A Cognitive Model of Epistemic Vigilance in Situations of Varying Competence, Consistency, and Utility

Chapter · September 2023

DOI: 10.1007/978-3-031-34920-1_2

4 authors:

Piotr Paweł Laskowski
University of Bialystok

Ivan Puga-Gonzalez
NORCE

F. LeRon Shults
Universitetet i Agder

Konrad Talmont-Kaminski
University of Bialystok

A Cognitive Model of Epistemic Vigilance in Situations of Varying Competence, Consistency, and Utility

Piotr Paweł Laskowski, Ivan Puga-Gonzalez, F. LeRon Shults, and Konrad Talmont-Kaminski

Abstract This paper outlines a computational, cognitive model representing how humans may use epistemic vigilance to evaluate socially-provided information in a way that reacts flexibly to differences in the reliability of content versus source vigilance strategies. Furthermore, the model explores how the system reacts in situations where the utility of the information provided is either unrelated to its accuracy or, even, is inversely proportional to it. We find that even a simple model is able to react flexibly to variation in these parameters, providing a basis for further exploration of the phenomenon.

Keywords Computational cognitive model · Cognition · Epistemic vigilance · Human cooperation · Bounded rationality

1 Introduction

According to epistemic vigilance theory [1], humans use two main methods in the attempt to ensure that the information they learn from others is trustworthy. The first of these consists in vigilance to the content of the information, i.e. the evaluation of the overall plausibility of the information, which can depend upon such things as its coherence with existing beliefs as well as its internal consistency. At the very least, this requires making a quick judgment as to the prima facie plausibility of the content, which may well be subject to any of a number of cognitive biases. The second concerns vigilance to the source of the information, i.e. whether the person providing the information is themselves plausible, which will depend upon

P. P. Laskowski (✉) · K. Talmont-Kaminski
Institute of Sociology, University of Bialystok, Białystok, Poland
e-mail: pio.laskowski@gmail.com

I. Puga-Gonzalez · F. L. Shults
Center for Modeling Social Systems, Norwegian Research Center, Kristiansand, Norway

F. Squazzoni (ed.), *Advances in Social Simulation*, Springer Proceedings in Complexity,
https://doi.org/10.1007/978-3-031-34920-1_2

considerations of their knowledge ability in the area, as well as their attitude or potential conflicts of interest. At a minimum, source vigilance relies upon the social status of the source, thereby making the assumption that social status is related to ability to provide accurate information.

Each of the methods has its own strengths and limitations and is therefore appropriate for use in particular contexts. It is highly significant, therefore, to what degree people are thought to be able to use these two forms of vigilance flexibly, applying them as needed in the situation they find themselves in. In particular, source vigilance is more appropriate where a person has very little ability to judge the content (low competence) and content vigilance is the preferable approach to take where there is little relationship between the apparent plausibility of the source and the quality of the information they offer (low consistency). Listening to one's doctor may be considered an example of the first kind of situation, while choosing one's life insurance on the basis of the actor who advertises it is an example of the latter. What makes it worse is that it is not always clear cut what kind of an environment one is in, so that it becomes necessary to judge on the basis of the outcomes of one's previous decisions.

The issue is further complicated by the question of the relationship between the accuracy of the information one is presented with and its utility. In general, it is important to act upon accurate information—it is no use arriving at the airport to catch the 11 am flight if it actually was scheduled to leave at 9 am. But the relationship between utility and accuracy of information need not be so straight-forward. In extreme cases, it may be that inaccurate information has a higher utility than accurate information. For example, it may be that false beliefs about potential punishment can motivate a group to cooperate to their mutual benefit [2].

We decided to model such a situation in order to determine how a cognitive model reacts to, on the one hand, variation in the relationship between accuracy and utility and, on the other hand, variation in its competence to judge the content of the information as opposed to the consistency between the plausibility of the sources and the quality of the information they offer [3]. The model represents the cognitive process of an agent that is being successively presented with pieces of information. Based on this cognitive process, the agent may provisionally accept or reject the information on the basis of source or content vigilance. If it provisionally accepts the information, it acts upon it, thereby generating utility that may or may not satisfice it [4]. If it satisfices, it accepts the piece of information and strengthens its preference for the form of vigilance it used to evaluate the information. If it does not satisfice, it rejects the piece of information, retaining its prior belief, and weakens its preference for the form of vigilance it used to evaluate it. Having built this model, we then examine how the agent deals with three environments, as relating to the relationship between accuracy and utility of beliefs: (1) where utility is equal to accuracy, (2) where utility is random and unrelated to accuracy, and (3) where utility has an inverse relationship to accuracy.

2 Methods

2.1 Model Description

The model represents an agent presented with a series of beliefs which they may choose to act upon, where each belief has a degree of accuracy (a value from a uniform distribution [0,1]). The decision whether to act upon the belief presented is made on the basis of either source or content vigilance.

When the belief is judged on the basis of content vigilance, the agent applies its competence (a value between 0 and 1) in order to arrive at an estimate of the accuracy of the content of the belief according to equation 1.

$$\text{EA} = \text{N}(0, 1, \mu = \text{Accuracy}, \sigma = 1 - \text{Competence}) \tag{1}$$

where EA represents the estimated accuracy, a value drawn from a truncated normal distribution at [0,1] with mean equal to the accuracy of the belief and standard deviation equal to one minus competence. Thus, the higher the competence of the individual, the more likely that the EA represents a value close to the real accuracy.

When the belief is judged on the basis of source vigilance, the agent relies upon the consistency, in their environment, of the relationship between source quality and information quality (also a value between 0 and 1). When consistency is high, the accuracy of the belief is closely connected to the plausibility of the source presenting it. Since the model does not include actual sources, EA is deemed to be equal to 'source plausibility' calculated 'backwards' from belief accuracy and the consistency of the environment, as per equation 2.

$$\text{EA} = \text{N}(0, 1, \mu = \text{Accuracy}, \ \sigma = 1 - \text{Consistency}) \tag{2}$$

where EA represents the estimated accuracy, a value drawn from a truncated normal distribution at [0,1] with mean equal to the accuracy of the belief and standard deviation equal to 1 minus the consistency of the environment. Thus, the higher the consistency of the environment the more likely that the EA represents a value close to the real accuracy.

At the beginning of the simulation the probabilities of the individual judging the belief on the basis of the content or source are both 0.5. These values may later change depending on which strategy (content or source) leads to higher utility beliefs.

After calculating EA on the basis of the content or the source, the individual tries the belief if the estimated accuracy is higher than the previous estimated accuracy minus an error; where error is kept constant at 0.1 and is necessary to allow the agent to try beliefs of similar but somewhat lower estimated accuracy. If the individual decides to try the belief, then the agent calculates the probability that the outcome satisfices it, according to equation 3:

$$Prob\_satisfaction = \frac{1}{1 + e^{-\lambda*(Utility-0.5)}} \quad (3)$$

This equation represents a sigmoidal curve, where lambda gives the shape of the curve and Utility − 0.5 gives the inflection point of the curve. Hence, if utility is equal to 1, the inflection point of the sigmoidal curve will be exactly at 0.5 on the x and y axes. We keep lambda fixed at 10.

We produce three different scenarios regarding the relationship between the value of utility and accuracy. In the first, utility is equal to accuracy, so that beliefs with high accuracy will also have a higher chance of satisficing if they are tried. In the second scenario, utility value is not tied to accuracy. In that case, utility is drawn from a uniform distribution [0,1] each time. Hence, no matter the level of accuracy of the belief, the satisficing level it produces is decided at random. In the third scenario, we reverse the relationship between accuracy and utility from that in the first scenario. Thus, beliefs with low accuracy will have higher utility and vice versa. This means that a low accuracy belief will have a high probability of satisficing.

Once the probability of satisficing is calculated, it is compared to a random number between [0,1], if it does satisfice (i.e. probability of satisficing is higher than the random number), then the agent accepts the belief and increases the probability of using the strategy (content or source) that produced this outcome. For instance, if the belief was judged on the basis of content, then the probability of using this strategy with future beliefs increases according to equation 4:

$$\text{Prob\_content}_{t+1} = \text{Prob\_content}_t + ((1 - \text{Prob\_content}_t) * \text{Learning}) \quad (4)$$

where learning is kept constant at 0.01.

Then, the probability of judging beliefs on the basis of the source is decreased by the same amount according to equation 5:

$$\text{Prob\_source}_{t+1} = 1 - \text{Prob\_content}_{t+1} \quad (5)$$

If the outcome does not satisfice, then the agent decreases the probability of using the strategy that produced this outcome and increases the probability of using the opposite strategy with future beliefs using equations 4 and 5. Note that both probabilities always add to 1.

If the individual decides not to try the belief, another belief is presented to the individual and the whole procedure is repeated. The agent keeps on trying beliefs for a specific number of time steps and the probability of using one or other strategy to judge beliefs may stabilize over time depending on the agents' competence and the environment's consistency.

2.2 Simulations and Data Collection

We ran simulations by varying the ability of the agent to judge the belief on the basis of its content or source. Hence, we assigned values to consistency and competence going from 0 to 1 by steps 0.01. By doing so, we ended up with a total of 10,100 combinations of values. All other parameters were kept constant. We ran 20 replications per combination of values and for each scenario representing the relationship between accuracy and utility. In total we ran 202,000 simulations per scenario. Simulations were stopped after the individual was presented with 1000, 3000 or 5000 beliefs, no matter whether the individual tried or not all the beliefs. For each combination of parameters and their 20 replications we calculated the percentage of replications where the agent tended to use the content strategy the most (i.e., probability of using content evaluation > 0.5) when evaluating the belief. The results presented here are the ones obtained after 1000 beliefs. Results for larger numbers of presented beliefs remain qualitatively the same as those of 1000 and are thus not presented.

3 Results

3.1 Scenario 1: Belief Utility Equals Belief Accuracy

Figure 1 presents the percentage of replications (n = 20) where content strategy was used by the agent, going from 0% (blue) to 100% (red) of the time. As expected, the agent used the content strategy the most when their competence was medium/high (i.e., > 0.5) and higher than the environment's consistency. On the contrary, when the environment's consistency was higher than the agent's competence and consistency was medium/high (i.e., > 0.5,) the agent relied more on the content strategy. When consistency and competence had a similar value (diagonal) or both of them had a medium/low value (i.e., < 0.5), the decision to use either of the strategies appears to be 50–50%.

3.2 Scenario 2: Belief Utility Equal to Uniform Distribution

The Fig. 2 presents the percentage of replications where content strategy was used by the agent, going from 0% (blue) to 100% (red) of the time. No clear pattern is observed, the prevalence of any of the strategies appears random for all combinations of consistency and competence values. Consistency and competence have a similar approximation value of 50% in most cases. We do not observe here any of the patterns seen in Figs. 1 and 3.

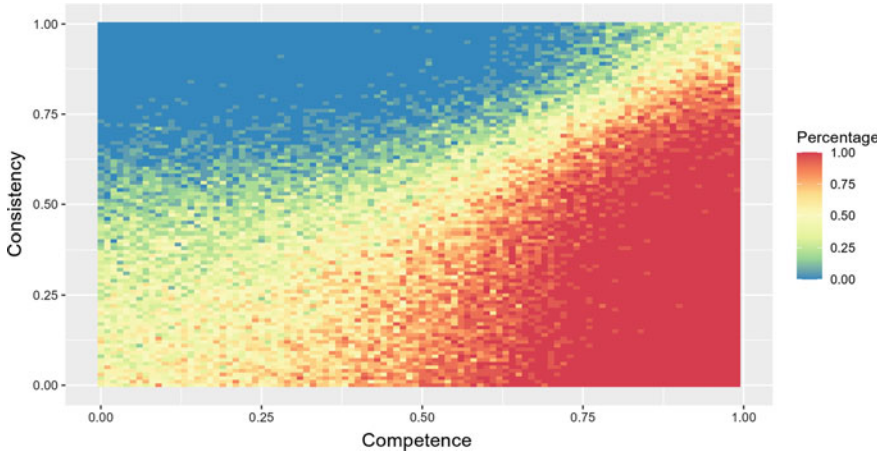

Fig. 1 Heatmap of the percentage of replications [5] where the content strategy was used 100% (red) or 0% (blue) of the time for each combination of values of competence (x-axis) and consistency (y-axis)

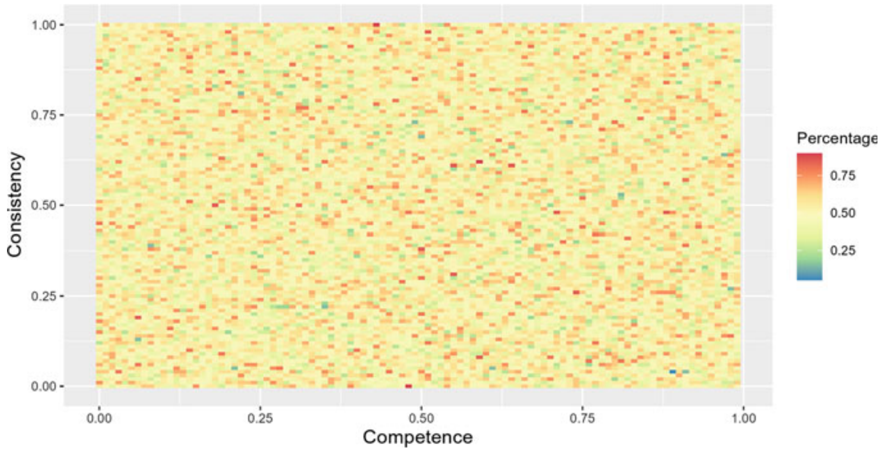

Fig. 2 Heatmap of the percentage of replications where the content strategy was used 100% (red) or 0% (blue) of the time for each combination of values of competence (x-axis) and consistency (y-axis)

3.3 Scenario 3: Belief Utility Equals 1—Belief Accuracy

Figure 3 again presents the percentage of replications where content strategy was used by the agent, going from 0% (blue) to 100% (red) of the time. In this case, we observe the reverse pattern of that from Fig. 1. The agent used the content strategy the most when its competence was lower than the environment's consistency and

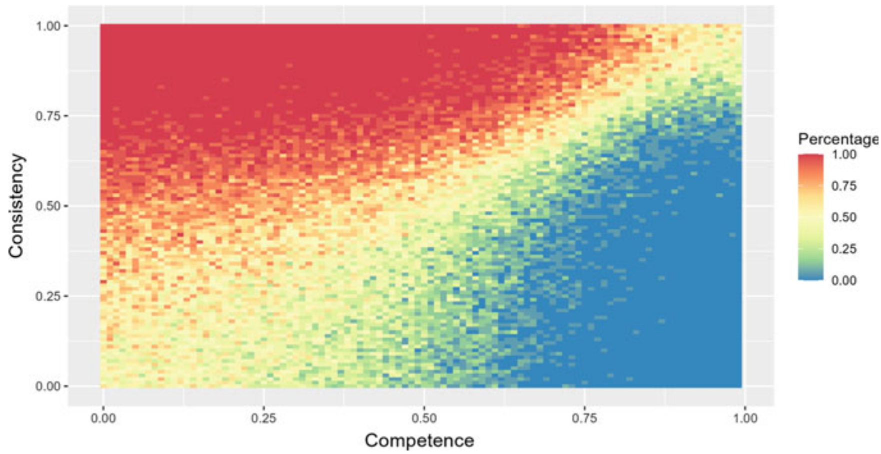

Fig. 3 Heatmap of the percentage of replications where the content strategy was used 100% (red) or 0% (blue) of the time for each combination of values of competence (x-axis) and consistency (y-axis)

when the environment's consistency was medium/high (i.e., > 0.5). On the contrary, the agent relied more on the source strategy when the agents' competence was higher than the environment's consistency and competence was medium/high (i.e., > 0.5). When consistency and competence had a similar value (diagonal) or both of them had a medium/low value (i.e., < 0.5), the decision to use either of the strategies appears to be 50–50%.

4 Discussion

In this model we have sought to understand how an agent can learn to change its epistemic vigilance strategy on the basis of two vital considerations. The first of these is the relation between its own competence in judging the content of the information that is presented and the degree to which the prima facie plausibility of the sources in its environment is consistently related to the accuracy of the information they present. The second is the relationship between the accuracy and utility of the presented information. What we have found is that the agent we modelled was able to flexibly modify its epistemic vigilance strategy depending upon both considerations. Thus, when presented with an environment in which utility and accuracy were closely connected and either consistency or competence were high, it was able to learn to primarily rely upon the epistemic vigilance strategy that could identify highly accurate beliefs, which (in that model) also had high utility. This scenario is an idealisation of most situations in which epistemic vigilance might be used, in that utility and accuracy typically are positively related. However, the inverse case was particularly interesting to examine both because of the potential role that it could have in cases where beliefs function to motivate human cooperation, and because of the additional difficulty the agent faced within it. The second of these considerations is due to the fact that the inverse relationship between utility and accuracy meant that neither source nor content vigilance could reliably produce beliefs with high utility.

In this scenario, the best approach would be to avoid the epistemic vigilance strategy that was the more likely to lead to accurate beliefs and rely upon the other strategy to accept beliefs some of which, by accident, had high utility. Indeed, when presented with the scenario where utility and accuracy had an inverse relationship, we saw that the agent was able to avoid the epistemic vigilance strategy that was likely to generate highly accurate beliefs that, in that scenario, lacked in utility. Thus, the agent was able to react appropriately to the inverse relationship between accuracy and utility. Finally, we examined the scenario where there was no relationship between accuracy and utility of beliefs. In this case we also did not observe any relationship between the choice of epistemic strategy used and the levels of competence or consistency, showing that where there was no relationship between utility and accuracy, the agent had no particular tendency to prefer source over content vigilance.

This study showed that an agent can be capable of learning to alter its epistemic vigilance strategy in a way that reacts flexibly to its abilities, the epistemic structure of its environment and the relationship between the accuracy and utility of beliefs. However, it was quite limited in that it only considered these variables in very simple terms, did not explore other parameters, and did not look at the speed with which the agent was able to modify its behaviour to suit its conditions. In future work, we will explore these aspects of the studied phenomenon in order to achieve insight into how humans are capable of reacting flexibly to a range of situations.

In addition, in this model, only the cognitive processes of the agent were simulated, in order to determine regularities in its behaviour. The same variables have also been examined using a multi-agent model that allowed agent interaction to be focussed upon in the analysis [6].

The research leading to these results has received funding from the Norwegian Financial Mechanism 2014-2021 (Project number 2019/34/H/HS1/00654).

References

1. Sperber, D., Clément, F., Heintz, C., et al.: Epistemic vigilance. Mind Lang. **25**, 359–393 (2010). https://doi.org/10.1111/j.1468-0017.2010.01394.x
2. Talmont-Kaminski, K.: For god and country, not necessarily for truth: the nonalethic function of superempirical beliefs. Monist **96**, 447–461 (2013). https://doi.org/10.5840/monist201396320
3. Talmont-Kaminski, K.: Epistemic vigilance and the science/religion distinction. J. Cogn. Cult. **20**, 88–99 (2020). https://doi.org/10.1163/15685373-12340075
4. Simon, H.A.: The Sciences of the Artificial. Cambridge, M.I.T. Press (1969)

5. Wickham, H.: ggplot2: Elegant Graphics for Data Analysis, 2nd edn. Springer, New York, NY (2016)
6. Rybnik, M., Puga-Gonzalez, I., Shults, L., Dąbrowska-Prokopowska, E., Talmont-Kaminski, K.: An agent-based model of the role of epistemic vigilance in human cooperation (under review)