Teaching Informal Logical Fallacy Identification with a Cognitive Tutor

Nicholas Diana
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
ndiana@cmu.edu

John Stamper
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
john@stamper.org

Kenneth R. Koedinger
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
koedinger@cmu.edu

ABSTRACT

In this age of fake news and alternative facts, the need for a citizenry capable of critical thinking has never been greater. While teaching critical thinking skills in the classroom remains an enduring challenge, research on an ill-defined domain like critical thinking in the educational technology space is even more scarce. We propose a difficulty factors assessment (DFA) to explore two factors that may make learning to identify fallacies more difficult: type of instruction and belief bias. This study will allow us to make two key contributions. First, we will better understand the relationship between sense-making and induction when learning to identify informal fallacies. Second, we will contribute to the limited work examining the impact of belief bias on informal (rather than formal) reasoning. We discuss how the results of this DFA will also be used to improve the next iteration of our fallacy tutor, how this tutor may ultimately contribute to a computational model of informal fallacies, and some potential applications of such a model.

Keywords

Cognitive Tutors, Informal Logical Fallacies, Informal Reasoning, Cognitive Task Analysis, Difficulty Factors Assessment

1. INTRODUCTION

Despite the recognized importance of critical thinking in traditional education, critical thinking is largely absent from the educational technology space (e.g., online courses/MOOCs, cognitive tutoring systems, etc.). Some of the recent work on critical thinking in educational technology has focused on comparing critical thinking in face-to-face and computer-mediated interactions. Researchers often use content-analysis to identify instances of critical thinking in online and face-to-face discussions [3, 10]. In this work, critical thinking is not the primary focus of the course, but rather an epiphenomenon.

Other work, particularly in the domains of philosophy, writing and law, has addressed critical thinking more directly. For example, some recent work has demonstrated that argument diagramming using a graphical interface improved argumentative writing skills [6] as well as critical thinking skills more generally [5]. However, similar gains are seen using paper-and-pencil argument diagramming as well, suggesting the software may be more of a convenience than a necessary factor [4].

Despite the challenges of working in an ill-defined domain [8], another intersection of critical thinking and e-learning has been in intelligent tutoring systems (ITS). For example, Ashley and Aleven [1] built an ITS to teach law students to argue with cases more effectively. The study we propose extends this work on critical thinking in the ITS space to a more general population. We will build a cognitive tutor that teaches users to identify several common informal logical fallacies. We chose informal fallacies because they offer a degree of structure to the otherwise ill-defined domain of informal reasoning, making the content more amenable for use in a cognitive tutor. Using this tutor, we will conduct a difficulty factors assessment (a type of a cognitive task analysis) [7] to evaluate the impact of two factors on the user's ability to identify logical fallacies.

The first factor explored will be *type of instruction*. The Knowledge-Learning-Instruction (KLI) framework lists three types of learning processes, and suggests that the best instruction for teaching a specific skill depends on the type of process used to learn that skill. The purpose of the *type of instruction* manipulation is to better understand the learning processes that underpin the identification of logical fallacies. Specifically, we are interested in whether this skill is more efficiently learned using induction (e.g., showing many examples of the fallacy) or sense-making (e.g., providing detailed descriptions of the fallacy's mechanics). Textbooks used to teach logical fallacies often take both approaches, giving readers an explanation of a fallacy followed by some small number of examples. As this skill may consist of multiple, more fundamental skills (or knowledge components), the mixed approach used by textbooks may prove to be the most efficient. Nevertheless, the proportion of time to devote to each learning process remains an open question that this experiment may help answer.

The second factor that may negatively impact a student's ability to identify logical fallacies is *belief bias*, the tendency

Table 1: Breakdown of the problems used in the tutor. Note that *(F)*, *(A)*, *(C)*, and *(L)* correspond to *for*, *against*, *conservative* and *liberal*, respectively. For example, in the first cell of the table, we see an *apolitical* prompt, which *fallacy 1* is used to argue *for*.

	Apolitical	Political	Apolitical	Political	Apolitical	Political
Fallacy 1	(F)	(C)	(A)	(L)	(F)	(C)
Fallacy 2	(A)	(L)	(F)	(C)	(A)	(L)
Fallacy 3	(F)	(C)	(A)	(L)	(F)	(C)
Fallacy 4	(A)	(L)	(F)	(C)	(A)	(L)
Fallacy 5	(F)	(C)	(A)	(L)	(F)	(C)
Fallacy 6	(A)	(L)	(F)	(C)	(A)	(L)

to judge arguments more favorably if we agree with the conclusion. Early work on belief bias explored its effect on formal reasoning using syllogisms [9, 2], but there is some evidence that suggests that belief bias may operate differently in informal reasoning [11]. The proposed study builds on and contributes to this research by empirically testing the effect of belief bias on learning to identify informal fallacies.

2. FUTURE RESEARCH PLANS

2.1 Difficulty Factors Assessment

We will use a Difficulty Factors Assessment (DFA) to identify the factors (if any) that make it more or less difficult for students to learn how to identify logical fallacies. The proposed experiment will explore the impact of two primary factors as well as several secondary factors.

2.1.1 Type of Instruction

The proposed experiment will explore the impact of *type of instruction* by randomly assigning each participant to one of three conditions. In each condition, when the participant is given a problem and asked to identify the logical fallacy, they will be given a set of possible answers and the option to view more information about each of the answers. In the first condition, when participants ask for more information they will be shown a brief, but detailed description of the mechanics of each fallacy (sense-making). In the second condition, participants will be shown two examples of each fallacy (induction). In the the third condition, participants will be shown a description and one example for each fallacy (mixed).

In addition to comparing the effect of increased examples between groups, we will be able to compare this effect within groups by treating completed problems as viewed examples. This analysis will help us pinpoint the average number of examples needed to be able to identify the fallacies used in the experiment, and compare that number to the average numbers seen in common textbooks.

2.1.2 Belief Bias

The proposed experiment will explore the impact of *belief bias* on a student's ability to identify logical fallacies by altering the political orientation of problem content and comparing performance on those problems with the participant's personal political orientation. Of the 36 problems presented, half will be apolitical (i.e., politically neutral) and half will be political. Of the political problems, half will have a conservative orientation, half a liberal orientation. The apolitical problems are also split into two categories (for an issue or against an issue) for balance. Problems can be broken down into three subcomponents: the prompt (either political or apolitical), the fallacy, and the conclusion (either for/against or conservative/liberal). Table 1 shows the breakdown of each problem.

2.1.3 Secondary Factors Explored

In addition to the main effects of *type of instruction* and *belief bias*, our design also allows us to explore several secondary factors. We can test whether *type of instruction* has a differential effect on specific fallacies. For example, sense-making may be more important for learning to identify a circular argument, while examples may be sufficient for learning to identify a Post Hoc fallacy. We can also test whether participants are more likely to identify a fallacy given the nature of the prompt (political vs. apolitical) or the valence of the conclusion (for/against or conservative/liberal).

2.2 Towards a Computational Model of Logical Fallacies

The ultimate goal of this work is to develop a computational model of logical fallacies. Achieving this goal requires overcoming several large challenges.

2.2.1 Lack of Labeled Examples

First, to train a model to detect such a nuanced use of language will most likely require a large number of labeled examples. Furthermore, these examples will most likely have to be varied and authentic (perhaps unlike many of the purposefully illustrative examples used in textbooks). To solve this shortage of labeled examples, we propose using our cognitive tutor to train crowd workers to identify fallacies in real-world media sources. The quality of those labels can be evaluated using traditional crowdsourcing methods (e.g., consensus of the crowd). High quality labels can then be automatically integrated into the tutor training system, increasing the number of potential examples crowd workers can use to achieve mastery. This increase in the number of examples may be especially important if our DFA reveals that learning to identify informal fallacies is a primarily inductive skill. Figure 1 shows the feedback loop relationship between crowd workers and the cognitive tutor.

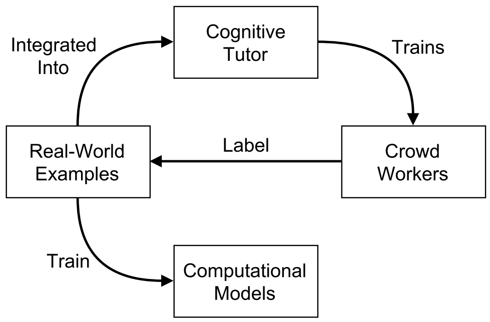

Figure 1: Feedback loop relationship between the cognitive tutor and crowd workers. The real-world examples labeled by crowd workers can be used to both improve the cognitive tutor and train computational models.

2.2.2 Modeling the Semantic Nature of Fallacies

Informal Logical Fallacies is an umbrella term that encompasses a diverse array of fallacies. Some of these fallacies may be easier for a machine learning model to detect. For example, the *Slippery Slope* fallacy often has the generic structure: "First X, pretty soon there'll be Y too!" These kinds of syntactic features will likely be easier to detect than the semantic features necessary to identify a fallacy like *Circular Reasoning*. Finding the right method for approaching these more difficult cases will be one of the key challenges of this work.

2.2.3 Potential Applications

If we meet these challenges and are able to detect logical fallacies in real-world text, there are potential applications in media (both traditional and social), politics, and education. One could imagine a plugin for your favorite word processor that underlines an *Appeal to Ignorance* just as it would a misspelled word. Similarly, one could imagine how broadcasts of presidential debates in the future might be accompanied by a subtle notification anytime a candidate uses *Moral Equivalence*.

In conclusion, we propose a plan to develop a computational model of informal logical fallacies. The first, and most concrete, step of this process is developing a better understanding of the factors that promote and hinder how we learn to identify informal fallacies. We propose a difficulty factors assessment to explore the impact of sense-making versus induction support, as well the impact of belief bias. Discovering how these factors regulate learning will not only allow us to build a better tutor, but will improve our understanding of how we learn informal logical fallacies in general.

3. REFERENCES

[1] K. D. Ashley and V. Aleven. Toward an intelligent tutoring system for teaching law students to argue with cases. In *Proceedings of the 3rd international conference on Artificial intelligence and law*, pages 42–52. ACM, 1991.

[2] J. S. Evans, J. L. Barston, and P. Pollard. On the conflict between logic and belief in syllogistic reasoning. *Memory & cognition*, 11(3):295–306, 1983.

[3] J. Guiller, A. Durndell, and A. Ross. Peer interaction and critical thinking: Face-to-face or online discussion? *Learning and Instruction*, 18:187–200, 2008.

[4] M. Harrell. No computer program required: Even pencil-and-paper argument mapping improves critical-thinking skills. *Teaching Philosophy*, 31(4):351–374, 2008.

[5] M. Harrell. Assessing the efficacy of argument diagramming to teach critical thinking skills in introduction to philosophy. *Inquiry: Critical Thinking Across the Disciplines*, 27(2):31–39, 2012.

[6] M. Harrell and D. Wetzel. Improving first-year writing using argument diagramming. *Proc. of the 35th Annual Conf. of the Cognitive Science Society*, (1987):2488–2493, 2013.

[7] K. R. Koedinger and A. Terao. A cognitive task analysis of using pictures to support pre-algebraic reasoning. In *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*, pages 542–547. Citeseer, 2002.

[8] C. Lynch, K. Ashley, V. Aleven, and N. Pinkwart. Defining ill-defined domains; a literature survey. In *Proceedings of the workshop on intelligent tutoring systems for ill-defined domains at the 8th international conference on intelligent tutoring systems*, pages 1–10, 2006.

[9] J. J. B. Morgan and J. T. Morton. The distortion of syllogistic reasoning produced by personal convictions. *Journal of Social Psychology*, 20(1):39–59, 1944.

[10] D. R. Newman, B. Webb, and C. Cochrane. A content analysis method to measure critical thinking in face-to-face and computer supported group learning. *Interpersonal Computing and Technology*, 3(September 1993):56–77, 1995.

[11] V. Thompson and J. S. B. T. Evans. Belief bias in informal reasoning. *Thinking & Reasoning*, 18(3):278–310, 2012.