

Developing a model for dialectical arguments

Antonis Bikakis



3 June 2024
Information Studies Research Open Day

Project team



- Antonis Bikakis
 - Department of Information Studies, UCL
 - Knowledge Representation, Computational Argumentation



- Luke Dickens
 - Department of Information Studies, UCL
 - Machine Learning



- Miriam Koschate-Reis (Exeter)
 - Department of Psychology
 - Computational Social Psychology

Talk outline

- Argumentation & Argumentation Theory
- Formal models of argument
- Research gaps in the study of argumentation
- Relevant theories in social psychology
- Research aims

What is argumentation?

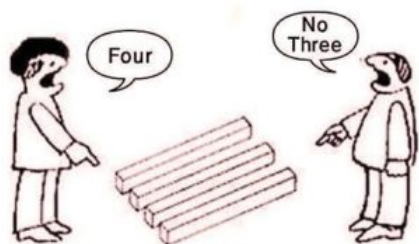
- An everyday human activity
- Exchange of arguments on a topic
- Resolving conflicts of opinion
- Influencing the thoughts or views of others
- “*The ability to consider, for a given question, the elements that are useful to persuade someone*” (Aristotle)
- A way of thinking
- A cognitive process
- Drawing conclusions based on evidence, which may be incomplete, uncertain or contradictory

A formal definition

“ a verbal, social, and rational activity aimed at convincing a reasonable critic of the acceptability of a standpoint by putting forward a constellation of propositions justifying or refuting the proposition expressed in the standpoint ”

(van Eemeren & Grootendorst, 2004)

A discursive activity



A cognitive activity



Types of information

- **Certain** (non-disputable)
 - *“Dogs are animals”*
 - *“Animals have been used in medical testing”*
- **Uncertain** (disputable)
 - *“Animal testing may be the best tool to defeat viruses like COVID-19”*
- **Objective** (can be observed, measured or verified)
 - *“Mice share more than 98% DNA with humans”*
- **Subjective** (based on beliefs or opinions)
 - *“Testing on animals is unethical”*
- **Hypothetical**
 - *“Animal testing will be banned within the next decade”*

Argumentation Theory

“ The study of argumentation in all its manifestations and varieties, irrespective of the intellectual backgrounds, primary research interests and angles of approach of the theorists”

(van Eemeren et al., 2002)

Argumentation is studied in several disciplines:

- Philosophy
- Communication studies
- Cognitive Psychology
- Linguistics
- Artificial Intelligence

However, there is a lack of interdisciplinary studies!

Argumentation in AI

- Formal models of argumentation
- Computer programs that model or support argumentative tasks
 - Identifying arguments, evaluating arguments, drawing conclusions, etc.
- Most research has focused on argumentation-based inference
- Research on argumentative dialogues is less mature
 - Game-theoretic models
 - Protocols
 - Strategies

Abstract Argumentation Frameworks

- A simple but elegant model for argument evaluation based on two notions: argument and attack

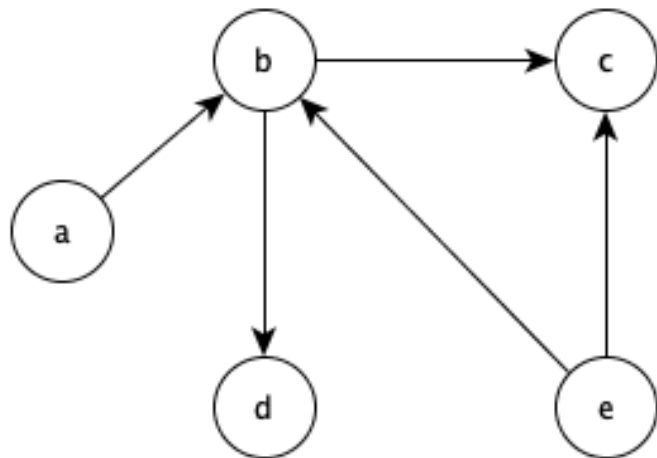
An argumentation framework is a pair $AF = \langle A, R \rangle$ where

- *A is a set of argument*
- *$R \subseteq A \times A$ is a binary relation on A*

(Dung, 1995)

Abstract Argumentation Frameworks

- An argumentation framework can be modelled as a directed graph where nodes denote arguments and directed edges denote attacks

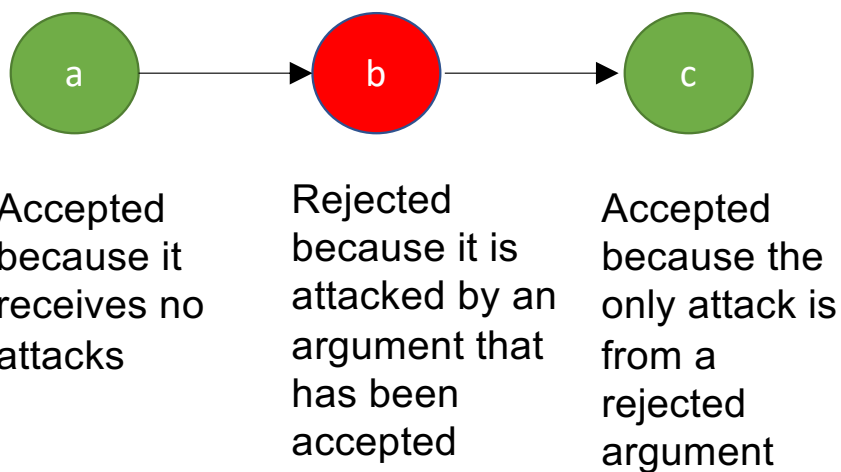


$$A = \{a, b, c, d, e\}$$

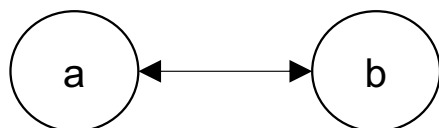
$$R = \{(a,b), (b,c), (b,d), (e,b), (e,c)\}$$

Evaluation of arguments

- An argument is
 - **accepted** if it does not receive any attacks from arguments that have not been **rejected**
 - **rejected** if there is a counter-argument that has been **accepted**



A more interesting case



- Arguments that are in conflict cannot be both accepted
- Should we accept neither or either of them?
- Scenario 1:
 - a: The weather in Cuba is great, let's go there for our holidays.
 - b: The tickets to Cuba are expensive, let's go somewhere else.
- Scenario 2:
 - a: Alice: Bob committed the murder. I saw him in the crime scene.
 - b: Bob: I didn't do it. Alice did it. She hated the victim!

Accept either

Accept neither

Acceptability semantics

- An **extension** of an argumentation framework $AF = \langle A, R \rangle$ is a set of arguments $E \subseteq A$ that we can reasonably accept
 - **Admissible**: E is conflict-free and defends all its members
 - **Complete**: E is admissible and contains all arguments it defends
 - **Grounded**: E is the minimal complete extension
 - **Preferred**: E is the maximal complete extension
 - **Stable**: E is conflict-free and attacks all arguments it does not contain
- Each semantics corresponds to a different form of reasoning

Extended models of argument

- **Bipolar**: integrate the notion of support among arguments
- **Preference-based**: use preferences to resolve conflicts among arguments
- **Value-based**: associate arguments with (social/ethical/..) values
- **Weighted**: assign weights to arguments and/or attacks
- **Social**: integrate arguments with social votes
- ...

Gaps in current research

- Much more focus on *logos* than *ethos* and *pathos* (Aristotle)
- **Credibility of the speaker**
 - “Scientists are 95% certain that humans are the “dominant cause” of global warming since the 1950s (source: IPCC, 5th assessment report)”
- **Emotions (threat, reward, fear, guilt)**
 - “If you get a good degree, I will buy you a new car”
- **Peer-pressure**
 - “All my friends have a smartphone so I should have one too”

Gaps in current research

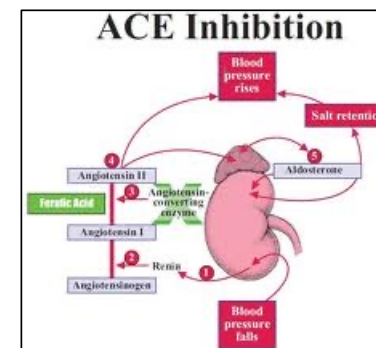
• Motives

- “Don’t consume too much salt because it could increase blood pressure”
- Hypertensive vs. hypotensive audience



• Language

- “The angiotensin-converting enzyme (ACE) inhibitors decrease the production of angiotensin and, in turn, that helps lower blood pressure”



Gaps in current research

- **Argument dynamics**

- Change of stance/opinion/behaviour in the course of a dialogue
- When new information becomes available
- When previously presented information is invalidated/outdated
- When the behaviour of other participants changes

- **Lack of empirical studies**

- Models based on intuition
- Focus on formal properties rather than the ability to explain features of real dialogues

Relevant theories in social psychology

- **Social influence**

- Influence based on factual (“objective”) information vs. normative influence based on reward and punishment (Deutsch & Gerard, 1955): conversion vs. compliance
- Cultural transmission (Tomasello, 2001)
- Self-categorization theory & referent informational influence (Turner et al., 1987)
- Social identity & linguistic style (Koschate, 2021)

Relevant theories in social psychology

- **Persuasion**

- Elaboration Likelihood Model (Petty & Cacioppo, 1980): central vs. peripheral route
- Cognitive dissonance theory (Festinger, 1962)
- Principles of persuasion (Cialdini, 1984)
 - Reciprocity
 - Commitment/Consistency
 - Social "proof"
 - Authority
 - Liking
 - Scarcity

Relevant theories in social psychology

- **Social Identity theory** (Tajfel and Turner, 1986)
 - how individuals derive a sense of self and identity from the groups to which they belong
 - the group's norms, values, and behaviours become an integral part of their identity
 - People tend to favour their own group (in-group) over others (out-groups)
- **Social impact theory** (Latané, 1981)
 - how individuals can be "*sources or targets of social influence*"
 - $i = f(S, I, N)$
 - i : social impact
 - S : status/power of influencer
 - I : immediacy (recency, physical proximity)
 - N : amount of sources of influence

Our research aims & methods

- **Aims**

- Investigate how useful current models of argument are in explaining how people argue in reality
- Identify psychological processes they fail to capture
- Study the dynamics of argumentation using empirical studies
- Develop a new model informed by empirical studies and relevant social psychology theories

- **Methods**

- Investigation based on annotated argumentation datasets and data from online fora
- Controlled experiments with human participants
- Extend/refine existing models
- Computational argumentation, NLP, Statistical Machine Learning, psychological experiments



Questions?
(or does anyone want to argue?)