

## Four Ways to Evaluate Arguments According to Agent Engagement

Pierre Bisquert, Madalina Croitoru, Florence Dupin de Saint Cyr - Bannay

► **To cite this version:**

Pierre Bisquert, Madalina Croitoru, Florence Dupin de Saint Cyr - Bannay. Four Ways to Evaluate Arguments According to Agent Engagement. Yike Guo; Karl J. Friston; A. Aldo Faisal; Sean L. Hill; Peng Hanchuan. BIH: Brain Informatics and Health, Aug 2015, London, United Kingdom. Springer, 8th International Conference, BIH 2015, London, UK, August 30 - September 2, 2015. Proceedings, 9250, pp.445-456, 2015, LNCS - Lecture Notes in Computer Science. <10.1007/978-3-319-23344-4\_43>. <lirmm-01245121>

**HAL Id: lirmm-01245121**

**<https://hal-lirmm.ccsd.cnrs.fr/lirmm-01245121>**

Submitted on 16 Dec 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Four Ways to Evaluate Arguments According to Agent Engagement\*

Pierre Bisquert<sup>1</sup>      Madalina Croitoru<sup>2</sup>

Florence Dupin de Saint-Cyr<sup>3</sup>

<sup>1</sup>INRA, France    pierre.bisquert@supagro.inra.fr

<sup>2</sup>Univ. Montpellier, France    croitoru@lirmm.fr

<sup>3</sup>IRIT, France    florence.bannay@irit.fr

October 3, 2015

## Abstract

In this paper we are interested in the computational and formal analysis of the persuasive impact that an argument can have on a human. We present a preliminary account of the listener mental process (representation and reasoning mechanisms of the dual process cognitive model) as well as her engagement based on the ELM model. This engagement determines the reasoning process that the agent will adopt in order to evaluate and incorporate the uttered argument.

**Keywords:** Cognitive Bias, Dual Process, Argumentation, Artificial Intelligence.

## 1 Introduction

Several decades worth of empirical studies have clearly shown that human responses often deviate from the expected rational response on many reasoning tasks. We study the link between attitudes and persuasion, a central point of social psychology [2, 18]. While important methodological and theoretical issues regarding the consistency between attitudes and behaviors were made early on [1, 8] no agreement between different modalities was achieved. The ELM model [16, 15] is one of the first models to generalise a theory of attitude change in persuasion. ELM distinguishes between two “routes” that govern the reception of persuasive communications and the outcomes concerning attitude changes: the *central route*, that involves a large amount of cognition and that has the tendency to be more rational concerning the logical quality of the received argument, and the *peripheral route*, that involves little cognition and where arguments are more likely to be evaluated thanks to simple cues such as the political view on the advocated position, the liking of the speaker, etc.

The art of persuasion is intimately connected with the psychological process of perception. In [5] the author presents the relationship between mental processes that psychologists label cognitive biases and legal arguments. Arguments built on this relationship contain what philosophers label informal fallacies. By using legal arguments that contain informal fallacies, the lawyer can play upon the listeners inherent cognitive biases to persuade them to see things in a way that favors its goals. Argumentation in a large sense is not to be identified with rational persuasion as some arguments are not addressing the rational part of the listener [3].

To bridge the different communities working on argumentation, in [10] the authors distinguish between the three meanings of the word argument: argument as a reason, argument as a structured sequence of reasons and claims, and argument as a social exchange. All three meanings are integral to a complete understanding of human reasoning and cognition. We follow in their foot steps by integrating different

---

\*This is a draft version the paper was published in *Brain Informatics and Health*, Volume 9250 of the series Lecture Notes in Computer Science pp 445-456.

aspects of what an argument is in our preliminary model. The core of our approach is to propose a model of cognitive biases and their role in persuasion and argumentation.

In their seminal article [22] Tversky and Kahneman explain how supposedly “rational” judgments are based on data with limited validity, processed according to heuristic rules. They illustrate their thesis with a number of biases empirically demonstrated (such as the illusion of validity, retrievability of instances, anchoring, framing, etc.). The ancient idea that cognitive processes can be partitioned into two main families (traditionally called intuition and reason) is now widely embraced under the general label of dual-process theories [6, 7, 19, 12]. According to Kahneman [22], human reasoning takes place in two systems: System 1 (S1) and System 2 (S2). The first system deals with quick and instinctive thoughts and is based on associations such as cause-effect, resemblance, valence etc. The second system, used as little as possible, is a slow and conscious process that deals with what we commonly call reason.

In this paper, after defining a new cognitive model and two reasoning processes based on [12] as well as [15] in Section 2, we present how an argument might be evaluated and its effect on the agent’s mind in Section 3. Finally, some properties are shown in Section 4.

## 2 Towards a Computational Model of Cognitive Evaluation

### 2.1 Cognitive Model

In this paper, our aim is to define a computational cognitive model of the evaluation of an argument. Based on Kahneman’s theory we propose to define an agent cognitive model on two components  $AT$  (an association table linking a formula to an ordered set of formulae and to a flag encoding an appreciation) and  $KB$  (a logical knowledge base) in order to encode S1 and S2 respectively.<sup>1</sup> Formally, we consider a logical language and we denote by  $\mathcal{L}$  the set of well formed formulae of this language given the usual connectives  $\wedge, \vee, \rightarrow, \neg$  and the constants  $\perp$  and  $\top$ . The set of terms in the language is denoted by  $\mathcal{T}$ .  $\vdash$  denotes classical inference. The fact that a term  $t$  appears in a formula  $\varphi$  is denoted by  $t \in \varphi$ .

**Definition 1 (Association table)** *An agent’s association table  $AT$  is a set of triples of the form  $(\varphi, (S, \succ_S), f)$  where:*

- $\varphi \in \mathcal{L}$  is a well formed formula representing a piece of knowledge,
- $S \subseteq \mathcal{L}$  is a set of well formed formulae associated to  $\varphi$  endowed with a total strict order  $\succ_S \subseteq S \times S$ , the pair  $(S, \succ_S)$  is called a stack (when there is no ambiguity, the total ordered will be omitted),
- $f \in \{acc, rej, \emptyset\}$  is a flag stating that  $\varphi$  is respectively accepted, rejected or not informed in the association table.

The set of all well formed formulae in the association table is denoted by  $L_{AT}$ , i.e.,  $L_{AT} = \bigcup_{(\varphi, S, f) \in AT} \varphi$ . Given a formula  $\varphi \in L_{AT}$ , the stack  $S$  associated with  $\varphi$  in  $AT$  will be denoted by  $AT(\varphi)$ , and the top element of this stack is denoted  $\text{Top}(\varphi)$ . Formally,  $\text{Top}(\varphi) = \varphi_0$  s.t.  $\forall \varphi' \neq \varphi_0 \in AT(\varphi), \varphi_0 \succ_S \varphi'$ . The flag  $f$  associated to  $\varphi$  is denoted by  $\text{flag}(\varphi)$ . If  $f$  is a flag then  $-f$  is a flag such that  $-acc = rej, -rej = acc$  and  $-\emptyset = \emptyset$ . Note that  $AT$  is implicit in the definitions of  $\text{Top}$  and  $\text{flag}$ .

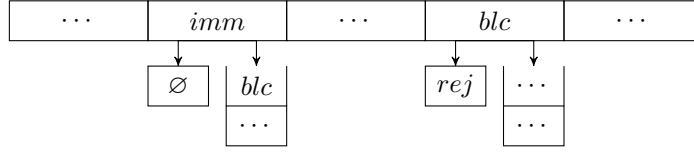
A cognitive model is based on  $AT$  and on a knowledge base  $KB$  which contains Beliefs, Appreciations (i.e. associations of formulae to flags) and a set of Appreciation Rules<sup>2</sup> called *a-rules*. An a-rule has the form  $\varphi \rightsquigarrow (\psi, f)$  where  $\varphi, \psi \in \mathcal{L}$  and  $f \in \{acc, rej\}$ . Due to space constraints, we will not detail how an operator  $\vdash$  is built such that  $KB \vdash (\varphi, f)$  iff the appreciation of  $\varphi$  is  $f \in \{acc, rej, \emptyset\}$  wrt  $KB$ .

**Definition 2 (Cognitive model)** *A cognitive model is a tuple  $\kappa = (KB, AT, \lambda, i)$ :*

- $KB = (K, F, R)$  s.t.  $K \subseteq \mathcal{L}$ ,  $F \subseteq \mathcal{L} \times \{acc, rej, \emptyset\}$  and  $R$  is a set of a-rules.
- $AT$  is an association table such that  $\forall \varphi \in K, \forall t, t' \in \mathcal{T}, \forall f \in \{acc, rej, \emptyset\}$   
 $((t, t' \in \varphi) \Rightarrow t \in AT(t'))$  and  $((\varphi, f) \in F \Rightarrow \text{flag}(\varphi) = f)$  (1)
- $\lambda \in \mathbb{N}$  is an integer value representing the threshold above which the agent feels to be enough aware about the topic of a formula to be able to reason rationally,

<sup>1</sup>Note that S1 and S2 are linked as we will see in (1) of Definition 2.

<sup>2</sup>Inspired from the Desire-Generation rules (of Rahwan and Amgoud [17]).



**Figure 1:** Partial representation of Alex's associative table.

- $i : \mathcal{L} \rightarrow \{0, 1, 2\}$  is a three value marker that gives the interest level of the agent relatively to a formula.

In other words, (1) expresses the link between  $KB$  and  $AT$ , more precisely, every pair of terms belonging to a given formula in  $K$  are associated in  $AT$  and the flags in  $AT$  comply with  $F$ . In case of ambiguity about the current cognitive model, the symbols  $AT$ ,  $Top$ ,  $flag$  will be indexed by the cognitive model  $\kappa$  they refer to.

**Example 1** Let us consider a fictional human agent called Alex. Alex has a keen interest in politics and he has started to learn more about immigration ( $imm$ ). He thinks that immigration will bring the much required workforce in domains that lack attractiveness. He also thinks that it will mean a better living for everybody as the local economy will be relaunched. His  $KB$  contains rules such as  $imm \rightarrow extraWork$ ,  $extraWork \rightarrow moreMoney$ ,  $moreMoney \rightarrow \neg lJobs$ .

Recently, Alex has also watched a documentary about immigration that presented the horrible living conditions of certain immigrants. While he still does not know whether to accept or reject the notion of immigration, the first thing he now associates immigration with is the bad living condition ( $blc$ ) of immigrants, something he disapproves of. This is represented in Fig. 1.

## 2.2 System 1 and System 2 Reasoning

Let us see how to use this representation framework in order to reason. In this paper, we call *reasoning* the process of evaluating the acceptability of a formula  $\varphi \in \mathcal{L}$ , i.e., mapping  $\varphi$  to a flag in  $\{acc, rej, \emptyset\}$ . The reasoning is not the same in S1 and S2. In S1, reasoning is based on the association table  $AT$  while in S2 it is based on an inference principle. We propose to encode S1-reasoning as follows: if the current formula has a non-empty flag, then this flag is returned; else, the head of the stack associated to the current formula becomes the current formula. The agent does this procedure until she is not able to reach new formulae in her S1 system anymore.

**Definition 3 (S1-reasoning)** Given a cognitive model  $\kappa = (KB, AT, \lambda, i)$ , we define S1-reasoning, about a formula  $\varphi$ , denoted  $eval_1(\varphi, \kappa)$ , as follows:

$$eval_1(\varphi, \kappa) = evalr(\varphi, \{\varphi\})$$

$$evalr(\varphi, R_\varphi) = \begin{cases} flag(\varphi) & \text{if } flag(\varphi) \neq \emptyset, \\ evalr(Top(\varphi), R_\varphi \cup \{Top(\varphi)\}) & \text{if } flag(\varphi) = \emptyset \text{ and } Top(\varphi) \notin R_\varphi, \\ \emptyset & \text{otherwise.} \end{cases}$$

where  $R_\varphi$  is the set of formulae visited by  $evalr$ .

**Example 2** Given the association table shown in Fig. 1, the result of  $eval_1(imm)$  is  $rej$ . Indeed, since the formula  $imm$  has the flag  $\emptyset$ , the S1-reasoning gets the top formula of the stack associated to  $imm$ , which is  $blc$ . The flag of  $blc$  being different than  $\emptyset$ , it is the result of the evaluation.

**Definition 4 (S2-reasoning)** Given a cognitive model  $\kappa = (KB, AT, \lambda, i)$ , S2-reasoning is defined by:

$$eval_2(\varphi, \kappa) = f \text{ s.t. } KB \vdash (\varphi, f)$$

### 3 Argument Evaluation

#### 3.1 Argument and Profiles

We first give a (restrictive<sup>3</sup>) definition of an argument: it is an entity containing a source (which is a formula and represents the speaker), a premise (which is a formula and can be itself a conjunction of more basic premises), a warrant (an *a-rule*) and a conclusion associated with an acceptance flag.

**Definition 5 (Argument)** *An argument is a tuple  $(s, h, w, (c, f))$  where  $s$  is the speaker enunciating the argument,  $h$  is a formula (the hypothesis of the argument),  $w$  (the warrant) is an *a-rule*,  $c$  is a formula (the conclusion) and  $f \in \{acc, rej\}$  is a flag stating that the argument conclusion should be accepted or rejected.*

In the ELM model [15], the determination of the “route” is made thanks to two main factors: the *interest* in processing the message and the *ability* (wrt knowledge and cognitive availability) to process it. We use these factors in order to define different profiles of agents (note that we leave the cognitive availability for future work). Such profiles represent typical (and extreme) dispositions wrt the evaluation of an argument.

- **Interest**, given by the function  $i$  (see Def. 2). An agent may be not interested by a formula  $\varphi$  ( $i(\varphi) = 0$ ), interested ( $i(\varphi) = 1$ ) or “fanatic” ( $i(\varphi) = 2$ ).
- **Knowledge**, represented by the size of the stack related to  $\varphi$  in  $AT$ . This size is compared to the agent’s threshold  $\lambda$  (see Def. 2) in order to link the quantity of information the agent has to his feeling about being sufficiently aware on  $\varphi$ .

These two parameters are used to determine the agent profile. We introduce four level of engagement: unconcerned, enthusiastic, quiescent or engaged with increasing involved level of cognition (see Definitions 8, 9, 10, 11).

**Definition 6 (Profile)** *The profile of an agent is a function that maps a formula  $\varphi \in \mathcal{L}$  and a cognitive model  $\kappa = (KB, AT, \lambda, i)$  to an element of  $\{unc, ent, qui, eng\}$ :*

$$\text{profile}(\varphi, \kappa) = \begin{cases} unc & \text{if } i(\varphi) = 0 \\ qui & \text{if } i(\varphi) = 1 \text{ and } |AT(\varphi)| < \lambda \\ eng & \text{if } i(\varphi) = 1 \text{ and } |AT(\varphi)| \geq \lambda \\ ent & \text{if } i(\varphi) = 2 \end{cases}$$

The following postulate expresses that if an agent is **enthusiastic** about a formula  $\varphi$ , then she has an opinion about  $\varphi$ .

**Postulate 1**  $\text{profile}(\varphi, \kappa) = ent$  implies  $\text{flag}_\kappa(\varphi) \neq \emptyset$

The next section details the value of the function  $\text{evalarg}$  defined below.

**Definition 7** *Given a cognitive model  $\kappa = (KB, AT, \lambda, i)$ , an argument  $a = (s, h, w, (c, f))$  and a profile  $p = \text{profile}(c, \kappa)$ , let  $\text{evalarg}$  be a function that maps  $a$  and  $p$  to an evaluation of the argument in  $\{acc, rej, \emptyset\}$ , denoted as  $\text{evalarg}(a, p)$ .*

#### 3.2 Argument Evaluation According to Profiles

In this section, we introduced the four profiles and we define formally how the evaluation is done with respect to a profile.

**Unconcerned** As its name implies, the unconcerned profile represents the fact that no interest is given by the agent in the received argument. Hence, an unconcerned agent will not bother trying to evaluate this argument and will just discard it.

**Definition 8** *Given an argument  $a = (s, h, w, (c, f))$ , the evaluation of  $a$  by an unconcerned agent  $unc$  is never done.*

<sup>3</sup>Here, we only consider arguments in favor of appreciations and not in favor of beliefs.

**Enthusiastic** The enthusiastic profile represents the fact that an agent is already convinced. As such, she does not feel the need to evaluate rationally the argument and will just check if the flag of the argument's conclusion correspond to the flag in her  $AT$ .

**Definition 9** Given an argument  $a = (s, h, w, (c, f))$ , the evaluation of  $a$  by an enthusiastic agent  $\text{evalarg}(a, \text{ent}) = \text{acc}$  iff  $\text{eval}_1(c) = f$  else  $\text{evalarg}(a, \text{ent}) = \text{rej}$ .

**Quiescent** A quiescent profile represents an “ideally instinctive” agent evaluating an argument thanks to her S1. More precisely, when receiving an argument, the agent evaluates the argument's conclusion and the speaker. She will accept the argument if she agrees with the conclusion and does not reject the speaker, or vice-versa.

**Definition 10** Given an argument  $a = (s, h, w, (c, f))$ , the evaluation of  $a$  by a quiescent agent with a cognitive model  $\kappa$  is defined as follows:

$$\text{evalarg}(a, \text{qui}) = \begin{cases} \text{acc} & \text{if } (\text{eval}_1(c, \kappa) = f \text{ and } \text{eval}_1(s, \kappa) \neq \text{rej}) \text{ or} \\ & (\text{eval}_1(c, \kappa) \neq -f \text{ and } \text{eval}_1(s, \kappa) = \text{acc}), \\ \text{rej} & \text{if } (\text{eval}_1(c, \kappa) = -f \text{ and } \text{eval}_1(s, \kappa) \neq \text{acc}) \text{ or} \\ & (\text{eval}_1(c, \kappa) \neq f \text{ and } \text{eval}_1(s, \kappa) = \text{rej}), \\ \emptyset & \text{otherwise} \end{cases}$$

In future work, we plan to take into account the extra sources of persuasion such as the context created by the source of information including trustworthiness and charisma of the source, the contextual mood of the agent, etc.

**Example 3** After a hard day's work, when Alex's wife said that “since immigration leads to less jobs, which would be dreadful, immigration has to be rejected”, he did not have the cognitive availability to rationally consider this argument. While he would not have agreed with a deeper analysis, he instead relied on his S1, where immigration is associated with something he rejects (see Fig. 1), and therefore accepted the argument.

**Engaged** An engaged profile represents an “ideally rational” agent evaluating an argument exclusively thanks to her knowledge base. In this work, we suppose that an engaged agent evaluates an argument in a way similar to critical questions (see [23]). More precisely, given an argument  $a = (s, h, (w_h \rightsquigarrow (w_c, w_f)), (c, f))$ , an engaged agent has to use her S2 in order to check that:

- The premises are consistent with the knowledge of the agent ( $K \not\vdash \neg h$ ),
- There is no exception to the warrant in the context  $h$  ( $KB \cup \{h \wedge w_h\} \not\vdash (w_c, -w_f)$ ),
- The premises and the warrant entail the conclusion ( $(\{h\}, \emptyset, \{w\}) \sim (c, f)$ ).

Note that we also have to check that the conclusion of the argument is not already deducible in order to ensure that the premises and the warrant are well required and are well related to the conclusion. In our particular case, it amounts to check if it is not always accepted or rejected (i.e.  $(\emptyset, \emptyset, \emptyset) \not\vdash (c, f)$ ). This leads us to the following:

**Definition 11** Given an argument  $a = (s, h, (w_h \rightsquigarrow (w_c, w_f)), (c, f))$ , the evaluation of  $a$  by an engaged agent with a cognitive model  $\kappa = (KB, AT, \lambda, i)$  with  $KB = (K, F, R)$  is defined as follows:

$$\text{evalarg}(a, \text{eng}) = \begin{cases} \text{acc} & \text{if } \left\{ \begin{array}{l} K \not\vdash \neg h \text{ and} \\ \text{eval}_2(w_c, KB \cup \{h \wedge w_h\}) \neq -w_f \text{ and} \\ (\text{eval}_2(c, (\{h\}, \emptyset, \{w\}))) = f \text{ and } \text{eval}_2(c, (\emptyset, \emptyset, \emptyset)) = \emptyset \end{array} \right. \\ \text{rej} & \text{if } \left\{ \begin{array}{l} KB \vdash \neg h \text{ or} \\ \text{eval}_2(w_c, KB \cup \{h \wedge w_h\}) = -w_f \text{ or} \\ (\text{eval}_2(c, (\{h\}, \emptyset, \{w\}))) = -f \text{ or } \text{eval}_2(c, (\emptyset, \emptyset, \emptyset)) \neq \emptyset \end{array} \right. \\ \emptyset & \text{otherwise} \end{cases}$$

**Example 4** Several days after the discussion with his wife, Alex thought of her argument again. The argument premise contains  $\text{imm} \wedge (\text{imm} \rightarrow \neg l\text{Jobs})$ . Now that he is able to analyse the argument more rationally, he logically deduces from his knowledge that immigration does not lead to less jobs (i.e.  $K \vdash \text{imm} \rightarrow \neg \neg l\text{Jobs}$ ), which means that the premises of the argument are inconsistent with Alex's knowledge. Hence, Alex rejects the argument.

profile( $c, \kappa$ )	evalarg( $(s, h, w, (c, f))$ )		
	$\emptyset$	$rej$	$acc$
$unc$	$\text{push}(\kappa, (c, h))^4$	$\times$	$\times$
$ent$	$\times$	$\text{push}(\kappa, (c, h))$ $\text{setflag}(\kappa, s, rej)$	$\text{push}(\kappa, (h, c))$ $\text{push}(\kappa, (c, h))$ $\text{setflag}(\kappa, s, acc)$
$qui$	$\text{push}(\kappa, (c, h))$	$\text{push}(\kappa, (c, h))$ $\text{setflag}(\kappa, c, -f)$ $\text{setflag}(\kappa, s, rej)$	$\text{push}(\kappa, (h, c))$ $\text{push}(\kappa, (c, h))$ $\text{setflag}(\kappa, c, f)$ $\text{setflag}(\kappa, s, acc)$
$eng$	$\times$	$\times$	$\text{addrule}(\kappa, w)$ $\text{addrule}(\kappa, (h \rightsquigarrow (c, f)))$ $\text{setflag}(\kappa, c, f)$

**Table 1:** Update of a cognitive state  $\kappa$ .

### 3.3 Argument Influence on the Agent's Mind

Once the argument has been evaluated by an agent, her cognitive model may have to be modified to account for the persuasive impact of the argument. Such modifications can either be the change of a flag value, the addition of a new association or the addition of an appreciation rule. Def. 12 gives the functions representing these modifications.

**Definition 12 (Update operations)** Given two cognitive states  $\kappa = (KB, AT, \lambda, i)$  with  $KB = (K, F, R)$  and  $\kappa'$ , two formulas  $x, y \in \mathcal{L}$  and a flag  $f \in \{acc, rej, \emptyset\}$ , we define:

- $\text{setflag}(\kappa, x, f) = \kappa'$  where  $\kappa' = ((K, F', R), AT', \lambda, i)$  with
  - $L_{AT'} = L_{AT}$ ,
  - $\forall \varphi \in L_{AT}$  s.t.  $\varphi \neq x$ ,  $\text{flag}_{\kappa'}(\varphi) = \text{flag}_{\kappa}(\varphi)$  and  $AT'(\varphi) = AT(\varphi)$ ,
  - $\text{flag}_{\kappa'}(x) = f$  and  $F' = F \cup \{(x, f)\}$  and  $AT'(x) = AT(x)$ .
- $\text{push}(\kappa, (x, y)) = \kappa'$  where  $\kappa' = (KB', AT', \lambda, i)$  with
  - if  $x \notin L_{AT}$  then  $AT' = AT \cup \{(x, S_x, \emptyset)\}$  with  $S_x = \{y\}$ ,
  - else
    - \*  $\forall \varphi \in L_{AT}$  s.t.  $\varphi \neq x$ ,  $\text{flag}_{\kappa'}(\varphi) = \text{flag}_{\kappa}(\varphi)$  and  $AT'(\varphi) = AT(\varphi)$ ,
    - \*  $\text{flag}_{\kappa'}(x) = \text{flag}_{\kappa}(x)$  and  $AT'(x) = AT(x) \cup \{y\}$  with  $\text{Top}(x) = y$ ,
- $\text{addrule}(\kappa, r) = \kappa'$  where
  - $\kappa' = \kappa$  if  $r \in R$ ,
  - else  $\forall t, t' \in r$ ,  $\kappa' = \text{push}(((K, F, R \cup \{r\}), AT, \lambda, i), (t, t'))$ .

Depending on the profile, the cognitive model will be modified in different ways. These differences aim at representing the fact that the persuasion may be deeper depending on the cognitive involvement of the agent. Table 1 gives the functions to apply to  $\kappa$  in order to update it, according to the possible evaluations of an argument by an agent and her profile. The “ $\times$ ” in the *ent* and *unc* lines corresponds to impossible cases due to, respectively, Postulate 1 and Def. 8. Concerning the *eng* line, “ $\times$ ” indicates the fact that an engaged agent is not influenced by what she hears when she does not accept the argument.

## 4 Properties

Let us first notice that after receiving an argument, the knowledge of an agent can only increase: more precisely, among the formulae that were already present, the number of flags that are not empty decreases (however some new formula may be added with an empty flag) and the number of associations grows. Moreover some rules can also be added in the case of an engaged profile.

**Proposition 1** Let  $\kappa = ((K, F, R), AT, \lambda, i)$ ,  $\kappa' = ((K', F', R'), AT', \lambda', i')$  such that  $\kappa'$  is the cognitive model obtained from  $\kappa$  after the utterance of an argument. It holds that  $L_{AT} \subseteq L_{AT'}$ ,  $\forall \varphi \in L_{AT}, AT(\varphi) \subseteq AT'(\varphi)$ , and  $K = K'$  and  $R \subseteq R'$  and  $\lambda = \lambda'$  and  $i = i'$ .

*The flag values are non-monotonic since a formula can obtain either an accepted, rejected or empty flag depending on the engagement profile.*

According to [24], the model of how information is transformed in public opinion follows four axioms mentioned below. Our proposal satisfies these axioms:

**Reception axiom:** *The greater the level a person’s level of cognitive engagement with an issue the more likely he / she will be exposed to and comprehend political messages concerning that issue.* It holds since an unconcerned agent does not evaluate the argument, an enthusiastic agent takes it into account if she agrees with the conclusion, a quiescent agent evaluates it with S1-reasoning and an engaged agent evaluates it with S2-reasoning. Hence, the more engaged an agent is, the more information she takes into account (in the following order: unconcerned, enthusiastic, quiescent, engaged).

**Resistance axiom:** *People tend to resist arguments that are inconsistent with their political predispositions but they do so only to the extent that they possess the contextual information necessary to perceive a relationship between the message and their predispositions.* Unconcerned, enthusiastic and engaged agents may resist an argument since they are not influenced by its flag. A quiescent agent resists arguments that are against her opinion or uttered by a source she rejects (see Definition 10).

**Accessibility axiom:** *The more recently a consideration has been called to mind, or thought about, the less time it takes to retrieve that consideration or related considerations from memory and bring them to the top of the head for use.* This axiom is satisfied concerning the association table *AT* since every kind of profile add the new piece of information at the top of the stack (see Table 1).

**Response axiom:** *Individuals answer survey questions by averaging across the considerations that are immediately salient or accessible to them.* It holds for quiescent and enthusiastic: a quiescent agent evaluates a formula by considering the most immediately accessible information and an enthusiastic agent evaluates only the immediate value of a formula. However, it does not hold for unconcerned and engaged agents: one does not evaluate the formula, and the other evaluates the formula with her knowledge base.

## 5 Conclusion

In this paper, we have shown a preliminary formal account of dual process theory and its link with persuasion. Based on a classic model of persuasion, ELM, we defined four profiles evaluating an argument in different ways. In particular, one of the profiles aims at representing the cognitive biases that may arise with error prone reasoning. Moreover, each of these profiles implies different types of update of the cognitive model. This allows us to represent that, when an argument is accepted, the more cognition was involved in its evaluation, the more persuasive content will take root in the mind of the agent.

### 5.0.1 Related Work

To tackle general artificial intelligence problems several cognitive architectures inspired by dual process theories have recently been implemented. [9] presents an extension of the CLARION problem solving architecture that relies on two modules: a bottom-level module allowing to handle implicit knowledge and a top-level module managing explicit knowledge, which recall respectively the S1 and S2 systems of dual process theory. More directly connected to dual process theory, [21] proposes a general intelligence cognitive architecture based on [11]. This architecture is composed of a long-term memory that is independent of specific tasks and a working memory that is a capacity-limited list of elements of the long-term memory in accordance with the context. The S1 and S2 systems are represented thanks to two binary relations on the element of the long-term memory and two propagation processes. This allows them to distinguish between perception and imagination. In an approach similar to ours, some works aim at explaining purely human processes. For instance, [14] studies the emergence of emotions thanks to a three-levels cognitive architecture based on an extension of dual process theory [20]. In this architecture, S1 (the *reactive* level) is responsible for fast and instinctive behaviours while S2 is subdivided into the *algorithmic* level, used for cognitive control, and the *reflective* level, handling rational behaviour. [13] proposes a cognitive architecture called mReasoner aiming at reasoning on syllogistic arguments by relying on three systems. The S0

---

<sup>4</sup>An argument is never evaluated by an unconcerned agent. We represent the fact that, like enthusiastic and quiescent agents, she is unconsciously influenced by what she hears.



computes *intentions* which are representations of the meaning of the premises. S1 uses these intentions to create a model representing the different possible subjects the premises may refer to; the S1 then reasons heuristically by drawing a conclusion that is true in this model. Finally, the S2 looks for other models that may contradict the initial model, which is a slow process that may confirm or invalidate the S1 conclusion. It is interesting to note that S1 and S2 of [13] are similar to S1 and S2 of [22]. However, none of the existing work are specifically interested in the analysis of persuasion through cognitive biases.

### 5.0.2 Perspectives

Our work is a preliminary step in the study of the link between persuasion and cognitive biases, and, as such, opens numerous perspectives. In particular, we are interested in taking into account the cognitive availability in order to determine which system will be used to reason. Such study would benefit from the definition of a profile corresponding to a continuum between the quiescent and the engaged profiles. This new profile might better represent human reasoning and might allow us to study the link between persuasion and bounded rationality, together with the influence of S1 on S2-reasoning. Moreover, we do not take explicitly into account influence between agents such as social proof, authority, etc. Yet, being instinctive, a quiescent agent is naturally drawn to follow non-rational cues. As such, work such as [4] and rhetoric could prove very promising in our formalization. Finally, even if we have shown the correspondence between our work and the public opinion axioms of [24], a study in collaboration with psychologists would indubitably be useful to check the soundness of our approach.

## References

- [1] I. Ajzen and M. Fishbein. Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, 84(5):888, 1977.
- [2] G. Allport. *Attitudes*. Handbook of Social Psychology. Clark Univ. Press, 1935.
- [3] J. A. Blair. Argumentation as rational persuasion. *Argumentation*, 26(1):71–81, 2012.
- [4] R. Cialdini. *Influence: Science and Practice*. Allyn and Bacon, 2001.
- [5] C. S. Clements. Perception and Persuasion in Legal Argumentation: Using Informal Fallacies and Cognitive Biases to Win the War of Words. *BYU Law Review*, 319, March 2013.
- [6] S. Epstein. Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, 49(8):709, 1994.
- [7] J. S. B. Evans, D. E. Over, et al. *Rationality and reasoning*. Psychology Press, 2013.
- [8] R. H. Fazio and M. P. Zanna. Direct experience and attitude-behavior consistency. *Advances in experimental social psychology*, 14:161–202, 1981.
- [9] S. Hélie and R. Sun. Incubation, insight, and creative problem solving: a unified theory and a connectionist model. *Psychological Review*, 117(3):994–1024, 2010.
- [10] J. Hornikx and U. Hahn. Reasoning and argumentation: Towards an integrated psychology of argumentation. *Thinking & Reasoning*, 18(3):225–243, 2012.
- [11] D. Kahneman. A perspective on judgment and choice: Mapping bounded rationality. *American psychologist*, 58(9):697–720, 2003.
- [12] D. Kahneman and S. Frederick. Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and biases: The psychology of intuitive judgment*, 49, 2002.
- [13] S. Khemlani and P. N. Johnson-Laird. The processes of inference. *Argument & Computation*, 4(1):4–20, 2013.

- [14] O. Larue, P. Poirier, and R. Nkambou. Emotional emergence in a symbolic dynamical architecture. In A. Chella, R. Pirrone, R. Sorbello, and K. R. Jóhannsdóttir, editors, *BICA 2012*, volume 196 of *AISC*, pages 199–204. Springer Berlin Heidelberg, 2013.
- [15] R. Petty and J. Cacioppo. The Elaboration Likelihood Model of Persuasion. *Advances in Experimental Social Psychology*, 19(C):123–205, 1986.
- [16] R. E. Petty, J. T. Cacioppo, and R. Goldman. Personal involvement as a determinant of argument-based persuasion. *Journal of personality and social psychology*, 41(5):847, 1981.
- [17] I. Rahwan and L. Amgoud. An argumentation based approach for practical reasoning. In *Proc. of Int. J. conf. on Autonomous Agents and MultiAgent Systems*, pages 347–354, 2006.
- [18] E. A. Ross. *Social psychology: An introduction and source book*. Macmillan, 1908.
- [19] S. A. Sloman. The empirical case for two systems of reasoning. *Psychological Bulletin*, 119(1):3, 1996.
- [20] K. Stanovich. *Rationality and the Reflective Mind*. Oxford University Press, 2010.
- [21] C. Strannegard, R. von Haugwitz, J. Wessberg, and C. Balkenius. A cognitive architecture based on dual process theory. In K.-U. Kühnberger, S. Rudolph, and P. Wang, editors, *AGI 2013*, volume 7999 of *LNCS*, pages 140–149. Springer Berlin Heidelberg, 2013.
- [22] A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974.
- [23] D. Walton, C. Reed, and F. Macagno. *Argumentation Schemes*. Cambridge University Press, Cambridge, 2008.
- [24] J. Zaller. *The Nature and Origins of Mass Opinion*. Cambridge Studies in Political Psychology Series. Cambridge University Press, 1992.