A Serious Game for Studying Blame Attributions

Keywords: Serious Games; Blame; Attribution; Artificial Intelligence; Inference

Games that serve a purpose beyond entertainment are known as serious games. Such games may be utilized for delivering interventions, educating, or gathering data for scientific investigation (Dörner, Göbel, Effelsberg, & Wiemeyer, 2016). Although the term "serious games" was coined in the 1970s (Abt, 1970), there has been a rapid increase in their use for research purposes. Laamarti, Eid, & El Saddik, 2014 propose a serious games taxonomy that includes five dimensions: 1) the player's activity type; 2) the sensory modalities experienced during gameplay; 3) the player's interaction with the game; 4) the digital game's environment; and 5) the application domain, which pertains to the game's objective, such as education.

Research in cognitive science has embraced the serious games approach towards delivering interventions and collecting data in *massive online experiments*. Sander van der Linden, along with his co-authors, have used the approach to develop interventions targeted at reducing the spread of misinformation (Roozenbeek & Van der Linden, 2019). This includes the *Fake News Game* in which participants created news articles using misleading tactics (Roozenbeek & Van Der Linden, 2019), the *Harmony Square Game*, in which players learn how political misinformation is produced and spread (Roozenbeek & van der Linden, 2020), and the *Bad News Game*, in which players learn about six common misinformation techniques used in fake news (Basol, Roozenbeek, & van der Linden, 2020). Playing all three games significantly improved people's ability to identify misinformation.

An illustration of the potential of serious games for social science research, particularly in the context of AI perception, can be found in the Moral Machine Experiment (Awad et al., 2018). The Moral Machine Experiment utilized serious game elements to collect data on people's moral decision-making regarding self-driving cars. Participants were presented with moral dilemmas that self-driving cars might face and were asked to make a judgment on which outcome they considered more acceptable. A significant aspect of the study was its use of a generative approach to scenario development. Each scenario drew upon a larger pool of components, such as a man, a pregnant woman, or a cat on the left side of the road. Randomization of these components allowed for exploration of how factors such as species, social value, gender, age, fitness, and utilitarianism might impact decision-making. The components were presented visually, with participants having access to text-based descriptions as well.

Inspired by the Moral Machine example, we've designed a serious game, *The Blame Game*, to collect data on people's blame attributions towards artificial and human agents. The objective of The Blame Game is for players to assume the role of an investigator investigating a case where an agent has committed a crime. Players aim to gather information about the case and make a determination about who should be held responsible for the wrongdoing. We we outline four separate criteria for using a causal cognitive approach for serious games towards gathering data to develop a framework of blame. Our approach utilizes the generative game design used by the Moral Machine Experiment (Awad et al., 2018)

First, the game will be presented in a visual novel format, which is an interactive game that incorporates both text and image-based storytelling. Visual novels require minimal gameplay, as players simply click to progress through the story while making some choices along the way. This approach offers several advantages, including ease of development as it only requires image and text, and enhanced engagement compared to purely text-based games. The choices made by players throughout the game will provide valuable data on their decision-making and

judgment. This data will enable us to investigate factors such as the sequence in which players seek information and the extent to which a specific piece of information impacts their decision-making process.

Second, we use a generative approach towards developing the visual novel scenarios. Specifically, this involves generating images and text that describe different factors and agents. This will include desire, intent, capability, foresight, willingness to learn, willingness to prevent, agent type, and context. They will be expressed in a binary way to avoid unimaginable combinatorics, thus expressing themselves as "on" or "off", or "high" or "low" with the exception of Agent, which is either human or AI and context, which can be various different contexts of AI use. For example, someone can either intend or not intend a certain outcome, and their capability can either be high or low. Players can receive information about factors either in the form of evidence they can analyze or text-based descriptions they can read. Thus, we are able to explore a larger exploratory space rather than focusing on specific hypotheses (Griffiths, 2015).

Third, players have the freedom to choose which information they want to learn and in what order they want to do so. They are also allowed to make a judgment about who is responsible at any point in the game, even if they have not examined all available information. This approach yields valuable data on the types of information people consider when making a responsibility judgment, which can vary depending on the case they are evaluating. Additionally, it provides an opportunity to investigate how unique combinations of inspected factors impact responsibility judgments. By affording players this level of autonomy, the game also serves an educational purpose by encouraging engagement with AI ethics-related questions.

Lastly, to capture the effect of context, the game will feature a variety of domains such as healthcare, self-driving cars, and algorithmic trading. Moreover, players will interact with scenarios that involve both A-bot and human characters. This approach facilitates an investigation into factors that uniquely influence people's judgments toward AIs compared to humans.

References

Abt, C. (1970). Serious games. *New York City, New York, USA, 1st edition*.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., ... Rahwan, I. (2018). The moral machine experiment. *Nature*, *563*(7729), 59–64.

Basol, M., Roozenbeek, J., & van der Linden, S. (2020). Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of cognition*, *3*(1).

Dörner, R., Göbel, S., Effelsberg, W., & Wiemeyer, J. (2016). *Serious games*. Springer.

Griffiths, T. L. (2015). Manifesto for a new (computational) cognitive revolution. *Cognition*, *135*, 21–23.

Laamarti, F., Eid, M., & El Saddik, A. (2014). An overview of serious games. *International Journal of Computer Games Technology*, *2014*.

Roozenbeek, J., & Van Der Linden, S. (2019). The fake news game: actively inoculating against the risk of misinformation. *Journal of risk research*, *22*(5), 570–580.

Roozenbeek, J., & Van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, *5*(1), 1–10.

Roozenbeek, J., & van der Linden, S. (2020). Breaking harmony square: A game that "inoculates" against political misinformation. *The Harvard Kennedy School Misinformation Review*.

Figure 1: The Blame Game Logo