Explaining Autonomy

Risk Communication & Robotics

Aaron Mannes, PhD
Senior Policy Advisor
ISHPI Information Technologies

This paper was written with the support of the Apex Data Analytics Engine in the Department of Homeland Security (DHS) Science and Technology Directorate (S&T).
In no way should anything stated in this paper be construed as representing the official position of DHS S&T or any other component of DHS. Opinions and findings expressed in this paper, as well as any errors and omissions, are the responsibility of the author alone.

All the brilliant engineers and workers in the world won't matter if the people don't really care. As the old saying goes, "People don't care what you know until they know that you care."

Tim Murphy, then Chair of the House Subcommittee on Oversight and Investigation in his opening statement in the 2014 hearing on the Chevrolet Cobalt ignition-switch recall[1]

The story of the Ford Pinto, as told by Malcolm Gladwell in *The New Yorker,* should provide a cautionary tale for the robotics industry. This 1970s sub-compact was vulnerable to catching fire when rear-ended. Ford had known about this issue, but only issued recall notices after numerous instances of accidents, some of which resulted in horrifying deaths (as well as terrible publicity for the company and expensive lawsuits). In his article, Gladwell finds that the engineering team responsible for issuing recall notices had not ignored the problem. Rather, after evaluating it, the recall team noted that Pintos catching on fire in collisions was still unlikely (and Pintos were no more vulnerable than any other subcompacts). Given a huge range of potential life-threatening accidents, recall orders could only be issued to address pressing issues, and the Pinto explosions were very rare. As engineers, the recall team saw automobiles as inherent compromises, a balancing of risks and benefits. Designing a small car to be safe in such a catastrophic, but

[1] Quoted in Malcom Gladwell, "The Engineer's Lament: Two ways of thinking about automotive safety," *The New Yorker*, May 4, 2015, https://www.newyorker.com/magazine/2015/05/04/the-engineers-lament

unlikely, accident was not realistic and the same resources could be applied to other issues that would save more lives. The public, distressed by the stories of people being burned alive, saw the situation quite differently. In his *New Yorker* article, Malcolm Gladwell highlights a series of controversies around public frustration with reportedly faulty vehicles that followed the same pattern of miscommunication between the technical experts and the public. Similar scenarios have played out in numerous other industries, most notably nuclear power, which has stagnated since the accident at Three Mile Island.

The robotics industry – which will be particularly vulnerable to the disconnect between experts and laypersons – should heed these lessons and invest in risk communication to foster productive human and robot collaboration, mitigate risks, and build resilience for when, as they inevitably must, accidents and failures occur.

Robots for the purpose of this paper are defined as "physically embodied systems capable of enacting physical change in the world."[2] Often directed by non-deterministic algorithms, robots (or autonomous systems – the terms will be used interchangeably) are going, at times, to act in unpredictable, and potentially dangerous ways. There has already been extensive discussion about ensuring the safety of self-driving vehicles.[3] Robots, however, will come in many forms and be increasingly ubiquitous in homes, places of business, and public spaces. There will also be a variety of autonomous systems less visible to the public in factories and infrastructure.[4] Further, many systems, even if not autonomous themselves, may include autonomous subsystems. All of these devices will have the potential for failures, of both mundane and

---

[2] A more expanded definition continues: "They enact this change with effectors, which can move the robot (locomotion), or objects in the environment (manipulation). Robots typically use sensor data to make decisions. They can vary in their degree of autonomy, from fully autonomous to fully teleoperated, though most modern systems have mixed initiative, or shared autonomy. More broadly, robotics technology includes affiliated systems, such as related sensors, algorithms for processing data, and so on." Riek, L.D. (2017). "Healthcare Robotics". *Communications of the ACM*, Vol. 60, No. 11. pp. 68-78.
Also: Riek, L.D. "Robotics Technology in Mental Healthcare". In D. Luxton (Ed.), *Artificial Intelligence in Behavioral Health and Mental Health Care.* Elsevier, 2015. pp. 185-203.

[3] See Ashley Halsey III, "We're listening, Department of Transportation says on the future of driverless cars," *Washington Post*, Mach 1, 2018
Also: Jeremy Hsu, "When It Comes to Safety, Autonomous Cars Are Still 'Teen Drivers,'" *Scientific American*, January 18, 2017, https://www.scientificamerican.com/article/when-it-comes-to-safety-autonomous-cars-are-still-teen-drivers1/

[4] Besides robots there are also a vast range of systems classified as cyber-physical systems (IoT) and intelligent agents (disembodied autonomous systems such as chatbots). While not considered robots, these systems share some characteristics with them and at least some of the discussion below will apply.

catastrophic varieties, that can cause harm. As systems affecting the material world, robots may do physical harm. But they are also cyber-systems that can be compromised in ways that injure privacy or security. Because robots will have a measure of autonomy, people interacting with them may infer agency so that certain types of failure may frighten people and do psychological damage. Finally, unwise and reckless human interactions might also lead to damaging accidents. Effective risk management can reduce the probability of these failures, but it cannot eliminate them. It will be essential to communicate with decision makers and the public, both to inform them of these possibilities and to develop resilience and prevent public or regulatory backlash in the face of these failures.

Risk communication is "the term of art used for situations when people need good information to make sound choices."[5] Seemingly straightforward, risk communication is a complex, multi-layered process with an extensive literature of theory and practice. It is not "spin." The field represents a set of tools that can be used in a crisis, but also in everyday situations (such as considering to undergo a medical procedure). The term is incomplete because risk communication also discusses the benefits of a decision.[6] This is particularly critical for robotics, where the potential for new kinds of accidents must be balanced against a reduction of many other kinds of hazards, as well as a range of other benefits. Effective risk communication also incorporates and addresses the critical issue of risk perception – that (as the episode of the Ford Pinto illustrates) experts and publics do not understand risk in the same way. If the public comes to see robots as dangerous it could lead to litigation, regulation, or simply a preference to not purchase or interact with them.

Given the potential for robots to be perceived as dangerous, risk communication will be an essential task for private and public sector entities that produce, market, use, and regulate robots. Risk communication is part of the risk management strategy in that people aware of risks will be better able to act to reduce these risks (for example, by engaging with a robot appropriately or purchasing the robot that meets their needs.) The process of risk communication can elicit perceived risks from communities and the public that experts may not consider, address these

[5] Baruch Fischhoff, Noel Brewer, Julie Downs, "Introduction," *Communicating Risks and Benefits: An Evidence-Based User's Guide*, eds. Baruch Fischhoff, Noel Brewer, Julie Downs, Food and Drug Administration, 2011, 1 https://www.fda.gov/AboutFDA/ReportsManualsForms/Reports/ucm268078.htm
[6] Ibid, 1

risks, and build relationships so that isolated incidents are less likely to trigger backlash. In the event of a large-scale failure (such as failures by robots integrated into critical infrastructure), effective risk communication will be essential to reduce panic.

For many companies in many industries, risk communication is little more than compliance. One observer, writing about warnings on drug labels with their tiny print on folded paper, states: "Taken as a whole, it fairly shouts: 'Don't read me!'"[7] Similarly, privacy rights are guaranteed by lengthy, jargon-filled statements that most people simply agree to, effectively signing away their privacy.[8] This paper argues that the robotics industry would be wise to incorporate robust risk communication capabilities into the field, rather than viewing it as a mere afterthought necessary for compliance, to foster productive human-robot partnerships to avoid, mitigate, and overcome the inevitable accidents.

The first section of the paper is an overview of risk communication. It is a vast, interdisciplinary field that has studied a variety of domains including public health and medicine, disaster preparedness and response, and financial planning. The second section of the paper will explain why risk communication is critical for the robotics industry. The final section of the paper will outline an agenda to incorporate risk communication into the field of robotics.

Part 1: Overview of Risk Communication

Risk is the possibility of injury, damage, or a "negative impact on some endeavor."[9] Risk analysis is the study of risk. Risk assessment is about estimating the scale and probability of risk. This effort is conducted, as much as possible, on a scientific basis. Risk management, in turn, is using risk assessment to develop policies and make decisions to reduce risk. This may sound

---

[7] Noel Brewer, "Goals," Ibid, 4

[8] This topic has been researched extensively. For an overview see: Omri Ben-Shahar, "The Failure of Transparency," Testimony Before House Committee on Energy and Commerce, November 29, 2017, http://docs.house.gov/meetings/IF/IF17/20171129/106659/HHRG-115-IF17-Wstate-Ben-ShaharO-20171129.pdf
Shahar's work is built on the work of many others, including:
McDonald, Aleecia M., and Lorrie Faith Cranor. "The cost of reading privacy policies." ISJLP 4 (2008): 543.
Jensen, Carlos, Colin Potts, and Christian Jensen. "Privacy practices of Internet users: self-reports versus observed behavior." *International Journal of Human-Computer Studies* 63.1-2 (2005): 203-227.
Mcdonald, Aleecia M., et al. "A comparative study of online privacy policies and formats." *International Symposium on Privacy Enhancing Technologies Symposium*. Springer, Berlin, Heidelberg, 2009.

[9] Lionel Galway, "Quantitative Risk Analysis for Project Management: A Critical Review," RAND Working Paper, February 2004, https://www.rand.org/content/dam/rand/pubs/working_papers/2004/RAND_WR112.pdf

anodyne and relatively straightforward, but assessing risks can be an extremely complex technical issue that is both art and science. Risks may involve a range of social and economic factors beyond the science. Further, efforts to manage risk can involve a vast number of complex technological, economic, legal, political, and social issues. Risk communication then "is the two-way exchange of information, concerns, and preferences about risks between decision-makers and the public."[10]

Risk communication emerged in the mid-1980s as risk assessment and management matured as fields and its practitioners needed to gain public support for their policies. Environmental and public health were the initial areas of focus for risk communication, and the Environmental Protection Agency has been at the forefront of the field.[11] There is also an extensive literature of practice and theory on risk communication and disasters, as well as the closely linked field of crisis communication.[12] Besides a range of technical disciplines, the field of risk communication draws on a range of social sciences including psychology, decision science, anthropology, sociology, and communication.

The rest of this section will provide a brief overview of the practice of communicating risk, followed by an examination of risk perception and how this can limit the effectiveness of risk communication. This section will end with discussion of communication and how it can be used to foster trust, which underpins effective risk communication. This section cannot hope to provide more than a glimpse into the vast field of risk communication, nonetheless it will hopefully be sufficient to highlight the potential importance of risk communication to robotics.

---

[10] Paul R. Portnoy, "Forward," *Readings in Risk*, eds. Theodore Glickman and Michael Gough (New York: Resources for the Future 2004), xi

[11] Alonzo Plough and Sheldon Krimsky, "The Emergence of Risk Communications Studies: Social and Political Context," in *Readings in Risk*, 223-231

[12] For overviews see:
Sheppard, Ben, Melissa Janoske, and Brooke Liu. "Understanding Risk Communication Theory: A Guide for Emergency Managers and Communicators," Report to Human Factors/Behavioral Sciences Division, Science and Technology Directorate, U.S. Department of Homeland Security. College Park, MD: START, 2012 http://www.start.umd.edu/sites/default/files/files/publications/UnderstandingRiskCommunicationTheory.pdf
Janoske, Melissa, Brooke Liu, and Ben Sheppard. "Understanding Risk Communication Best Practices: A Guide for Emergency Managers and Communicators," Report to Human Factors/Behavioral Sciences Division, Science and Technology Directorate, U.S. Department of Homeland Security. College Park, MD: START, 2012 http://www.start.umd.edu/start/publications/UnderstandingRiskCommunicationBestPractices.pdf

Communicating Risk

In its simplest paradigm, risk communication is about ensuring that information about risk and the actions that can be taken to counter these risks are presented clearly. Risk communication can be efforts to inform, such as explaining the risks and benefits of various medical procedures so the patient can choose the option that meets their needs. Risk communication can be intended to change attitudes, such as outlining the environmental impacts of a new development. Most significantly, risk communication can be intended to change behavior: persuading people to buckle seat-belts, stop smoking, or prepare for highly probably natural disasters.[13]

To be effective and achieve any of these goals risk communication must:

- Include the information that users need
- Reach those who need it
- Be understood by the recipients[14]

Each of these tasks is complex and requires an understanding of the intended audiences (different groups will receive and process information differently), which must in turn be rooted in well-defined goals for the communication.

Effective risk communication includes a determination about what information to present. This is not a matter of concealing information, but rather ensuring that audiences are not overwhelmed with extraneous information and have the critical data needed to be properly informed to make a decision. Decision theory offers tools for identifying the most critical information to communicate such as value-of-information analysis, which prioritizes information based on how much each item enables the user to choose the best option.[15] Carrying out these kinds of analyses requires knowing what the intended audience values.[16] An aging community might not find

[13] Noel Brewer, "Goals," FDA 2011, 4-7
https://www.fda.gov/downloads/AboutFDA/ReportsManualsForms/Reports/UCM268069.pdf
[14] Baruch Fischhoff, "Duty to Inform," FDA 2011, 19
https://www.fda.gov/downloads/AboutFDA/ReportsManualsForms/Reports/UCM268069.pdf
[15] See: Clemen, R.T., & Reilly, T. *Making Hard Decisions* (Boston: Duxbury 2003)
Raiffa, H. (1968) *Decision analysis* (Belmont, MA: Addison-Wesley 1968)
vonWinterfeldt, D., & Edwards, W. *Decision Analysis and Behavioral Research* (New York: Cambridge University Press 1986)
[16] Fischhoff, "Definitions," FDA 2011, 41-43
https://www.fda.gov/downloads/AboutFDA/ReportsManualsForms/Reports/UCM268069.pdf

information about a proposed industrial development's impact on pregnancies useful. Retired patients and busy parents might have different priorities and need different information to evaluate medical options.

This process of identifying critical information for the audience has numerous pitfalls. Research indicates that people over-estimate their grasp of the perspectives of others, creating a large number of assumptions that can undermine effective communication. Experts may neglect to inform audiences about items that they consider basic information about their field or fail to mention risks or aspects of risks, believing them to be insignificant when the audience would in fact consider them an important factor. At the same time, risk communicators should not continue to discuss information that is clear and known to the target audience, because that will also reduce the audience's receptivity to additional information.[17]

Risk communicators also face the challenge of successfully disseminating necessary information. This can vary depending on the type of information being shared and the community that needs to be reached. In a public health emergency all available mechanisms would be deployed with an emphasis on speed – mailings might be too slow to be effective. To discuss the environmental impact of a planned development, meetings with community leaders might be sufficient. Risk communicators may partner with organizations that can reach constituencies on their behalf. For health issues, one on one discussions with medical personnel may be most appropriate. Most issues that require risk communication campaigns will affect many different communities, which will require multiple communications channels. Messages will have to be tailored to the recipients and the channel, but also not contradict one another.[18]

Once the critical information is identified, and the appropriate mechanisms for delivery are identified, it needs to be presented clearly and effectively. The potential issues for effective presentation are vast, ranging from efficient layout of print or internet publications, to the recognition that very subtle changes in word use can change perceptions of risk. Because defining the probabilities of potential risks is central to the field, there is an extensive literature of best practices for the specific questions around communicating probabilities to the general

[17] Fischhoff, "Duty to Inform," FDA 2011, 22
[18] Ibid 23-24

public. There are a vast range of best practices to surmount some of these problems. One of the basic best practices is to quantify risk with numbers rather than words. “Rare” or “unlikely” can have an ambiguous meanings, whereas 10% is much more specific.[19]

All numerical communication is not equal. In some cases audiences may suffer from innumeracy. Even when that is not the case, Fischhoff describes several ways in which general audiences tend to misinterpret numbers and methods to ensure that the public understands them:

- The choice of unit by which risk is expressed (economic losses in dollars versus time at work) reflect underlying values and should reflect the decision-maker’s preferences.
- Relative and absolute terms for risk (“one out of” vs “*x* times as likely”) are interpreted differently. Because they are incomplete, best practices recommend not using relative terms.
- People underestimate cumulative risk and often do not see the equivalence between rates and ratios. Fischhoff notes, “Communications should do the math for them.”[20]

Risk communication has risks of its own. Attempts to reduce concerns about risks that people were not previously worried about, may in fact lead to greater concern. In some cases efforts to reduce risk may increase risk. Concerned about potential computer failures due to Y2K, individuals purchased generators and fuel and in some cases firearms. An increase in the number of people keeping flammable material and guns in their homes created new risks. Similarly, after 9/11 significant numbers of people chose to drive rather than fly. In the United States, flying is significantly safer than driving, so the increase in drivers may have contributed to more accidents and deaths.[21]

Best practice calls for building an evaluation process to systematically analyze the results of risk communication efforts. This can include a vast range of research techniques from quantitative surveys, open-ended interviews, mental models, and usability testing of proposed materials to

---

[19] Baruch Fischhoff, “Definitions,” FDA 2011, 45
[20] Baruch Fischhoff, “Communicating about analysis,” in *Intelligence Analysis: Behavioral and Social Scientific Foundations*, Baruch Fischhoff and Cherie Chauvin, eds, (Washington, DC: National Academies Press, 2011), 237
[21] Donald MacGregor, “Public response to Y@K: social amplification and risk adaption: or, “how I learned to stop worrying and love Y@K,” eds. Nick Pidgeon, Roger Kasperson, Paul Slovic, *The Social Amplification of Risk* (Cambridge UK: Cambridge University Press 2010), http://www.macgregorbates.com/uploads/1/5/8/4/15840810/13-public_response_to_y2k_full_text.pdf

full-scale randomized controlled trials to measure outcomes. Much risk communication is developed based on intuitions, which, while powerful, may be misguided and need to be rigorously examined to ensure effectiveness.[22]

The audience for risk communication may not be the public. Decision-makers in the government, industry, and other large organizations must understand potential risks in making decisions and enacting policies. There is an extensive literature on applying the axioms of communications in the intelligence community to allow analysts to more effectively communicate their findings. In this situation the audience is usually very small, a single client or a small staff, and quite often a voracious consumer of information. Nonetheless, the same basic principles apply. Advisors to leaders must carefully consider what information to present and how best to organize it, enact a rigorous process of evaluating the effectiveness of the communications, and learn to meet the client's preferred methods of processing information.[23]

A risk communication program is a complex research project in which the participants need to do research on what they hope to achieve and how they might achieve it, followed by hypotheses about what methods will be effective to inform a community about a risk and/or change its behavior to reduce risk. Then the program needs to be implemented and evaluated to test whether or not the hypotheses were correct. It is a significant undertaking that requires a range of skills. It is a demanding process, but it is not the entirety of effective risk communication.

[22] Julie Downs, "Evaluation," FSA 2011, 12
https://www.fda.gov/downloads/AboutFDA/ReportsManualsForms/Reports/UCM268069.pdf
[23] Fischhoff, "Communicating about Analysis," NAS 2011, 227-247

Risk Perception

Familiarity with the technical risk analysis can breed contempt for those who don't share the same views of risk.

Donald MacGregor, PhD of MacGregor-Bates Applied Decision Concepts[24]

Traditional risk communication is a rational enterprise, with cost-benefit analysis (CBA) at its core. That is not, however, how people actually make decisions. The case of the Ford Pinto, described above, illustrates this gap. For the designers and engineers who oversaw the recall process, the risk of fire after a rear end collision was known, but not the most serious risk to the vehicle. It represented a tiny fraction of the scores of people who die in accidents daily. To the public the deaths in this accidents were particularly horrible and the reaction was visceral.

Paul Slovic, having done extensive research on risk perception, has found that there is vast gap between expert perception of risk, which focuses on probabilities and fatalities, and layperson perception of risk which is shaped by heuristics over reason. There are two major factors that shape people's risk perception: the extent to which the risk is unknown and the extent to which the risk is dreaded. The known/unknown factor includes whether or not a risk is observable and well-known to science. Examples of known and observed risks are home swimming pools or bicycles. While both of these things are objectively dangerous, they do not occupy tremendous space in risk perception. Genetic engineering, on the other hand, is a much less known risk and its consequences would not be immediate. The other factor, dread, distinguishes between risks that are controllable and limited as opposed to risks that are uncontrolled and viewed as potentially catastrophic. Nuclear war or terrorism are examples of catastrophic risks that inspire great senses of dread.

Incidents that trigger worries of catastrophic impact, particularly from poorly understood origins, even if the mishap is not particularly serious, can create what Slovic calls "signals" that have an impact well beyond the actual harm done. The impact can include indirect costs such as a change in the public perception of an industry or technology and lead to litigation and regulation. The

[24] Author conversation with Donald MacGregor PhD, February 14, 2018

Three Mile Island nuclear accident, which had no fatalities and minimal broader health impacts, nonetheless had an enormous effect on the nuclear energy industry in the United States. Increased public opposition along with regulation led to a dramatic decrease in the construction of nuclear plants and raised doubts among the general public about other complex technologies. Research shows that people will consider these risks when the benefits are great, but if the benefits are viewed as marginal, uncertain and dread risks will shape public consciousness. Nuclear power was seen as a great and dreaded risk, but few people saw significant direct benefits from it.[25]

The work of Slovic (who built on the work of many other pioneers of the study of risk perception)[26] draw on the work of Nobel Laureates Daniel Kahnemann and Amos Tversky, who found that much of human decision-making is not rational but made based on heuristics developed from individual experience and knowledge. There are a vast range of factors that can skew perceptions of risk. Personal experience, either directly or through an acquaintance, can skew estimations of that risk's probability. The media amplifies rare, but dramatic, risks so that they are seen as more likely and of greater concern. This dynamic works the opposite way as well. Most people drive regularly and do not have accidents, thus the probability of automobile accidents is usually under-estimated. In fairness to this heuristic approach, it has its virtues. The reality is that there can be a great deal of uncertainty and unconscious bias in risk assessment, particularly in new technology. Heuristics may capture risks that experts do not.[27]

Discussions ostensibly about risk may in fact be about something else altogether. How a group or society defines and evaluates risk can reflect deeply held cultural mores or important social relations. New technology can create issues of inequality and isolation, while threatening power and influence structures. These very real and much deeper risks are social and not easily

[25] Paul Slovic, "Perception of Risk," *Science*, April 17, 1987, 280-285
[26] This paper is not a history of risk analysis and communication, but it would be remiss in not mentioning Chauncy Starr author of the seminal, "Social Benefit versus Technological Risk," *Science*, September 19, 1969.
[27] Paul Slovic, "Perceived Risk, Trust, and Democracy," in Paul Slovic ed. *The Perception of Risk*, (London: Earthscan 2000), 316-326

quantifiable and hence are often not included in technical risk assessments.[28] Nonetheless, these issues require sensitivity and consideration.

Slovic concludes by noting that the typical CBA deployed by experts does a poor job of taking account of risk perception – focusing strictly on fatalities, injury, and property damage and downplaying fear and dread. He writes:

> Perhaps the most important message from this research is that there is wisdom as well as error in public attitudes and perceptions. Lay people sometimes lack certain information about hazards. However, their basic conceptualizations of risk is much richer than that of experts and reflects legitimate concerns that are typically omitted from expert risk assessments. As a result, risk communication and risk management efforts are destined to fail unless they are structured as a two-way process. Each side, expert and public, has something valid to contribute. Each side must respect the insights and intelligence of the other.[29]

Communicating Trust

Literature on risk communication emphasizes that it is, ideally a two-way process, not simply an effort by risk communicators to clarify risks so that communities are more willing to accept them. Beyond crafting messages that specific communities can grasp, risk communicators must show compassion and interest in the community and its concerns. Critically, risk communication is an opportunity to elicit how the community perceives risks and identify what the community views as risks, as opposed to what the risk assessment models indicate. An effective risk communication program engages the public as a partner and gives them a role in the decision-making process.[30]

Research on risk communication has consistently emphasizes the importance of trust between those communicating risks and the intended audience. A high-level of trust in the communicator

---

[28] Mary Douglas and Aaron Wildavsky, *Risk and Culture: An Essay on the Selection of Technical and Environmental Dangers* (Berkeley: University of California 1982)
[29] Paul Slovic, "Perception of Risk," *Science*, April 17, 1987, 284-285
[30] Covello, V., Sandman, P. & Slovic, P. "Guidelines for Communicating Information About Chemical Risks Effectively and Responsibly," in Mayo, D.G. & Hollander, R.D., eds. *Acceptable Evidence: Science and Values in Risk Management* (Oxford: Oxford University Press 1991), 66–90

and in the organization the communicator represents can reduce a community's anxiety, increase their willingness both to accept risks and consider benefits. Low levels of trust will have the opposite effect. Inconsistency between statements and actions by people and organizations will destroy trust. Building trust requires taking the intended audiences concerns about risk seriously and providing them truthful and reliable information.[31] Skepticism and tough questions however, do not necessarily imply a lack of trust. Building the needed trust is a difficult and time-consuming process. Destroying it can happen very quickly and distrust creates a self-reinforcing loop in which information that fosters distrust is sought out and quickly assimilated, while exposure to those with alternative points of view becomes less frequent.[32]

Slovic illustrates the importance of trust by highlighting that the public is extremely skeptical and distrustful of the nuclear power and chemical industries, but regularly accepts doses of chemicals and radiations as part of medical treatment. In the case of the former the risks are seen as high and the benefits are low, in the case of medical treatments the benefits are perceived as high and the risks as low. This reflects, fundamentally, that medical professionals are trusted and viewed as having the patient's and communities best interests at heart. Leaders from the chemical and nuclear power industry, on the other hand, are not trusted.[33] This may be a cautionary tale for the robotics industry.

[31] For an overview of research on trust and risk communication see Janoske, Melissa, Brooke Liu, and Ben Sheppard. "Understanding Risk Communication Best Practices: A Guide for Emergency Managers and Communicators," 4-6
[32] Slovic, "Perceived Risk, Trust, and Democracy," 319-320
[33] Ibid, 316

Part 2: Risk Communication & Robotics

There are known knowns. These are things that we know we know. There are known unknowns. That is to say, there are things that we know we don't know. But there are also unknown unknowns. There are things we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tend to be the difficult ones.

Secretary of Defense Donald Rumsfeld, February 12, 2002

The Ford Pinto discussed at the beginning of the paper, a car manufactured in the 1970s, had limited and reasonably well understood dimensions of failure. It could suffer a failure from its design or parts and there was also the potential for human error and unsafe driving conditions. There were known knowns and known unknowns. Nonetheless, a failure, which was known to the manufacturer but judged less significant than other potential points of failure led to a public outcry. Autonomous systems will have a vast range of potential points of failure that, at least to some extent, can be quantified. But autonomous systems will also create potential unknown unknowns, both through failures and acting in unpredicted ways, but also in their interactions with people and society.

Unknown and difficult to predict risks over which people have little control are the ingredients for a "signal" event that triggers a backlash against robots.

This section begins by describing the kinds of risks robots may present, followed by a discussion of the role risk communication programs can play in mitigating these risks by reducing the scope of unknown unknowns while the building resiliency and risk tolerance to manage them.

Robotics Risk Assessment

Table 1 provides a taxonomy of potential robot crises incorporating types of robot behavior (or misbehavior) and the dimensions of the crises (how many people does it affect or how disturbing are the incidents.)[34] While it is interesting to speculate about potential robot risks and crisis, it is

[34] This paper is not seriously considering the possibility of artificial intelligence revolting against humankind in the foreseeable future. Table 1 was originally published in Aaron Mannes, "Anticipating Autonomy: Institutions &

likely that actual incidents will not be predicted or even predictable. Managing these incidents will entail solving complex technical problem but also reassuring the public. A weak response could trigger a "signal" event that leads to a backlash either in public perception about autonomous systems or in the regulatory environment.

TABLE 1: A TAXONOMY OF ROBOT CRISES

Crisis Type / Crisis Dimensions / Crisis Description	**Quantity** How many people are affected	**Ubiquity** How common are the devices or systems	**Strangeness** How odd is the crisis	**Violence** Are people hurt or killed
Malfunction Bugs and defects	A smart medical device improperly measures a user's condition and delivers a dangerous dose of medication. This would be a malfunction with low **quantity** but high potentially high **ubiquity.** The **strangeness** factor would be moderate and if the person were injured the **violence** level would be high.			
Misfunction Robots act in unexpected and upsetting ways	Autonomous cars interact with one another to create unusual traffic patterns that disturb motorists and pedestrians. This would have high **quantity** and **ubiquity** (since most people use cars) and high **strangeness.** Assuming no one was hurt, the **violence** would be low.			
Dysfunction Robots acting as expected distress people	Home health service robots monitoring the ill and elderly accompany people into restrooms and other personal situations in order to ensure their safety. This could have moderate **quantity** and **ubiquity** (depending on how many people were using these devices), high **strangeness**, but probably low **violence**.			
Mis-Use Robots used in harmful manner	Terrorists use autonomous drones to carry out multiple simultaneous attacks would be a case in which the incidence has high **quantity** but perhaps only moderate **ubiquity** and **strangeness**. The violence level would be **high**.			

Politics of Robot Governance," presented at the American Political Science Association Annual Conference, September 2, 2016, https://ssrn.com/abstract=3025193

In some ways, Table 1 is incomplete because it does not express the number of potential vectors of failure. Besides its autonomous processing capability, a robot is also a set of physical parts and software – all of which can experience failures. Further, as computational systems, robots can be cyber-security risks. As collections of sensors, robots can create privacy risks. Many autonomous systems will rely on large data-sets to learn models, this could create the kinds of risks related to big data. Human factors represent a very large set of risks in using autonomous systems, as people may act in surprising and unpredictable ways interacting with the autonomous system.

This raises several potential issues. An autonomous vehicle is still a car, and could experience faulty brakes or any of the myriad software and hardware defects. Similarly, a defective sensor could lead to minor errors in data collection, or a flaw in the data curation scheme, could skew the autonomous system's decision-making. Strictly speaking, these problems would not reflect a failure of the autonomy component of the system, but the public might not make this distinction.

These issues may interact with one another in complex and unpredictable ways. A cyber-security breach could be misinterpreted by the autonomous system leading to unexpected mishaps. Unexpected activity by people recorded by an autonomous system might not be recognized as sensitive and released in a way that compromises their privacy.

Threats to physical safety or property may be the top priorities, it is important to remember that dignity is a critical human value – something people treasure.[35] If autonomous systems are operating safely, but in a manner that undermines an individual's or community's dignity it may trigger public reactions in a way that mundane accidents do not. Dignity touches on social relationships and mores, the areas that technical risk assessment addresses least effectively, but also areas that can trigger very deep reactions.

The sense of dread, with its attendant feelings of helplessness, is exacerbated by a "deliberate and calculated intention" to do harm.[36] This underpins why terrorism looms large as a perceived

[35] On a personal note, having studied terrorism for the past two decades, a sense of humiliation is a common theme among those who turn to terrorism all over the world.

[36]Norman A. Milgram, "An Attributional Analysis of War-Related Stress: Modes of Coping and Helping," in Norman A. Milgram, ed., *Stress and Coping in Time of War: Generalizations from the Israeli Experience* (New York: Brunner/Mazel Publishers, 1986), 11-12.

risk when its likelihood is (at least in the United States) far lower than more mundane risks such as traffic accidents.[37] Robots could be used by terrorists and criminals, but also, because of their autonomy, they could be perceived as having agency and of acting with malicious intent. This is significant because, unlike in a natural disaster, when faced with an accident caused with intent the public may seek to strike back in an effort gain a sense of control.[38]

Finally, robots may not only be involved in small accidents, they may also trigger large-scale breakdowns. An autonomous vehicle may be involved in an accident with another car, pedestrian, or bicyclist. Alternately, autonomous vehicles interacting on a large scale with other autonomous systems in an urban transportation network might accidentally experience a large-scale city-wide gridlock. As robots become increasingly common and are given increasingly critical roles, this possibility for large-scale failures also needs to be considered.

Robots represent an enormous range of potential risks, some will be known unknowns, but many of will be difficult to identify beforehand will be unknown unknowns. The unique characteristics of autonomous systems will exacerbate perceived risks. A failure to address this reality will have significant consequences for the industry when, inevitably, some of these risks manifest themselves and become crises.

Risk Communication and Protecting Autonomy

The large-scale adoption of autonomous systems requires that people trust the system (just as they must trust the risk communicator) to do what is expected and appropriate.[39] This paper takes it as an article of faith that autonomous systems will on occasion fail in this regard and that risk communication will play an important role in maintaining broader public trust that autonomous

[37] Again reaching to my past work studying terrorism, I often found myself frustrated that a relatively rare risk received such extensive attention from policy-makers, the media, and academia. Reading research on risk perception has been extremely useful and enlightening.

[38] Joshua Pollack and Jason Wood, "Enhancing Public Resilience to Mass-Casualty WMD Terrorism in the United States: Definitions, Challenges, and Recommendations," Defense Threat Reduction Agency, June 2010, 3 https://fas.org/irp/agency/dod/dtra/resilience.pdf

[39] Andrew R. Lacher, Robert Grabowski, Stephen Cook, "Autonomy, Trust, and Transportation," *The Intersection of Robust Intelligence and Trust in Autonomous Systems: Papers from the AAAI Spring Symposium*, 2014, https://www.aaai.org/ocs/index.php/SSS/SSS14/paper/viewFile/7701/7728

systems and their builders are worthy of trust. Otherwise, failures may become "signal" events such as those that struck the nuclear power industry.

As discussed in the overview of the field of risk communication, by explaining risks and benefits of robotics clearly, individuals can have deeper knowledge of the system and better manage the risks through interacting with the robots in a safe manner. Understanding the robots will reduce feelings of dread around them.

The role of risk communication is much deeper than this however. Communication is not simply a broadcast of safety information, it is a two-way process. It will allow organizations producing and deploying autonomous systems to glean community concerns. These concerns, if ignored (or if the organization is simply unaware of them), could spark backlash and frustration. Learning of these issues beforehand can enable accommodations to them. The recent incident in which the San Francisco ASPCA was subject to public outrage after it deployed a security robot is a cautionary tale. Besides the awful publicity resulting from the security robot appearing to target the homeless, the issue tapped into deeper social issues facing the city.[40] It is easy to imagine this type of conflict occurring in many contexts as autonomous systems are deployed with increasing frequency and an increasing variety of roles.

Most significantly, the risk communication process, at least has the potential to build trust for the inevitable unknown unknowns. If the communicator is trusted, hears and responds to concerns, and replies honestly and consistently to them, communities will be more inclined to accept risks, tolerate some failures, and be open to potential benefits.

Risk communication creates partnerships between those making and deploying robots and those interacting with them. In these partnerships all of the stakeholders will have a voice and a role in decision-making. Because robots have autonomy, they will be perceived as impinging on human control. As discussed above, risks are perceived as greater when people do not feel that they have control. An effective risk communication program should help protect and even expand people's autonomy, thereby fostering productive partnerships between humans and robots.

---

[40] David Morris, "San Francisco Security Robot Fired After Public Outcry," *Fortune*, December 16, 2017, http://fortune.com/2017/12/16/san-francisco-security-robot-fired-after-public-outcry/

Part 3: Agenda for Risk Communication and Robotics

Sometimes companies won't spend $100,000 to stop something impeding a $100 million project.

> Donald MacGregor, PhD of MacGregor-Bates Applied Decision Concepts[41]

Given the vast potential for autonomous systems to perform in unexpected ways and the expectation that the public will take on some risk in adopting this technology, it is urgent that the field establish a risk communication plan sooner, rather than later. The good news on this front is that risk communication is a well-established field. There are academics who study it, companies that provide risk communication services, and an extensive literature about how to conduct risk communication campaigns. The bad news is that it is not an exact science. Risk communication involves having difficult discussions, sometimes with dissatisfied and frustrated stakeholders. Successful risk communication relies on trust, which is hard to build and easy to destroy.

This author is under no illusions about the challenges of organizational change. Businesses exist to make money and, in the short term, risk communication is a cost – not a revenue source for robotics manufacturers. Organizations deploying robotics will likely do so because they believe the autonomous systems will make it easier to accomplish their mission and not be inclined to develop new risk communication capabilities. While it is important to develop risk communication capabilities in order to build trust for accidents and mishaps, this endeavor is not only preparation for an emergency. It will also contribute to better robots and relations with customers, users, and other stakeholders. The in-depth engagement required for risk communication will yield insights about concerns, needs, and fears, enabling the companies and organizations that produce and deploy robots to better partner with the public. Thus this section is written in a spirit of hope that the case for risk communication for robotics having been made it is appropriate to outline some steps to bring it into being.

The first stage in this plan would be surveying the various existing fields of risk communication and adapting their best practices for robotics. Since robotics is a diverse field there will not be a one-size fits all approach. Different types of robots will require different terms and framing.

---

[41] Author conversation with Donald MacGregor PhD, March 6, 2018

Domain knowledge of the specific sector in which the system was deployed would need to be blended with technical expertise in robotics. There may be some commonality in addressing the fundamental issues surrounding autonomy, but concerns and sense of risk will vary tremendously between for example, autonomous vehicles, law enforcement robots, smart homes, and specialized medical robots. From this, spadework, the technology industry can establish and support a research agenda, perhaps bringing together the fields of risk communication and Human Robot Interaction. Researchers would need to identify effective methods of communicating the new kinds of risks robots represent, and how to gauge public feelings about them.

Research should be supported by practice. Businesses in the robotics industry (as manufacturers, distributors, or users) would be wise to engage risk communication professionals who can help companies obtain an objective picture of how the public perceives risks of an autonomous system. Besides crafting appropriate and informative materials, it will require identifying key stakeholder communities and developing relationships with them. Once these relationships are established, concerns about risk can be heard and addressed. This process is not a quick one, but it will build trust, and should be done proactively – as systems are developed.

Technology creates a range of communications platforms and feedback mechanisms, like social media and online videos, to disseminate information and receive concerns. Depending on the situation, the robot itself could be a source of information and channel to register concerns. However, in using these channels, companies should apply best practices and rigorously test and evaluate the content to ensure it is effective and the dissemination means to ensure that the appropriate stakeholders are being reached.

In short, robotics companies must embrace risk communication as part of their culture. Market research should include gathering risk perceptions. Designers and builders should incorporate risk perception into their work. Sales teams, while obviously focused on revenue generation, should discuss risks and how autonomous systems fit with the needs of the users. Customer service departments can be trained as frontline risk communicators. If corporate leadership views risk communication as a critical process of building relationships and trust, this attitude will pervade the company.

Enacting a serious risk communication program will be a significant endeavor, and may strain the finances of start-ups. Because effective risk communication should be an industry-wide concern, programs could be fostered through industry associations that could then provide risk communication support and training to smaller companies.

Ideally, risk communication will be part of the customer support manufacturers offer their clients. The manufacturer of law enforcement robots, for example, would work with police departments to engage with communities to ensure that the law enforcement robots were deployed in a manner that the public was comfortable with and met the community's needs.

Government agencies that regulate and deploy robots will also need to grow their risk communication capabilities. Some government agencies, such as the Food and Drug Administration and the EPA have been leaders in the field. Others agencies, particularly smaller agencies at the state and local level, may have very limited capabilities. Even those agencies that have long practiced risk communication will face a learning curve addressing the concerns raised by autonomous systems. If industry embraces risk communication as a critical program, government agencies can collaborate, and also benefit from the general research they have sponsored.

Baruch Fischhoff, a professor at Carnegie-Mellon and a leading scholar or communication and risk communication writes:

> Communication is sometimes seen as a tactical step, transmitting results to clients. However, unless communication also plays a strategic role, those analyses may be off target and incompletely used. If an analytical organization makes a strategic commitment to communication, behavioral science can help with its execution, overcoming some of the flawed intuitions that can lead people to exaggerate how well they understand one another.[42]

This is perhaps the critical point of this paper. Communication in general, and risk communication in particular, should be not be an afterthought, but rather should be a central

[42] Baruch Fischhoff, "Communicating about analysis," in *Intelligence Analysis: Behavioral and Social Scientific Foundations*, Baruch Fischhoff and Cherie Chauvin, eds, (Washington, DC: National Academies Press, 2011), 245

component for many endeavors – particularly in emerging technologies such as robotics. The institutional challenge is how to create the incentives to enable this approach for the robotics industry. Organizational change requires changing incentives. One mechanism is legislation. Companies could be required to make these investments in risk communication, but under this regime risk communication could become a compliance process rather than an organizational value. At this point, the better – although not necessarily likely – approach is to point to the lessons of history and persuade technology leaders that with foresight they can break this cycle of signal events which are followed by regulatory, legal, and public backlash.

A Concluding *Cri de Coeur*

Instead of spending all day worrying, why don't you wait until there's a near miss… Let's not translate that worry into premature constraints on the innovators…"

> Eric Schmidt, former CEO of Google, speaking at MIT on regulating AI.[43]

Schmidt's concern about ill-considered regulation hampering the development of AI, which is closely linked to robotics, is understandable. As discussed above, the nuclear power and automotive industries, among others, would sympathize. But it is hopefully evident to any reader of this paper that Schmidt's remark was insensitive risk communication. Schmidt dismisses public concerns about risks while demanding that the public assume these risks on the basis of the industry's expertise. Given Schmidt's standing in the technology sector, this worldview is probably not an outlier. It assumes a vast reservoir of trust that the industry can draw upon when the inevitable failures occur. This reservoir may be far shallower than the industry realizes. It may not exist.

It is in this circumstance that the near miss Schmidt describes may be become, like Three Mile Island, a signal event that leads to broad public and regulatory backlash. (It is also not certain that the failure will be a near miss, it could in fact result in a tragedy.) To avoid this eventuality, industry, government, and any organization that deploys robots to carry out its mission, would be

[43] Quoted by Jack Clark of *ImportAI* and @jackclarkSF, March 1, 2018

wise to invest in risk communication programs. Carried out effectively, risk communication can identify sources of public anxiety and reduce them. Most importantly, risk communication can help create a relationship based on trust.

Risk communication are not a panacea, they are expensive ongoing efforts that will require commitment and engagement from organization and industry leaders. Nonetheless it is a prudent investment. The public will increasingly be asked to trust robots, but they are unlikely to do so if they do not trust their makers.

Coda

Risk communication, and communication more broadly, is an inherently interdisciplinary field that brings together qualitative and quantitative research across a range of areas from anthropology and psychology to neuroscience. It can wax philosophical and be extremely practical. Its practice is a blend of art and science. For all of the reasons discussed above, the robotics field would be wise to invest in risk communication now, when the field is still emerging. It is also a useful bridge for a broader issue. There has been extensive discussion about bringing the humanities and social sciences together with engineering and hard sciences.[44] An important part of this type of collaboration is for individuals and organizations to be smart consumers of other fields. Right now, engineers and roboticists, by training might not be particularly well-informed consumers of social science and will thus either ignore their insights or not make effective use of them. Besides the direct benefits of a robotics risk communication field, this endeavor might also serve as a bridge between disciplines and enable more effective and fruitful collaboration.

[44] Kate Crawford and Ryan Calo, "There is a blind spot in AI research," *Nature*, October 13, 2016, http://www.nature.com/news/there-is-a-blind-spot-in-ai-research-1.20805.