9858 of *LNCS*, pages 184–198. Springer, 2016.

[Hun16c] A. Hunter. Two dimensional uncertainty in persuadee modelling in argumentation. In *Proceedings of the 22nd European Conference on Artificial Intelligence (ECAI'16)*, pages 150–157. IOS Press, 2016.

[LAWW17] S. Lukin, P. Anand, M. Walker, and S. Whittaker. Argument strength is in the eye of the beholder: Audience effects in persuasion. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL'17): Volume 1, Long Papers*, pages 742–753. Association for Computational Linguistics, 2017.

[Mac79] J. Mackenzie. Question begging in non-cumulative systems. *Journal of Philosophical Logic*, 8:117–133, 1979.

[MBL16] J. Murphy, E. Black, and M. Luck. Heuristic tratego for persuasion. In *Proceedings of the 6th International Conference on Computational Models of Argument (COMMA'16)*, pages 411 – 418. IOS Press, 2016.

[MBL+18] J. Murphy, A. Burdusel, M. Luck, S. Zschaler, and E. Black. Deriving persuasion strategies using search-based model engineering. In *Proceedings of the 7th International Conference on Computational Models of Argument (COMMA'18)*, pages 221–232. IOS Press, 2018.

[MC87] R. McCrae and P. Costa. Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52(1):81, 1987.

[ME98] N. Maudet and F. Evrard. A generic framework for dialogue game implementation. In *Proceedings of the 2nd Workshop on Formal Semantics & Pragmatics of Dialogue*, page 185–198. University of Twente, 1998.

[MP02a] P. McBurney and S. Parsons. Dialogue games in multi-agent systems. *Informal Logic*, 22:257–274, 2002.

[MP02b] P. McBurney and S. Parsons. Games that agents play: A formal framework for dialogues between autonomous agents. *Journal of Logic, Language and Information*, 11:315–334, 2002.

[MvEPA03] P. McBurney, R. van Eijk, S. Parsons, and L. Amgoud. A dialogue-game protocol for agent purchase negotiations. *Journal of Autonomous Agents and Multi-Agent Systems*, 7:235–273, 2003.

[NME07] H. Nguyen, J. Masthoff, and P. Edwards. Persuasive effects of embodied conversational agent teams. In *Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments: 12th International Conference, HCI International 2007*, volume 4552 of *LNCS*, pages 176–185. Springer, 2007.

[PH18] S. Polberg and A. Hunter. Empirical evaluation of abstract argumentation: Supporting the need for bipolar and probabilistic approaches. *International Journal of Approximate Reasoning*, 93:487–543, 2018.

[PHT17] S. Polberg, A. Hunter, and M. Thimm. Belief in attacks in epistemic probabilistic argumentation. In *Proceedings of the 11th International Conference on Scalable Uncertainty Management (SUM'17)*, volume 10564 of *LNCS*, pages 223–236. Springer, 2017.

[Pra05] H. Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, 15(6):1009–1040, 2005.

[Pra06] H. Prakken. Formal sytems for persuasion dialogue. *Knowledge Engineering Review*, 21(2):163–188, 2006.

[RK16] A. Rosenfeld and S. Kraus. Providing arguments in discussions on the basis of the prediction of human argumentative behavior. *ACM Transactions on Interactive Intelligent Systems*, 6(4):30:1–30:33, December 2016.

[RL08] I. Rahwan and K. Larson. Pareto optimality in abstract argumentation. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence (AAAI'08)*. AAAI Press, 2008.

[RLT09] I Rahwan, K Larson, and F Tohmé. A characterisation of strategy-proofness for grounded argumentation semantics. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI'09)*, pages 251–256. AAAI Press, 2009.

[RMB+10] I. Rahwan, M Madakkatel, J. Bonnefon, R Awan, and S. Abdallah. Behavioural experiments for assessing the abstract argumentation semantics of reinstatement. *Cognitive Science*, 34(8):1483–1502, 2010.

[RPSD09] I. Rahwan, P. Pasquier, L. Sonenberg, and F. Dignum. A formal analysis of interest-based negotiation. *Annuals of Mathematics and Artificial Intelligence*, 55:253–276, 2009.

[RTO13] T. Rienstra, M. Thimm, and N. Oren. Opponent models with uncertainty for strategic argumentation. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI'13)*, pages 332–338. AAAI Press, 2013.

[Thi12] M. Thimm. A probabilistic semantics for abstract argumentation. In *Proceedings of the 20th European Conference on Artificial Intelligence (ECAI'12)*, volume 242 of *Frontiers in Artificial Intelligence and Applications*, pages 750–755. IOS Press, 2012.

[Thi14] M. Thimm. Strategic argumentation in multi-agent systems. *Künstliche Intelligenz*, 28:159–168, 2014.

[VCJ+17] S. Villata, E. Cabrio, I. Jraidi, S. Benlamine, M. Chaouachi, C. Frasson, and F. Gandon. Emotions and personality traits in argumentation: An empirical evaluation. *Argument & Computation*, 8(1):61–87, 2017.

[VSM+13] J. Vargheese, S. Sripada, J. Masthoff, N. Oren, P. Schofield, and V. Hanson. Persuasive dialogue for older adults: promoting and encouraging social interaction. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 877–882. ACM Press, 2013.

[VSMO16] J. Vargheese, S. Sripada, J. Masthoff, and N. Oren. Persuasive strategies for encouraging social interaction for older adults. *International Journal of Human Computer Interaction*, 32(3):190–214, 2016.

A List of the arguments

8miles People cycle 8 miles on average to go to work.

pollution Pollution is everyone's problem.

further I live further than 8 miles away.

weight I am not fit enough because of my weight.

parking Many companies now have secure bike parking for the employees not to waste time on looking for one.

home My home is too small to store a bike.

clothes You can have work clothes in a bag and wear cycling clothes.

stamina I lack the stamina for cycling far.

relaxing Even if cycling does take more time, it is a more relaxing way to travel if you avoid busy roads.

unhealthy Cycling in a city is unhealthy because of the exposure to the pollution.

savestime Cycling saves time because there is no need to find a car parking space.

lanes The city can build cycle-only lanes to remove cyclists from the traffic.

taxes It is unfair everybody has to pay the infrastructures with taxes for the use by only some.

heart Cycling is good for your heart and muscles.

hassle It is a hassle having to change clothes.

inexpensive Plenty of shops sell inexpensive cycling clothes and equipment.

cannot I cannot physically cycle.

fitness Cycling improves your fitness while you are commuting or doing errands.

facilities Many companies have shower facilities.

lessshopping Cycle-only lanes will decrease shopping because of hassle to drivers.

wastedtime Time is wasted looking for a secure place to lock the bike.

encourage Traffic jams encourage people to cycle.

expensive Buying cycle clothing is too expensive.

toofar My workplace is too far away from my home.

bikedangerous Cycling in the city is dangerous for the cyclists.

dontknow But many cyclists don't know how to ride in traffic.

killmore Cars kill more pedestrians than bikes.

expensivelanes Building cycle lanes is too expensive for the city.

lesspollution Studies show that cyclists are subject to lower pollution levels than drivers.

bright Wearing bright cycling clothes makes cycling much safer.

courses The city can arrange free cycle courses to improve road safety.

shower Changing clothes does not remove the need to shower.

footdanger Cycling in the city is dangerous for the pedestrians.

lessdrivers More cyclists means less drivers and less pollution.

cheaper Cycling is cheaper for the citizens than driving or public transportation.

shopmore Studies show that cyclists come to the shops more frequently than drivers and so are good customers.

weightloss Cycling improves weight loss.

rain You are at risk of rain and your clothes can get wet.

carrymore Cars can carry more shopping than bikes and so contribute more to the local economy.

uncomfortable Work clothes are uncomfortable for cycling.

trafficjams Cycle lanes create traffic jams.

lessparking Creating bike parking means less car parking is available and that can affect the local economy.

moretime It takes more time to travel by bike.

infrastructures Evidence shows that infrastructures for cyclists favour the local economy generating more taxes for the city to use.

habit Changing clothes is just a habit that you do quickly at the start and end of the working day.

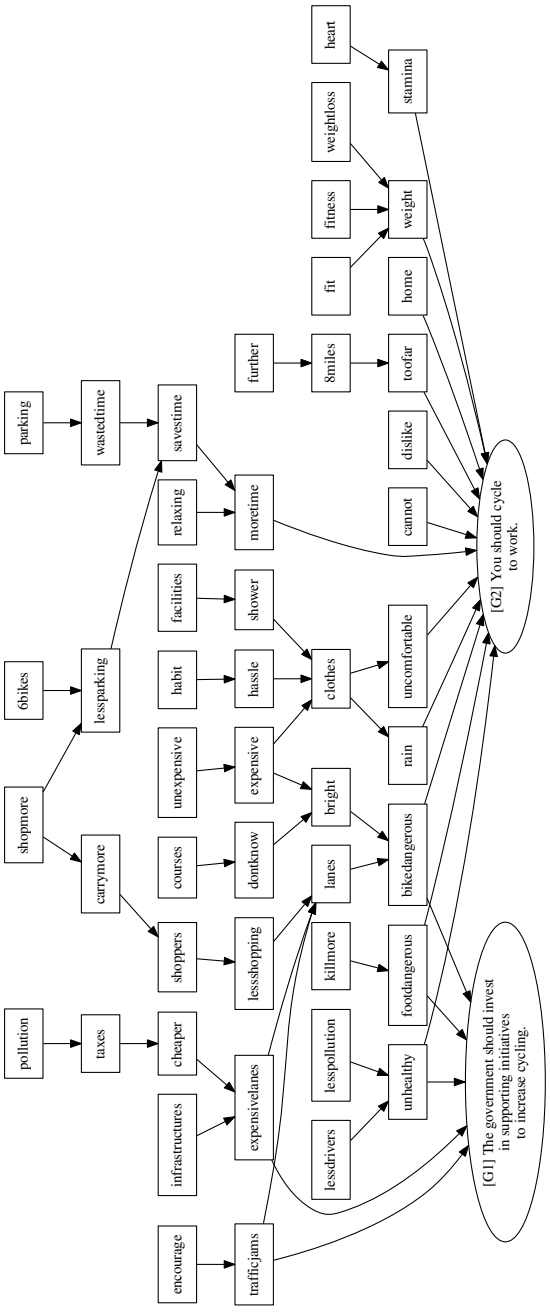
6bikes Each car parking space can accommodate 6 bikes which means more people can be catered for and that can boost the local economy.

shoppers Many cyclists are also shoppers.

fit Cycling is an easy way to get fit.

dislike I do not like to cycle.

B Argument graph



C Association of types of concern to arguments

Table 4 shows the results regarding the associations. The first column is the short argument name as specified in Appendix A. The types in bold are the types kept for the following steps of the experiments. To select them, we kept all the types having strictly more than half of the votes of the most voted type. For instance, if the most voted type gathered 8 votes over 18, all types with strictly more than 4 votes were kept in the association.

Arguments	Types					
8miles	**Fitness**	8	**Time**	7	Env.	3
pollution	**Env.**	14	Health	4		
further	**Time**	11	Fitness	2	P. Eco.	2
	Comfort	2	Env.	1		
weight	**Fitness**	13	Health	5		
parking	**Time**	12	Safety	3	Comfort	2
	C. Eco.	1				
home	**P. Eco.**	11	Comfort	5	Env.	2
clothes	**Comfort**	15	P. Eco.	3		
stamina	**Fitness**	10	**Health**	6	Comfort	2
relaxing	**Health**	6	**Comfort**	5	**Time**	5
	Env.	1	P. Eco.	1		
unhealthy	**Health**	14	Env.	4		
savestime	**Time**	17	P. Eco.	1		
lanes	**Safety**	11	**C. Eco**	7		
taxes	**P. Eco.**	10	**C. Eco.**	8		
dislike	**Comfort**	12	P.Economy	4	Health	1
	Fitness	1				
lesspollution	**Health**	11	Env.	5	C. Eco.	2
bright	**Safety**	17	Health	1		
courses	**Safety**	15	C. Eco.	3		
fit	**Fitness**	16	Health	2		
shoppers	**C. Eco.**	12	P. Eco.	6		
shower	**Comfort**	12	Health	3	P. Eco.	2
	Time	1				
footdangerous	**Safety**	14	C. Eco.	2	Env.	2
6bikes	**C. Eco.**	15	P. Eco	1	Comfort	1
	Env.	1				
lessdrivers	**Env.**	15	Health	3		
cheaper	**P. Eco**	15	C. Eco.	3		
shopmore	**C. Eco.**	13	P. Eco.	4	Env.	1
weightloss	**Fitness**	11	Health	7		
rain	**Comfort**	13	Health	3	Env.	2
carrymore	**C. Eco.**	12	P. Eco.	3	Comfort	3

Table 4: Results for the argument/types association experiment

Arguments	Types					
uncomfortable	**Comfort**	17	P. Eco.	1		
trafficjams	**Time**	7	**Env.**	4	**C. Eco.**	4
	Safety	2	Comfort	1		
lessparking	**C. Eco.**	12	P. Eco.	2	Comfort	1
habit	**Time**	10	**Comfort**	7	P. Eco.	1
heart	**Health**	11	**Fitness**	7		
hassle	**Time**	9	**Comfort**	7	P. Eco.	2
unexpensive	**P. Eco.**	15	C. Eco.	2	Comfort	1
cannot	**Health**	7	**Fitness**	6	Comfort	2
	P. Eco.	2	Safety	1		
fitness	**Fitness**	14	Health	4		
facilities	**Comfort**	10	P. Eco.	4	Time	1
	Health	1	C. Eco.	1	Fitness	1
lessshopping	**C. Eco.**	11	Env.	3	P. Eco.	1
	Safety	1	Health	1	Comfort	1
wastedtime	**Time**	18				
encourage	**Time**	11	Env.	3	C. Eco.	2
	Safety	1	Health	1		
expensive	**P. Eco.**	18				
toofar	**Time**	11	Fitness	3	Comfort	2
	Env.	1	P. Eco.	1		
bikedangerous	**Safety**	18				
dontknow	**Safety**	15	Time	1	Health	1
	C. Eco.	1				
killmore	**Safety**	15	Health	2	C. Eco.	1
expensivelanes	**C. Eco.**	17	Env.	1		
moretime	**Time**	18				
infrastructures	**C. Eco.**	18				

Table 4: Results for the argument/types association experiment (cont'd)

D Argument/types association survey

Title: "Associate arguments about cycling with topics"

Description: "The purpose of this task is to associate each argument with the topic that you find is the most closely related. Please note that everything is anonymous and no information can be used to personally identify you. Moreover you can choose to quit this survey at any point in time."

Step 1: "Please enter your Prolific ID."

Step 2: "For each argument, please select the most appropriate topic to categorize it." For 9 arguments and 1 attention check.

Step 3: Identical for 9 other arguments and 1 attention check.

Step 4: Identical for 10 other arguments.

Step 5: Identical for 10 other arguments.

Step 6: Identical for the 11 remaining arguments.

We broke the list of arguments in 5 different steps in order for them to look less daunting to the participants. Note that the arguments inside each step were presented in a random order to each participants. However, the split of the list amongst the steps was the same.

E Pairwise preferences survey

Title: "Preferences on argumentation topics"

Description: "We first ask you some questions in order to create a profile. Please note that everything is anonymous and no information can be used to personally identify you. Moreover you can choose to quit this survey at any point in time."

Step 1: "Please enter your Prolific ID."

Step 2: "Please select your age in the list." "18-24", "25-34", "35-44", "45-54", "55-64", "65+"

Age group	18-24	25-34	35-44	45-54	55-64	65+
Proportion	10.2%	47.0%	14.3%	20.4%	6.1%	2.0%

Table 5: Age distribution

Step 3: "Please select the category of your occupation."

1. "Non working"
2. "Students"
3. "Semi-skilled and unskilled manual workers"
4. "Skilled manual workers"
5. "Supervisory or clerical and junior managerial, administrative or professional"
6. "Intermediate managerial, administrative or professional"
7. "Higher managerial, administrative or professional"
8. "Other"

The results from this question are given in Table 6.

Category	1	2	3	4	5	6	7	8
Proportion	10.2%	16.3%	12.2%	4.1%	26.5%	22.5%	4.1%	4.1%

Table 6: Occupation repartition

Step 4: "Please select your sex." "Female", "Male", "Other" The results from this question are given in Table 7.

Category	Female	Male	Other
Proportion	69.4%	30.6%	0%

Table 7: Gender distribution

Category	0	1	2	3+
Proportion	61.2%	20.4%	10.2%	8.2%

Table 8: Number of children

Step 5: "Please select your number of children who are below 18 years old." "0", "1", "2", "3+" The results from this question are given in Table 8.

Step 6: "Please select your home location." The results from this question are given in Table 9.

1. "Town with 100,000+ inhab."
2. "Suburban area of a town with 100,000+ inhab."
3. "Town with more than 10,000 inhab."
4. "Smaller town"
5. "Countryside"

Step 7: Idem but for the workplace instead of the home. The results from this question are given in Table 9.

Category	1	2	3	4	5
Home	36.7%	22.5%	24.5%	12.2%	4.1%
Workplace	47.0%	20.4%	18.3%	14.3%	0%

Table 9: Home and workplace location

Step 8: "Please enter an estimation of the distance in miles. For information, a mile is approximately 1.6 km."

If we remove the 4 very long distances of 4 participants (30, 22, 20 and 20), the average distance is 3.6 miles.

Step 9: "I cycle from my home to my workplace."

The result was that 9 participants answered yes, with an average of 3.4 miles.

Step 10: "I experience a condition that is making it hard for me to move/walk/cycle."

The result was that 11 participants answered yes, most of them being above the participants' average age.

Step 11: We asked for the Ten-Item Personality Inventory (TIPI) in the exact same way it is described in the original work [GSRS03].

Step 12: "For each pair of topics, please state if you think that the left/right hand side topic is a more important priority for you, if both are equal or if you cannot compare them."

- "I prefer the first"
- "I prefer the second"
- "They are equal to me"
- "I cannot compare them"

Two attention checks are inserted into the set of pairs to evaluate where we asked explicitly to choose the first or the second answer.

F Dialogue with a chatbot

Title: "Cycling in the city"

Description: "You are about to take part in a survey concerning cycling from home to work. This is completely anonymous and you can decide to stop your participation at any point in time. In addition, there is no right or wrong answer. We are only interested in what you think."

Step 1: "Please enter your Prolific ID."

Step 2: "Please select your sex from the list." "Female", "Male", "Other" The results from this question are given in Table 10.

Category	Female	Male	Other
Proportion	64.6%	35.4%	0%

Table 10: Gender distribution

Step 3: We asked for the Ten-Item Personality Inventory (TIPI) in the exact same way it is described in the original work [GSRS03].

Step 4: "What is your position towards cycling from home to work?" The participant could select a value between -5 and 5 with a 0.01 step.

Step 5: "What is your position on the government investing in initiatives and infrastructures for cyclists?" The participant could select a value between -5 and 5 with a 0.01 step.

Step 6: "Please read and interact with the chat box below. You will not be able to type in your answer. However, you will be able to click on messages when asked to do so." This is the part of the study where the dialogue is undertaken with the participant. Each dialogue is a sequence of arguments, and menus of counterarguments, that are presented to a participants, and the participant is asked to select counterarguments from each menu (as illustrated in Figure 1.

Step 7: "What is your position towards cycling from home to work?" The participant could select a value between -5 and 5 with a 0.01 step. Step 7 is the same as Step 4 except it is asked after the dialogue.

Step 8: "What is your position on the government investing in initiatives and infrastructures for cyclists?" The participant could select a value between -5 and 5 with a 0.01 step. Step 8 is the same as Step 5 except it is asked after the dialogue.

Step 9: "Do you have any additional argument concerning cycling in the city?"

Note that this survey is identical whether it is using the preference-based chatbot or the random one. Only the arguments played in Question 6 are chosen differently.