Appearing in Weijers, Dan (in press-due late 2020). "Chapter 15. "Just wrong," "disgusting," "grotesque": How to Deal with Public Rejection of New Potentially Life-saving Technologies", in Steff, R., Soare, S., & Burton, J. (eds.), *Emerging Technologies and International Security: Machines, the State and War*. Routledge, pp. 254-272.

Chapter 15

"Just wrong", "disgusting", "grotesque": How to Deal with Public Rejection of New Potentially Life-saving Technologies

Dan Weijers (Senior Lecturer, University of Waikato)

Introduction: "Just wrong", "disgusting", "grotesque"

Dear researchers and policymakers, imagine that you have just finalized plans for the rollout of a new game-changing technology, one that could significantly improve security and save lives. At the first press conference for your project, some reporters start asking questions you had not anticipated. The next morning you awake to a deluge of media alerts about your project. Excited, you hurriedly begin to read them…

"just wrong"
"ridiculous… grotesque"
"absurd… disgusting"
"morally reprehensible"
"despicable… callous… retarded"[1]

Oh dear; a tidal wave of vitriol, a veritable flood of moral repugnance! This was certainly not what you had hoped for. Why would the public respond like this to your new life-saving technology? Should you ignore this moral outrage, or engage with it? Can you turn public perception back around, or is it too late? Unfortunately for the team behind Policy Analysis Market (PAM), they did not have the liberty of considering these questions before the program was terminated at the highest levels of the US government. PAM was a proposed prediction market funded by the United States (US) Defense Advanced Research Projects Agency (DARPA) in 2003 that had the potential to prevent terrorist attacks. After an unexpected public outcry, PAM was cancelled without even sending a "please explain" to the team behind it (Hanson, 2007).

This chapter begins by explaining how PAM was supposed to work and how public expressions of moral repugnance led to the project being shut down. It then introduces and explains the psychology of moral repugnance and discusses a taxonomy of kinds of moral repugnance that might result from the roll out of a new or disruptive technology for security purposes. A major contribution of this chapter is the discussion of what these different kinds of repugnance should mean for new disruptive technology proposals. Using PAM as the central example, the chapter works through a taxonomy of kinds of repugnance to identify the responses that would be appropriate for teams to use when researching and implementing new technologies. The chapter closes with some words of advice about managing public perception of disruptive new security technologies.

Prediction Markets and the Policy Analysis Market

[1] The "just wrong" comment was made by Senator Thomas Daschle in Congress (Congressional Record Vol. 149, No. 114, July 29, 2003: S10082–S10083). The "ridiculous… grotesque" comments were made by Senator Ron Wyden (Wyden & Dorgan 2003a). The "absurd… disgusting" comment was made by Senator Byron Dorgan the next day (Wyden & Dorgan 2003b). The "morally reprehensible" comment was made by the editors in "Pentagon drowns in its own 'dead pool' ", *The Virginian Pilot*, July 31, 2003, B10. The "despicable… callous… retarded" remark was made by blogger Greg Saunders (See Meirowitz and Tucker (2004) for more detail on the reactions to PAM).

The terrorist attack on New York's World Trade Center in 2001 reportedly cost the US $3.3 trillion (Carter & Cox, 2011) and, far worse, the horrifying loss of 2,977 lives (Amadeo, 2019). In the wake of this devastating attack, DARPA, an agency tasked with generating breakthrough technologies and capabilities to aid US national security efforts, began investigating new methods to better understand international security threats, including terrorism (Hanson, 2007). A small part of this investigation was the plan to trial a prediction market, called PAM, based on the work of Robin Hanson (Hanson, 2007).

As explained by Weijers (2013a), modern prediction markets (PMs) are electronic marketplaces that facilitate the purchase and sale of shares in predictions about real-world outcomes. Participants browse a list of specific predictions and buy or sell shares in predictions they think are priced too high or too low. The price of a share in a PM varies between $0.00 and the payout price for predictions that turn out to be true. For example, in a PM that pays out $1.00 for each share in an accurate prediction, prices will vary between about $0.01 and $0.99. The most recent sale price of a share in a PM is an indication of how likely the prediction is to be true according to the collective wisdom of the participants in the market. For example, PredictIt is currently running a PM on "Who will win the 2020 U.S. presidential election?", with several specific predictions available to trade on, such as "Donald Trump will win the 2020 U.S. presidential election". As at September 26, 2019, the collective wisdom about Trump's re-election chances is that he has a 42% chance.[2] Any participant in the market who believes Trump has a much lower chance of a second term as president would sell (or short) shares in the prediction, hoping for a pay-out if Trump is not elected president in 2020.[3] If Trump is elected president in 2020, then participants holding shares in the prediction will be paid out $1.00 per share, netting them $1.00 minus the average price they paid for their shares for each share they hold.

Prediction markets have proven themselves to be very well-calibrated prediction machines (Surowiecki, 2004). In other words, predictions with high share prices usually end up being true and those with low prices usually end up being false. For example, the predictions facilitated by the Iowa Electronic Markets regularly outperform polls in predicting the outcomes of national elections (Berg, Nelson, & Rietz, 2008). Encouraged by the success of PMs in many domains, several researchers have argued that carefully constructed PMs could be used to fight terrorism (Hanson, 2006a; Looney, 2004; Surowiecki, 2004; Weijers & Richardson, 2014a; 2014b; Weijers, 2016; Yeh, 2006). It has also been argued that an anti-terrorism PM, such as PAM, could be run for about the same cost as increasing the number of active spies by one (Weijers, 2016). These considerations doubtless spurred on DARPA and Hanson to work towards a trial of PAM. The main idea behind PAM was to use technology to create a new avenue to gather information that may not have been collected by existing intelligence and security measures (Hanson, 2007). A PM that anyone from around the world could participate in might harvest the ambient wisdom of globally dispersed crowds and insights from individuals with concrete information that they would not share through existing channels (Hanson, 2006a; Weijers & Richardson, 2004b).[4]

In 2003, PAM was presented to a small audience to explain the basics of how the trial and the intended PM would work (Hanson, 2007). As was widely documented at the time, a slide showing example predictions included the assassination of Yasser Arafat and a North Korean missile attack (Hanson, 2006b). These predictions were cited by Senators Ron Wyden and

[2] See: https://www.predictit.org/markets/detail/3698/Who-will-win-the-2020-US-presidential-election.

[3] For more information on PMs, see: Cowgill, Wolfers, and Zitzewitz, (2009), Surowiecki, (2004), Wolfers and Zitzewitz (2004), or Weijers (2013; 2018).

[4] For more information on how a PM could be set up to effectively help fight terrorism, see: Hanson (2006a), Weijers (2016), and Weijers and Richardson (2014a).

Byron Dorgan (2003a) in their rebuke of PAM the day after the presentation.[5] The following media cycle was a flood of moral outrage, laced with a spattering of reasons why PAM was morally problematic and unlikely to work (Weijers & Richardson, 2014b). PAM was cancelled two days later.

Moral Repugnance – More than a Feeling?

When someone criticizes a disruptive, possibly life-saving, security technology very strongly and very emotionally, they are probably expressing moral repugnance about it. *Moral* repugnance is potentially much more threatening to a new technology than *regular* repugnance. Garden-variety, or regular, repugnance is best thought of as intense revulsion – finding something extremely disgusting. Moral repugnance, as discussed by Weijers and Keyser (2016), has the added complexity of the disgust being explicitly or implicitly linked to one or more *moral* reasons why the target of the disgust is worthy of further contempt. Part of the complexity of moral repugnance is the often nebulous or ill-defined interplay between moral emotion and moral reason that usually occurs within people as a morally-charged intuition – a feeling that the proposed technology is abhorrent in the extreme.

As any good modern politician knows, reasons are often not as powerful as emotions – knowledge that is backed up by various experiments, including those by Taber and Lodge (2006), which demonstrated political reasoning is biased by moral emotions. Unfortunately, winning over a crowd, or a populace, seems much more easily achieved by eliciting positive emotions toward yourself than by expressing a sound line of reasoning (Lodge & Taber, 2013). Leon Kass (the former chair of President G.W. Bush’s Council on Bioethics) might support this feature of politics. He has argued that emotions and intuitions, even when they are not backed by explicit reasons, are often a better guide to morality and policy than reason-based arguments (Kass, 1998). Discussing moral repugnance,[6] Kass (1998, p. 687) claims that it can be the “emotional expression of deep wisdom”, and that lack of a good moral reason to explain the intuition does not impede its authority. Kass assumes there is some good moral reason for widespread moral repugnance, even if none come to mind (for any of those people!). If Kass is right about the deep wisdom of widespread moral repugnance even when it lacks explicit moral reasoning, then this is how moral repugnance might be founded upon *implicit* moral reasons.

It is reasonable to assume that most philosophers and any people inspired by the goals of the enlightenment would likely find it easier to engage in debates about the moral acceptability of new security technologies if the alleged reasons for the technologies’ immorality were made explicit. Consider moral debates at home or at work. Reasons offered for opposing views provide the opportunity to gain a new perspective, to learn and possibly refine one’s own view on the issue. But, amenability to debate is not the main difficulty for those like Kass who find it acceptable to win a moral argument with feelings alone. A huge amount of research in social and moral psychology and behavioral economics has demonstrated the many ways in which our intuitions, including moral intuitions, are subject to pernicious biases.[7]

[5] Senators Ron Wyden and Byron Dorgan (2003a) were the prominent and very vocal leaders of the pushback against PAM, which they suggested would be “offensive to almost everyone” (no page). See Meirowitz and Tucker (2004) for more detail on the reactions to PAM.

[6] Kass calls it “repugnance”, but makes it clear he views the term as morally loaded (Kass, 1998).

[7] For psychology, see especially the work of Jonathan Haidt (2001; 2007). For behavioural economics, see especially the work of Kahneman and Tversky, summarised in Kahneman (2011). For a summary of moral biases and their potential impact on policy and law, see Sunstein (2005).

Consider the following example of moral bias.[8] Security forces are planning a rescue mission in three remote locations. Traditional methods can be employed (Plan A) or a new and relatively untested technology, such as fully autonomous drones, can be used (Plan B). Here are the best predictions of the outcomes for each plan:

1) Plan A results in 200 people being saved
2) Plan B results in a one third chance of 600 people being saved and a two thirds chance of saving no one

Which plan is morally preferable? Most people think Plan A is preferable because it seems morally better to have the guarantee of saving lives or to avoid the possibility of saving no one.[9]

Imagine now another remote rescue situation. Again, you are presented with two options, the second of which employs a new technology. The best predictions of the outcomes for each plan are:

1) Plan C results in 400 deaths
2) Plan D results in a one third chance of no deaths and a two thirds chance of 600 deaths

Which plan is morally preferable? Most people prefer Plan D in this second remote rescue scenario because it seems morally repugnant to choose a policy that results in 400 deaths when there was an option that included a chance of no one dying.[10]

The reason this is an example of how moral judgments can be biased is revealed by noticing that people tend to prefer the riskier new technology plan in the second scenario, but not in the first, despite both scenarios and sets of plans being identical in substance. The only difference between the first and second scenarios is the framing of the plans. It appears that framing a plan as saving lives or permitting deaths affects our moral intuitions about the plans. People tend to *feel* that one plan is better than another, but rational inspection of the details reveals that if those feelings are based on deep wisdom, then deep wisdom seems like a poor guide for national security decision making in this context.

Since moral intuitions can be biased, and therefore unhelpful, policymakers should use widespread moral repugnance as a *red flag* – a warning that a thorough moral investigation may be warranted. The purpose of the investigation is to search for reasons which indicate that pushing ahead with the technology would be immoral, and to carefully tease those out from negative intuitions and feelings of disgust that are not based on important moral considerations. For example, the ice cream cone is a technology that allows people to walk around while eating ice cream without generating wasteful packaging. Kass revealed in his book about nourishing the soul through eating that he finds this technology morally repugnant: "…licking an ice cream cone… offends those who know eating in public is offensive. ... This doglike feeding… ought to be kept from public view, where, even if *we* feel no shame, others are not compelled to witness our shameful behavior" (Kass, 1999, p.148–149; emphasis in original). Despite the strong moral language used by Kass to describe eating an ice cream, a thorough moral investigation of his repugnance may well reveal no compelling moral reasons for his repugnance.

[8] This case was inspired by similar cases by Kahneman and Tversky's famous flu case (1984). See Sunstein (2004) for other examples.

[9] This assumption about what most people would prefer is based on the complete structural similarity between my scenario and Kahneman and Tversky's famous flu case (1984).

[10] Again, this assumption about what most people would prefer is based on the complete structural similarity between my scenario and Kahneman and Tversky's famous flu case (1984).

The Public is outraged! What Should I do?

So, you have discovered that the public is up in arms about your new disruptive technology. What should you do? Depending on whether you are a researcher or a policymaker, you will have several important interests to balance, especially ethics, professional expectations, political implications, geopolitical imperatives (such as keeping up with expected developments by strategic rivals),[11] and to a lesser extent prudence. The best way to respond to moral repugnance depends very much on whether or how the repugnance is or could be justified. The next section sets out a taxonomy of kinds of moral repugnance that make reference to relevant contextual aspects from the point of view of researchers and policymakers proposing disruptive new technologies.

Before turning to the taxonomy, the general procedure for action should be noted. Figure 1, below, outlines the appropriate steps. Exactly how to carry out each step is highly context dependent, and so will only be discussed briefly here. This procedure for action is, as noted, a general one, so application of it should be attempted with an open mind and a decent serving of common sense. This procedure should be followed whenever moral repugnance is expected or detected. Especially for disruptive new technologies, being prepared to deal with potential moral repugnance before it occurs could easily make the difference between your idea never getting off the shelf and your new technology saving lives. So, it is best to run this procedure within your team and then with a focus group well before planning any public events or disseminating any information.

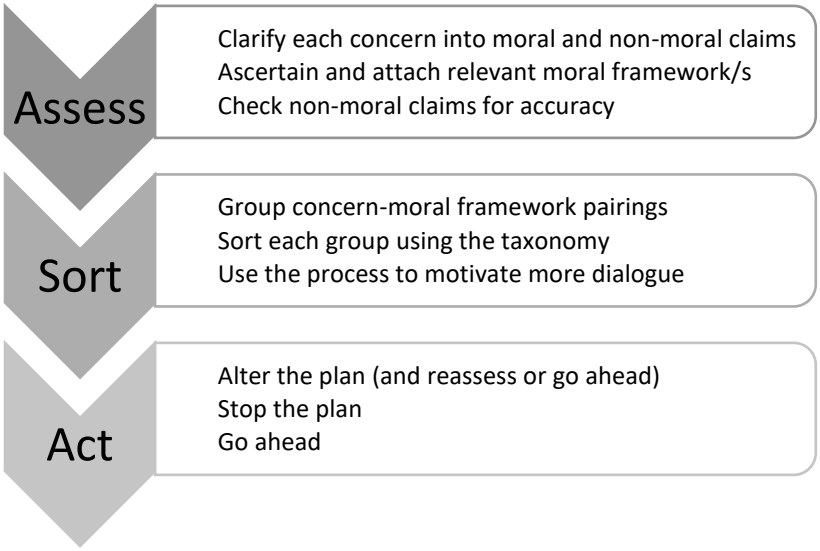

Figure 1: A General Procedure for Action when Moral Repugnance is Expected or Detected

Gather and Assess the Potential Moral Repugnance

The first step is to assess any apparent moral repugnance. Combining in-house brainstorming and primary and secondary research, gather all of the instances of potential moral repugnance. Once all of the actual and likely utterances of moral repugnance are gathered, they need to be clarified. Clarifying the concern requires identifying what the reason is for the repugnance and whether that reason is moral. It is also useful to clarify how strong and widespread the

[11] A mounting concern in the US is that potential authoritarian adversaries do not face the same domestic or normative constraints when it comes to developing and deploying, for civilian or military purposes, new technologies.

repugnance based on this reason is. Trending and similar search engine and social media metrics may help with this.

When seeking to clarify potential moral repugnance, it is important to check for records and read all of what a complainant has said. Focusing on just the attention-grabbing headlines or snippets of what someone has said can cause an unhelpful lack of clarity. The context around the snippets is where the potential moral reasons are likely to be stated or at least implied. The expressions of moral repugnance made by Senators Ron Wyden and Byron Dorgan (2003a, no page) during their news conference on the "terror financing scheme", PAM, are an illuminating example of this. Amidst calling PAM "ridiculous… grotesque" (Wyden) and "absurd… disgusting" (Dorgan), the Senators also presented several moral claims and arguments. These comments from Dorgan, for example, strongly imply reasons for his moral repugnance:

> This betting parlor on the Internet will include wagers, for example…: Will Mr. Arafat be assassinated? Will there be missile attacks from North Korea? … And those predictive bets will then give intelligence presumably to the Department of Defense. I think this is unbelievably stupid. … It is a tragic waste of taxpayers' money. … It is offensive. And it is, in my judgment, nearly useless.

Dorgan clearly believes that PAM will not be effective in helping intelligence forces learn anything they do not already know. Dorgan also claims that betting on death is offensive. But, clarifying the potential moral concern is not always as easy as reading a transcript. In other instances, influential people may need to be sought out and offered the opportunity to clarify their comments themselves.

In the process of clarifying each moral concern, care should be taken to distinguish between moral and non-moral claims. Moral claims link some action or state of the world to a moral value, thereby imbuing that action or state with a special force that is thought to be weightier than nearby non-moral claims. For example, "betting on death is morally wrong" is a moral claim that links the action of betting on death to a negative moral value, one that is more powerful than a mere violation of etiquette or convention. Non-moral claims do not link moral values with the world in this way. For example, the statement from earlier that the events at New York's World Trade Centre on September 11, 2001 caused 2,977 people to die does not link a moral value to any action or state of the world.[12]

Consider Dorgan's concern about PAM's ineffectiveness, which can be captured by the following premises:

1) PAM will not produce any new anti-terrorist information
2) PAM costs a lot of tax-payers' money
3) Spending a lot of tax-payers' money on something that doesn't work is wasteful of tax-payers' money
4) Being wasteful of taxpayers' money is morally wrong
5) Therefore PAM is morally wrong

Premises 1 and 2 are non-moral empirical claims – claims about how the world is or will be. If either of these claims are false, then PAM would not fit the definition of "wasteful of tax-payers' money" in premise 3. Premise 4 is the moral claim in the argument – it links being wasteful to a moral value, in this case a negative one.

[12] To be clear, this claim likely arouses negative feelings and possibly moral outrage within us, but the claim *itself* does not even mention a moral value.

The separation of moral and non-moral claims is important because the truth of non-moral claims is usually established in a different way to the truth of moral claims. Indeed, philosophers are still debating whether it is possible to establish the truth of moral claims, or even sensibly apply a truth value (e.g., true or false) to them (e.g., Nolan, Restall, & West, 2005). Given this important difference, moral and non-moral claims are treated differently in the taxonomy, which requires their prior separation.

Also while in the process of clarifying each moral concern, it is important to ascertain the relevant complainants' moral frameworks and ensure the concerns are coupled with the relevant frameworks during the process. As discussed by Weijers and Keyser (2016, p. 101), a moral framework is a "systematic set of moral beliefs that can be reasoned through". The framework needs to be accessible via introspection, such that the complainant in question could ask themselves, "do I believe in any moral principles that are relevant to this situation?", and articulate those principles. Noting the moral framework linked to each potential complainant is important because the moral frameworks feed into the sorting process and can cause different responses to be specified to the repugnance.

The final task in the "assess" step is to investigate whether the non-moral claims are true, plausible, or false. Many non-moral empirical claims can be checked directly against current scientific findings, or if they are claims about the future, deemed plausible or otherwise based on a similar check. For example, several experts on prediction markets and security have argued that PAM would plausibly have gathered some new and useful information not otherwise available to security and intelligence forces (e.g., Hanson, 2006b; Looney, 2004; Yeh, 2006).

Non-moral conceptual claims should also be checked. For example, a disruptive new technology may be labelled expensive or wasteful, like PAM was. Complete understanding of these concepts usually relies on some background knowledge about the domain they are being applied to. It is important to check whether the concept is being employed in a sense that is appropriate for the relevant domain of inquiry. For example, "expensive" in the domain of federal security policy means something quite different in dollar terms to the same concept used in the domain of grocery shopping. Thorough investigations of the non-moral conceptual claims should check for the inappropriate use of concepts, be they context-blind or otherwise idiosyncratic. For example, the budget for the trial or full roll-out of PAM could be compared to similar projects to support or undermine claims about how "expensive" the program is. Weijers (2016) did exactly this, finding that PAM would have cost about US$1m to set up and run, which is approximately equal to the cost of deploying one more spy in the field.

Remember to go through this process with members of your team very early on in the development process. If the team finds your new technology morally repugnant, then it is unlikely to stand a chance once the public discovers it, and should not be pursued. In April 2018, over 3,000 Google employees successfully revolted against Google's involvement with Project Maven (Conger, 2018). A Pentagon initiative, Project Maven, involves harnessing machine learning to help drones distinguish between humans and objects (Work, 2017). This case demonstrates the potential for security-related technologies to cause moral outrage and the importance of (at least) getting your own team onboard before developing the new technology. It is entirely possible that the outrage over Project Maven, and recognition that the Pentagon will seek to deploy other AI-enhanced military technologies in the near future, has contributed to the US Department of Defense's newly formed Joint Artificial Intelligence Center seeking to hire an ethicist to guide development and deployment of AI technologies. After all, the Center's Director, Jack Shanahan, led Project Maven (Lopez, 2019).

Sort the Potential Moral Repugnance using the Taxonomy

As indicated in Figure 1, the first task in the “sort” step is to group the potential expressions of moral repugnance such that within groups both the concern and the relevant moral framework are the same. This sorting will likely result in several groups that can each be considered a singular concern-framework pairing from here on. Resource-based decisions may need to be made at this point; you may wish to proceed with only the groups of concerns that seem popular or contain compelling moral reasons.

The resulting groups are then subjected to the taxonomy of moral repugnance discussed in the next section. Figure 2, below, presents a flow chart to make using the taxonomy easier.

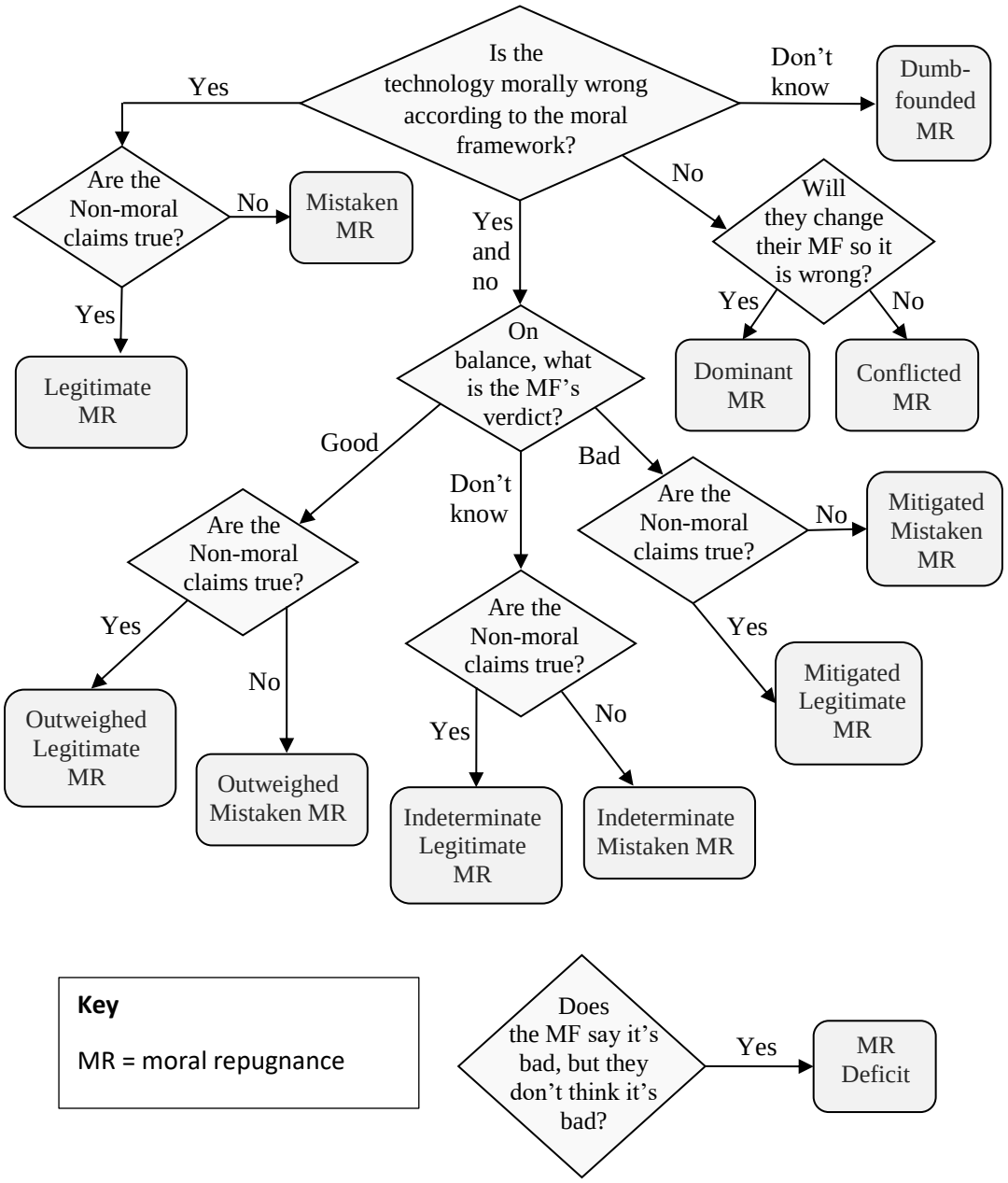

Figure 2: A Flowchart for the Taxonomy of Moral Repugnance

Ideally, this process should also be worked through in anticipation of potential public outrage as well as in response to any actual public outrage. Engaging with the process ahead of time makes it easier to find the time to enter a dialogue with people that may hold the identified concerns. As will be discussed in the next section, the taxonomy recommends ignoring some concerns. By explaining this to potential complainants, it may motivate them to provide more reasons and fewer unjustified expressions of disgust. This is helpful for your team because reasons can be considered, and changes to the plan more nuanced than “shelve it forever” or “just go ahead” can be considered.

Reassess the Plan and Put it Into Action

Correct use of the taxonomy will result in a general recommendation for action for each of the concern-framework pairings. Some recommendations will be to ignore the concern, but others will suggest changing the plan or even scrapping the whole project. Scrapping a project halfway or further through development can be heartbreaking, but if it is recommended by the taxonomy, then it is the right thing to do. In many cases, the process of using the taxonomy of moral repugnance on a disruptive technology will produce mixed recommendations. Here, as elsewhere, common sense should be applied.

If use of the taxonomy was in response to actual expressions of concern, then action should be as swift as possible. The plan should be changed in accordance with the recommendations and the changes should be communicated to the relevant stakeholders. If use of the taxonomy was in anticipation of potential expressions of concern, then the recommendations should also inform purely cosmetic and communicative decisions about how the new technology is presented to the public.

A Taxonomy of Moral Repugnance

The taxonomy of moral repugnance presented in this section is based on the detailed conceptual work of Weijers and Keyser (2016). A major contribution of this chapter is the discussion of what these different kinds of repugnance should mean for new disruptive technology proposals. The first subsection discusses the simpler kinds of potential moral repugnance to introduce readers to the main issues. The second subsection discusses more complex, but also more realistic kinds of potential moral repugnance. The final subsection briefly discusses some less common kinds of moral repugnance.

Some Basic Varieties of Moral Repugnance and how to Deal with Them

This subsection discusses three basic varieties of moral repugnance: Legitimate Moral Repugnance, Mistaken Moral Repugnance, and Dumbfounded Moral Repugnance.

Legitimate Moral Repugnance (Legitimate MR) is the most concerning of all the types of repugnance for the team behind the new technology. Legitimate MR is repugnance backed by a logically valid argument that contains only true or plausible non-moral claims and moral claims that cohere with the complainant’s moral framework (Weijers & Keyser, 2016). In other words, Legitimate MR does not result from any empirical mistakes or conceptual peculiarities. Rather, Legitimate MR is the result of a well-formed moral concern.

For example, Wyden and Dorgan might argue that:

1) PAM would encourage and enable betting on death (non-moral claim)
2) Encouraging and enabling betting on death is always immoral (moral claim)
3) Therefore, PAM is immoral

Defenders of PAM might quibble a little about Premise 1, urging for the use of "trading" or "investing" instead of "betting". However, prediction markets tend to operate more effectively when some traders have little or no new information (Wolfers & Zitzewitz, 2004) and trade in a way that is difficult to distinguish from betting. So, Premise 1 seems highly plausible. The logical form of the argument is valid, so the conclusion can be confirmed if Premise 2 is true.

For the purposes of the taxonomy, the initial check of any moral premise is done earlier in the process, when the concern is clarified with the relevant complainant. The next check on any moral premise is against the moral framework of the complainant. This check tests for whether the proposed moral claim is in tension with any other moral commitments that the complainant might have. If the moral claim coheres with the complainant's other moral commitments, then we have an example of Legitimate MR. Presumably, Wyden and Dorgan endorse the moral claim that encouraging and enabling betting on death is always immoral, and that claim coheres with their other moral beliefs.

This description of Legitimate MR allows for a range of possible moral concerns. Even if only one person expresses Legitimate MR for a particular reason, a deeper investigation is warranted. The key questions to answer in such an investigation are: What proportion of people are likely to agree with this criticism if they heard it? And, does the team behind the new technology agree with it?

In the case above, it seems likely that many people would consider betting on death to be morally bad (but not "almost everyone" as Dorgan (2003a, no page) implies).[13] Furthermore, it's not at all clear what proportion of people believe betting on death is something that should *never* be permitted, meaning that it cannot be done, even if it produced some greater good, such as saving lives from terrorist attacks. The moral claim in the argument attributed to Wyden and Dorgan above is absolute, meaning it permits no exceptions. But, as Surowiecki (2004) mentions in his discussion of PAM, national security is a fairly brutal business. As such, good people regularly have to resolve moral dilemmas by trading off various bad actions and consequences (e.g. invasions of privacy, cost to taxpayers) in order to achieve a more important good (e.g., saving lives by keeping innocent people safe). So, it is unlikely the team behind PAM would agree with the absolute moral claim that betting on death is *always* morally wrong. Weijers and Richardson (2014a) investigate this issue in depth, arguing that it is unreasonable to believe that betting on death is always morally wrong, especially when it could save lives. So it seems that, although possibly legitimate and widespread, this concern of Wyden and Dorgan should be thought of as an issue to manage rather than a reason to shut the program down. However, the more reasonable version of this moral claim, that betting on death is morally bad *but not absolutely bad*, is complex and will be revisited in the next subsection.

After investigations are complete, widespread Legitimate MR that the team behind the new technology also find compelling, requires a thorough reworking of the product to avoid the moral concern. If the product cannot be reworked in a way that avoids the moral concern, then the project should be scrapped.

In cases where Legitimate MR is possibly widespread, but apparently based on the widespread belief in a moral claim that the team thinks is wrong, then the team has to decide what balance

[13] Contrary to Dorgan's implication, the popularity of death pools seems to show that many people do not find betting on death at all offensive (Weijers & Richardson, 2014a). "Death pools" are competitions in which participants, for fun or financial reward, pick a list of celebrities that they think will die in the next year (Weijers & Richardson, 2014a). Unfortunately, several of these death pools already allow or even encourage play or real betting on when current and past presidents will die (e.g., ptrradio.com/deathpool, stiffs.com, etc.).

of changing the product and the communication strategy will best achieve their goal of saving lives. Changes that alleviate the legitimate moral concern without detracting from the technology's main benefit should probably be made, but this is not always possible.

For example, the team behind PAM could not have removed the betting element at the heart of their technology,[14] but they might have pointed out to reporters and politicians at that fateful presentation that although betting on death is not nice, it is not so morally bad that it should never be permitted, for example, when that betting can save lives. This may have changed key communicators' beliefs about the absolute nature of the moral claim in Premise 2, and possibly pushed them towards Outweighed Moral Repugnance (discussed below).

Finally, some investigations will reveal that, although one person finds the new technology legitimately morally repugnant, it is based on a highly idiosyncratic moral framework and a moral claim that would be rejected as immoral or amoral by nearly all other moral frameworks. For example, believing that prediction markets are immoral because making predictions is morally wrong, may well fit into someone's moral framework somewhere, but it would be rejected by the vast majority. The team behind the new technology should not treat cases like these as reasons to stop or change their project.

Mistaken Moral Repugnance (Mistaken MR), is morally charged repugnance that results from the combination of a moral claim and a false non-moral claim (that, if true, would trigger the moral claim) (Weijers & Keyser, 2016). In other words, someone who has Mistaken MR, would see the technology as morally permissible if they understood all of the relevant non-moral facts. This kind of repugnance is likely to be common, especially given the emerging prevalence of technologically incompetent talking heads (Dawkins, 1998) and purposeful misinformation via news and social media (Lazer et al., 2018).

As mentioned earlier, a major concern with PAM, is whether it would have been successful at generating security-relevant information that would not have otherwise been gathered by existing methods (e.g., Richey, 2005; Stiglitz, 2003; Wyden & Dorgan, 2003a). However, many subject experts have argued that anti-terrorist prediction markets would likely be effective (Hanson, 2006a; Looney, 2004; Surowiecki, 2004; Weijers & Richardson, 2014a; 2014b; Weijers, 2016; Yeh, 2006). Given the extent of disagreement over this empirical issue, careful investigation of such non-moral premises is important. Depending on the severity of the moral claim attached to the non-moral claim under investigation, feasibility studies can be a good option for resolving such disagreements.

In other cases, it will be easy to establish whether the non-moral claim is too implausible to distract the team behind the new technology. For example, a Flat-Earther's moral repugnance at a round-the-world yachting race based on the non-moral premise, "since the Earth is flat, a round-the-world yachting race would result in the deaths of all the competitors", could be dismissed without so much as a Google search.

Assuming the non-moral premise can confidently be dismissed as false or implausible, then the project should proceed. However, depending on the depth and scope of the misinformation among stakeholders, a communications effort may have to be launched alongside the technical development of the product.

Dumbfounded Moral Repugnance (Dumbfounded MR) occurs when someone very strongly believes that the new technology is morally wrong, but they cannot find within their moral framework a moral principle that justifies their belief (Weijers & Keyser, 2016). Based on the

[14] Note that removing the financial incentives of betting by play money or "bragging points" instead would be unlikely to alleviate the moral concerns about the frivolous and disrespectful nature of betting on death.

findings of Haidt, Bjorklund, and Murphy (2000) it seems that a person experiencing Dumbfounded MR might say that they don't know *why* it's morally wrong, but they are sure that it *is*.

If Dumbfounded MR is widespread, then the team should consider investigating potential causes for the dumbfounding that explain but not justify it. For example, someone with an unconscious racist bias may have inadvertently acquired it through an upbringing devoid of real contact with people of that race, but replete with negative media coverage of them. This conditioning account explains how the racism came to be, but it does not morally excuse racist behavior in an age rich with information about racial equality.

The likelihood of the cause of the intuition being irrelevant to any reasonable moral framework, such as in the racism case, seems high for any intractable Dumbfounded MR. So, contrary to Kass's advice, if despite best efforts, no plausible reason for the Dumbfounded MR can be found, the team can ignore such complainants. Further, I would use Kass's moral repugnance at eating ice cream in public as an example of the possibility of atavistic notions inadvertently polluting our current moral intuitions. This is especially true of discourse about new technologies, since our intuitions are less trustworthy when they are about things unlike anything we have previously experienced (Lieberman, 2000; Myers, 2004; Weijers, 2013b; Woodward & Allman, 2007). Wisdom about new technologies ultimately comes from reasons, not feelings.[15]

Some Complex Varieties of Moral Repugnance and how to Deal with Them

In many real life instances of following the procedure outlined here, individuals that have considered the proposed technology, and their own moral framework, carefully will often answer "yes and no" to the question from Figure 2: "Is the technology morally wrong according to [your] moral framework?". Their answer to the follow up question – about how the moral concerns balance out – will decide which of the following three categories their moral repugnance falls into: Outweighed Moral Repugnance, Indeterminate Moral Repugnance, or Mitigated Moral Repugnance.

Outweighed Moral Repugnance (Outweighed MR) occurs when a complainant, after consulting their own moral framework, views the technology as having both up- and down-sides, morally speaking, and sees the moral up-sides as outweighing the down-sides (Weijers & Keyser, 2016). In other words, the new technology is seen as overall morally positive, despite the perception of one or more moral problems. The first step is to investigate any non-moral claims, including those that form part of the argument *in favor* of the new technology, to see whether they are mistaken. Any mistaken aspects of the moral repugnance should be dealt with as recommended in the subsection above.

The second step is to treat all of the perceived downsides that do not include mistakes as individual instances of Legitimate MR. Depending on the moral assessment of the team and the expected level of stakeholder agreement, these instances of Legitimate MR may result in changes to, and even the cessation of, the project (despite them being outweighed in one or more complainants). In many cases, this process will result in something like the example of betting on death used in the subsection above; the moral concern checks out as legitimate and is shared by others, but is outweighed by a more pressing moral concern. By encouraging and enabling betting on death, PAM was being disrespectful to human life, which is morally bad, but not so bad that it overrides all other considerations, such as PAM's potential to save

[15] To be clear, Dumbfounded MR should certainly be investigated during the process to see whether good moral reasons can justify it (even if the complainant could not see those reasons themselves). *Intractable* Dumbfounded MR, however, does not constitute a reason to stop or change the plan for the new technology.

innocent lives from terrorist attacks. Of course, once a legitimate moral concern has been identified, minor changes to the new technology that can reduce that concern without jeopardizing the main benefit should be investigated.

Indeterminate Moral Repugnance (Indeterminate MR) occurs when someone can’t decide whether the moral up-sides they see in the technology outweigh the moral down-sides they see in it (Weijers & Keyser, 2016). Indeterminate MR should be dealt with in much the same way as Outweighed MR. Although, it is likely that the team behind the new technology will have to consider changes to their product more seriously for Indeterminate MR than for Outweighed MR.

Mitigated Moral Repugnance (Mitigated MR) occurs when someone believes that the moral up-sides they see in the technology make it more acceptable to them, but that the moral down-sides are so significant that they cannot, on balance, support the new technology (Weijers & Keyser, 2016). Again, this kind of repugnance should be dealt with by first checking the non-moral moral claims for mistakes, and then investigating any Legitimate MR as outlined in the subsection above. This investigation may result in the team having to dramatically rework the project or even abandon it.

After investigating the legitimate concerns for popularity and resonance with the team behind the new technology, careful thought needs to be given to whether public perception could be shifted from Mitigated MR to Outweighed MR by framing the moral issues in a clearer or more relatable way. Trading off different moral values is a difficult task, often with no explicit rules. As simple a thing as publishing one or two photographs can considerably tilt public perception on something as major as a war or a humanitarian crisis.[16]

Some Unusual Varieties of Moral Repugnance and how to Deal with Them

In this subsection, three more categories of moral repugnance, and what to do about them, are briefly discussed.

Dominant Moral Repugnance (Dominant MR) occurs when someone’s moral repugnance about a technology is so strong that they will change their existing moral framework to accommodate it (Weijers & Keyser, 2016). In response to Dominant MR, the team behind the new technology should take the issue very seriously. It is possible that the disruptive technology is so new that members of the public initially overreact out of fear of change or the unknown. The team behind the technology should focus this investigation on the argument behind the Dominant MR, including the moral and non-moral claims. Mistaken non-moral claims and peculiar moral claims may not change the team’s plans, but any reasonable and widespread Dominant MR, probably means the project needs to be radically changed or stopped.

Conflicted Moral Repugnance (Conflicted MR) occurs when someone finds a new technology morally repugnant, but when consulting their own moral framework, can only find reasons that the technology is morally good (Weijers & Keyser, 2016). Conflicted MR should be treated in a very similar way to Dumbfounded MR, except more effort could be put into communicating the moral benefits of the new technology in order to encourage these people to accept their own moral reasons for why the new technology is morally good.

Moral Repugnance Deficit (MR Deficit) occurs when someone does not feel like a technology is morally bad, but when consulting their own moral framework, finds overwhelming reasons

[16] For example, see Astor’s (2018, no page) discussion of photographs that “Changed the Course of the Vietnam War”.

to consider the technology morally bad.[17] If MR Deficit is discovered within the team, the team should take seriously the moral concerns captured in the person’s moral framework, if not their feelings, because members of the team are very well-versed with how the technology will work. When dealing with MR Deficit, the concerns should be investigated as though they were Legitimate MR. Going ahead as planned with a major project that is unethical is a morally, and most likely in the long-run, prudentially bad idea.

Planning and Framing Public Perceptions

So many social and political issues are debated nearly exclusively within echo chambers, such as groups on social media (Garrett, 2009). This complicating factor should be taken into account when developing disruptive security technologies. The framing of the initial message about a new morally praiseworthy technology can go a long way to ensuring it is accepted by the public.

For example, if PAM had been called Anti-Terrorism Markets (even though the original purpose was wider than that), then reporters at the original presentation may not have been so quick to think that the technology was disrespectful. A more convincing discussion of how the markets might attract novel information might also have prevented moral repugnance about the technology being a waste of money.

Ethics screening for disruptive security technologies should not be dominated by cherry picking advisors expected to rubber stamp whatever is proposed. By using the process outlined here, teams behind disruptive life-saving technologies might more effectively work out whether their idea is morally positive and how best to communicate that to others. In other words, the taxonomy on moral repugnance could help organizations like DARPA deal with the information they receive when “engaging a variety of experts and stakeholders with varying points of view – both to hear what they and their professional communities of practice have to say and to help convey to those communities DARPA’s insights about what technology can and cannot do” (DARPA, no date, no page).

Bibliography

Amadeo, K. (2019). How the 9/11 Attacks Affect the Economy Today?: What Is Their Lasting Damage? *The Balance*, Jan 26, 2019. Accessed on Feb 20, 2019 from

[17] This kind of moral repugnance does not appear in Weijers and Keyser (2016).

https://www.thebalance.com/how-the-9-11-attacks-still-affect-the-economy-today-3305536.

Astor, M. (2018). A Photograph That Changed the Course of the Vietnam War, New York Times, Feb 1, 2018. Retrieved February 20, 2019 from https://www.nytimes.com/2018/02/01/world/asia/vietnam-execution-photo.html.

Berg, J. E., Nelson, F. D., & Rietz, T. A. (2008). Prediction Market Accuracy in the Long Run. *International Journal of Forecasting*, *24*(2): 285–300.

Carter, S., & Cox, A. (2011). One 9/11 Tally: $3.3 Trillion, *New York Times*, Sep 8, 2011. Retrieved February 20, 2019 from https://archive.nytimes.com/www.nytimes.com/interactive/2011/09/08/us/sept-11-reckoning/cost-graphic.html?_r=1.

Conger, K. (2018). Google Plans Not to Renew its Contract for Project Maven, A Controversial Pentagon Drone AI Imaging Program. *Gizmodo*. Retrieved September 27, 2019 from https://www.gizmodo.com.au/2018/06/google-plans-not-to-renew-its-contract-for-project-maven-a-controversial-pentagon-drone-ai-imaging-program/.

DARPA (no date). *Ethics & Societal Implications*. Accessed on September 27, 2019 from https://www.darpa.mil/program/our-research/ethics.

Dawkins, R. (1998). What’s Wrong with Cloning? In: Nussbaum M. C., & Sunstein, C. R. (eds.), *Clones and Clones. Facts and Fantasies Human Cloning*. New York and London: W.W. Norton & Company, pp. 54–66.

Garrett, R. K. (2009). Echo Chambers Online?: Politically Motivated Selective Exposure Among Internet News Users. *Journal of Computer-Mediated Communication*, 14(2), 265–285.

Haidt, J. (2001). The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment. *Psychological Review*, 108(4), 814.

Haidt, J. (2007). The New Synthesis in Moral Psychology. *Science*, *316*(5827), 998–1002.

Haidt, J., Bjorklund, F., & Murphy, S. (2000). Moral Dumbfounding: When Intuition Finds No Reason. (Unpublished manuscript, University of Virginia). Retrieved February 20, 2019 from http://faculty.virginia.edu/haidtlab/articles/manuscripts/haidt.bjorklund.working-paper.when%20intuition%20finds%20no%20reason.pub603.doc.

Hanson, R. (2006a). Designing Real Terrorism Futures. *Public Choice*, 128(1), 257–274.

Hanson, R. (2006b). Decision Markets for Policy Advice. In: Gerber, A. S., & Patashnik, E. M. (eds.), *Promoting the General Welfare: New Perspectives on Government Performance*, pp. 151-173, Brookings Institution Press.

Hanson, R. (2007). The Policy Analysis Market (A Thwarted Experiment in the Use of Prediction Markets for Public Policy). *Innovations: Technology, Governance, Globalization*, *2*(3), 73–88.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Straus and Giroux.

Kahneman, D., & Tversky, A. (1984). Choices, Values, and Frames. *American Psychologist*, 39, 341–50.

Kass, L. (1998). The Wisdom of Repugnance: Why We Should Ban the Cloning of Humans. *Valparaiso University Law Review*, *32*(2): 679–705.

Kass, L. (1999). *The Hungry Soul: Eating and the Perfecting of Our Nature*. University of Chicago press.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... & Schudson, M. (2018). The Science of Fake News. *Science*, *359*(6380), 1094-1096.

Lieberman, M. (2000). Intuition: A Social Cognitive Neuroscience Approach, *Psychological Bulletin*, 126: 109–137.

Lodge, M., & Taber, C. S. (2013). *The Rationalizing Voter*. Cambridge University Press.

Looney, R. E. (2004). DARPA's Policy Analysis Market for Intelligence: Outside the Box or Off the Wall? *International Journal of Intelligence and Counterintelligence*, *17*(3), 405–419.

Lopez, T. C. (2019). *DOD Seeks Ethicist to Guide Artificial Intelligence Deployment*. US Department of Defense. Retrieved October 1, 2019, from

https://www.defense.gov/explore/story/Article/1950724/dod-seeks-ethicist-to-guide-deployment-of-artificial-intelligence/source/GovDelivery/

Meirowitz, A., & Tucker, J. A. (2004). Learning from Terrorism Markets. *Perspectives on Politics*, *2*(2), 331–336.

Myers, D. G. (2004). *Intuition: Its Powers and Perils*. Yale University Press.

Nolan, D., Restall, G., & West, C. (2005). Moral Fictionalism Versus the Rest. *Australasian Journal of Philosophy*, 83(3), 307–330.

Richey, M. (2005). Thoughts on the Theory and Practice of Speculative Markets Qua Event Predictors. *Essays in Philosophy*, 6(1), Article 26, no page. Retrieved March 26, 2013, from http://ssrn.com/abstract=1670777

Sunstein, C. R. (2004) Lives, Life-Years, and Willingness to Pay. *Columbia Law Review*, 104, 205– 52.

Sunstein, C. R. (2005). Moral Heuristics. *Behavioral and Brain Sciences*, *28*, 531-573.

Stiglitz, J. (2003). Terrorism: There’s No Futures in it, *Los Angeles Times*, 31 July, 2003. Retrieved October 1, 2019, from http://mason.gmu.edu/~rhanson/PAM/PRESS2/LATstiglitz-7-31-03.htm.

Surowiecki, J. (2004). *The Wisdom of Crowds*. New York: Doubleday.

Taber, C. S., & Lodge, M. (2006). Motivated Skepticism in the Evaluation of Political Beliefs. *American Journal of Political Science*, *50*(3), 755–769.

Weijers, D. (2013). Prediction Markets. *Observatory for Responsible Research and Innovation in ICT*. Retrieved October 1, 2019, from http://www.danweijers.com/pdf/Technology-Prediction_Markets(Dan_Weijers).pdf

Weijers, D. (2013b). Intuitive Biases in Judgements about Thought Experiments: The Experience Machine Revisited, *Philosophical Writings*, 41(1), 17–31.

Weijers, D. (2016). Prediction Markets as an Alternative to One More Spy. In: Galliott J. & Reed W. (eds.), *Ethics and the Future of Spying: Technology, Intelligence Collection and National Security*. Routledge, pp. 80–92.

Weijers, D. & Keyser, V. (2016). The Varieties and Dynamics of Moral Repugnance: Prediction Markets and Betting on Matters of Life and Death. *The Humanities and Technology Review*, 3, 91–129.

Weijers, D. & Richardson, J. (2014a). Is the Repugnance about Betting on Terrorist Attacks Misguided? *Ethics and Information Technology*, 16(3), 251–262

Weijers, D. & Richardson, J. (2014b). A Moral Analysis of Effective Prediction Markets on Terrorism. *International Journal of Technoethics*, 5(1), 28–43

Wolfers, J., & Zitzewitz, E. (2004). Prediction Markets. *Journal of Economic Perspectives*, *18*(2), 107–126.

Woodward, J. & Allman, J. (2007). Moral Intuition: Its Neural Substrates and Normative Significance, *Journal of Physiology-Paris*, 101(4-6), 179–202.

Work, R. (2017). *Establishment of an Algorithmic Warfare Cross-Functional Team (Project Maven)*. Government Executive. Retrieved September 27, 2019 from https://www.govexec.com/media/gbc/docs/pdfs_edit/establishment_of_the_awcft_project_maven.pdf

Wyden, R., & Dorgan, B. (2003a). Senators Ron Wyden and Byron Dorgan Hold News Conference on a Terror Financing Scheme. Retrieved February 9, 2019, from http://hanson.gmu.edu/PAM/govt/senator-wyden-dorgan-pressconf-7-28-03.txt.

Wyden, R., & Dorgan, B. (2003b). U.S. Senator Ron Wyden Holds News Conference Following Closed Policy Luncheon. Retrieved February 9, 2019, from http://hanson.gmu.edu/PAM/govt/Senator-Wyden-Dorgan-pressconf-7-29-03.txt.

Yeh, P. F. (2006). Using Prediction Markets to Enhance US Intelligence Capabilities. *Studies in Intelligence*, *50*(4), 137–149.