

Identifying Factors that Influence Trust in Automated Cars and Medical Diagnosis Systems

Michelle S. Carlson

The MITRE Corporation, 202 Burlington Road, Bedford, MA 01730
ohm@mitre.org

Munjal Desai

Google Inc., 1600 Amphitheater Parkway, Mountain View, CA 94043
munjaldesai@google.com

Jill L. Drury

The MITRE Corporation, 202 Burlington Road, Bedford, MA 01730
jldrury@mitre.org

Hyangshim Kwak

United States Military Academy, West Point, NY 10997
hyangshim.kwak@usma.edu

Holly A. Yanco

The MITRE Corporation, 202 Burlington Road, Bedford, MA 01730
holly@cs.uml.edu

Abstract

Our research goals are to understand and model the factors that affect trust in automation across a variety of application domains. For the initial surveys described in this paper, we selected two domains: automotive and medical. Specifically, we focused on driverless cars (e.g., Google Cars) and automated medical diagnoses (e.g., IBM's Watson). There were two dimensions for each survey: the safety criticality of the situation in which the system was being used and name-brand recognizability. We designed the surveys and administered them electronically, using Survey Monkey and Amazon's Mechanical Turk. We then performed statistical analyses of the survey results to discover common factors across the domains, domain-specific factors, and implications of safety criticality and brand recognizability on trust factors. We found commonalities as well as dissimilarities in factors between the two domains, suggesting the possibility of creating a core model of trust that could be modified for individual domains. The results of our research will allow for the creation of design guidelines for autonomous systems that will be better accepted and used by target populations.

1. Introduction

The past two decades have seen an increase in the number of robot systems in use, and the trend is still continuing. According to a survey, 2.2 million domestic service robots were sold in 2010 and the number is expected to rise to 14.4 million by 2014 (IFR 2011). Not only is the number of robots in use increasing, but the number of application domains that utilize robots is also increasing. For example, self-driving cars have been successfully tested on US roads and have driven over 300,000 miles autonomously (e.g., Thrun 2011, Dellaert and Thorpe 1998). Telepresence robots in the medical industry constitute another example of a new application domain for robots (e.g., Michaud et al. 2007, Tsui et al. 2011).

There is also a push to introduce or add additional autonomous capabilities for these robots. For example, the Foster-Miller (now QinetiQ) TALON robots used in the military are now capable of navigating to a specified destination using GPS. The unmanned aerial vehicles (UAVs) deployed by the military are also becoming more autonomous (Lin 2008); the Global Hawk UAV, for

example, completes military missions with little human supervision (Ostwald and Hershey 2007).

Robots are not the only examples of automated systems. For example, IBM's intelligent agent Watson is now being used as an aid for medical diagnosis (Strickland 2013). Additionally, many of the trading decisions in the stock and commodities markets are being made by automated systems. Automation has been in use for decades as autopilot systems in airplanes and as assistants for running factories and power plants.

As autonomous systems become more commonly used, it is necessary to understand how people will trust these systems. Without an appropriate level of trust or distrust, depending upon the circumstances, people may refuse to use the technology or may misuse it (Parasuraman and Riley 1997). When people have too little trust, they are less likely to take full advantage of the capabilities of the system. If people trust systems too much, such as when challenging environmental conditions cause the systems to operate at the edge of their capabilities, users are unlikely to monitor them to the degree necessary and therefore may miss occasions when they need to take corrective actions.

Thus it is important to understand trust prior to designing these increasingly capable autonomous systems. Without understanding the factors that influence trust, it is difficult to provide guidance to developers of autonomous systems or to organizations who are commissioning their development. In contrast, a knowledge of the way particular factors influence trust can allow a system to be designed to provide additional information when needed to increase or maintain the trust of the system's user in order to ensure the correct usage of the system.

We have seen that trust of autonomous systems is based on a large number of factors (Desai et al. 2012). In our prior work (Desai et al. 2013, Desai et al. 2012), we have found that the mobile robotics domain introduces some different trust-related factors than have been found in the industrial automation domain. There is some overlap, however; a subset of trust factors appear in both domains. Given our prior results, we believe that there is a core set of factors across all types of autonomous system domains which has yet to be codified. Further, it may be necessary to identify factors specific to each application domain.

Our ultimate goal is to understand the factors that affect trust in automation across a variety of application domains. Once we have identified the factors, our objective is to develop a core model of trust. We have embarked on this project by identifying the factors that most influence people's trust in two domains: automated cars and medical diagnosis systems. Our goal was to determine the factors influencing trust for these domains and compare them to determine the degrees of overlap and dissimilarity. This paper describes our research methodology and findings from this initial phase of this research.

2. Background

Utilizing autonomous capabilities can provide benefits such as reduced time to complete a task, reduced workload for people using the system, and a reduction in the cost of operation. However, existing research in the field in plant, industrial, and aviation automation highlights the need to exercise caution while designing autonomous systems, including robots. Research in human-automation interaction (HAI) shows that an operator's trust of the autonomous system is crucial to its use, disuse, or abuse (Parasuraman and Riley 1997).

There can be different motivations to add autonomous capabilities; however, the overall goal is to achieve improved efficiency by reducing time, reducing financial costs, lowering risk, etc. For example, one of the goals of the autonomous car is to reduce the potential of an accident (Guizzo 2011). A similar set of reasons was a motivating factor to add autonomous capabilities to plants, planes, industrial manufacturing, etc. However, the end results of adding autonomous capabilities were not always as expected. There have been several incidents in HAI that have resulted from an inappropriate use of automation (Sarter et al. 1997). Apart from such incidents, research in HAI also shows that adding autonomous capabilities does not always provide an increase in efficiency. The problem stems from the fact that, when systems or subsystems become autonomous, the operators that were formerly responsible for manually controlling those systems are relegated to the position of supervisors. Hence, such systems are often called supervisory control systems.

In supervisory control systems, the operators perform the duty of monitoring and typically only take over control when the autonomous system fails or encounters a situation that it is not designed to handle. A supervisory role leads to two key problems: loss of skill over time (Boehm-Davis et al. 1983) and the loss of vigilance over time in a monitoring capacity (Endsley and Kiris 1995, Parasuraman 1986). Due to these two reasons, when operators are forced to take over manual control, they might not be able to successfully control the system.

As such systems are developed, it is important to understand how people's attitudes about the technology will influence its adoption and correct usage. A key factor shaping people's attitudes towards autonomous systems is their trust of the system; hence, we are striving to learn the factors that influence trust, whether for all autonomous systems or for particular domains.

3. Methodology

For the research reported in this paper, we identified two automation domains to study: automotive automation (e.g., self-driving cars), and the use of IBM's Watson in the

medical sector. We chose these two domains for several reasons. The successful completion of over 300,000 miles by Google's driverless car, as well as the rulings in three states and the District of Columbia legalizing the use of driverless cars (Clark 2013), holds much promise for these cars becoming commonplace in the near future. Watson, a question-answering agent capable of referencing and considering millions of stored medical journal articles, is also promising. Little research has been conducted about the public's opinion on IBM's Watson, so the relationship between humans and medical diagnosis agents is uncharted territory.

We felt that the general public could be expected in the future to interact with both automated cars and Watson (in conjunction with their physicians). Thus, we developed computer-based survey instruments that could be administered over the internet to a wide audience. The surveys resided in Survey Monkey and were accessed via Amazon's Mechanical Turk so that respondents could be paid for their time. The surveys were administered in two rounds, with the first round being exploratory. After making improvements to the surveys, including the addition of factors identified by the initial participants in the "other" category, we released the second round, the results of which are reported in this paper.

Each round of surveys consisted of eight different variations: four for each domain. All of the surveys began with the same demographics questions, including gender, age, computer usage, video game playing, and tendencies to take risks. Then, each survey variant had a unique scenario designed to capture differences in public opinions depending on the seriousness of the situation ("safety critical" versus "non-safety critical") and the brand of the automated machine (well-known brand from a large organization versus a brand from an unknown startup). Thus there are four variations for each domain: safety critical and well-known brand ("branded"); safety critical and unknown brand ("non-branded"); non-safety critical and branded; and non-safety critical and unbranded.

In the automotive safety critical scenarios, the environment was described as high-speed, with lots of traffic. In the non-safety critical scenarios, the environment was described as low-speed with little traffic. While one might argue that all driving is safety critical, clearly it is more difficult to ensure safe travels at higher speeds and with more traffic. It is also more difficult to imagine oneself taking over control from such an autonomous system at high speeds in difficult driving conditions.

In the medical safety critical scenarios, the task is to determine diagnoses and treatments of three possible types of cancer. In the non-safety critical scenario, the respondent is given ample information to be certain that the affliction is not life threatening. The three possible afflictions in the non-safety critical scenario include

mononucleosis, influenza, or the common cold. Cancer in denotes a greater level of importance and urgency whereas the latter situation seems less dire.

In addition to the severity of the situation, we wanted to see whether the brand of the automated machine affected people's trust level as well. For the automotive domain, we explicitly described the automated system as being a Google Car for the two branded surveys. For the medical domain, we specified that Watson was a product of IBM in two survey variants. In the remaining survey variants, we did not label the automated machine as either IBM's Watson or the Google Car; instead, we said that the machines were developed by a small, startup company. In this way, we hoped to identify the extent to which the reputation of the company influences trust in an automated machine system in these domains.

Each survey in the automotive domain presented a list of 29 factors that could influence a person's trust of an automated system; surveys in the medical domain presented 30 factors. This list of factors was determined initially from a literature search, including the factors from Desai (2012) discussed below in the results section. We started with a shorter list in the initial design of our surveys; we released each of these initial surveys to small sample sizes (25 per survey; 100 in each domain, for a total of 200). Based upon these preliminary results, we added some additional factors, which were identified by respondents in a free-text "other" field. This process resulted in the full list factors for each domain used in the second version of the surveys, some of which were specific to the particular automation domain and others that were common to the two. The results presented in this paper are from the second version of the surveys, with 100 respondents for each of the eight survey variants.

The surveys also included three test questions used to ensure that respondents were actually reading the survey and answering to the best of their ability: "this sentence has seven words in it," "most dogs have four legs," and the influence of the color of one's shirt on their trust of an autonomous system. If a respondent answered one or more of these test questions incorrectly, they were not paid for the survey and the data was removed from the dataset.

We created each survey on Survey Monkey and utilized Amazon Mechanical Turk to disseminate them to the public. We narrowed our pool to residents from the United States and Canada with the minimum age of 18. We paid each respondent \$0.90.

4. Results and Discussion

We released 100 HITs (Human Intelligence Tasks) on Mechanical Turk for each of the versions of our surveys. Each domain's survey had 83 questions, similar except for

the wording that pertained to branding/not and safety critical/not. After discarding responses that had one or more of the test questions answered incorrectly, we had 382 responses in the medical diagnosis domain (231 male, 151 female; mean age 31.1 (9.0)) and 355 in the car domain (191 male, 164 female; mean age 35.6 (12.6)).

As mentioned in the methodology section, some of the responses were rejected if a respondent gave an incorrect answer to one or more of the three test questions in the survey. For the medical domain, we had 91 valid responses for the branded and safety critical version, 100 for branded and not safety critical, 97 for non-branded and safety critical, and 93 for non-branded and not safety critical. For the automotive domain, we had 90 valid responses for the branded and safety critical version, 92 for branded and not safety critical, 82 for non-branded and safety critical, and 91 for non-branded and not safety critical. The gender and age demographics were not significantly different between the survey versions for each domain and are not reported here for length considerations.

For each of the trust factors, respondents were asked to rank how the factor would influence their trust in the system on a 7 point Likert scale, with 1 meaning “strongly disagree” and 7 meaning “strongly agree.” The results for the trust factors were aggregated for the automotive domain and for the medical domain. In Tables 1 and 2, we present the list of factors sorted on the mean score from the Likert scale; while a Likert scale is not a continuous scale and averaging the responses is not strictly correct, it does allow us to see which factors have greater influence on trust across the respondents. Due to this limitation of a Likert scale, we discuss our results in terms of the top, middle and bottom thirds, rather than a strict ordering based upon the mean. We will be issuing follow-up surveys that will ask respondents to rank the factors (see the discussion of future work below).

In both domains, the ability of a system to stay up-to-date, statistics about its past performance, and the extent of the research on the system’s reliability are important factors for influencing trust in the system, appearing in the top third in both domains. In the middle third, both domains included the person’s own past experience with the system, the reputation of the system, the effectiveness of the system’s training and prior learning, and observing a system failure. These common factors could form the basis of a model of trust for automated systems; of course, we need to expand our work to many other domains in order to discover the true core.

In the bottom third, both domains include the system’s possibility of being hacked, the system’s user-friendliness, its ability to communicate effectively, the popularity of the system, and the aesthetics of the system. These factors are being judged as unimportant to trust by respondents in both domains. However, there may be some domains where

issues related more to user interface and the usability of the system could come into play. For example, in a social robot domain such as companion robots for the elderly, the way the system looks could have a greater influence on the user’s trust of the system: a pet-like robot covered in fur might be more trusted than a more machine-like system showing metal and wires, for example.

We found that there are domain specific factors present in the top third of the list. For the medical domain, respondents ranked the accuracy of the diagnosis, verification of the diagnosis, and the doctor’s ability to use the machine in the top third. In the automotive domain, reliability also ranked in the top third through several of the factors. In our survey design, we elected to have a number of questions about reliability to determine if there were different aspects of reliability. While we did see some differences, the list of factors could be reduced by using reliability in place of this group of factors; we will do this when we move to the next phase where we ask respondents to rank trust factors in order of importance.

Of note is where the responsibility of system verification and understanding lies between the two domains. In the top third of the factors in the medical domain, we see that people are looking to the doctor to mediate the results of the automated system. However, respondents are relying more on themselves in the automotive domain. This responsibility can be demonstrated by the fact that “your understanding of the way the [system] works” ranks in the top third for the automotive domain, but in the bottom third for the medical domain. Models of trust for automated systems will need to take into account whether the system is used directly by an end-user or whether it is utilized by a mediator or the end-user. Other such domains might include automated stock trading systems.

Desai (2012) also utilized Amazon’s Mechanical Turk to determine factors that would influence human-robot interaction for novice robot users. To obtain these results, Desai created a series of videos showing robots moving in a hallway environment, which were watched by the survey respondents. Test questions included the color of the robot shown in the video to ensure that the video had been watched. There were 386 valid responses received.

Desai (2012) reports the top six factors that influence trust of a robot system are reliability, predictability, trust in engineers that designed the robot, technical capabilities of the robot, system failure (e.g., failing sensors, lights, etc.), and risk involved in the operation. The factors in the middle third were error by automation, reward involved in the operation, interface used to control the robot, lag (delay between sending commands and the robot responding to them), and stress. The factors in the bottom third were training, situation awareness (knowing what is happening around the robot), past experience with the robot, size of the robot, and speed of the robot.

Automotive Domain					
	Rank	Ref	Influence Factor	Mean	Std dev
Top Third	1	A	Statistics of the car's past performance	5.98	1.32
	2	B	Extent of research on the car's reliability	5.87	1.39
	3		My own research on the car	5.82	1.33
	4		Existence of error/problem indicators	5.79	1.49
	5		Possibility that the hardware or software may fail	5.69	1.70
	6		Credibility of engineers who designed the car	5.69	1.53
	7	C	The car's ability to stay up-to-date	5.64	1.53
	8		Technical capabilities of the car	5.55	1.55
	9		Your understanding of the way the car works	5.54	1.48
	10		Your past experience with the car	5.53	1.66
Middle Third	11	D	Reputation of the car	5.49	1.57
	12		Level of accuracy of the car's routes	5.49	1.48
	13		Amount of current roadway information available to the car (e.g., weather, traffic, construction, etc.)	5.43	1.65
	14	E	Effectiveness of the car's training and prior learning	5.41	1.63
	15		Amount of information that the car can access	5.41	1.64
	16	F	Observing a system failure (e.g., making a wrong turn, running a stop light)	5.37	2.00
	17		Accuracy of the route chosen	5.36	1.60
	18		User's familiarity with the car	5.29	1.58
	19		The reputation of the car manufacturer	5.27	1.66
Bottom Third	20		Agreement of routes between car and my knowledge	5.26	1.55
	21		The car's methods of information collection	5.24	1.52
	22	G	Possibility of the car being hacked	5.09	1.94
	23	H	The user-friendliness of the car	5.04	1.69
	24		Amount of verification by your friend of the car's proposed route and driving ability	4.73	1.71
	25		Your friend's training to use the car effectively	4.68	1.99
	26	I	The car's ability to communicate effectively (e.g., accurate grammar, breadth of vocabulary)	4.60	1.88
	27	J	Popularity of the car	3.38	1.74
	28	K	Aesthetics of the car	3.01	1.72

Table 1. Rankings of the factors that can influence trust of an automated system in the automotive domain. Factors ranked in the same thirds for both the automotive (this table) and medical (Table 2) domains are cross-referenced with letters in the "Ref" column. These common factors appearing in the same third of the rankings give evidence that a core model of trust factors could be developed. The other factors, which are common to both domains but ranked in different thirds or which are domain specific, would be the domain specific factors used to customize the core trust model for a particular domain.

While our surveys presented 30 possible factors to respondents and Desai's had a total of 17, we see some similarities between the factors in the top third, most notably reliability (although, as discussed above, our

surveys presented several questions about aspects of reliability). We also found that trust in the engineers who designed the system was important to our respondents, largely through the different surveys presented for

Medical Domain					
	Rank	Ref	Influence Factor	Mean	Std dev
Top Third	1		Accuracy of the diagnosis	6.33	1.04
	2		Level of accuracy of the machine's diagnosis	6.07	1.16
	3	A	Statistics of machine's past performance	6.04	1.20
	4	C	The machine's ability to stay up-to-date	5.97	1.17
	5		Amount of your information available to the machine (e.g., x-rays, physicals, cat scans, etc.)	5.85	1.26
	6		Amount of verification by your doctor of the machine's suggestions	5.84	1.19
	7		Agreement of diagnoses between doctor and machine	5.83	1.32
	8		Doctor's training to use the machine effectively	5.80	1.24
	9		Amount of information that the machine can access	5.79	1.25
	10	B	Extent of research on the machine's reliability	5.79	1.30
Middle Third	11	E	Effectiveness of the machine's training and prior learning	5.63	1.35
	12		Technical capabilities of the machine	5.62	1.31
	13		Existence of error/problem indicators	5.52	1.49
	14	D	Reputation of the machine	5.50	1.41
	15		The machine's methods of information collection	5.46	1.29
	16		Possibility that the hardware or software may fail	5.34	1.64
	17		Credibility of engineers who designed the machine	5.31	1.52
	18		Your past experience with the machine	5.25	1.44
	19	F	Observing a system failure (e.g., making an incorrect diagnosis)	5.15	1.88
Bottom Third	20		User's familiarity with the machine	5.07	1.52
	21	G	Possibility of the machine being hacked	5.06	1.88
	22		My own research on the machine	5.04	1.52
	23		Your understanding of the way the machine works	5.03	1.59
	24		The reputation of the machine's manufacturer	4.87	1.65
	25		Amount of time the doctor consults other doctors	4.74	1.72
	26	I	The machine's ability to communicate effectively (accurate grammar, breadth of vocabulary)	4.61	1.68
	27	H	The user-friendliness of the machine	4.07	1.66
	28	J	Popularity of the machine	3.77	1.72
	29	K	Aesthetics of the machine	2.64	1.74

Table 2. Rankings of the factors that can influence trust of an automated system in the medical domain. Factors ranked in the same thirds for both the automotive (Table 1) and medical (this table) domains are cross-referenced with letters in the in the "Ref" column.

branded vs. non-branded automated systems.

For both application domains, we found a significant difference in people's trust of the system based upon whether the system was made by a well-known company

(Google for the automotive domain; IBM's Watson for the medical domain) vs. a "small, startup company." Our surveys had two questions about branding, to which participants answered on a 7 point Likert scale, with 1

	Questions		Mean		T Value
	Brand (Watson)	Non-Brand	Brand	Non-Brand	
Safety Critical	I trust the machine's capabilities because it was created by IBM.	I trust the machine's capabilities because it was created by a small, upstart company.	3.04	2.54	0.006
	My trust in a fully-autonomous system similar to this machine would decrease if it was created by a lesser-known company.	My trust in a fully-autonomous system similar to the machine would decrease if it was created by a more established company such as IBM.	3.42	2.81	0.003
Non Safety Critical	I trust the machine's capabilities because it was created by IBM.	I trust the machine's capabilities because it was created by a small, upstart company.	3.24	2.67	0.002
	My trust in a fully-autonomous system similar to this machine would decrease if it was created by a lesser-known company.	My trust in a fully-autonomous system similar to the machine would decrease if it was created by a more established company such as IBM.	3.46	2.87	0.003

Table 3. Branded vs. Non-branded Technology: Medical Domain.
Reputation Matters: Significant differences were seen for responses for branded automated systems in the medical domain.

	Questions		Mean		T Value
	Brand (Google)	Non-Brand	Brand	Non-Brand	
Safety Critical	I trust the car's capabilities because it was created by Google.	I trust the car's capabilities because it was created by a small, upstart company.	3.19	2.41	<0.001
	My trust in a fully-autonomous system similar to cars would decrease if it was created by a lesser-known company.	My trust in a fully-autonomous system similar to cars would decrease if it was created by a more established company such as Google.	3.67	2.89	0.001
Non Safety Critical	I trust the car's capabilities because it was created by Google.	I trust the car's capabilities because it was created by a small, upstart company.	3.33	2.55	<0.001
	My trust in a fully-autonomous system similar to cars would decrease if it was created by a lesser-known company.	My trust in a fully-autonomous system similar to cars would decrease if it was created by a more established company such as Google.	3.88	2.82	<0.001

Table 4. Branded vs. Non-branded Technology: Automotive Domain
Reputation Matters, Part II: Significant differences were also seen for responses for branded automated systems in the automotive domain.

meaning “strongly disagree” and 7 meaning “strongly agree.” In the first, participants were asked to rate the statement “I trust the machines’ capabilities because it was created by [‘IBM’, ‘Google’, or ‘a small, startup company’].” The second statement asked if the participant’s “trust in a fully-autonomous system similar to this machine would decrease if it was created by [‘a lesser-known company’ for the IBM and Google versions or ‘a more established company’ such as Google or IBM].”

Clearly, given these findings, it will require additional work for designers of automated systems to convince users to trust the systems made by small companies. However, one could note that Google was a small, startup company not long ago. Other factors such as past performance of the system can also be used to assist with the trust of a non-branded automated system.

5. Conclusions and Future Work

Our ultimate goal is to build models of the factors that influence people’s trust in automated systems, across many

domains, building a common core model of trust for automated systems and identifying factors specific to particular domains. Such models will serve to inform the designers of automated systems, allowing the development of systems that address the key factors for developing and maintaining a person’s trust of an automated system.

This paper presents some of our initial work towards this goal, identifying the factors that most influence people’s trust of automated cars and medical diagnosis systems. We are still analyzing the data for differences between safety-critical situations and ones that are not. We are also exploring possible differences between respondents of different ages, willingness to take risk, amount of computer use, education level, and other factors.

It is important to note that given the fact that the users of Amazon’s Mechanical Turk skew towards having more education than the average population (Ross et al. 2010), the responses reported in this paper might not be applicable to the general population but instead are only applicable to the population with an undergraduate degree or greater. We need to conduct an analysis of the data with respect to

education level to determine if there are differences between responses for different levels of education. However, despite this potential limitation of our survey population, we believe surveys like ours can identify factors that will influence trust.

Our next step will be to conduct surveys asking people to choose the top factors which influence their trust, ranking them from most to least important. We will also explore the influence that these factors have upon each other; for example, a system's ability to explain its action influences the system's understandability.

We are also expanding this research to other automated system domains. Our methodology will need to change for some of these domains, as we have been relying on people from the population of Mechanical Turk workers. While such people are well qualified to answer questions about cars and doctor's visits, they will be less qualified to answer questions about the use of automated systems in very specialized domains such as the military or power plants. However, we believe that the use of surveys, whether completed by "average" people or people working in specialized domains, will allow us to identify the top factors influencing trust in automated systems in each domain. As we explore more domains, we will be able to identify those factors that are common to many domains; these factors will form the common core of a trust model.

6. Acknowledgements

We acknowledge the support of The MITRE Corporation Innovation Program, project 51MSR661-CA (Approved for Public Release; Distribution Unlimited; 13-3767).

Portions of this research were supported by the National Science Foundation (IIS-0905228) at the University of Massachusetts Lowell.

Hyangshim Kwak participated in this work during an internship at The MITRE Corporation. Cadet Kenneth Voet also assisted with the project.

All product names, trademarks, and registered trademarks are the property of their respective holders.

7. References

- Boehm-Davis, D. A.; Curry, R. E.; Wiener, E. L.; and Harrison, L. 1983. Human Factors of Flight-Deck Automation: Report on a NASA-Industry Workshop. *Ergonomics*, 26(10):953–961.
- Clark, M. 2013. States Take the Wheel on Driverless cars. *USA Today*. <http://www.usatoday.com/story/news/nation/2013/07/29/states-driverless-cars/2595613/>.
- Dellaert, F., and Thorpe, C. 1998. Robust Car Tracking Using Kalman Filtering and Bayesian Templates. In *Proceedings of the Intelligent Transportation Systems Conference*, 72–83.
- Desai, M. November 2012. Modeling Trust to Improve Human-Robot Interaction. Ph. D. diss., Computer Science Department, University of Massachusetts Lowell, Lowell, MA.
- Desai, M.; Medvedev, M.; Vazquez, M.; McSheehy, S.; Gadea-Omelchenko, S.; Bruggeman, C.; Steinfeld, A.; and Yanco, H. 2012. Effects of Changing Reliability on Trust of Robot Systems. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, Boston, MA.
- Desai, M.; Kaniarasu, P.; Medvedev, M.; Steinfeld, A.; and Yanco, H. 2013. Impact of Robot Failures and Feedback on Real-Time Trust. In *Proceedings of the Eighth Annual ACM/IEEE International Conference on Human-Robot Interaction*, Tokyo, Japan.
- Endsley, M., and Kiris, E. 1995. The Out-of-the-loop Performance Problem and Level of Control in Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(2):381–394.
- Guizzo, E. 2011. How Google's self-driving car works. <http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/how-google-self-driving-car-works>.
- International Federation of Robotics (IFR). Accessed 12/30/2011. Statistics about Service Robots. <http://www.ifr.org/service-robots/statistics/>.
- Lin, P. 2008. Autonomous Military Robotics: Risk, Ethics, and Design. Technical Report. Defense Technical Information Center.
- Michaud, F.; Boissy, P.; Corriveau, H.; Grant, A.; Lauria, M.; Labonte, D.; Cloutier, R.; Roux, M.; Royer, M.; and Iannuzzi, D. March 2007. Telepresence Robot for Home Care Assistance. In *Proceedings of the AAAI Spring Symposium on Multidisciplinary Collaboration for Socially Assistive Robotics*.
- Ostwald P., and Hershey, W. April/May 2007. Helping Global Hawk Fly with the Rest of Us. In *Proceedings of the Integrated Communications, Navigation, and Surveillance Conference*.
- Parasuraman, R. 1986. Vigilance, Monitoring, and Search. Boff, K.; Thomas, J.; and Kaufman, L. eds. *Handbook of Perception and Human Performance: Cognitive Processes and Performance*, 1–29. John Wiley & Sons.
- Parasuraman, R., and Riley, V. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors*, 39(2):230–253.
- Ross, J.; Irani, I.; Silberman, M.; Zaldivar, A.; and Tomlinson, B. 2010. Who are the Crowdworkers?: Shifting Demographics in Amazon Mechanical Turk. In *CHI EA 2010*. (2863-2872).
- Sarter, N.; Woods, D.; and Billings, C. 1997. Automation Surprises. *Handbook of Human Factors and Ergonomics*, 2:1926–1943. Wiley.
- Strickland, G. E. January 2013. Watson Goes to Med School. *IEEE Spectrum*, 50(1): 42-45.
- Thrun, S. 2010. What We're Driving At. <http://googleblog.blogspot.com/2010/10/what-were-driving-at.html>.
- Tsui, K.; Norton, A.; Brooks, D.; Yanco, H.; and Kontak, D. 2011. Designing Telepresence Robot Systems for Use by People with Special Needs. In *Proceedings of International Symposium on Quality of Life Technologies*.