The role of cognitive biases in reactions to bushfires

Maël Arnaud
Univ. Grenoble Alpes, LIG, F-38000
Grenoble, France

Carole Adam
Univ. Grenoble Alpes, LIG, F-38000
Grenoble, France

Julie Dugdale
Univ. Grenoble Alpes, LIG, F-38000
Grenoble, France
University of Agder, Norway

ABSTRACT

Human behaviour is influenced by many psychological factors such as emotions, whose role is already widely recognised. Another important factor, and all the more so during disasters where time pressure and stress constrain reasoning, are cognitive biases. In this paper, we present a short overview of the literature on cognitive biases and show how some of these biases are relevant in a particular disaster, the 2009 bushfires in the South-East of Australia. We provide a preliminary formalisation of these cognitive biases in BDI (beliefs, desires, intentions) agents, with the goal of integrating such agents into agent-based models to get more realistic behaviour. We argue that taking such "irrational" behaviours into account in simulation is crucial in order to produce valid results that can be used by emergency managers to better understand the behaviour of the population in future bushfires.

Keywords

Multi-agent modelling, social simulation, cognitive biases, BDI paradigm, Victoria bushfires

INTRODUCTION

The bushfires in Victoria, Australia, on February 7, 2009 (also known as Black Saturday) caused 173 deaths and 414 injuries. Most of the victims were badly prepared to face a fire of such severity and were caught out by surprise (*Victorian 2009 Bushfire Research Response Final Report* 2009; Thornton, 2010). The population's behaviour before and on that day is still not fully understood and is sometimes referred to as being irrational (*i.e.* people did not behave according to what would objectively be in their best interests). This was surprising since most of the victims had lived in high fire-risk areas for many years and were aware of the safety issues. The emergency services expected most of the population in the affected area to evacuate before the fire arrived. This was not the case, highlighting a problem with Victoria's crisis management plan, including the communication with the public.

Agent-based modelling and simulation (ABMS) is a technique from Artificial Intelligence (AI) that provides us with a tool to model human behaviours and to test the effect of several parameters. By specifying the attributes of each individual (*e.g.* stress, exhaustion, knowledge, and also fire plans, etc.) at the micro-level, we can observe the overall behaviour during fires at the macro-level (*e.g.* the number of people who escape the fire, number of deaths). The proposed approach is to apply ABMS to the field of crisis management in order to improve evacuation plans. The application of such a technique has been used previously (Pan et al., 2007; Cardon, 1998; Dugdale et al., 2010) and it is still a very active area of research. One of the most recurrent issues is the level of complexity of the agents; should humans be modelled with simple rules, following the KISS (Keep It Simple, Stupid) principle advocated by (Axelrod, 1997), or is it better to create more complex agents that are closer to reality, following the KIDS (Keep It Descriptive, Stupid) principle advocated by (Edmonds and Moss, 2004)?

The specific difficulty of choosing an agent's level of complexity when trying to model complex human-like agents has already been discussed (Adam and Gaudou, 2016; Dugdale, 2010). These authors advocate that an agent

trying to mimic human behaviour can not be efficiently modelled with simplistic rules. They concluded that the Belief-Desire-Intention (BDI) agent architecture offers some appreciable features such as: "adaptability, robustness, abstract programming and the ability [for agents] to explain their behaviour."

These features are an undeniable asset in order to model and understand the behaviour of the Victorian population during a fire event. Nonetheless, as stated by (Norling, 2004), the BDI architecture does not capture many aspects of human behaviour and reasoning. It therefore needs some additional work in order to better mimic human behaviour and serve as a grounding model for social agents (Dignum et al., 2014). Throughout this paper we argue that part of this additional work should deal with cognitive biases. We advocate that they play an important role in human behaviour and need to be taken into account when developping a model, especially during crisis situations.

Cognitive biases are known to be mechanisms widely used by the human brain and which are sometimes unadapted to the situation, leading to mistakes or inaccuracies. Cognitive biases are believed to be found in any strategic decision process (Das and Bing-Sheng, 1999). They are also found in decision-making processes in uncertain situations (Tversky and Kahneman, 1974). Strategic decision making and deciding under uncertainty are typical cognitive reasoning mechanisms undertaken by people in crisis situations such as bushfires (Adam et al., 2015). In addition, crisis situations impose a time constraint on reasoning. Therefore strategic decision making and reasoning with imperfect information in limited time are common, making cognitive biases more likely to happen.

Our overall goal is to build a BDI model (Adam et al., 2016) of human behaviour in crisis in order to improve social simulations by better mimicking human behaviour. In light of the information above, we can see that cognitive biases strongly affect human behaviour during crisis situations. Specifically, they can be the cause of biased perceptions and/or judgements, leading to dangerous behaviours for oneself and reliant others. As a result, we claim that their integration in agent models will improve social simulations of crisis situations. The first goal of this paper is to prove this impact of cognitive biases in crisis situations by identifying some well-known cognitive biases in the victims' testimonies of the Black Saturday event. The second goal is then to formalise these cognitive biases in terms of BDI algorithms, in order to make a first step towards their integration in an agent model for social simulation.

Similar work of integrating cognitive biases in a model has already been performed in the domains of health (Voinson et al., 2015) and national defense (Kulik and Davis, 2002). However none of these models use the BDI paradigm, which is central in our approach.

The following section presents our data set and the methodology used for data extraction. A presentation of the BDI paradigm follows, with an explanation about the required architecture to implement our BDI model of cognitive biases. We then describe our formalisation of three cognitive biases in the form of algorithms, and illustrate these biases with examples from our data set. Finally the conclusion discusses limits of this work and compares our approach to the literature.

FINDING AND EXTRACTING DATA

The first task in order to ascertain whether or not cognitive biases played a part in Black Saturday victims' behaviour is to find relevant data. Black Saturday was the worst fire event Victoria ever suffered, consequently casualties and material losses were severe. As a result the state of Victoria created several work groups to understand what happened. Two of them are the *2009 Victorian Bushfires Royal Commission* (VBRC) and the *Bushfire Cooperative Research Centre* (Bushfire CRC). These groups drew data from what they could find (*e.g.* flora, fauna, victims interviews stating their feelings and memories, police hearings, and general state of the land) and wrote detailed reports (*Victorian 2009 Bushfire Research Response Final Report* 2009; *Final report* 2009). The data contained in these reports were extensively used in the scope of this work.

Post Black Saturday reports

Particular attention has been given to the fourth volume of the VBRC final report (*Final report - vol. IV : the statements of lay witness* 2009). It contains 100 statements from lay witnesses of the Black Saturday events, ordered alphabetically by the witness' name. Statements are heterogeneous in terms of content and source (*e.g.* testimonies from firefighters who helped during the events, information about the physical and mental state of victims from nurses who came later to help, testimonies of individuals focusing on their struggle with the insurance companies, testimonies of Black Saturday events from family members who suffered a human loss, and detailed testimonies of the day from people who survived the events.) They are narrative transcriptions of oral statements with attached evidence (*e.g.* photos, maps). Statements are divided into numbered paragraphs and attempts have been made by the interviewer to create consistent sections throughout the statements when possible (*e.g.* fire plan, the property, views previous to Black Saturday, views during Black Saturday).

The bushfire CRC Final report (*Victorian 2009 Bushfire Research Response Final Report* 2009) has also been studied in depth. Sections about human behaviour helped in the formulation of the hypothesis relative to the importance of cognitive biases (*e.g.* by highlighting the importance of environmental cues in alerting Victorians and thus the unconscious concealment of other cue types). This report was also used for the statistics that it contains about the Victorians intentions and actions during Black Saturday. These statistics were compared to those found in a 2010 survey about the behaviour that Victorians would adopt on Code Red (*i.e.* worst conditions for bushfires) days (Research, 2010), in order to form a more accurate idea of how Black Saturday affected the Victorians representations of a fire event. This second report concludes by stating that only 3% of Victorians would leave on a Code Red day, even though they aware that authorities' advice is to leave early on those days. Only 15% would stay and defend. This leaves 82% of Victorians in a "wait and see" state, which is the reason why so many Victorians were trapped during Black Saturday fires.

More on cognitive biases

Cognitive bias is a broad and poorly defined concept that encompasses roughly a hundred effective cognitive biases (see (Benson, 2016) for a large but non-exhaustive list). Some are well known, while others are merely suspected. Some are known to be groups of cognitive biases (*e.g.* the Risky Shift Phenomenon is a cognitive bias discovered in 1960, stating that decisions made as a group are less conservative than decisions made by the average group member (Shaw et al., 1976); this can in fact be caused by other, more specific, cognitive biases). Some are more concerned with affection (loving) (Reyna and Brainerd, 2008; Kahneman and Frederick, 2002))than cognition (thinking). Nonetheless all cognitive biases have two things in common.

First they all serve a common purpose. They help the brain to do its work faster and/or longer, often by using low-cost mechanisms to make decisions or to filter the amount of information gathered by the brain at any time. For example, the "Attribute substitution" mechanism enables us to give the answer to an easy question when confronted with a difficult one (Kahneman and Frederick, 2002) (*e.g.* when asked the question "How far is the mountain ?", if one does not know the answer, one will answer the question "How blurred is the mountain compared to the ambient haziness?" instead.)

Second they can be the source of mistakes. Since their purpose is to help the brain to make a decision faster or to ignore information, it is not surprising to find that these biases can sometimes lead to inaccurate decisions. Nonetheless it is worth noting that despite the denomination of "cognitive biases" inaccuracies seem to occur only in specific, rather uncommon, situations (Pohl, 2004).

Examining the testimonies

Cognitive biases are a wide and heterogeneous family; hence the goal was to find out more about which specific cognitive bias could be found in Black Saturday testimonies. For this purpose we went manually through 30 of the testimonies gathered by the VBRC. Since little attention has been given in the past to identifying cognitive biases in testimonies of populations during crisis situations, no tool exists yet to automatically analyse such interviews. Therefore a new method had to be designed.

Half of testimonies were ignored, since they could not be of any help (*i.e.* the interviewee was not talking specifically about the Black Saturday events). The other half were searched for situations where survivors or deceased people put themselves at risk. Each such dangerous situation was isolated and we looked for the decision which could have led to it. Then this decision was interpreted according to a set of well-established cognitive biases. For example Edward Cherry is in a dangerous situation (p.16) when outside of his house during an ember attack (ember attacks precede the fire front, which is the most dangerous part of a bushfire event). He took the decision to water the roof of his house a bit longer , while his wife was "shrieking at him to come inside", before suddenly becoming aware that everything around him was on fire. This is likely to be caused by an Optimism bias (*i.e.* a cognitive bias which lead people to under-estimate the risk of a given situation they are involved in compared to the estimation they would make if not involved). It is worth noting that at least one cognitive bias or set of cognitive biases potentially leading to each dangerous situation were found. Sometimes several unique sets were found.

In the following example, several cognitive biases could led to the decision. Brett Savage in Michelle Buntine's testimony (*Final report - vol. IV : the statements of lay witness* 2009) evacuates his property after having tried to defend it (p.31-32). Earlier, he did not listen to his partner in life, Michelle Buntine, when she told him repeatedly to evacuate (p.26, 28). He also received a call from a friend who is a firefighter and who, in order to make him leave, explained to Brett what would happen during the next hours if he stayed (p.29-30). He stayed and fought the fire for some time before having to evacuate during the fire event.
In terms of cognitive biases it is hard to detect what played a part in Brett's decision to stay and fight. It could be

the Conservatism bias (*i.e.* a slow update of his beliefs that he would be safe while defending his property), the Semmelweis Reflex (i.e he clung to his belief, even with compelling evidence that it was false), an Anchoring effect (*i.e.* he was committed to the idea to defend his house), the Optimism bias (*i.e.* he though he was not at risk in this situation while someone else would have been) or the Overconfidence effect (*i.e.* he considered his beliefs as more accurate than the ones from his partner and friend). The most simple, robust, and consistent with the psychology literature sets were kept. However, we acknowledge that some may have been missed.

Retained cognitive biases

The use of our method revealed that some types of cognitive biases were found more often than others: biases about 1) beliefs (*i.e.* biases acting in favour of, or against, the update of a previous belief), 2) affect (*i.e.* biases favouring the option in which something one likes could be preserved over the option in which one is in a safe place) and 3) bad probability estimation (*i.e.* some options are seen as far more/less probable than they really are).

Three cognitive biases were chosen, according to this finding and for two other reasons: biases chosen are relevant to each other (*i.e.* the same tools can be used to model the three of them, and their interaction creates a logical model of information processing); they are easily adaptable to the BDI paradigm, which makes an extensive use of the notion of belief. The three biases are: 1) The Neglect of Probability bias, which belongs to the bad probability estimation category, 2) the Semmelweis Reflex and 3) the Illusory Truth effect. The last two biases belong to the belief updates category. No cognitive bias found in the affection category satisfied the conditions and so have not been included.

PROPOSED BDI MODEL OF COGNITIVE BIASES

The BDI paradigm

The BDI paradigm describes agents in terms of their mental attitudes (Beliefs, Desires, Intentions), which are "folk psychology concepts that straightforwardly match human reasoning as people understand it, making the models easier to design and to understand" (Adam and Gaudou, 2016).

1. **Beliefs** represent an agent's knowledge about the world (*e.g.* road 45 is blocked, ember attacks precede the fire front).
2. **Desires** are the goals of an agent, *i.e.* their preferred states (*e.g.* be close to one's children). They can be inconsistent (*e.g.* during an ember attack, the agent might both want to defend its house and to flee the fire). Choosing between desires is the role of a reasoning engine (Norling and Sonenberg, 2004), that remains to be chosen in this work.
3. **Intentions** are what an agent is committed to do. They are basically refined, practical desires. They have to be consistent with each other, and should not be dropped easily. The agent usually has a library of plans (sequences of actions) allowing it to achieve its intentions.

Required agent model

In this section we describe the required minimal architecture for an agent to be able to implement the cognitive biases. We do not provide a conceptual agent model, but give a minimal set of requirements for an agent architecture to be able to implement the cognitive biases algorithms in this paper.

- A belief operator with associated subjective probability, ranging from 0 (certainly wrong) to 100 (certainly true). The addition of subjective values of a belief and its opposite sum to 100. A belief with a probability of 50 means that the agent is completely unsure.
- A belief base that stores beliefs.
- Psychological attributes, such as level of risk aversion.
- A table of occurrences of received information (counting how many times the same information have been received)
- A function **getBeliefProbability (Belief φ)** that returns the probability of a given belief.

- A function **saveBeliefProbability (Belief φ, Number α)** that updates the probability of a given belief in the agent's belief base.
- A function **incrementNumberOfOccurrences(Information φ)** that increments by 1 the number of times a given information has been gathered by the agent.
- A function **getAcquiredInfoOccurrences(Information φ)** that returns the number of times the agent has gathered a given information.
- A function **acquireBelief(Belief φ, Number α)** that saves a given belief and associated probability to the agent's belief base.
- The functions **dramaticallyIncreaseBeliefProbability (Belief φ, Number α)** and **decreaseBeliefProbability (Belief φ, Number α)** that modifies the probability of a given belief.

This minimal required architecture may be provided by PLEIAD (an affective agent coded in Prolog (Adam and Lorini, 2014)) or by GAMA (an agent-based simulation platform (Grignard. et al., 2013) recently enriched with a BDI architecture for the agents (Taillandier et al., 2016)).

MODEL OF THE THREE BIASES

In this section we provide a generic model of three cognitive biases using the architecture described above. Only pseudo-code is described so as not to commit to a particular implementation language.

Neglect of Probability

We formalise the Neglect of Probability bias in algorithm 1. We begin by explaining what we mean when using the expression "are perceived to be...", then we detail this bias and its mechanisms. Uses of the expression "are perceived to be ..." in the formalization of the Neglect of Probability bias are shortcuts for two statements.

- The consequences implied by this belief could be dire or extremely favourable (*e.g.* for agent A believing that road R is on fire implies that he could not reach his home, this will leave his home unprotected and at risk of probable burning; which is a dire event).
- The risk aversion or risk seeking value of the agent makes him perceive the consequences as extremely favourable/unfavourable, as described in (Kahneman and Tversky, 1979).

The Neglect of Probability bias is composed of three mechanisms. Two of them target low probability beliefs while the third one targets high and medium probability beliefs.

- The first mechanism (described in (Reyna and Brainerd, 2008)) is a way to ignore what is perceived both as unlikely to happen and as having no consequences. It is formalized on lines 9-10 of algorithm 1. It reads: if the agent estimates something only has a low chance of being true and has no consequences, then the agent considers it as false.
- The second mechanism (described in (Sunstein and Zeckhauser, 2010)) is a way to anticipate an event that would greatly impact one's life and thus enables one to prepare for it. It is formalized on lines 11-12 of algorithm 1. It reads: if the agent estimates that something has a low chance of being true but either desires or dreads it, then the agent considers that it has a substantial chance of being true.
- The third mechanism (described in (Tversky and Kahneman, 1983; Kahneman and Tversky, 1979)) reduces the perception of high and medium probabilities. It is formalized on lines 13-14 of algorithm 1. It reads: if the agent estimates something has a high or medium chance of being true, then the agent under-estimates its probability of being true.

Algorithm 1 Pseudo-code for the Neglect of Probability bias

```
1:  procedure UPDATEBELIEFPROBABILITY (info, perceivedProbability )
2:      ancientBeliefProbability ← getBeliefProbability (info)
3:      newBeliefProbability ← ancientBeliefProbability + perceivedProbability
4:
5:      if newBeliefProbability > 100 then
6:          newBeliefProbability ← 100
7:      end if
8:
9:      if newBeliefProbability is small and consequences are not perceived to be dire and consequences are not
    perceived to be extremely favourable then
10:         newBeliefProbability ← 0
11:     else if beliefProbability is small and (consequences are perceived to be dire or consequences are perceived
    to be extremely favourable) then
12:         newBeliefProbability ← dramaticallyIncraseBeliefProbability (info, newBeliefProbability)
13:     else    # newBeliefProbability is medium or high
14:         newBeliefProbability ← decreaseBeliefProbability (info, newBeliefProbability)
15:     end if
16:     saveBeliefProbability (info, newBeliefProbability)
17: end procedure
```

Alice Barber

Alice Barber is a woman who survived the events but was injured. She lost her house. No children or partner in life are recorded. She likes to garden and therefore her house is surrounded by vegetation. Prior to the event she possessed a fire plan, *i.e.* go to a place that she scouted earlier and identified as a potential shelter, and a fire plan trigger, *i.e.* act according to the fire plan as soon as the power is cut in the house.

- **The dangerous situation** is the fact that she lives in a high risk area with a weak fire plan trigger.
- **The decision leading up to the situation** is to choose a weak fire plan trigger.
- **The Neglect of Probability** is highlighted when she testifies about her mental attitude: "what will happen, will happen" (p. 7). She does not want to bother herself with costly fire preparations and does not consider a fire event as dire. Plus she has good reasons to believe that her fire plan trigger is good. Thus neglecting the probability of such an event.

This situation could be modelled by one agent as follows:

- **Beliefs:** Fire plan will be triggered before a fire is detected (100)
- **Desires:** Be safe, Be close to the house
- **Intentions:** Stay close to the house until the fire plan triggers, then follow fire plan
- **Risk aversion:** Low

Andrew Paul Trenwith Berry

Andrew Berry is a married man who survived the events with his wife, Nicole, and his father. Although Andrew and Nicole made relevant decisions regarding this case, they lost their house. Prior to the events Andrew possessed a house with a sprinkler-based defense system (considered as the best anyone can do (p. 14) by some Country Fire Authority (CFA) members) and a bunker that was close to the house that was large enough to accommodate a dozen people. The fire plan was to stay and defend. The family also experienced fire in 2006. This experience raised their risk aversion. This example is reversed, since the Neglect of Probability saved the family.

- **The safe situation** is the fact that Andrew and Nicole were able to shelter inside the bunker while their house burnt.

- **The decision leading up to the situation** is to have a the bunker built in the yard, despite a good house defense system.
- **The Neglect of Probability** underlies the decision of building the bunker. The 2006 fires raised Nicole and Andrew's father risk aversion. Thus the probability associated to the belief of experiencing a fire strong enough to threaten their lives while sheltered in their apparently well-defended house were overestimated (weather conditions during Black Saturday were unique and very unlikely).

This situation can be modelled with a single agent as follows:

- **Beliefs:** House is safe (60), Bunker is comfortable (0), Bunker is safe (70)
- **Desires:** Be safe, Be comfortable
- **Intentions:** Stay safe in the house before and during the fire front, then defend it. If not safe in the house, go to the bunker.
- **Risk aversion:** High

Fiona Wallace

Fiona has a partner in life named Heather and they both survived Black Saturday. No children are recorded as being involved. They are both Victorians, living in low fire-risk areas. They went to a Bed and Breakfast (B&B) hostel located in a high fire-risk area during Black Saturday. They both were aware of the warning issued by the authorities during the days prior to Black Saturday.

- **The dangerous situation** occurs when the fire reaches the B&B they are staying in and they have to defend it.
- **The decision leading up to the situation** is the decision to go on a trip in a high fire risk area while knowing about the hazardous weather conditions.
- **The first Neglect of Probability** underlying this decision occurs when Fiona and Heather hear the news about the weather: "the comparisons to Ash Wednesday emphasised the severity of the risk and we knew that it was going to be an extremely bad fire danger day." (p. 3). The probability of a fire being high in the information received is being neglected and is considered as medium rather than high.
- **The second Neglect of Probability** underlying this decision occurs when Fiona and Heather reason about the probability of having their particular B&B being attacked by fire in such a large area: "We reasoned that even when there is a major fire, a huge amount of rural Victoria does not get burnt." (p.4). This probability ends up being low. Since their aversion to risk is also low the probability associated to the belief of being caught in a fire event is neglected.

This situation can be modelled with a single agent, as follows (beliefs in italic can be omitted if a reasoning engine allows agents to build their own beliefs this way):

- **Beliefs:** B&B is relaxing (80), B&B is prone to fires (45)*, B&B will experience a fire event during the trip (0)*
- **Desires:** Be safe, Be relaxed
- **Intentions:** Go to a safe and relaxing place.
- **Risk aversion:** Low

Semmelweis Reflex / Belief Perseverance

The Semmelweis Reflex, also called Belief Perseverance, is a cognitive bias in which people cling to their beliefs even when they face proofs they are not true (Anderson, 1983). The Semmelweis Reflex is formalized in algorithm 2, based on an explanation given in (Savion, 2009); stating that a certainty is harder to update than a belief one thinks is only probable. One of the ways to "overcome" this bias has been suggested (p. 85 of (Savion, 2009)). This has been modelled in order for the agents to eventually update their certainty. The way to overcome the bias is based on the repetition of information (*e.g.* if an agent is certain of φ it needs to get the information $\neg\varphi$ a certain number of times before it will be able to update its former certainty). This is consistent with the Illusory Truth formalization.

Algorithm 2 Pseudo-code for the Semmelweis Reflex

```
1: function testForSemmelweisReflex(info)
2:     if getBeliefProbability (info) = 0 and acquiredInfoOccurrences(info) is not enough then
3:         return True
4:     else   # getBeliefProbability (info) > 0 or acquiredInfoOccurrences(info) is enough
5:         return False
6:     end if
7: end function
```

John Benett

John lives alone, he survived and, with other people, he successfully defended his house. During the whole afternoon, John felt a North wind blowing. At 3pm he received a call from a friend, warning him that a fire was coming from the West to his property. John ignored this warning. Later, he saw people driving dangerously near his property. He did not pick up on these cues either.

- **The dangerous situation** occurs at 5pm when a fire caught John by surprise. It came from the West while a North wind was blowing.
- **The decision leading up to the situation** is to ignore cues going against his belief that no fire could come from the West while the wind was blowing from the North.
- **The Semmelweis Reflex** underlying this decision prevents him from considering cues contrary to his belief that the fire will come from the North, as he states: "I thought that the North wind would prevent the fire from reaching Kinglake West and that it would direct it somewhere else." (p. 7)

This can be modelled in one agent as follows:

- **Beliefs:** House is safe (100), Fire goes in the same direction as the wind (100)*, Fire will come from [current wind direction] (100)*
- **Desires:** Be safe, Defend the house
- **Intentions:** Monitor *[current wind direction]* side of the land to gather environmental cues about an approaching fire, then proceed to defend the house.

Jim Baruta

Jim is married but his wife was not involved in the events. He survived but lost his house.

- **The dangerous situation** occurs at 5pm when when he drove on a road during a fire event (p.15).
- **The decision leading up to the situation** is to to keep driving even though he saw smoke and fire spots on the road.
- **The Semmelweis Reflex** occurs after at least one other cognitive biases was triggered (*e.g.* Optimism Bias) to build a certainty about the fact that he would be safe travelling home even though he knows he should not be doing it: "[in case of fire] you can't get on the roads, that is the last thing you should do" (p. 10). The Semmelweis Reflex prevents him considering cues that are contrary to his belief that he will be safe on the road.

This can be modelled in one agent as follows:

- **Beliefs:** The road home is safe (100), Do not be on the road during a fire event (80)
- **Desires:** Be safe, Defend the house
- **Intentions:** Go home then proceed to prepare and defend the house.
- **Risk aversion:** Extremely low

Illusory Truth

The Illusory Truth bias was shown for the first time in 1977 (Hasher et al., 1977) and has proven to be quite robust. Its effect is straightforward: the more one hears some information, the more one is inclined believe that the information is true. The Illusory Truth effect also ignores prior knowledge (Fazio et al., 2015), which makes it easier to formalize. It is formalized in algorithm 3, on lines 8-12. It reads: if the current agent already possesses a belief relative to the last information it gathered, then reinforce the belief according to the credit it puts in this information, multiplied by the number of times it gathered it. This has been implemented regarding *processing fluency* (Fazio et al., 2015). According to the author the Illusory Truth would come from the fact that repetition makes information easier to understand. Processing fluency states that the easier something is to understand the more truthful it appears. In our algorithm, on line 10, the probability associated with an information perceived is linked to the number of times this same information has already been perceived.

Algorithm 3 Pseudo-code for the Illusory Truth effect

```
1:    # φ is an information given by the environment to an agent
2:  info ← perceive(φ)
3:  perceivedProbability ← perceiveProbability (φ)
4:  incrementNumberOfOccurrence(info)
5:
6:  if not hasBelief(info) then
7:      acquireBelief(info, perceivedProbability )
8:  else    # hasBelief(info)
9:      if not testForSemmelweisReflex(info) then
10:         illusoryProbability ← perceivedProbability * acquiredInfoOccurences(info)
11:         updateBeliefProbability (info, illusoryProbability)
12:     end if
13: end if
```

Peter Ross Brown

Peter Brown is a married man with three children (aged 20, 18 and 16). He lives with his wife and children in their own house with a pool. The pool is intended to serve as a water tank in case of a fire event. During Black Saturday Peter's oldest child was away. Peter also listened to ABC radio during the entire day. The whole family survived the events but lost the house.

- **The dangerous situation** occurs when the family members are gathered in and around the pool and the fire unexpectedly strikes.
- **The decision leading up to the situation** is for the family not to evacuate and not to prepare the house before the fire arrived. Peter was trying to connect the fire pump to the pipes when it did.
- **The Illusory Truth** makes Peter believe in the accuracy of ABC announcements, since the ABC announcer repeated that they provided the most up-to-date source of information concerning the bushfire (p. 16). Next Peter reasons that since there was no warning from ABC for his area it means that his house and family are not threatened.

This situation can be modelled with a single agent as follows:

- **Beliefs:** ABC Radio provides up-to-date information (100)*, No fire is detected in my area (100)*
- **Desires:** Be safe, Defend the house, Do not prepare the house if not needed
- **Intentions:** Monitor for a threat by listening to the best radio channel found and looking at the environment, proceed to prepare the house if a fire is detected.
- **Risk aversion:** Medium

Michael Halls

Michael is the father of Natasha Davey. His daughter died during the Black Saturday along with her family. The family was sheltering inside their house, which had been well prepared and was defended by Natasha's husband.

- **The dangerous situation** occurs when the family is in the house while a fire is coming.
- **The decision leading up to the situation** is to stay and defend the house.
- **The Illusory Truth** makes Natasha believe that a well-prepared and defended house is a safe place in case of a fire event. Michael Halls testifies that his daughter's family has been reading brochures, comparing what they were going to do in the face of a fire to what their neighbours were going to do (which was more), and went to CFA meetings (p.7 - 13, 29). Furthermore, the CFA brochures they consulted implied that if one is well prepared, one has nothing to fear from a fire (p. 38, "Why does the bushfire guide show pictures of children protecting a property?").

This situation can be modelled with a single agent as follows:

- **Beliefs:** The house is well prepared (80), I can defend the house (80), My husband can defend the house (80), A well-prepared and defended house is safe (100), The house is comfortable (80)
- **Desires:** Be safe, Have the children in a safe area, Have the children being comfortable
- **Intentions:** Monitor fire threat from inside the house. In case of a fire event, shelter in the house then defend it.
- **Risk aversion:** Medium-High

CONCLUSION AND FUTURE WORK

Psychology and cognitive science related works show that cognitive biases affect human behaviours in crisis situations, where uncertainty or stress factors are present. Examples of individuals putting themselves, their friends or their family at risk by acting in a way that was not anticipated by decision-makers have been extracted from a concrete set of data. These behaviours were then connected to cognitive biases. Three specific cognitive biases have been formalized into algorithms, which are consistent with their related bias description in the literature. We argue that implementing these algorithms in BDI agents can make them more realistic by addressing some of BDI core issues (as identified by (Norling, 2004)) in terms of 1) decision making, 2) inaccuracies and 3) situation awareness of agents.

This work though should be considered preliminary. We provided algorithms to ease implementation of the biases, which constitutes a good basis for a computational model, but we have not implemented and tested them yet. These algorithms, although not yet implemented constitute a good basis for a computational model and ease implementation.

The methodology used to find occurrences of cognitive biases in our data set should also be treated as a work in progress. Since it is not based on any scientifically-grounded model of cognitive biases (we could not find any both exhaustive and reliable), it relies on disparate scientific evidence. Results are therefore based on a subjective interpretation of the testimonies.

Scarce examples of models integrating cognitive biases can be found. This work can be compared to two of them. The first is in the field of national defense (Kulik and Davis, 2002), while the other is in the field of health (Voinson et al., 2015).

The national defense model aims to predict a target's reaction to "Effect-Based Operation" (EBO) (*e.g.* the 9/11 events in the U.S.). It is based on a "Synthetic Cognitive Modelling" (SCM) approach, which takes several factors into account and gives probabilistic results. While their work aims to predict how an EBO would affect a target, we are trying to understand why a population reacts the way it does in a crisis situation. This explains the use of different paradigms. We make use of a white box (*i.e.* something we can see into) paradigm (BDI), in order to explain behaviours. Eventually we would like to build a serious game to experiment with the impact of different factors on the population. This work in national defense uses a black box (SCM) paradigm, looking only at inputs

(data), and outputs (behaviours). Their approach of cognitive biases also differs from ours, since they do not offer any framework to integrate identified cognitive biases or mind mechanisms in their model.

The model in the health field is based on mathematical functions and thus uses a white box approach. It aims to help understanding why individuals in developed and developing regions can be reluctant to vaccination, even though "vaccination has greatly reduced the burden of infectious diseases worldwide". The authors advocate that models should not assume individuals to be rational, which bring our work close to theirs. This irrationality is modelled using using a standard model made for capturing the disease transmission process, augmented by a belief model and two functions representing two distinct cognitive biases. However, the belief model is far simpler than the one implemented in BDI and only takes into account two parameters for each individual (epidemiological status and opinion). Our work is closer to this model than to the previous one, even though this model is mathematically based. However this work uses the KISS principle, in which things are as simple as possible, while ours is more descriptive.

Two areas for future work are planned. First, more work is needed on the conceptual side of our cognitive biases model in order to refine how it integrates with the underlying formal semantics of BDI agents. Second, we want to implement and test the biases that we have presented in order to validate our hypothesis. The implementation may then be enhanced in several ways. The Belief Perseverance Effect implementation will be complemented by the addition of other ways to overcome the fallacy as described in (Savion, 2009), along with a way to slow down the belief-updating process (*i.e.* Conservatism bias). The Illusory Truth Effect will be widened by the implementation of the processing fluency mechanism. In a more general way, this work would benefit from the formalisation of trust, which can affect perceived probabilities. The affective heuristic (Slovic et al., 2002) will also be integrated, along with work on an "affective weight" agents give to their environment. This will help us understand how agents integrate and contextualize objects in the world.

ACKNOWLEDGEMENTS

This research was funded by the University Grenoble-Alpes through AGIR project SWIFT (2016-2019), and by Grenoble Pôle Cognition (2017).

REFERENCES

Adam, Carole and Gaudou, Benoit (2016). “BDI agents in social simulations: a survey”. In: *The Knowledge Engineering Review* 31 (3), pp. 207–238.

Adam, Carole and Lorini, Emiliano (2014). “A BDI emotional reasoning engine for an artificial companion”. In: *A-Health workshop at PAAMS*.

Adam, Carole, Beck, Elise, and Dugdale, Julie (2015). “Modelling the tactical behaviour of the Australian population in a bushfire”. In: ISCRAM-Med, Tunisia.

Adam, Carole, Danet, Geoffroy, Julie, Dugdale, and Thangarajah, John (2016). “BDI modelling and simulation of human behaviours in bushfires”. In: ISCRAM-Med, Madrid.

Anderson, Craig A. (1983). “Abstract and Concrete Data in the Perseverance of Social Theories: When Weak Data Lead to Unshakable Beliefs”. In: *Journal of experimental social psychology* 19, pp. 93–108.

Axelrod, Robert (1997). *The complexity of cooperation: agent-based models of conflict and cooperation*. Princeton University Press.

Benson, Buster (2016). *Cognitive bias cheat sheet*. https://betterhumans.coach.me/cognitive-bias-cheat-sheet-55a472476b18.

Cardon, Alain (1998). In: *Information Modelling and Knowledge Basis IX*. Ed. by P.-J. Charrel, H. Jaakkola, H. Kandassalo, and E. Kawaguchi. IOS Press. Chap. A multi-agent model for cooperative communications in crisis management systems: The act of communication.

Das, T. K. and Bing-Sheng, Teng (1999). “Cognitive Biases and Strategic Decision Processes: An Integrative Perspective”. In: *Journal of Management Studies* 36.

Dignum, Frank P M, Prada, Rui, and Hofstede, Gert Jan (2014). “From Autistic to Social Agents”. In: *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), Paris, France, May 5-9, 2014* Aamas, pp. 1161–1164.

Dugdale, Julie (2010). “Human behaviour modelling in complex socio-technical systems”. Habilitation a diriger des recherches (HDR). Joseph Fourier University.

Dugdale, Julie, Bellamine-Ben Saoud, Narjes, Pavard, Bernard, and Pallamin, Nico (2010). "Simulation and Emergency Management". In: *Information Systems for Emergency Management*. Advances in Management Information Systems.

Edmonds, Bruce and Moss, Scott (2004). "From KISS to KIDS - an 'anti-simplistic' modelling approach". In: *Multi-Agent Based Simulations Conference*, pp. 130–144.

Fazio, Lisa, Brashier, Nadia, Payne, Keith, and Marsh, Elizabeth (2015). "Knowledge Does Not Protect Against Illusory Truth". In: *Journal of Experimental Psychology: General* 144, pp. 993–1002.

Final report (2009). 2009 Victorian Bushfires Royal Commission.

Final report - vol. IV : the statements of lay witness (2009). 2009 Victorian Bushfires Royal Commission.

Grignard., A., Taillandier, P., Gaudou, B., Huynh, N.Q., Vo, D.-A., and Drogoul, A. (2013). "GAMA v. 1.6: Advancing the art of complex agent-based modeling and simulation". In: *PRIMA*.

Hasher, Lynn, Goldstein, David, and Toppino, Thomas (1977). "Frequency and the Conference of Referential Validity". In: *Journal of verbal learning and verbal behavior* 16, pp. 107–112.

Kahneman, Daniel and Frederick, Shane (2002). "Representativeness Revisited: Attribute Substitution in Intuitive Judgment". In: *Heuristics and biases: the psychology of intuitive judgement*, pp. 49–81.

Kahneman, Daniel and Tversky, Amos (1979). "Prospect theory: an analysis of decision under risk". In: *Econometrica* 47, 263—292.

Kulik, Jonathan and Davis, Paul (2002). "Modeling Adversaries and Related Cognitive Biases". In: *SPIE's AeroSense Conference*.

Norling, Emma (2004). "Folk Psychology for Human Modelling: Extending the BDI Paradigm". In: *Third International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

Norling, Emma and Sonenberg, Liz (2004). "Creating Interactive Characters with BDI Agents". In: *Australian Workshop on Interactive Entertainment*.

Pan, Xiaoshan, Han, Charles, Dauber, Ken, and Law, Kincho (2007). "A Multi-agent Based Framework for the Simulation of Human and Social Behaviors during Emergency Evacuations". In: *AI and society* 22, pp. 113–132.

Pohl, Rudiger (2004). In: *Cognitive illusions : a handbook on fallacies and biases thinking, judgement and memory*. Ed. by Rudiger Pohl. Psychology press. Chap. Introduction.

Research, Strahan (2010). *Behaviour and intentions of households on code red days*. Country Fire Authority.

Reyna, Valerie and Brainerd, Charles (2008). "Numeracy, ratio bias, and denominator neglect in judgments of risk and probability". In: *Learning and Individual Differences* 18, 89—107.

Savion, Leah (2009). "Clinging to discredited beliefs: the larger cognitive story". In: *Journal of the Scholarship of Teaching and Learning* 9, pp. 81–92.

Shaw, Marwin, Robbin, Rhona, and Belser, James R. (1976). *Group Dynamics: The Psychology of Small Group Behavior*. New York : McGraw-Hill.

Slovic, Paul, Finucane, Melissa, Peters, Ellen, and Macgregor, Donald G (2002). "The Affect Heuristic". In: pp. 397–420.

Sunstein, Cass and Zeckhauser, Richard (2010). "Dreadful possibilities, neglected probabilities". In: *The Irrational Economist: Making Decisions in a Dangerous World*, pp. 116–123.

Taillandier, Patrick, Bourgais, Mathieu, Caillou, Philippe, Adam, Carole, and Gaudou, Benoit (2016). "A situated BDI agent architecture for the GAMA modelling and simulation platform". In: *MABS at AAMAS*.

Thornton, R. P. (2010). "Short Communication on Research Response to the Black Saturday (7th February 2009) Victorian Bushfires, Australia". In: *Fire Technology*.

Tversky, Amos and Kahneman, Daniel (1974). "Judgment under Uncertainty: Heuristics and Biases". In: *Science* 185, pp. 1124–1131.

Tversky, Amos and Kahneman, Daniel (1983). "Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment". In: *Psychological Review* 90.

Victorian 2009 Bushfire Research Response Final Report (2009). Bushfire CRC.

Voinson, Marina, Billiard, Sylvain, and Alvergne, Alexandra (2015). "Beyond Rational Decision-Making: Modelling the Influence of Cognitive Biases on the Dynamics of Vaccination Coverage". In: *PLOS ONE* 10.