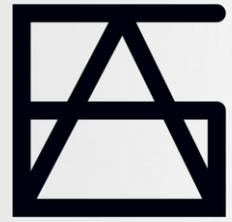


March 2023



ELSA LAB
DEFENCE

No. 1

Research memo series

Experiences in Participatory Design Using the
BetterBeliefs Tool During the REAIM 2023
Conference

AUTHORS

Marc Steen, Jurriaan van Diggelen, Kate Devitt

www.elsalabdefence.nl

1 Introduction

One of the aims of the ELSA Lab Defence (<https://elsalabdefence.nl/background/>) is to develop and deploy methods to enable people with different roles and backgrounds to engage in dialogues. Notably, the aim is to enable various people to discuss Ethical, Legal, and Societal Aspects of the development and deployment of AI systems in military context.

At the REAIM 2023 conference (<https://reaim2023.org/>), we had the opportunity to collaborate with BetterBeliefs, to use and explore the BetterBeliefs tool (Devitt, 2021) within this context to facilitate such conversations: in a workshop setting, with approximately 30 participants (16 February).

Please note that this had been a relatively small and exploratory experiment, and so is our reporting: tentative and exploratory. In the following sections, we report on the following:

2. Experiment and questions
3. Preparations and workshop
4. Outcomes and discussion
5. Conclusions

2 Experiment and questions

In the context of the ELSA Lab Defence, we are interested in all sorts of methods to facilitate experts to discuss and collaborate. Notably, we are interested in the following types of discussion and collaboration:

- › **Between experts with different disciplinary backgrounds:** Ethical, Legal, and Societal Aspects, with experts on AI technology, with experts on the military and operations.
- › **Between different types of organizations, in the *Quadruple Helix*** (Carayannis and Campbell 2009): academia, government, industry, and society—where we understand that the latter can be difficult, due to the secrecy inherent to military applications.

Given the sensitive domain (military) and the different disciplines and organizations, these discussions and collaborations can be complex or difficult. Consequently, we are looking for methods or tools that help to promote a **careful approach**, e.g., methods that enable participants to express their ideas or claims with some precision and care, and, e.g., to provide supporting (or refuting evidence) for their and other participants' ideas or claims.

This is why we were interested in using and trying-out the **BetterBeliefs** system; it is 'an online platform for sharing and evaluating ideas, using collective intelligence to make evidence-based decisions' (<https://betterbeliefs.com.au/>). The system was pioneered and developed by Kate Devitt (adjunct associate professor Human-Centred Computing, and CEO BetterBeliefs), who has a research interest in epistemology: a branch of philosophy that aims to better understand how people develop and deal with knowledge (hence the key role for adding and evaluating evidence, see below).

This system offers the following key functionalities (and more, but less relevant for our current report):

- › People can **input hypotheses**, in the form of a statement, with some background
- › Others can **vote on these hypotheses**; they can agree (up-vote) or disagree (down-vote)
- › People can **add pieces of evidence** to each hypothesis, either supporting it or refuting it
- › Others can **evaluate each piece of evidence** in terms of quality (on a 10-point scale)

At the REAIM conference, we had the opportunity to use and try-out this system. We organized a workshop (<https://reaim2023.org/events/elsa-labs-applying-participatory-methods-to-design-complex-military-technologies/>) to explore the following questions:

1. How does this tool enable people to **express their ideas** on a difficult topic such as 'weaponized drones'? (see next section for this focus and other preparations) (*this question relates to process*)
2. How does this tool enable people to **engage in conversations about these ideas**, e.g., reacting to each other, listening to (or reading) other people's ideas? (*this question relates to process*)
3. How does this tool enable researchers (like the authors) to collect and develop **knowledge about a topic** such as 'weaponized drones' (*this question relates to content, to substantive findings*)

Please note that this was a relatively small and explorative study—aiming to draw only relatively lightweight and tentative conclusions.

3 Preparations and workshop

Supported by the people of BetterBeliefs, we made the following preparations:

- › **We created a scenario.** In order to enable a focused conversation, participants need a scenario, as a starting point, to promote shared understanding (Steen, Buijs, and Williams 2014)—without such a scenario, people will very likely miscommunicate and diverge
- › **And prepared a series of hypotheses.** In theory, participants would be able to come up with hypotheses during the workshop, but this would **not be practically** feasible during the workshop: you need time for it, and some support to make ‘good’ hypotheses

The ELSA Lab project team members entered these hypotheses into the BetterBeliefs system, with relevant pieces of evidence—critically, one can only enter a hypothesis together with at least one piece of evidence (to promote thinking in terms of evidence)

Hypotheses articulated by the ELSA Lab Defence project team, with support of Kate Devitt:

1. Ethical: We should only use weaponized drones that cannot use force on citizens
2. Ethical: Weaponized drones must have a ‘human in the loop’, to make decisions to use lethal force
3. Ethical: The use of weaponized drones is morally reprehensible
4. Legal: The employment of weaponized drones leads to accountability gaps.
5. Legal: The laws of armed conflict are sufficient to regulate the use of weaponized drones.
6. Legal: The obligation of an Article 36 Additional Protocol I weapons review provides sufficient legal apparatus to lawfully deploy weaponized drones.
7. Societal: The public believes that weaponised drones are a danger to humanity
8. Societal: The public will support the use of weaponised drones in a just war
9. Societal: The public will support the use of weaponised drones in a war that ‘we’ are winning

Hypotheses added by others, before or during the workshop

10. It is irresponsible to place humans in or on the loop of weaponized drones
11. The use of data-driven algorithms for the identification of human targets poses serious risks of violations of International Humanitarian Law (IHL)
12. It is possible to use weaponized drones while abiding to legal and ethical standards.



The workshop's duration was 100 minutes. We first gave a short presentation about the ELSA Lab, the BetterBeliefs system, and the scenario that we/they would be working with.

After that were three slots of 20-25 minutes, to discuss Ethical, Legal, and Societal Aspects respectively, with 2-3 minutes of introduction and the remaining 20 minutes for the participants to engage in conversations. Participants were sitting in groups of 6 around 6 round tables, to facilitate also face to face conversations. We invited participants to **engage both face to face, and via the BetterBeliefs system**. Regarding the latter, participants were invited to prioritize voting on hypotheses (that were already in the system) (up and down) and evaluating pieces of evidence (that were already in the system) (in terms of quality). Formulating and inputting new hypotheses and finding and inputting new pieces of evidence was possible but was not focused upon—mostly for practical reasons; we did not want to make people spend most of the time articulating hypotheses and finding evidence.

4 Outcomes and discussion

All of the participants' activities were logged. We will briefly report on and discuss several key outcomes and findings:

Hypotheses that people **agreed** with most ('up-votes', resulting in 'high degree belief'):

- 11. The use of data-driven algorithms for the identification of human targets poses serious risks of violations of IHL (highest belief)
- 12. It is possible to use weaponized drones while abiding to legal and ethical standards
- 2. Ethical: Weaponized drones must have a 'human in the loop', to make decisions to use /lethal force
- 8. Societal: The public will support the use of weaponised drones in a just war

Our interpretation: Participants believe that weaponized drones may violate IHL; yet they believe it is possible to use them within legal and ethical norms; they also believe that there must be a human in the loop, and that the public will support their use in a just war.

Hypotheses that people **disagreed** with most ('down-votes, resulting in 'low degree belief'):

- 1. Ethical: We should only use weaponized drones that cannot use force on citizens
- 5. Legal: The laws of armed conflict are sufficient to regulate the use of weaponized drones.
- 10. It is irresponsible to place humans in or on the loop of weaponized drones
- 3. Ethical: The use of weaponized drones is morally reprehensible (lowest belief)

Our interpretation: Participants believe that the use of weaponized drones is not necessarily reprehensible, provided that these drones cannot use force on citizens; moreover, they believe that the current laws of armed conflict are sufficient to regulate their usage.

Please note that it is easier to understand an agreement with a positively phrased hypothesis (e.g., "It is possible ...") than to understand a disagreement with a negatively phrased hypothesis (e.g., "It is irresponsible ..."). For the latter, e.g., we **cannot** reason *a contrario*, like: "It is it responsible to place humans in or on the loop". This illustrates the importance of adequate support in the articulation of hypotheses; please note that this particular hypothesis was entered by a participant—not by the project team members.

Hypotheses that people **engaged with most** (either in agreement or in disagreement):

- 4. Legal: The employment of weaponized drones leads to accountability gaps (17 up-votes; 11 down-votes)
- 9. Societal: The public will support the use of weaponised drones in a war that 'we' are winning (17 up-votes; 9 down-votes)
- 2. Ethical: Weaponized drones must have a 'human in the loop', to make decisions to use lethal force (18 up-votes; 5 down-votes)
- 7. Societal: The public believes that weaponised drones are a danger to humanity (14 up-votes; 10 down-votes)

Looking at these hypotheses (above), it would be interesting to further discuss these, precisely because they attracted people's engagement—especially when they received both up-votes and down-votes in somewhat equal measure, like: '7. The public believes that weaponized drones are a danger to humanity". Such a further discussion could help to clarify, e.g., what danger(s) people think

of (e.g., being killed, being wounded, etc.) and about which part(s) of humanity people are concerned (e.g., own soldiers, enemy soldiers, civilians, etc).

We did not look into how people evaluated individual pieces of evidence (although these data are available). We did, however, have a look at how people on average rated the quality of pieces of evidence behind each hypothesis, mainly for illustrative purposes.

Evidence for the following hypotheses was evaluated, on average, as **high quality**:

10. **It is irresponsible to place humans in or on the loop of weaponized drones** (4,7 on a 5-point scale), with one piece of evidence: '... Remote drone pilots are merely geographically remote from their actions, while pilots that oversea or are gate-keepers of weaponized drones are both geographically remote and "interactionally" remote' ..., source: <https://www.amazon.com/Killing-Remotely-Psychology-Drones/dp/0316628298>
4. **Legal: The employment of weaponized drones leads to accountability gaps** (4,0 on a 5-point scale), with two supporting and two refuting pieces of evidence; an example of supporting evidence: 'Look at the analogy of landmines: they kill without human intervention, and no person can be held accountable ...', source: https://en.wikipedia.org/wiki/Land_mine#Impacts

Evidence for the following hypotheses was evaluated, on average, as **low quality**:

5. **Legal: The laws of armed conflict are sufficient to regulate the use of weaponized drones** (2,1 on a 5-point scale), with one piece of evidence: '... no reason to believe that advanced, AI-supported weapons systems will not be capable of reliably distinguishing between civilians and combatants ...', source: <https://www.japcc.org/articles/autonomous-weapon-systems-in-international-humanitarian-law/>
9. **Societal: The public will support the use of weaponised drones in a war that 'we' are winning** (2,6 on a 5-point scale), with two supporting pieces of evidence, e.g., this one: 'This 2013 study, shows the majority of Americans support drone warfare. This is in stark contrast with the rest of the world ... source: <https://www.pewresearch.org/fact-tank/2013/10/23/report-questions-drone-use-widely-unpopular-globally-but-not-in-the-u-s/>

The BetterBeliefs system can create a diagram (Figure 1) that plots all hypothesis in a two-dimensional grid: on the x-axis the degree of belief, i.e. agreement; and on the y-axis the weight of evidence, i.e. the evaluation of the quality of the evidence behind that hypothesis. Decision makers can shift the threshold for WoE and DoB with a slider to make the coloured quadrants incorporate different numbers of hypotheses depending on the risk of the decision. E.g. low weight of evidence threshold may be sufficient for low risk decisions. A belief greater than 50% may be sufficient to enable an action.

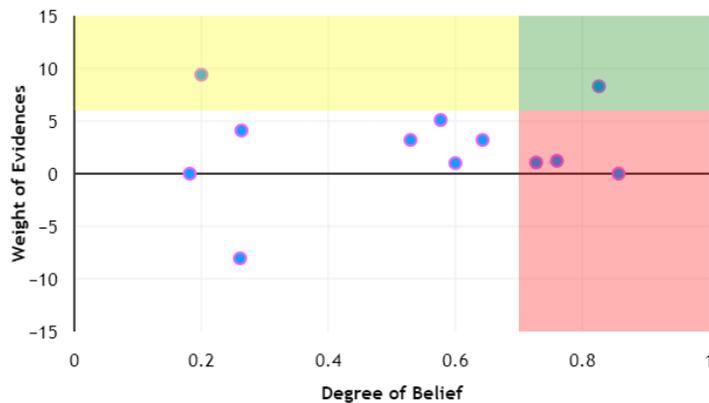


Figure 1: Plot made by BetterBeliefs, with hypotheses plotted on two axes: degree of belief, and weight of evidence.

We can look, e.g., at the **green area**, for a hypothesis that many people agree with (high on x-axis) and with evidence that many people rate as high quality:

12. It is possible to use weaponized drones while abiding to legal and ethical standards.

Using a 'race' metaphor, we can argue that this hypothesis ('It is possible ...') is the best from this workshop: most people agreed with it, backed up by evidence they value.

In addition, we can look at the **red area**, for hypotheses with a **high** degree of belief, **but** relatively **low quality** evidence—many people agree with these statements, but the evidence is somewhat lacking and may need to be improved:

11. The use of data-driven algorithms for the identification of human targets poses serious risks of violations of IHL

2. Ethical: Weaponized drones must have a 'human in the loop', to make decisions to use lethal force

8. Societal: The public will support the use of weaponised drones in a just war

Regarding the overall process at this workshop, please note that our workshop had ad-hoc participants with no explicit commitment or shared goal: visitors to a two-day conference, choosing from a program with 10 parallel sessions, presumably rather randomly. In addition, please note that the BetterBeliefs system is probably better used in a more dedicated context, e.g., a two-day project or task group meeting, with dedicated project group or task group, with a shared goal, with a series of sessions with sufficient time to engage in conversations with more depth, and sufficient time to conduct research, e.g., to find and evaluate evidence.

5 Conclusions

We experimented with using the BetterBeliefs system to facilitate conversations about Ethical, Legal, and Societal Aspects of the design and application of AI systems. We did that with a group of 30 people during a 100 minutes workshop at the REAIM 2023 conference.

This has been a relatively small and explorative study, which allows us to draw several tentative conclusions—**which only says something about this particular workshop** (see end of previous section):

1. The BetterBeliefs tool can enable people to **express their ideas** on a difficult topic such as ‘weaponized drones’. First, even with little time and effort, they can vote-up and vote-down hypotheses (statements) in order to express their agreement or disagreement. Second, with a bit more time and effort, they can articulate and add hypotheses (statements)—although this requires some guidance (e.g., avoid negative phrases)
2. The BetterBeliefs tool potentially enables people **to engage in conversations**, e.g., by adding new pieces of evidence, either supporting or refuting. However, it is not possible to add remarks or questions, which could be useful, e.g., to clarify or question a hypothesis. Conversations can, of course, be done face to face—but then they are not documented in the system (and probably not documented at all).
3. The BetterBeliefs system can be used as a poll, by only looking at voting-up (agree) and voting-down (disagree) of hypotheses. This gives substantive (**and tentative**) **findings** (section 4), even in little time with little effort. Of course, we need to appreciate the small sample size, i.e. group of participants (and potentially biased, as visitors to the REAIM conference; although not sure in which direction ... ‘pro’ or ‘contra’ weaponized drones).

Overall, the BetterBeliefs tool is interesting. Crucially, if participants have more time and—probably more importantly,—if they have a shared goal, e.g., when they are working together in a project or task group, and need to deliver, they can do all of the above with more quality.

Bibliography

- Carayannis, Elias, and David F. J. Campbell. 2009. "'Mode 3' and 'Quadruple Helix': Toward a 21st century fractal innovation ecosystem." *International Journal of Technology Management* 46 (3-4):201-234.
- Devitt, S. K., Pearce, T. R., Chowdhury, A. K., & Mengersen, K. (2021). A Bayesian social platform for inclusive and evidence-based decision making. arXiv preprint arXiv:2102.06893.
- Steen, Marc, Jan Buijs, and Doug Williams. 2014. "The role of scenarios and demonstrators in promoting shared understanding in innovation projects." *International Journal of Innovation and Technology Management*.