Cognitive biases impact on social simulations (in reactions to bushfires)

Pierre Blarre
June 2018, Grenoble-Alps University, France

Abstract

The key intent of this work is to produce a crisis simulation model that implements the Belief-Desire-Intention (BDI) paradigm and cognitive biases, in order to test their impact on social simulations.

The original contribution of this paper is two-fold:

(1) Present the different concepts needed to build an agent-based BDI model with cognitive biases, that will faithfully mimic human behavior in reaction to crisis situations. We will define what is agent-based modeling, the BDI paradigm, what are cognitive biases and use testimonials of real bushfires to define the requirements of our model.

(2) Present an Agent Based Modeling and Simulation tool (GAMA). Define a model structure that could be reused for future crisis simulation. Provide a working prototype that meets our previous requirements. Demonstrate the effects of BDI implementation and cognitive biases in human behavior simulation. Show how they not only improve the realism of the simulation, but also make it possible to the create tangible new scenarios.

Keywords

Social Simulation, Crisis Situation, Bushfires, Cognitive Biases, Belief-Desire-Intention (BDI), Agent Based Modeling and Simulation (ABMS), GAMA Platform

Introduction

Despite recurring occurrences of bushfires and prevention campaigns, the population's behavior during bushfires is difficult to foresee. People do not behave according to what would objectively be in their best interests (Arnaud, Adam, & Dugdale, 2017).

These systematic patterns of deviation from rationality in judgment are referred to as cognitive biases (Martie G. Haselton, Daniel Nettle, & Paul W. Andrews, n.d.).

The notion of cognitive biases was introduced by Amos Tversky and Daniel Kahneman in 1972, who later published a book (Kahneman Daniel, 2011) that became a reference in psychology of human behaviors in the decision making processes. The book describes our minds as machines for jumping to biased conclusions.

We argue that cognitive biases are an appropriate mechanism for describing people's behavior in crisis situations like bushfires, where uncertainty and stress factors are important.

Multi-agent system (MAS) and Agent-Based Modeling and Simulation (ABMS, a sub-branch of MAS) are techniques from Artificial Intelligence that provides us with tools to model human behaviors and to test the effect of several parameters (Arnaud et al., 2017, p. 13).

The proposed approach is to apply ABMS to the field of crisis management, specifically to bushfires. The goal is to improve evacuation plans, prevention campaigns and to better understand what went wrong during past events where testimonials and reports are available.

The Belief-Desire-Intention (BDI) model is one of the software model developed for programming intelligent agents. It implements the principal aspects of Michael Bratman's theory of human practical reasoning (also referred to as Belief-Desire-Intention, or BDI). The BDI model has been proved as a dominant view in contemporary philosophy of human mind and action (Jo & Einhorn, 2005).

In essence, it provides a mechanism for separating the activity of selecting a plan (from a plan library or an external planner application) from the execution of currently active plans. Consequently, BDI agents are able to balance the time spent on deliberating about plans (choosing what to do) and executing those plans (doing it) ("Belief–desire–intention software model," 2018).

We argue that the BDI architecture is appropriate to model agents in crisis situations. Agents have to decide which plans of action they will execute (ex: Go home. Evacuate. Fight fire.) regarding to their beliefs (ex: I am in danger. I can stop this fire.) and desires (ex: I want to stay alive. I want to defend my house. I want to save my family).

The bushfires in Victoria, Australia, on February 7, 2009 (also known as Black Saturday) caused 173 deaths, 414 injuries and severe material losses. Although bushfires are common in Victoria and most victims had lived in high fire-risk areas, people were badly prepared and did not behave rationally during these events.

After assessing interviews from people who survived the Black Saturday, three main cognitive biases were identified in people's behavior: the Neglect of Probability, the Semmelweis Reflex and the Illusory of Truth effect (Arnaud et al., 2017). BDI pseudo code algorithms were also proposed for these cognitive biases implementation in (Arnaud et al., 2017).

The aim of this work is to test the effects of cognitive biases in social simulations. The model will be implemented with the GAMA platform using the BDI architecture.

State of the art

This work is mostly a follow up on The role of cognitive biases in reaction to bushfires by (Arnaud et al., 2017). It should be consulted for detailed explanations on the retained cognitive biases and proposed algorithms.

The overall goal is to build a BDI model of human behavior in crisis, in order to improve social simulations by better mimicking human behavior (Arnaud et al., 2017).

A BDI approach on integrating cognitive biases in models was presented, but no model was ever implemented. We will address this issue by presenting a prototype that follows the proposed approach on implementing cognitive biases within a BDI model.

1 GAMA Platform and BDI

GAMA is a modeling and simulation development environment for building spatially explicit agent-based simulations (*GAMA Platform*, 2018). It allows to create and configure multi-agent environment.

In 2016, (Taillandier, Bourgais, Caillou, Adam, & Gaudou, 2016) proposed a BDI plugin (also called BDI architecture) for GAMA called Simple BDI.

By creating sets of beliefs, desires and plans, the agents will see their intentions changing and in return execute different plans, thus simulating the BDI logic. The agents have the capacity to perceive other agents with conditions (ex: perceive a fire if it's within a given distance)

An explanation on how to implement this architecture in a GAMA model, and incidentally the concepts used to develop our model can be found in the GAMA documentation. (Julien Mazars, 2016)

2 The retained cognitive biases

Three cognitive biases were identified from the interviews of the Black Saturday's victims. We will implement and test them in our model. The following descriptions of these cognitive biases are shortened as it has been extensively explained in (Arnaud et al., 2017)

2.1 The neglect of probability

It is the tendency to disregard probability when making a decision under uncertainty and is one simple way in which people regularly violate the normative rules for decision making. Small risks are typically either neglected entirely or hugely overrated. The continuum between the extremes is ignored. The term probability neglect was coined by Cass Sunstein (Kahneman Daniel, 2011).

2.2 The Semmelweis reflex or Semmelweis effect

It is a metaphor for the reflex-like tendency to reject new evidence or new knowledge because it contradicts established norms, beliefs or paradigms (Mortell, Balkhy, Tannous, & Jong, 2013).

2.3 The illusory truth effect

Also known as the validity effect, truth effect or the reiteration effect. It is the tendency to believe information to be correct after repeated exposure (Chris, 2007).

3 Seven behavior profiles

After extracting data from the Black Saturday victims' interviews, seven behaviors were identified. These behaviors will be implemented in our model and various level of cognitive biases influence will be experimented for each behavior.

The seven archetypes of population behavior identified by (Commissioner F.S., 2013; Rhodes A., 2014) are the following:

- Can-do defenders: are determined to protect their house, have good knowledge of the area, previous experience and skills, are action-oriented, self-sufficient, and confident;
- Considered defenders: are strongly committed to stay and defend their house, are aware of the risks and made deliberate efforts to prepare and train;
- Livelihood defenders: are committed to stay and defend what they consider as their livelihood (farm, hotel, etc.), are therefore well-prepared year-round;
- Threat monitors: do not intend to stay in front of a serious threat, but do not intend to leave until it feels necessary, will wait and see;
- Threat avoiders: are conscious of the risk, feel vulnerable, and plan to leave early before any real threat, but therefore have no plan if caught by surprise;
- Unaware reactors: are unaware of the risk, feel they are not concerned by fires, have no knowledge of how to react (e.g. tourists), therefore totally unprepared;
- Isolated and vulnerable: are physically or socially isolated (e.g. elderly people) which limits their ability to respond safely.

Each archetype therefore has a different behavior, with its own underlying motivation, level of risk-awareness and knowledge, response strategy, and associated information needs. (Adam, Beck, & Dugdale, 2015)

We argue that each archetype must have different belief, desires and intentions, at least in the strength value of their beliefs.

The Model

A previous bushfires GAMA model was designed as part of a university project but it was not implementing the BDI paradigm nor cognitive biases. Some parts of it have been used as the base for our new model, since it included some useful features like fire spreading, firefighting, communication actions and more. The model has been completely refactored and the code cleaned and commented as part of this work.

In this paper we cannot cover every details of the model. We are going to present the general structure and operations, then focus on the residents, the BDI and the cognitive biases attributes. You can find a more detailed documentation on the model as well as the commented code at the model's public repository. (Pierre Blarre, 2018)

1 Structure

Simple simulations do not necessary require files separation. A single model file can contain all species and environment attributes.

In this case we are aiming at building a complex model. It involves multiple agents, environment variables and experiments. A clear files separation is necessary for ease of understanding, team work, code sharing and future simulations.

The structure and naming convention was chosen to be as self-explanatory as possible.

We argue that this structure should be reused in future simulations to provide the basis of a common framework for social simulations :

- The assets folder contains the images and shape files used for visualizations. Any external asset should be placed in this folder.
- The doc folder is for the documentation.
- The experiments folder contains the different experimentations of the simulation. We propose to create a main experimentation within the main model file (described below). By doing this, all experiments within the experiments folder can inherit the user interface of the main experiment. This reduces code duplication. For our model, the experiments files' naming convention is using CB for Cognitive Biases and BP for behavior profiles in order to quickly identify the experiments variations from the filenames.
- The models folder contains the following files:
    - Main: contains all the global variables, the world initialization and the main experiment. We argue that the creation of a main file that imports all the others helps readability and cohesion, as it is common practice with other programming languages.
    - Environment: contains all environment related agents: Fires, Plots (The grid), Buildings, Roads, City Exits, Fire Watch Towers and Waterways.
    - People: contains the main attributes and actions of all other derived "human" agents.

   - Emergency Services: contains the firefighters and policemen species.
   - Residents contains the general physical and psychological attributes of the residents.
   - The behavior profiles folder contains the residents variations regarding the behavior profiles defined in the previous section.
- The results folder should contain all form of results : screenshots, csv files, etc. Its files can be used within the documentation.

2 General scenario

A typical bushfires crisis situation unfolds like this:

- Fires start somewhere, at some point they are detected.
- Firefighters take action.
- People who see smoke or receive a warning and are aware of the risks can already start to trigger their fire plan.
- If the fires become uncontrollable (too many fires or not enough resources to control them), global alerts are sent to the people (through the radio, global phone messages, etc) to trigger a city evacuation.
- Policemen take action to alert residents and help them evacuate.
- Depending on their current situation and behavioral pattern, people are going to ignore the warnings, trigger their fire plan or try to evacuate.
- People who ignore the warnings, do not have a fire plan or take too long to trigger it risk to get hurt, lose their life or their material possessions. The same goes if they take too long to evacuate.

3 The agents (also called species)

The model is composed of:

- Nature, roads, buildings and plots agents, which compose the environment. Each plot has a burning capacity or resistance.
- The fire (which is just a starting point that spreads to neighboring plots).
- The people, which is the mother specie of:
   - The fire fighters.
   - The policemen (whom send global evacuation alerts and help residents to evacuate).
   - The residents, whom are the mother specie of the seven behavior archetypes we discussed in part 3 of the state of the art.
   - The people specie is the one containing the cognitive biases algorithms.

We will focus mainly on the residents. All residents are defined by physical attributes (energy, is alive, home location, in a shelter, etc.) and psychological attributes (motivation, risk awareness, knowledge, training, fear of fire).

These physical and psychological attributes will be distributed differently between the different behaviors profiles. The distribution for each profiles remain a personal appreciation but were chosen in trying to reflect the seven profiles. These could be refined in a future research.

The psychological attributes will be used to influence the agent's probability to react to a given situation and their decision process.

The cognitive biases will also be used to influence the agent's probability to react, and hopefully make the agents more realistic by mimicking irrational judgments and uncertainty.

4 The BDI

For the residents, we propose six beliefs and five desires. The intentions will be derived from the desires.

More beliefs and desires could be added of course, but we argue that this base is enough to cover most cases and to build a model complex enough.

4.1 Beliefs

- There is no danger, I feel safe.
- There is a potential danger, I have seen smoke or I have been alerted and I do not feel safe.
- There is an immediate danger, I can see the fire, I have been hurt or I am convinced by the alert messages that I am not safe.
- There is a risk of fire today (which is different from a potential danger, as it will only increase the risk awareness)
- I can defend my building. Used for defenders profiles.
- I can escape (as in, I believe I still have an escape route to avoid the fire).

4.2 Desires

- I want to go to work.
- I want to go home.
- I want to call emergency services.
- I want to defend my building (work or home depending of the profile).
- I want to escape.

4.3 Intentions and plans

The intentions and plans are directly derived from the desires.

The GAMA BDI architecture includes a reasoning engine that manages agents' intentions regarding their beliefs and desires and make them decide which plans should be executed.

4.4 Rules

As per explained in the GAMA BDI architecture documentation, a set of rules can be defined to manipulate desires regarding agents' beliefs. We define 4 rules in our model:

- The no danger belief will create the desire to go somewhere and will remove the desire to escape.
- The potential danger belief will create the desire to call for help and remove the work and escape desires.
- The immediate danger belief will create the desire to escape and remove the work and home desires.
- The can defend belief will create the desire to defend.

4.5 Perceptions

Two perceptions will be implemented for now:

- A distant fire detection (simulating seeing smoke or distant flames) that will add the belief of a potential danger to the agent (and incidentally the desire to call for help as defined in the previous rules)
- A close fire detection that will remove previous belief and add the immediate danger belief.

The perceptions and rules will actually differ for different behavior profiles, as detecting a fire for a "Can do defender" profile will for example trigger a defense fire plan, when a "Threat avoider" profile will decide to escape.

5 The cognitive biases

The three cognitive biases are implemented as three separated functions. They can be called within agents plans or actions. The model has a general option to turn them on or off.

When the cognitive biases option is activated, some agents will randomly receive a cognitive bias. Every time an agent is initialized, it has a probability of receiving one of the three cognitive biases:

```
if(flip(cognitive_biases_distribution)) {
    neglect_of_probability_cb_influence <- true;
    cognitive_biases_influence <- true;
}
```

The flip method returns true or false with the cognitive_biases_distribution given probability. If the method returns true, then one of the cognitive bias will be given to the resident (in this example only the neglect of probability is represented).

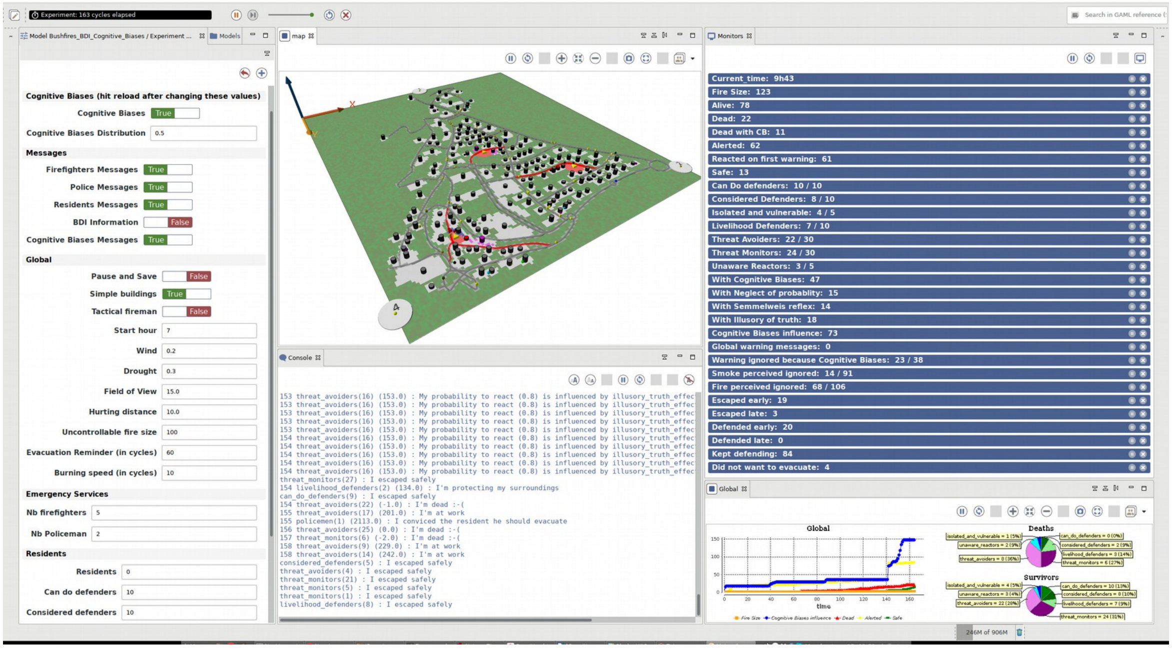
This method of distribution is somewhat hackish for there is no tangible statistics on the percentage of people affected by the different cognitive biases.

By default, we use a fifty percent probability for cognitive biases distribution. This value distributes cognitive biases to around fifty percent of the population. It is an interesting equilibrium to be able to compare agents with and without the effects of the cognitive biases within the same experimentation.

A resident must make a choice when he receives a warning message, perceives smoke or perceives flames. We added cognitive biases influences at this three breakpoints. Every time the resident's probability to react is involved, and if he has a cognitive bias, it goes through the cognitive biases algorithms.

6 Experiments and interface

The main experiment of the model includes behavioral profiles, cognitive biases, and many adjustable variables through the user interface. It also includes a real time 3d map and live graphs of the simulation.



The monitored values will allow us to extract the following data:

- The fire size.
- Information on residents : Total, Alive, Dead, Dead under cognitive bias influence, Alerted, Reacted on first warning, Safe.
- The number of residents alive and dead by behavioral profiles.
- The residents that are influenced by at least one cognitive bias and the number of residents influenced by each cognitive bias.
- The number of times a cognitive bias was triggered.
- The warning messages ignored because of cognitive bias influence.
- The ignored perceptions of smoke or flames.
- The number of times plans started early or late given the current agent's beliefs.
- The console displays agents messages concerning their beliefs, plans and when their decisions are influenced by cognitive biases.

The user interface also allows to turn on and off different messages for the console from firefighters, police, residents, BDI information and cognitive biases messages.

The significant and interesting part should be the monitoring of people who reacted, survived or escaped faster or slower regarding to the cognitive biases influences, for example:

- How many people escaped faster because they neglected the probability that the fire was far away from them and they weren't in danger? (positive influence).
- How many people who ignored the initial warnings survived because after hearing the alert message many time they ended up believing it? (positive influence).
- How many people died because even when they perceived a dangerous fire, they clung to their belief of being able to escape safely? (negative influence).

When a simulation is stopped or finished, all monitored data is saved in a CSV file which allows comparison between simulations.

Results

The presented results are meant to show that the cognitive biases algorithms have an impact on agents' decisions. They are not statistics meant to be reused, since the initial probabilities to react have been chosen through a personal appreciation. Moreover, since the model is complex and resource consuming, it would require a lot of processing power to run a significant amount of simulations.

The following values are average values extracted from the results of a dozen of simulations:

- 30% of warning messages are ignored for agents influenced by a cognitive bias. The Unaware Reactors and Isolated and Vulnerable profiles are the one mostly ignoring messages.
- 50% started an escape plan when the danger was only potential. Threat Avoiders and Threat Monitors are the profiles mostly escaping early.
- 90% of the Can Do, Considered and Livelihood Defenders triggered their defense plan early, when the danger was only potential.
- 20% did not want to evacuate with the police right away.
- 20% died because of biases' influence.

Reaction time differs between similar behavior profile agents with cognitive biases activated or not.

What is truly remarkable is that reaction times between agents having cognitive biases varied a lot. From one run to another, the number of cycles required to complete a simulation can be quadrupled.

The fact that the agents' decisions are not based on a simple probabilistic choice, but that their probability to react is influenced by cognitive biases algorithms make the agents' behavior more complex to foresee. The model is mimicking the differences between individuals and their random or irrational behaviors in the different steps of the simulation.

We observe that the agents' behaviors feel more natural than when the cognitive biases are not activated. This result meets our requirement to better mimic humans. The simulation brings us further towards a more realistic social simulation, specifically during a crisis situation.

Conclusion

It is always possible to create deterministic simulations that will output values similar to real events. By fine tuning parameters, we can always simulate a situation and results we would like to attain, with or without BDI or cognitive biases. This approach will allow us to observe what happened and experiment on specific events. But it will neither be a realistic representation nor a basis for future simulation.

By adding the BDI paradigm, a dominant view in contemporary philosophy of human mind and action, we bring a new way of reasoning to our agents. The process takes into account beliefs and desires and the time needed to take decisions. The BDI paradigm brings us closer into mimicking the human behavior.

To go further, we implemented the acts of making wrong decisions or taking too long to decide by using cognitive biases algorithms.

Decision processes are sped up, delayed or blocked. As a result, agents with similar physical and psychological attributes are reacting differently and within different time scales. Agents' behaviors feel more natural thanks to randomness and irrationality simulation.

We built a prototype based on a real set of events, the Black Saturday, but it can also be used to simulate future crisis. This model can be reused with different cities or terrains, but also different populations. Beliefs like "I am in danger" and desires like "I want to escape" are the same across cultures and countries.

We saw that even in high fire-risk areas like Victoria, where knowledge should be higher, people do not behave rationally. We can argue that the cognitive biases we implemented are shared across cultures and countries, whether the area is at risk of crisis situations or not.

The BDI paradigm and cognitive biases are important components that should be completed with adding emotions, norms, negotiations and more, in order to reach an even more faithful human behavior simulation.

In conclusion, we think this prototype is a good basis that should be refined, extended, formalized and standardized into a reusable framework for crisis situation simulations.

Possible future works

This work left out some topics and created new ones, which could be done in future works. Some possible ideas are :

- Refine the agents probabilities to react and how they are distributed among the different profiles. At the moment, there is only one probability to react per agent, which might not be enough.
- Refine the conditions used that add new beliefs and desires.
- Model optimization: use terminal experiments, as the visualization and graph generation are resource consuming and are slowing down the model. Optimize the fire spreading. Identify other resource consuming components and try to optimize them.
- Modifying the simple_bdi GAMA plugin to bring cognitive biases features into its core, such as:
    - A cognitive biases base, similar to the beliefs or desires bases. Each biases could have a strength value associated to change the degree of influence.
    - The cognitive biases as core functions of the plugin.
    - The possibility to easily add new cognitive biases.
    - Add psychological features to simulate stubbornness, ease to convince, degree of influence, etc.
- Study and define precise range of values for physical and psychological attributes regarding to the seven behavior profiles.
- Study and define cognitive biases distribution regarding to the seven behavior profiles.
- Identify and implement more cognitive biases.
- Bring emotions, norms, negotiations into the model.

References

Adam, C., Beck, E., & Dugdale, J. (2015). Modelling the Tactical Behaviour of the Australian Population in a Bushfire. In N. Bellamine Ben Saoud, C. Adam, & C. Hanachi (Eds.), *Information Systems for Crisis Response and Management in Mediterranean Countries* (Vol. 233, pp. 53–64). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-24399-3_5

Arnaud, M., Adam, C., & Dugdale, J. (2017). The role of cognitive biases in reactions to bushfires, 13.

Belief–desire–intention software model. (2018). Retrieved from https://en.wikipedia.org/wiki/Belief–desire–intention_software_model

Chris. (2007, September 18). The Truth Effect and Other Processing Fluency Miracles. Retrieved from http://scienceblogs.com/mixingmemory/2007/09/18/the-truth-effect-and-other-pro/

Commissioner F.S. (2013). *Review of the community response in recent bushfires.* Retrieved from http://goo.gl/wJcGn3

GAMA Platform. (2018). Retrieved from http://gama-platform.org/

Jo, C.-H., & Einhorn, J. M. (2005). A BDI Agent-Based Software Process, *4*(9), 21.

Julien Mazars. (2016, May 21). GAMA Platform - Using BDI. Retrieved from https://github.com/gama-platform/gama/wiki/UsingBDI

Kahneman Daniel. (2011). *Think, fast and slow.*

Martie G. Haselton, Daniel Nettle, & Paul W. Andrews. (n.d.). The evolution of cognitive bias (pp. 724–746).

Mortell, M., Balkhy, H. H., Tannous, E. B., & Jong, M. T. (2013). Physician 'defiance' towards hand hygiene compliance: Is there a theory–practice–ethics gap? *Journal of the Saudi Heart Association*, *25*(3), 203–208. https://doi.org/10.1016/j.jsha.2013.04.003

Pierre Blarre. (2018). Cognitive Biases BDI - GAMA Model (Version 1) [GAMA Platform]. Grenoble-Alps University, France. Retrieved from http://github.com/ierpe/Cognitive_Biases_BDI

Rhodes A. (2014). *Why don't they do what we think they should? Understanding People's Response to Natural Hazards.* (Emergency Management Victoria - AFAC.). Retrieved from http: //goo.gl/D0R8YC

Taillandier, P., Bourgais, M., Caillou, P., Adam, C., & Gaudou, B. (2016). A situated BDI agent architecture for the GAMA modelling and simulation platform, 14.