

Towards a Formal Framework for Motivated Argumentation and the Roots of Conflict

Tomasz Zurek^{1,**,†}, Adam Wyner^{2,*,†}

¹*T.M.C. Asser Institute, University of Amsterdam, Netherlands*

²*Department of Computer Science, Swansea University, United Kingdom*

Abstract

In computational argumentation, values adjudicate between conflicting arguments, where values hold of an argument as a whole rather than of any constituent parts; preference rankings between values determine the winning argument. We propose a novel formal framework towards an account for motivated reasoning, which is the widespread, natural observation that an agent constructs (instantiated) arguments from those propositions which are a selective subset of the set of all available propositions; and more specifically, an agent selects those propositions which accord with their values; as such, the propositions and arguments indirectly reflect an agent's values. Conflicts between arguments are grounded in conflicts between the values associated with the constituent propositions rather than with the arguments per se.

Keywords

argumentation, values, knowledge-base

1. Introduction

Values are widely acknowledged as a way to select amongst alternative arguments and resolve conflict in reasoning and decision-making. Kahneman discusses how values play a role in everyday motivated reasoning [1]; Perelman highlights values in judicial reasoning [2]; O'Callaghan Richards provides empirical evidence to support the view that judges decide relative to their values [3]. There are various models of values, e.g., Schwarz's model of values [4] such as *conformity*, *tradition*, and *security*. Searle [5] and Kahneman [1] relate values to "facts-of-the-world".

In computational argumentation, values adjudicate conflicts between arguments [6], where values are ordered in a hierarchy and are properties associated with arguments as a whole. Values are generally treated abstractly. Each value ordering can be taken to represent an *audience*. In abstract argumentation frameworks [7], values adjudicate attacks, where an attacking argument

CMNA'22: *Computational Models of Natural Argument*, Sept 12, 2022, Cardiff, UK

*Corresponding author.

**Tomasz Zurek received funding from the Dutch Research Council (NWO) Platform for Responsible Innovation (NWO-MVI) as part of the DILEMA Project on Designing International Law and Ethics into Military Artificial Intelligence. The authors thank anonymous reviewers for insightful comments and suggestions.

†These authors contributed equally.

✉ t.zurek@asser.nl (T. Zurek); a.z.wyner@swansea.ac.uk (A. Wyner)

🌐 <https://tomaszzurek.wordpress.com/> (T. Zurek); <https://www.swansea.ac.uk/staff/a.z.wyner/> (A. Wyner)

🆔 0000-0002-9129-3157 (T. Zurek); 0000-0002-2958-3428 (A. Wyner)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

with the value higher on the value ordering than the attacked argument “wins”. The winning argument can be said to “promote” the value.

While formally clear and well-developed, the relationship between an argument and associated value is opaque. In addition, a key aim of our work is account for *motivated reasoning*, where from a common pool of accessible information, agents use their values to select how to construct their arguments; e.g., in justifying what item to purchase such as a camera, each purchaser associates an item’s features with values relative to them, thus coming to a highly subjective justification [8]. Relatedly, individuals are often asked to assign values to statements as in surveys or in social media assessments. Furthermore, an agent may justify their position without fully arguing against a contrary position by another agent. Current computational approaches to argumentation do not address how arguments are constructed relative to an agent’s values.

To formalise motivated reasoning, we assume: each agent has a value profile that expresses what is abstractly important to the agent and to what degree, e.g., *family*, *freedom*, *security*, and others; each agent assesses the relation between propositions and values. This is grounded in the observation that, apart from logical contradictions (i.e., p vs $\neg p$) and commonsense incompatibilities (i.e., an agent cannot be in two places simultaneously) which are “objective relations”, there are conflicts which are subjective and related to values. An agent filters from the available propositions those in an agent’s knowledge base, which is a *subjective knowledge base*; from these available propositions, the agent constructs their arguments. As agents differ in their value profiles, then so too do their knowledge bases and arguments. As values profiles and relations of propositions to values can conflict, conflicts then arise between arguments.

In this paper, we initially develop a formal, novel approach to motivated argumentation as *value-based instantiated argumentation* (VIA), which represents how values intervene in argumentation. The present work is focused on the development of subjective knowledge bases, which can be used in instantiated theories of argumentation. In this regard, VIA is scoped and agnostic with respect to theories of abstract or instantiated argumentation theories (e.g., [9, 10, 11, 12, 13] among others); the presentation of the knowledge base and chains of reasoning (aka arguments) is a simplification and used to highlight some key aspects. We abstract from context as well as structure or inferences amongst values. Importantly, the work does not propose a novel argumentation framework, argumentation semantics, or treatment of dialogue.

The structure of the paper follows the layers of analysis. In Section 2, we provide the vocabulary, basic predicates, and filters which are built relations amongst agents, propositions, values, and weights. In Section 3, subjective knowledge bases and reasoning chains are detailed, where reasoning chains are an argumentation-theoretic agnostic approach to chaining together of rules. Section 4 presents a core, novel concept of conflict between arguments which is rooted in the conflict between values. A worked example illustrates the formalisations. We conclude with a discussion of related work in Section 5 and discussion in Section 6.

2. Agents, Propositions, Values, and Weights

In this section, we develop a language for knowledge bases and reasoning. We assume the following denotations.

- a finite set of agents $\text{Agent} = \{\text{agent}_0, \dots, \text{agent}_n\}$ of entities.

- a finite set `AtomicProposition` of atomic propositions (aka prop).
- a finite set `incompProp` containing pairs of elements of `AtomicProposition` which cannot co-occur. The relation is symmetric. The props of such a pair are called *objectively incompatible*, otherwise, they are *objectively compatible*.
- a finite set of values `Value` = {value₀, ..., value_n} of abstract objects.
- a totally ordered, finite set of scalar elements `Scale` = {weight₀, ..., weight_n}.
- a designated unordered element `?`, used where a weight is indeterminate.
- a set `Weight` = `Scale` ∪ {`?`}. There may be alternative interpretations of ‘weights’. Here, they reflect the relative ‘importance’ to an agent, e.g., *family* might be a very important value and *personal status* very unimportant.
- as `?` is unordered with respect to other elements of `Scale`, any prop comparing `?` to the other entity is false, e.g., `? > weight1` is false.

To quantify over any type in the basic vocabulary, we have: variables for each type, represented by Greek subscripts, e.g., `agentα`, `agentβ`,; and constants of each type represented by Latin subscripts `agenta`, `agentb`, In addition, we can quantify over variables and constants over tuples defined with respect to the basic vocabulary.

Basic Predicates With the following predicates, we construct the predicate `PropBaseCleanagenth` for some `agenth`, which contains all and only those props that are filtered according to an agent’s value profile. We show how Agents may “hold” different sets of props relative to their values such that values can be taken as the root of conflict.

`AgentValue` represents all the values an agent might consider.

$$\text{AgentValue} = \{ \langle \text{agent}_\alpha, \text{value}_\beta \rangle \mid \forall \text{agent}_\alpha \forall \text{value}_\beta (\text{agent}_\alpha \in \text{Agent} \wedge \text{value}_\beta \in \text{Value}) \}$$

From this, we construct an agent’s *value profile*, `AgentValueToWeight`, which indicates the degree of importance that the agent ascribes to the value, where the higher the weight, the more important and the lower the weight the less important. Given an agent and Value, the value can only have one weight in order to avoid conflicts.

$$\text{AgentValueToWeight} = \{ \langle \langle \text{agent}_a, \text{value}_b \rangle, \text{weight}_c \rangle \mid \forall \text{agent}_a \forall \text{value}_b \exists! \text{weight}_c (\langle \text{agent}_a, \text{value}_b \rangle \in \text{AgentValue} \wedge \text{weight}_c \in \text{Weight}) \}$$

`AgentValueToWeight` is a total function. We indicate each agent’s value profile with a subscript, e.g., `AgentValueToWeightagentk` for agent `agentk`. Given the `?` weight, the importance an agent associates with a value can be indeterminate, reflecting the view that the agent has no specific association with respect to the value. Note that any two agents may ascribe different weights to the same value, perhaps at opposite ends of the `Scale`; as such, this can be taken to represent antithetical assessments on the values, one agent viewing the value as more important and the other less important. Note that since differences in `AgentValueToWeight` of particular agents represent differences in the levels of importance the agents puts on those values, they can be seen as a subjective or personal value profiles of those agents.

To represent how an agent assesses an `AtomicProposition` with respect to a `Value` and a `Weight`, we first associate `Values` and `Weights`.

$$\text{ValueWeight} = \{ \langle \text{value}_\alpha, \text{weight}_\beta \rangle \mid \forall \text{value}_\alpha \forall \text{weight}_\beta (\text{value}_\alpha \in \text{Value} \wedge \text{weight}_\beta \in \text{Weight}) \}$$

Then, we have a functional relation that expresses an agent's disposition towards an AtomicProposition with respect to values and weights.

$$\begin{aligned} \text{AgentToAtomPropToValueWeight} = \\ \{ \langle \text{agent}_\alpha, \langle \text{prop}_\beta, \langle \text{value}_\gamma, \text{weight}_\delta \rangle \rangle \rangle \mid \\ \forall \text{agent}_\alpha \forall \text{prop}_\beta \forall \text{value}_\gamma \exists! \text{weight}_\delta (\text{agent}_\alpha \in \text{Agent} \wedge \\ \text{prop}_\beta \in \text{AtomicProposition} \wedge \text{value}_\gamma \in \text{Value} \wedge \text{weight}_\delta \in \text{Weight}) \} \end{aligned}$$

Conceptually, props with positive weights are “proponents” or positive instances of the value and with negative weights as “antagonists” or contrary instances. The basic notion of associating a prop with a particular value and weight appears in [14] and in papers devoted to case-based reasoning with values, where values are assigned to factors as in [15].

The predicates `AgentToValueWeight` and `AgentToAtomPropToValueWeight` are taken to be conceptually distinct. In particular, we interpret `AgentToValueWeight` to indicate the importance that the agent ascribes to the Value. In contrast, we interpret `AgentToAtomPropToValueWeight` to indicate an agent's assessment of the Value and Weight ascribed to a particular prop from `AtomicProposition`. In other words, `AgentToValueWeight` expresses an agent's “ideal” and “global” view on Values, while `AgentToAtomPropToValueWeight` expresses an agent's assessment of particular props in terms of their association with a value. In this approach, values are directly associated with props relative to agents and their value profile.

Filters on Sets of Relations To reflect an agent's value-based *world view*, all the props from `AgentToPropositionToValueWeight` are gathered which have a Value-Weight where the Weight on the Value is no less than the agent's assignment of the Weight on that Value in `AgentToValueWeight`. Propositions can be indeterminate about the Weight on Values.

$$\begin{aligned} \text{AgentMinFilterOwnClean}_{\text{agent}_h} = \\ \{ \langle \text{agent}_h, \langle \text{prop}_\alpha, \langle \text{value}_\beta, \text{weight}_\gamma \rangle \rangle \rangle \mid \text{agent}_h \in \text{Agent} \wedge \\ \forall \text{proposition}_\alpha \forall \text{value}_\beta \forall \langle \text{agent}_h, \langle \text{prop}_\alpha, \langle \text{value}_\beta, \text{weight}_\epsilon \rangle \rangle \rangle \\ \forall \langle \langle \text{agent}_h, \text{value}_\beta \rangle, \text{weight}_\gamma \rangle (\text{prop}_\alpha \in \text{AtomicProposition} \wedge \text{value}_\beta \in \text{Value} \wedge \\ \langle \text{agent}_h, \langle \text{prop}_\alpha, \langle \text{value}_\beta, \text{weight}_\epsilon \rangle \rangle \rangle \in \text{AgentToAtomPropToValueWeight} \wedge \\ \langle \langle \text{agent}_h, \text{value}_\beta \rangle, \text{weight}_\gamma \rangle \in \text{AgentValueToWeight} \wedge \neg(\text{weight}_\gamma > \text{weight}_\epsilon)) \} \end{aligned}$$

Values and weights discriminate amongst props. For a particular agent, a lower weight on a particular value implies, with respect to `AgentMinFilterOwnClean`, that there is a lower discriminatory threshold on the acceptability of props, which themselves are associated with that Value and the Weight. Simply put, if an agent has a lower weight on a particular value, then more props may pass the filter, as they have higher weights on the same value. The higher the weight means that an agent has higher standards with respect to the value; there is greater discrimination such that fewer props pass the filter. An intuitive example may help to clarify the relations between the value-weights in `AgentValueWeight` and `AgentToAtomPropValueWeight` as they appear in `AgentMinFilterOwn`. Suppose an agent is not much bothered by the quality of

coffee, so on the value *taste* the weight is low, and the agent acts accordingly. When they are served a coffee that on the value *taste* has weight that is also low, then the agent drinks the coffee; when served another coffee with value *taste* with weight high, then the agent also drinks the coffee. In effect, the taste makes little difference to this agent, as they don't discriminate. On the other hand, suppose the agent has on the value *taste* a high weight. In the first instance, the agent rejects the coffee as not upholding their higher standards on the value; in the second instance, the agent is satisfied and drinks the coffee. `AgentMinFilterOwnClean` represents an idealised view of an agent's assessment of a set of `AtomicPropositions`, where all `AtomicPropositions` must pass an agent's highest weighting; presentation of other, less stringent variations remain for future work.

Now we are prepared to abstract from consideration of Values and Weights and extract those `AtomicPropositions` which, in effect, represent the agent's value-based *world view*. A set `PropBaseCleanagenth` contains only all those props which pass the value-weight in the value profile for agent_h for all values. It is important to emphasise that this is a very strict, formally convenient definition, which assesses all props and creates a set of only those props that pass the agent's value filter. Matters are more complex and less strict; in other work we develop lenient, flexible definitions.

$$\begin{aligned} \text{PropBaseClean}_{\text{agent}_h} = \\ \{ \text{prop}_\alpha \mid \text{agent}_h \in \text{Agent} \wedge \forall \langle \text{agent}_h, \langle \text{prop}_\alpha, \langle \text{value}_\beta, \text{weight}_\gamma \rangle \rangle \rangle \\ \langle \text{agent}_h, \langle \text{prop}_\alpha, \langle \text{value}_\beta, \text{weight}_\gamma \rangle \rangle \rangle \in \text{AgentMinFilterOwnClean}_{\text{agent}_h} \} \end{aligned}$$

Any set of `PropBaseClean` may contain incompatible props. Moreover, two agents may each accept the same prop, yet for different settings of values and weights.

We assume `PropBaseClean` represents *static* (all at once) and *private* (inaccessible to others) associations of value-weights to props by an agent, which contrasts to the publicly reported reasoning chains, as discussed below. Note that we do not analyse whether the props in `PropBaseClean` are true or believed to be true; the set presents props which the agent can accept in the light of his/her value profile. The key point of creation of `PropBaseClean` is to distinguish props which are coherent with the agent's value profile, i.e. props which pass `AgentMinFilterOwnClean`. By the same token, the complement to `PropBaseCleanagenth` reflects all those props which are incompatible with the Agent's values and weights. Both notions are used in Section 4.

$$\begin{aligned} \overline{\text{PropBaseClean}_{\text{agent}_h}} = \\ \{ \text{prop}_\beta \mid \forall \text{prop}_\beta (\text{prop}_\beta \in \text{AtomicProposition} \wedge \text{prop}_\beta \notin \text{PropBaseClean}_{\text{agent}_h}) \} \end{aligned}$$

Consider the various ways props may appear in intersecting or complementing sets of two Agents' `PropBaseClean`. For intersecting: the Agents have the same value-weight profile and same value-weights on same prop; same value-weight profile and different value-weights on prop, but not sufficient to block; different value-weight profile and same value-weight on prop, but not sufficient to block; different both, but not sufficient to block. Where they have the same props, neither the value-weight profile nor value-weight on prop is sufficient to discriminate. For complementing: same value-weight profile and different value-weights on prop, and sufficient to block; different value-weight profile and same value-weight on prop, and sufficient to block; different both and sufficient to block. In other words, differences in sets of props arise where

Table 1AgentToAtomPropToValueWeight for $agent_A$ and $agent_B$

Agents	Propositions	Values	Weights
agent_A	p_X	value_Q	3
agent_A	p_X	value_P	1
agent_A	p_Y	value_Q	3
agent_A	p_Y	value_P	-1
agent _A	p _Z	value _Q	-1
agent _A	p _Z	value _P	1
agent_B	p_X	value_Q	3
agent_B	p_X	value_P	1
agent_B	p_Y	value_Q	3
agent_B	p_Y	value_P	1
agent_B	p_Z	value_Q	1
agent_B	p_Z	value_P	1

Table 2

PropBaseClean for each Agent

Agents	Propositions
agent _A	p _X , p _Y
agent _B	p _X , p _Y , p _Z

value-weight profiles or value-weights on props are sufficient to discriminate. Broadly speaking, where a prop appears in the intersection of the sets of PropBaseClean of two Agents, we can say the agents agree on that prop in one sense or another, while where the prop is in complementary distribution (in one set, but not the other), we say there is some sense of disagreement. Note that there may be different justifications for the agreement or disagreement as well as different extents of such justification, e.g., greater or lesser difference in weights associated with the value. Relatedly, two agents can have the same denotations for their respective PropBaseClean, yet different value-weight profiles or different value-weights on the same prop.

Example 1. [Creating PropBaseClean] To illustrate PropBaseCleans, suppose 2 agents ($agent_A$ and $agent_B$), 3 propositions $\{p_X, p_Y, p_Z\}$, and 2 values: $V = \{v_Q, v_P\}$, e.g., privacy or law enforcement, where

$$AgentValueWeight_{agent_A} = \{ \langle agent_A, value_Q, 3 \rangle, \langle agent_A, value_P, -2 \rangle \}$$

$$AgentValueWeight_{agent_B} = \{ \langle agent_B, value_Q, 2 \rangle, \langle agent_B, value_P, 1 \rangle \}$$

$agent_A$ has very high requirements concerning value Q , and very low requirements concerning value P , while $agent_B$ has a more balanced value profile.

For AgentToAtomPropToValueWeight, Table 1 is a tabular form for instances of agents, props, values, and weights. The props that pass the AgentMinFilterOwnClean are indicated in bold. In Table 2, we have PropBaseClean for each Agent (the props that pass AgentMinFilterOwnClean).

3. Subjective Knowledge Bases

Given PropBaseClean, we construct subjective knowledge bases, from props that are most important relative to an Agent's values; in this sense, a knowledge base is relativised to an Agent and their values. Given this, we then are in position to compare and relate alternative

justifications across Agents relative to each of their values. Given a subjective knowledge-based, we can construct reasoning chains, which here are intended to be theoretically agnostic about the construction of arguments (e.g., [9, 10, 11, 12, 13] among others) and for simplicity only have strict implication. The constructions are intended to be a common basis; the key point is the use of a subjective knowledge base in formulations of instantiated arguments. In future work, we intend to relate knowledge bases with different argumentation settings and semantics.

Definition 1 (Rules). We suppose a set $\mathcal{P} \subseteq \text{AtomicPropositions}$, a set \mathcal{R} of rules $r_i \dots r_j$ with antecedents $\mathcal{A} \subseteq \mathcal{P}$. Given $r_i \in \mathcal{R}$, and a single prop $p_i \in \mathcal{P}$ that is the consequent of r_i , the consequent p_i is inferred where all the props in \mathcal{A} and r_i hold.

Definition 2 (Subjective Knowledge Base). Given a set of atomic propositions $P_i \subseteq \text{AtomicProposition}$, a set of rules $R_j \subseteq \mathcal{R}$, and an Agent_A, a subjective knowledge base KB_{Agent_A} is $\langle P_i, R_j \rangle$ where $\forall p_\alpha$, s.t. $P_i \cup R_j \vdash p_\alpha$, $p_\alpha \in \text{PropBaseClean}_{\text{agent}_A}$

Informally, a subjective knowledge base represents what an Agent intuitively *accepts* relative to that Agent's values. Set-theoretic relations can hold between the Agents subjective knowledge bases. A subjective knowledge base can be inconsistent, as is widely accepted in argumentation theory [16, 17].

Finally, we define reasoning in subjective knowledge bases.

Definition 3 (Subjective Reasoning chain). Given an subjective knowledge base $KB_{\text{Agent}_A} = \langle P_{\text{Agent}_A}, R_{\text{Agent}_A} \rangle$ and a reasoning chain RC_{Agent_A} as $\langle KB_{\text{Agent}_A}, p_{\text{Agent}_A} \rangle$, where $KB_{\text{Agent}_A} = \langle P_{\text{Agent}_A}, R_{\text{Agent}_A} \rangle$, $P_{\text{Agent}_A} \subseteq P_{\text{Agent}_A}, R_{\text{Agent}_A} \subseteq R_{\text{Agent}_A}$:

- Every proposition derived from RC_{Agent_A} should be in $\text{PropBaseClean}_{\text{Agent}_A}$:
 $\forall p_\alpha$.s.t. $RC_{\text{Agent}_A} \vdash p_\alpha$ ($p_\alpha \in \text{PropBaseClean}_{\text{Agent}_A}$)
- KB_{Agent_A} is a minimal set of atomic propositions and rules necessary to derive p_i ;
- KB_{Agent_A} does not contain cycles;
- p_j (conclusion) is single atomic proposition such that $KB_{\text{Agent}_A} \vdash p_j$;
- $\neg(\exists p_Q, p_T$ s.t. $KB_{\text{Agent}_A} \vdash p_Q$ and $KB_{\text{Agent}_A} \vdash p_T$ and $\langle p_Q, p_T \rangle \in \text{incompProp}$).

In contrast to a subjective knowledge base, a reasoning chain must be *internally* consistent. While Def. 3 outlines a generic construction of a reasoning chain based on a subjective knowledge base, it says nothing about selection amongst the available alternatives. Nonetheless, given that the reasoning chains are constructed from a subjective knowledge base, they represent some principled selection from amongst reasoning chains otherwise constructed.

4. The Roots of Conflict in Argumentation

In our approach, the roots of conflict in argumentation can be found in the underlying differences between Agents' value profiles, which surface in the props found in the knowledge bases and reasoning chains. We identify several notions of incompatibility amongst reasoning chains, which may be taken as "attacks" in the argumentation-theoretic sense, which might not otherwise be apparent. Thus, in addition to attacks based in a logical notion of incompatibility, the attacks can be grounded in the Agents' different value-weight settings.

Definition 4 (Relations between Reasoning Chains). Given two agents, $agent_A$ and $agent_B$, and their subjective knowledge bases

$KB_{agent_A} = \langle P_A, R_A \rangle$ of a $agent_A$ and $KB_{agent_B} = \langle P_B, R_B \rangle$ of a $agent_B$, we can provide reasoning chains relative to KBs.

Suppose $RC_{agent_{A_a}}$ and $RC_{agent_{B_b}}$, where

$RC_{agent_{A_a}}$ is $\langle KB_{agent_{A_a}}, p_h \rangle$, where $KB_{agent_{A_a}} = \langle P_h \subseteq P_A, R_h \subseteq R_A \rangle$ and

$RC_{agent_{B_b}}$ is $\langle KB_{agent_{B_b}}, p_l \rangle$, where $KB_{agent_{B_b}} = \langle P_{agent_{B_b}} \subseteq P_B, R_{agent_{B_b}} \subseteq R_B \rangle$, and

$RC_{agent_{A_a}}$ is subjectively consistent to $agent_A$ and $RC_{agent_{B_b}}$ is subjectively consistent to $agent_B$.

1. $RC_{agent_{A_a}}$ and $RC_{agent_{B_b}}$ are objectively incompatible iff:
 $\exists_{p_\alpha, p_\beta} : (KB_{agent_{A_a}} \vdash p_\alpha) \wedge (KB_{agent_{B_b}} \vdash p_\beta) \wedge \langle p_\alpha, p_\beta \rangle \in \text{incompProps}$
2. $RC_{agent_{A_a}}$ and $RC_{agent_{B_b}}$ are subjectively symmetrical incompatible w.r.t. $agent_A$ and $agent_B$ if $\exists_{p_\alpha, p_\beta} : (KB_{agent_{A_a}} \vdash p_\alpha) \wedge (KB_{agent_{B_b}} \vdash p_\beta) \wedge$
 $(p_\alpha \in \overline{\text{PropBaseClean}}_{agent_B}) \wedge (p_\beta \in \overline{\text{PropBaseClean}}_{agent_A})$
3. $RC_{agent_{A_a}}$ and $RC_{agent_{B_b}}$ are subjectively asymmetrical incompatible if
 $\exists_{p_\alpha \text{ s.t.}} : KB_{A_a} \vdash p_\alpha, p_\alpha \in \overline{\text{PropBaseClean}}_{agent_B}$, and
 $\forall_{p_\gamma} : (KB_{B_b} \vdash p_\gamma \rightarrow p_\gamma \in \overline{\text{PropBaseClean}}_{agent_A})$

Subjective incompatibility reflects some difference between the value-weight settings of each Agent, respectively, though not specifically which setting, though we could do so. Def. 4(1) represents one way that reasoning chains may be construed to attack one or the other; it is consistent with the *context independence* property introduced by [18] in which, for two arguments and two knowledge bases, if argument A attacks argument B in one KB, then it also attacks in the second KB. Def. 4(2), where props are reciprocally absent in the PropBaseClean of each agent, imply argument attack between two reasoning chains in virtue of the underlying disagreement about the value preferences of the agents, even if “objectively” or logically the props are not incompatible; here, the presence or absence of props are proxies of an agent’s value profile. Def. 4(3) is the asymmetrical version of Def. 4(2), where the PropBaseClean of Agent B is a superset of Agent A. As such, Agent A finds some arguments of Agent B to be incompatible with Agent B’s value profile, but not vice versa. In such a case, Agent A would appear to be attacked in that Agent B can present reasoning chains that are not consistent with Agent A’s value profile. VIA does not uphold context independence, as conflicts between reasoning chains are Agent dependent.

Subjective incompatibility indirectly represents the subjective, value-based opinion of an agent and explains the root of a common real life situation in which: there is indirect conflict; and for one Agent, reasoning chains are in conflict, but not for the other one. Note that, subjective incompatibility indirectly represents conflict between the Agents’ value systems.

Example 2 (subjectively asymmetrical). Suppose a set of rules $\mathcal{R} = \{p_x \rightarrow p_y, p_y \rightarrow p_z\}$.

The subjective knowledge bases of agents: $KB_{agent_A} = \langle \{p_x, p_y\}, \{p_x \rightarrow p_y\} \rangle$

$KB_{agent_B} = \langle \{p_x, p_y, p_z\}, \{p_x \rightarrow p_y, p_y \rightarrow p_z\} \rangle$

On the basis of the above the agents can create following (incomplete list of) reasoning chains:

$RC_{agent_{A_2}} = \langle \langle \{p_x\}, \{p_x \rightarrow p_y\} \rangle, p_y \rangle$, $RC_{agent_{B_2}} = \langle \langle \{p_x\}, \{p_x \rightarrow p_y, p_y \rightarrow p_z\} \rangle, p_z \rangle$

Reasoning chains $RC_{agent_{A_1}}$ and $RC_{agent_{B_1}}$ are subjectively asymmetrical conflicting reasoning chains, because $p_z \in \overline{\text{PropBaseClean}}_{agent_B}$ but $p_x \in \overline{\text{PropBaseClean}}_{agent_A}$.

5. Related Work

While there is related work bearing on legal theory [3], preferences [19], and multi-agent systems [20], we focus on key proposals related to argumentation and values.

VIA is agnostic with respect to theories of instantiated or abstract argumentation (e.g., [9, 10, 11, 12, 13] among others), which could use VIA's subjective knowledge bases. However, VIA adopts rather than excludes context independence found in most prior work.

Researchers [2, 21, 19] point out that conflicts between particular agents' arguments can be rooted not only in the errors in logic, calculation, or different beliefs concerning facts, but they can also disagree on their preferences or values. A key difference between such prior approaches and VIA is *where* and *how* such conflicts appear and are used in the course of argumentation.

Preferences have been widely considered in AI for decision-making and choice [19]. Broadly speaking, elements in a set are subject to choice, where preference is a comparative ordering relation between elements. The specific type of elements and the ordering relations over them can be defined in a variety of ways. In argumentation, preferences are used to adjudicate "winning" arguments which would otherwise "tie". In our proposal, there is no direct comparison between elements and no ordering over them. Rather, values filter props according to the value profile of the agent. In future work, we aim to account for how arguments "inherit" the values of the component props and derivative concepts of argument attack.

In Value-base Argumentation Frameworks (VAF) of abstract argumentation [6, 21, 7], values adjudicate the "winning" abstract argument according to an ordering amongst values [21]; specific orderings are taken to represent an *audience*. Values are associated with abstract arguments themselves, where the argument "promotes" the value. As noted above, it is opaque how the value of the argument is determined. In VIA, an Agent can also be construed as an audience, but in the sense that the Agent's value profile filters the props used to construct the knowledge base and arguments rather than a way to adjudicate amongst arguments.

Action-based Alternating Transition Systems with values (AATS+V) represent arguments about actions in multi-agent systems [6]. Essentially, an AATS+V expresses actions as transition functions from state to state, where states are sets of props. Values are a set of abstract objects. A valuation function describes whether a state transition promotes or demotes the value; in this sense, it is a preference over actions, which associates values to actions as a whole.

Case-based reasoning associates factors, which are essentially props, and values [22, 6, 16, 15]. Bench-Capon and Sartor [22, p.103] write that: "... a factor favours an outcome is because deciding for that outcome in a case where that factor is present promotes or defends some value, which it held that the legal system should promote or defend." There may be degrees (dimensions) by which a factor is representative of the value [23]. VIA differs in three respects. First, agents associate values with props, so it is not an association that is independent of agents or a point of view of the legal system. Second, such an association creates a subjective knowledge base and related reasoning chains, which is not found in work on factor analysis. Third, VIA allows that agents make the same argument based on different values.

6. Conclusions

The main aim of our paper was to take a step towards VIA, a formal model for “motivated reasoning” in terms of value-based conflicts between reasoning chains. We observe (following [2, 21]) that disagreements between various agents are rooted not only in the logical or commonsense reasons, but also they are inherited from the differences of value systems of those agents. Our model allows for the representation of such conflicts and explains how differences in value systems influence individual knowledge bases of agents and, in consequence, explains which props can be intuitively accepted by particular agents. Such an approach results in individualisation of the relation of incompatibility by assigning it to a particular agent.

VIA is motivated and guided by several intuitions about the role of values in reasoning. We observe that individuals are often asked to assign a value to some statement as in surveys or in social media assessments. Just how individuals make such assignments is a matter for social science and psychology. In this regard, a statement can be taken as an instance of the abstract value. The approach does not preclude being asked for a value assignment to a whole, e.g., marking argumentative essays. Indeed, different sorts of value may be more or less directly associated with different types of props, which remains for future exploration. In addition, we observe that from a common pool of accessible information, agents select what is used to construct their arguments according to their values; agents may argue for their side without considering the full spectrum of information, alternative arguments, or attacking and defeating them. As well, argumentative discussions do not necessarily proceed in the manner of explicit attack and counter-attack; arguments (as in differences of opinion or analysis) appear in a variety of dialogue types [24], where defeat may not be key. There is also voting, where distinct, antagonistic arguments can put forward, sometimes without explicit rebuttal.

VIA provides a basis for a range of future explorations, which we briefly remark on here. Broadly, we can see how VIA relates to settings of argumentation. The current report does not explore the compositional relation of values associated with a “whole”, e.g., argument with values of the “parts”, e.g., props, and how they are combined. We believe there is a connection, but leave it for future work. In relation to voting, issues such as consortia, where agents gather around shared arguments, and confirmation bias, where agents do not challenge their positions, would need to be explored further. However, the approach taken would appear to offer a fruitful line of analysis. It would be worthwhile to examine how VIA does (or does not) impact on the semantics of argumentation, i.e., extensions, as well as relevant formal properties. As it is, VIA is static, so some development could be done with respect to deliberation and persuasion dialogues. While we have abstracted over context, clearly this should be articulated; that is, how shouting “fire” in one context but not another is an instance of protected speech. In our view, this may be tied to implications that follow from one context or another. Relatedly, we believe it worth exploring how values relate in terms of implication and hierarchy. VIA has some suggestive bearing on abduction [25]; while an agent makes some value-based selection of props from the “universe” of discourse, there are still alternatives presented with respect to an Agent’s knowledge base. More concretely, we can explore how VIA might be used to represent and reason with legal decisions and the case base as well as actions, particularly with reference to the concepts of value promotion/demotion. explanations of judicial behaviour.

References

- [1] O. Kahneman, *Judgment under Uncertainty: Heuristics and Biases*, Cambridge University Press, 1982. doi:10.1017/CBO9780511809477.
- [2] C. Perelman, L. Olbrechts-Tyteca, J. Wilkinson, P. Weaver, *The New Rhetoric: A Treatise on Argumentation*, University of Notre Dame Press, 1969. URL: <http://www.jstor.org/stable/j.ctvpj74xx>.
- [3] R. Cahill-O’Callaghan, B. Richards, Policy, principle, or values: An exploration of judicial decision-making, *Louisiana Law Review* 79 (2019) 397–418.
- [4] S. Schwartz, An overview of the schwartz theory of basic values, *Online Readings in Psychology and Culture* 2 (2012).
- [5] J. R. Searle, *Rationality in Action*, MIT, 2001.
- [6] K. Atkinson, T. J. M. Bench-Capon, Value-based argumentation, *FLAP* 8 (2021) 1543–1588. URL: <https://collegepublications.co.uk/ifcolog/?00048>.
- [7] P. M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, *Artif. Intell.* 77 (1995) 321–358. URL: [https://doi.org/10.1016/0004-3702\(94\)00041-X](https://doi.org/10.1016/0004-3702(94)00041-X). doi:10.1016/0004-3702(94)00041-X.
- [8] A. Z. Wyner, J. Schneider, Arguing from a point of view, in: S. Ossowski, F. Toni, G. A. Vouros (Eds.), *Proceedings of the First International Conference on Agreement Technologies, AT 2012, Dubrovnik, Croatia, October 15-16, 2012*, volume 918 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2012, pp. 153–167. URL: <http://ceur-ws.org/Vol-918/111110153.pdf>.
- [9] P. Besnard, A. Hunter, *Elements of Argumentation*, MIT Press, 2008.
- [10] H. Prakken, An abstract framework for argumentation with structured arguments, *Argument and Computation* 1 (2010) 93–124.
- [11] P. M. Dung, R. A. Kowalski, F. Toni, Assumption-based argumentation, in: G. R. Simari, I. Rahwan (Eds.), *Argumentation in Artificial Intelligence*, Springer, 2009, pp. 199–218. URL: https://doi.org/10.1007/978-0-387-98197-0_10. doi:10.1007/978-0-387-98197-0_10.
- [12] A. J. García, G. R. Simari, Defeasible logic programming: An argumentative approach, *Theory Pract. Log. Program.* 4 (2004) 95–138. URL: <https://doi.org/10.1017/S1471068403001674>. doi:10.1017/S1471068403001674.
- [13] G. Governatori, M. J. Maher, G. Antoniu, D. Billington, Argumentation semantics for defeasible logic, *Journal of Logic and Computation* 14 (2004) 675–702.
- [14] T. Zurek, Goals, values, and reasoning, *Expert Systems with Applications* 71 (2017) 442 – 456. URL: <http://www.sciencedirect.com/science/article/pii/S0957417416306303>. doi:<https://doi.org/10.1016/j.eswa.2016.11.008>.
- [15] T. Bench-Capon, H. Prakken, A. Wyner, K. Atkinson, Argument schemes for reasoning with legal cases using values, in: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law, ICAIL ’13*, Association for Computing Machinery, New York, NY, USA, 2013, p. 13–22. URL: <https://doi.org/10.1145/2514601.2514604>. doi:10.1145/2514601.2514604.
- [16] P. Besnard, A. Hunter, Argumentation based on classical logic, in: G. R. Simari, I. Rahwan

- (Eds.), *Argumentation in Artificial Intelligence*, Springer, 2009, pp. 133–152.
- [17] S. Modgil, H. Prakken, The aspic+ framework for structured argumentation: a tutorial, *Argument & Computation* 5 (2014) 31–62. doi:10.1080/19462166.2013.869766.
 - [18] P. M. Dung, An axiomatic analysis of structured argumentation with priorities, *Artificial Intelligence* 231 (2016) 107–150. URL: <https://www.sciencedirect.com/science/article/pii/S000437021500168X>. doi:<https://doi.org/10.1016/j.artint.2015.10.005>.
 - [19] G. Pigozzi, A. Tsoukiàs, P. Viappiani, Preferences in artificial intelligence, *Ann. Math. Artif. Intell.* 77 (2016) 361–401. URL: <https://doi.org/10.1007/s10472-015-9475-5>. doi:10.1007/s10472-015-9475-5.
 - [20] M. Winikoff, G. Sidorenko, V. Dignum, F. Dignum, Why bad coffee? explaining BDI agent behaviour with valuing, *Artif. Intell.* 300 (2021) 103554. URL: <https://doi.org/10.1016/j.artint.2021.103554>. doi:10.1016/j.artint.2021.103554.
 - [21] T. J. M. Bench-Capon, Persuasion in Practical Argument Using Value-based Argumentation Frameworks, *Journal of Logic and Computation* 13 (2003) 429–448. URL: <https://doi.org/10.1093/logcom/13.3.429>. doi:10.1093/logcom/13.3.429.
 - [22] T. Bench-Capon, G. Sartor, A model of legal reasoning with cases incorporating theories and values, *Artificial Intelligence* 150 (2003) 97 – 143. URL: <http://www.sciencedirect.com/science/article/pii/S0004370203001085>. doi:[https://doi.org/10.1016/S0004-3702\(03\)00108-5](https://doi.org/10.1016/S0004-3702(03)00108-5), *AI and Law*.
 - [23] K. Atkinson, T. J. M. Bench-Capon, H. Prakken, A. Z. Wyner, Argumentation schemes for reasoning about factors with dimensions, in: K. D. Ashley (Ed.), *Legal Knowledge and Information Systems - JURIX 2013: The Twenty-Sixth Annual Conference*, December 11-13, 2013, University of Bologna, Italy, volume 259 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, 2013, pp. 39–48. URL: <https://doi.org/10.3233/978-1-61499-359-9-39>. doi:10.3233/978-1-61499-359-9-39.
 - [24] D. Walton, E. Krabbe, *Commitment in dialogue: basic concepts of interpersonal reasoning*, University of New York Press, 1995.
 - [25] A. Aliseda-Llera, The Logic of Abduction: an Introduction, in: A. Aliseda (Ed.), *The Logic of Hypothetical Reasoning, Abduction, and Models*, Springer, 2016, pp. 219–230.