The Effect of Culture on Trust in Automation: Reliability and Workload

SHIH-YI CHIEN, National Sun Yat-sen University
MICHAEL LEWIS, University of Pittsburgh
KATIA SYCARA, Carnegie Mellon University
JYI-SHANE LIU, National ChengChi University
ASIYE KUMRU, Özyeğin University

Trust in automation has become a topic of intensive study over the past two decades. While the earliest trust experiments involved human interventions to correct failures/errors in automated control systems a majority of subsequent studies have investigated information acquisition and analysis decision aiding tasks such as target detection for which automation reliability is more easily manipulated. Despite the high level of international dependence on automation in industry and transport almost all current studies have employed Western samples primarily from the US. The present study addresses these gaps by running a large sample experiment in three (US, Taiwan and Turkey) diverse cultures using a 'trust sensitive task' consisting of both automated control and target detection subtasks. This paper presents results for the target detection subtask for which reliability and task load were manipulated. The current experiments allow us to determine whether reported effects are universal or specific to Western culture, vary in baseline or magnitude, or differ across cultures. Results generally confirm consistent effects of manipulations across the three cultures as well as cultural differences in initial trust and variation in effects of manipulations consistent with 10 cultural hypotheses based on Hofstede's Cultural Dimensions and Leung and Cohen's theory of Cultural Syndromes. These results provide critical implications and insights for enhancing human trust in intelligent automation systems across cultures. Our paper presents the following contributions: First, to the best of our knowledge, this is the first set of studies that deal with cultural factors across all the cultural syndromes identified in the literature by comparing trust in the Honor, Face, Dignity cultures. Second, this is the first set of studies that uses a validated cross-cultural trust measure for measuring trust in automation. Third, our experiments are the first to study the dynamics of trust across cultures.

Categories and Subject Descriptors: **H.1.2 [Information Systems]**: User/Machine Systems – Human factors, Human information processing, Software psychology

General Terms: Human Factors, Experimentation, Performance, Reliability

Additional Key Words and Phrases: Cross-cultural research; Trust in automation; Human-automation interaction; Empirical study; Imperfect automation; Workload

ACM Reference Format:

Shih-Yi Chien, Michael Lewis, Katia Sycara, Jyi-Shane Liu, and Asiye Kumru, 2016. The Effect of Culture on Trust in Automation. *ACM Trans. Interactive Intelligent Systems).* x, x, Article x (Nov 2016), 30 pages.
DOI: http://dx.doi.org/10.1145/0000000.0000000

This work is supported by AFOSR FA9550-13-1-0129.

Author's addresses: S.-Y. Chien, Dept. of Information Management, National Sun Yat-sen University; M. Lewis, School of Information Sciences, University of Pittsburgh; K. Sycara, Robotics Institute, Carnegie Mellon University; J.-S. Liu, Dept. of Computer Science, National ChengChi University; A. Kumru, Dept. of Psychology, Özyeğin University.

DOI: http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

As use of technology becomes increasingly globalized, there is a need to study factors that would aid in determining how users in different cultures will adopt and use technology. This is a difficult challenge that may involve many different facets of automation and factors that affect automation use, especially in critical circumstances, such as aviation, crisis response and military. Willingness to use automation is highly related to trust which may be influenced by prior and current experience. In particular, it has been observed that the human may fail to use automation when it would be advantageous to do so. This has been called *disuse (underutilization or under-reliance)* of the automation [Parasuraman and Riley 1997]. On the other hand, people have been observed to fail to monitor automation properly (e.g., turning off alarms) when automation is in use, or they accept the automation's recommendations and actions when inappropriate [Lyons et al. 2011; Parasuraman and Riley 1997]. This has been called *misuse, complacency, over-reliance, or automation bias*. Both misuse and disuse are associated with improper calibration of trust and have contributed to accidents. Misuse has led to mishaps in aviation and marine navigation [Funk et al. 1999]; while disuse has been shown to damage performance through behaviors such as ignoring safety alarms in air traffic control scenarios [Parasuraman and Riley 1997]. A growing literature suggests that trust significantly contributes to human decisions about the use of automation [Hoff and Bashir 2015; Lyons et al. 2016; Wang et al. 2016; Martelaro et al. 2016]. In other words, people tend to rely on automation they trust and not use automation they do not trust. For example, trust has frequently been cited [Lee and Moray 1992; Muir 1994] as a contributor to human decisions about monitoring and using automation. Indeed, within the literature on trust in automation, complacency is conceptualized interchangeably as the overuse of automation, the failure to monitor automation, and lack of vigilance [Billings et al. 1976; Llinas et al. 1998; Parasuraman and Manzey 2010]. For optimal performance of a human-automation system, *human trust in automation should be well-calibrated*, so that appropriate reliance can be achieved.

Trust has been studied in a variety of disciplines (including social psychology, human factors, and industrial organizational psychology) for understanding relationships between humans or between human and machine. The different context within which trust has been studied has led to definitions of trust as an attitude, an intention, or a behavior [Mayer et al. 1995; Moray 2000; Madsen and Gregor 2000]. It is generally agreed that trust is best conceptualized as a *multidimensional psychological attitude* involving beliefs and expectations about the trustee's trustworthiness derived from experience and interactions with the trustee in situations involving uncertainty and risk [Jones and George 1998]. Lee and See [2004] noted that "*trust (in automation) can be defined as the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability*." In other words, trust is derived from an expectation of help, which influences the users' willingness to rely on automation in uncertain situations.

The basis of trust can be considered as a set of attributional abstractions (trust dimensions) that range from the trustee's competence to its intentions. Although in the literature, the number and concepts in the trust dimensions vary [Lee and See 2004], there seems to be a convergence on three dimensions– *Ability, Integrity, and Benevolence* [Mayer et al. 1995] in the interpersonal relations literature and their corresponding notions of *Purpose, Process, and Performance* [Lee and See 2004] for trust in automation. Purpose is a person's knowledge of what the automation is supposed to do. Process is the way the automation functions and performance is how well it fulfills its design specifications. While such models seem plausible, support for the contribution of factors other than performance has typically been limited to correlation between questionnaire responses and automation use. Despite multiple studies in trust in automation, the conceptualization of trust and how it can be reliably modeled and measured is still a challenging problem. Three measures of trust in automation, Empirically Derived [Jian et al. 2000; Spain et al. 2008], Human-Computer Trust [Madsen and Gregor 2000] and SHAPE Automation Trust Index [Goillau and Kelly 2003] have benefited from systematic development and validation. However, these measures have been based on Western (mostly US) population for development and validation.

In order to measure trust across cultures more accurately we [Chien et al. 2014; Chien et al. 2015, Chien et al. 2017] have developed a new measure of trust in automation validated across large samples in three diverse cultures, US, Taiwan and Turkey, as representative of Dignity, Face, and Honor cultures [Leung and Cohen 2011]. The Cross-cultural measure of trust is consistent with the three (performance, purpose and process) dimensions of [Meyer 2001; Lee and See 2004] and contains two 9-item scales, one measuring the *propensity to trust* as in [Jian et al. 2000] and including additional items relating to context of use and the other measuring *trust in a specific system*. The second scale is designed to be administered repeatedly to measure the effects of manipulations expected to affect trust while the propensity scale is administered only once at the start of an experiment. The scales have been developed and validated for US, Taiwanese, and Turkish samples and are based on 773 respondees (propensity scale) and 1673 responses (specific scale). Ratings of trust are reported for each of the scales and as an average for the aggregate. An average (sum) was chosen as the simplest form of aggregation in the absence of a more specific theoretical basis. All data and analyses involved in the development of the instrument are archived in [Chien et al. 2017] at OpenICPSR to provide open data for conducting additional validation testing, assembling specialized scales based on item data or testing hypotheses related to the scales. English, Chinese and Turkish versions of the scales are available at [Chien et al.2017]. An English version of scale is provided in Appendix A. We use this measure to assess trust in automation in the experiments reported in this paper.

Various factors, such as automation reliability, the presence and severity of faults, level of automation, operator workload, and operator's propensity to trust have been studied in the literature as affecting trust (see a brief overview in the next section). However, only scant attention has been paid to how culture may influence human trust in automation, and therefore human reliance on automation. Most of the existing studies on trust in automation were performed within Western cultures. As globalization of automation use becomes omnipresent, there is an urgent need for studies of how trust in automation functions in different cultures. Our paper is the first to our knowledge to perform a systematic study of trust in automation across cultures. Our experiments manipulate factors known to affect trust in automation in Western societies as measured by subjective report and overt behavior in order to distinguish whether reported effects are universal or specific to Western culture, vary in baseline or magnitude, or differ across cultures. The answers to these questions address both the design of automation for cross cultural use and the universality and usefulness of trust as an intervening variable for predicting automation usage.

The present study addresses these gaps by running a large sample experiment in three (US, Taiwan, Turkey) diverse cultures using a modification of the RESCHU simulation [Boussemart and Cummings 2008]. Our 'trust sensitive task' contains both automated control and target detection subtasks designed to manipulate the primary factors found to affect trust in previous studies. Multiple subtasks were chosen in part because multitasking demands have been shown necessary [Lee and See 2004; Parasuraman et al. 1993; Parasuraman et al. 1994; Molloy and Parasuraman 1996] to produce overtrust and overreliance, effects we wished to study, in ways single tasks [Thackray and Touchstone 1989] do not. While [Keller and Rice 2009; Bean et al. 2011] have found evidence for "system-wide trust" in pairing reliable with unreliable gauges, studies with functionally distinct subsystems [Lee and See 2004; Lee and Moray 1992; Lee and Moray 1994; Muir and Moray 1996] suggest that trust can be associated with particular subsystems rather than the system as a whole, allowing us to investigate trust independently for the two subtasks if their independence can be established. The navigation automated control subtask allowed us to manipulate the level of autonomy as defined by either a single dimension [Sheridan and Verpank 1978] or stages of processing [Parasuraman et al. 2000] as well as control the transparency of action implementation processing. Results from the target detection subtask, reported in this paper, address another set of factors by manipulating automation reliability and task load. Due to constraints of sample size, manipulations of risk, uncertainty and other factors suspected of influencing trust in automation could not be incorporated into these experiments.

The article is organized as follows: In the Introduction factors affecting trust and use of automation manipulated in the study are introduced and discussed. Theories of Cultural Dimensions and Cultural Syndromes are then introduced and used to derive a set of hypothesized differences among cultures in trust and use of automation. The Experiment section presents the trust sensitive task and experimental design. Results presents the effects of the task load and reliability manipulations on trust, trust related behavior, and performance. The Discussion section contrasts current findings on task load with prior reports and agreement with findings for reliability. Hypotheses involving cultural differences in trust and trust related behaviors are compared with findings and found largely in agreement.

2. FACTORS INFLUENCING TRUST

The factors that are likely to affect trust in automation have generally been categorized as those pertaining to the *system*, the *operator*, and the *environment*. Most work on factors that have been empirically researched pertains to characteristics of the automation. Here we briefly present relevant work on the factors involved in our experiments (for a more expanded overview of these factors see [Lee and See 2004; Hoff and Bashir 2015]).

2.1 System Properties

The most important system property that affects trust has been *system reliability*, i.e. the rate of system error [Moray and Inagaki 1999; Riley 1994; Parasuraman and Manzey 2010; Parasuraman and Riley 1997; Parasuraman et al. 2008]. This and almost all subsequent research has shown that when reliability decreases, so does trust. Moreover, it appears that most recent experiences with the system are more influential than more distant ones [Goillau and Kelly 2003; Lee and Moray 1992]. Reliability is manipulated in these experiments through use of a likelihood display for an information acquisition and analysis joint target identification task. Other system properties known to affect trust including level of autonomy and transparency were manipulated for the navigation subtask only and are not reported in this paper.

2.2 Environmental Characteristics

Task load is a crucial factor manipulated in our experiments mediating the effects of trust on compliance and reliance. A common finding [McBride et al. 2011; Wang et al. 2011; Rajaonah et al. 2008; Willems et al. 2002; Biros et al. 2004] has been that increases in workload lead to higher reliance/compliance with automation. This is consistent with eutactic monitoring explanations [Moray 2003] that users balance the cost of monitoring with probability of errors so that as task load increases users are more likely to rely and comply with automation in order to keep up with task demands. Other studies [Biros et al. 2004; Spain and Bliss 2008; Wetzel 2005], however, have found no relation between workload and reliance or compliance. A related question is whether increased trust and reliance go hand in hand, are independent, or move in opposing directions with increases in workload. Here the answer is more equivocal. Recent reviews of research in trust in automation [Hoff and Bashir 2015; Schaefer et al. 2016] report that trust is unaffected or decreases with increased workload without distinguishing between experiments in which workload is manipulated from those in which it is merely measured. Where perceived workload is measured rather than manipulated automation has been found [Spain and Bliss 2008; Wang et al. 2011] to be negatively correlated with measures of workload such as NASA-TLX [Hart and Staveland 1988] or to have no effect [Rajaonah et al. 2008]. Another group of studies that actively manipulated task load have found variously; higher ratings of trust in low workload conditions [Willems et al. 2002; Biros et al. 2004; Karpinsky et al. 2016], higher ratings of trust in low workload conditions but only for low reliability automation [Daly 2002; Wetzel 2005] and higher ratings of trust in high workload conditions but only for high reliability automation [Wetzel 2005]. This variation among results suggests that the relation between workload and trust is complex and highly dependent on characteristics of the automation and task such as reliability and feedback. These questions are addressed in our experiment from a cross-cultural perspective.

2.3 Operator characteristics

One of the most important characteristics in interpersonal trust has been found to be a trustor's propensity to trust [Rotter 1967]. However, there is very little work on propensity to trust in the trust in automation literature. In Parasuraman and Riley's study [1997] it was found that an operator's overall propensity to trust was distinct from trust towards a specific system. In other words, an operator may have high propensity to trust automation, but she may not trust a given specific system. Guided by such findings, we have considered both propensity to trust in automation in general and also specific system trust in our cross cultural trust measure.

In the inter-personal trust literature, individual differences of the operator, such as self-esteem [Rotter 1967; Rotter 1971], secure attachment [Cassidy 1988], and motivational factors [Kruglanski and Thompson 2000] have been identified as affecting trust. We report elsewhere [Chien et al. 2016] on the relations between personality, Hofstede's cultural dimensions and propensity to trust in automation for the current sample. However, because these individual differences were not directly implicated in performance on the target-finding subtask they are not reported here.

Furthermore, socio-cultural factors have also been identified such as high power distance with authority [Carlson et al. 2004] (see next section on definition of culture and cultural dimensions). People in high power distance (PD) societies expect authority figures to be benign, competent and of high integrity. Thus people in high power distance societies will engage in less vigilance and monitoring for possible violations by authority figures. In contrast to the interpersonal trust literature, To date, only a handful of studies [Rau et al. 2009; Wang et al. 2010; Chien et al. 2016] consider cultural factors and potential differences in the context of trust in automation. The work described in this paper is an additional step towards filling this gap.

3. CULTURE AND TRUST IN AUTOMATION

Culture has been defined as the unique nature of a social group with regards to values, beliefs, norms, and practices [House et al. 2004]. Cultures can have a central theme or syndrome, which is a compilation of shared beliefs and practices. The majority of cross-cultural research has relied on the cultural themes of individualism and collectivism (e.g. Triandis [1994]) as well as Hofstede's cultural dimensions [Hofstede 1991]. However, recently researchers have focused on Honor, Face, and Dignity cultural syndromes as they provide a more refined framework on how people interact, form relationships and handle conflicts [Aslani et al. 2013]. We use the most well studied and discriminating of Hofstede's dimensions, as well as the 3 cultural syndromes in our studies. Below we present a brief review of Hofstede's dimensions and the cultural syndromes and discuss how these can affect trust.

3.1 Hofstede's cultural dimensions

To measure the cultural differences on trust in automation, three of Hofstede's cultural dimensions, which have been well studied in prior research, were used in our studies.

— *Power Distance* (PD) is defined as "*the fact that all individuals in societies are not equal, and it expresses the attitude of the culture toward these power inequalities amongst us"* [Hofstede 1991]. In societies with high PD, a less powerful person must accept instructions given by more senior and powerful members of the organization. This factor may affect the extent that an individual from a high PD culture perceives the automation as authoritative, and as a result, the operator will be quick to establish trust in the automated suggestions. On the other hand, people in high PD cultures should be slow to restore trust once violations have occurred [Brockner et al. 1992].

— *Individualism versus Collectivism* (IDV) is defined as "*the extent to which people from birth onwards are integrated into strong, cohesive in-groups, which throughout people's lifetimes continue to protect them in exchange for unquestioning loyalty"* [Hofstede 1991]. It represents an individual's self-image between "I" or "We" in a society. People from an individualistic culture tend to take care of only themselves and direct family members, whereas an individual from a collectivist society takes care of others in exchange for unquestioning loyalty. In other words, in a society with high IDV an individual focuses more on his/her own achievements rather than on group goals. Fulmer and Gelfand [2010] found the "black sheep" effect in a collectivist society, in which people from this culture became less trusting after experiencing violations from in-group rather than out-group members. To the extent that automation is perceived as a helper, individuals from highly collectivist societies, may form trust quickly but would be slower to regain trust, if automation is unreliable.

— *Uncertainty Avoidance* (UA) is defined as "*the extent to which the members of a culture feel threatened by uncertain or unknown situations"* [Hofstede 1991]. People in greater UA cultures look for structured formats and clear instructions to shun ambiguous conditions and make events more predictable. This dimension then would affect trust, especially if automation is unreliable and in variable ways.

3.2 Cultural syndromes

Although Hofstede's cultural dimensions have been well studied in the literature and continue to be relevant when examining the general effects of cross-cultural differences, recent research [Leung and Cohen 2011] has indicated that Hofstead's metrics failed to measure an individual's behaviors in terms of adherence to cultural norms in their interactions with various situations and consequently, the influence of their values by a particular member. To address the gaps, Cultural Syndromes [Leung and Cohen 2011] were also included in our study to provide complementary approaches to measuring cultural differences. Cultural syndromes encompass cultures of Dignity, cultures of Honor, and cultures of Face, which contrast with the meaning and importance that are given to norms of exchange, reciprocity, punishment, honesty, and trustworthiness. Recently, interest in the cultural syndromes of Dignity, Honor, and Face has resurfaced [Leung and Cohen 2011; Aslani et al. 2013; Aslani et al. 2016] with particular significance to antecedents of trust.

In *Dignity cultures,* prevalent in Western Europe and North America, one's self-worth is derived internally. It is not determined by the opinions and values of others and cannot be altered by other people; it is only evaluated by the individual's own standards [Leung and Cohen 2011]. Dignity cultures are high on individualism and low on power distance. The context that surrounds interactions is egalitarian, consisting of autonomous individuals who focus on personal, individual goals [Schwartz 1992], supported by an effective system of law that enforces contracts and rights [Leung and Cohen 2011]. In interactions, people are treated as equals and positive reciprocity occurs in the form of short-term tit-for-tat exchanges that signal integrity and trustworthiness [Leung and Cohen 2011]. In these cultures, people generally have a "swift trust" assumption: others deserve to be trusted until they prove otherwise [Dirks et al. 2009; Weber et al. 2004]. These characteristics would lead to the belief that operators from Dignity cultures will be quick to trust in automation.

Face cultures are prevalent in East Asian societies where one's self-worth is derived externally. Self-worth is the view that others have of the individual and is based on social interactions with others. It is stable so long as the social hierarchy in which the person interacts is stable [Leung and Cohen 2011]. So, self-worth is interdependent with a person's role in a stable social hierarchy, and on fulfillment of role obligations [Heine 2001]. In these cultures, people can lose face if others disapprove of their actions and behaviors [Leung and Cohen 2011]. Face cultures are high in collectivism and high in power distance. People interact in stable hierarchies, and social interactions are governed by norms imposed by social institutions, such as religion, family, community or the state. People's conformity to those norms is monitored and if necessary, managed by institutional sanctioning (see [Yamagishi et al. 1998; Gunia et al. 2011]). Because of this institutional monitoring and sanctioning, people can engage in smooth social interactions in the absence of trust [Yamagishi et al. 1998].

Honor Cultures can be found in the Middle East, Latin America, and Mediterranean countries. People's self-worth is dependent on interactions with others and one's perception of self. Accordingly, it is derived both internally and externally [Pitt-Rivers 1966]. Norms and values of Honor culture fall between Dignity and Face cultures, creating its own unique cultural prototype. Honor cultures are in the middle range on Hoffstede's dimensions of collectivism and power distance (Fig. 1). For instance, people in honor culture are relational and interdependent within their social groups. Yet, they are more likely to engage in direct confrontations. In these cultures, honor, linked to self-worth, must be claimed as well as paid to others [Leung and Cohen 2011]. Honor can also be taken away, thus it must be protected. Honor cultures manifest with a reputation for toughness in protecting the self and family and involve not letting others take advantage of you [Nisbett and Cohen 1997]. The social context of Honor cultures is unstable social hierarchies. Consequently, members of Honor culture tend to have slow trust (low trust at the beginning of interactions) and low trust in laws and institutions. Bohnet and Zeckhauser [2004] suggested that in Honor cultures, it is betrayal aversion (people's aversion toward risk caused by other people) not simply risk aversion, that affects people's trust decisions. A betrayal-averse individual would be more likely not to trust another individual at the beginning of a trust relation, would be more likely to monitor for trust violations, and would be more likely to make negative attributions if trust violations do occur. Indeed, Bohnet et al. [2009] found that people in the Persian Gulf required a higher level of trustworthiness before they were willing to trust other individuals than either Americans or Swiss. The socio-cultural factors of distrust that include surveillance and monitoring [Sitkin and Roth 1993], are found in Honor [Nisbett and Cohen 1997], and collectivistic culture [Triandis 1995]. This is very important to trust in automation since surveillance and monitoring has been shown to be relevant to misuse and disuse of automation. Therefore, cultural characteristics that may be linked to them were the subject of careful study in our studies.

The defining characteristics of Dignity, Face, and Honor cultures have elements that are also present in Hofstede's dimensions (especially in PD, IDV and UA). An interesting observation is that as regards the dimension of IDV, Dignity cultures are high on IDV, Honor cultures medium and Face cultures low. Dignity, Honor and Face cultures do not fall as neatly in the dimensions of PD and UA. While the Dignity cultures are low on PD, a few Honor and Face cultures are very close. Therefore, it seems that cultural syndromes could bring relevant elements in addition to the Hofstede's dimensions that may provide the basis for greater discriminatory power.

As the hypotheses based on Hofstede's cultural dimensions [1991] and a more recent theory of cultural syndromes [Leung and Cohen 2011; Triandis 1994] suggest, it is reasonable to expect culture to affect trust and use of automation in a variety of ways. These cultural characteristics that have been identified as influencing inter-personal trust will guide the proposed research in how cultural factors may influence trust and use of automation and help formulate research hypotheses.

4. RESEARCH QUESTIONS AND HYPOTHESES

In our studies, we selected the US as a prototypical Dignity culture, Taiwan as a prototypical Face culture and Turkey as a prototypical Honor culture. In addition, the contrasts that these countries provided on Hofstede's dimensions (Fig. 1) revealed some substantial cultural differences. For instance, Turkey is high on PD and UA but low on IDV, whereas US is high on IDV but low on PD and UA, with Taiwan in the middle among these three constructs.

There are three general questions we examine in our studies: (a) Do effects of trust that have been observed in prior research in Western cultures hold universally in all 3 cultural syndromes? Such effects include the increase in trust in automation with increase in reliability and increase in reliance with increase in task load. (b) Even if trust effects are universal across syndromes, do they differ in terms of magnitude in the different cultures? (c) Do some of the effects work in a one way in one culture and in a different way in another? In other words, are there interactions among the effects?

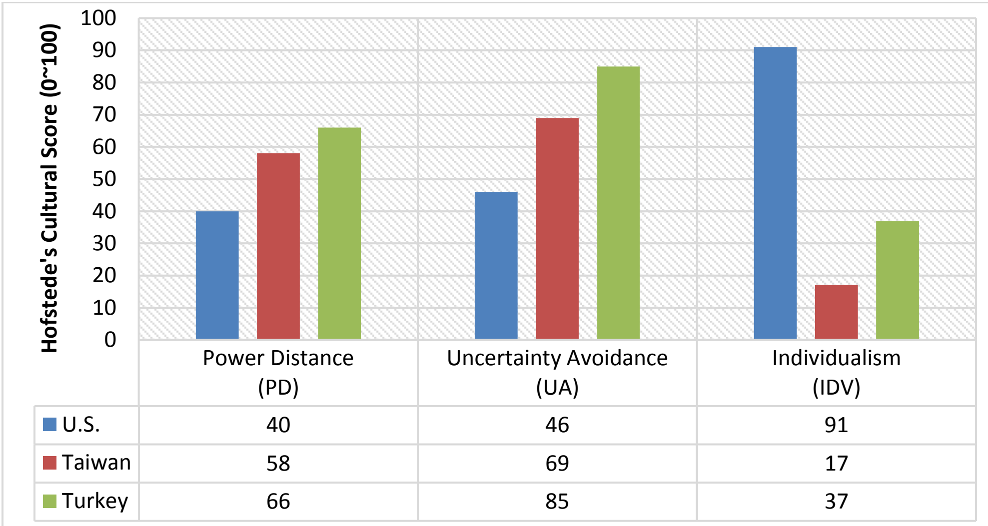


Figure 1. Cultural comparisons in Hofstede's dimensions [Hofstede 1973].

Based on the cultural characteristics of the 3 syndromes along with Hofstede dimensions, and in order to answer the three general research questions above, we form the following research hypotheses.

— *H1: Individuals from Dignity cultures are more likely to have higher level of initial trust in automation than those from Honor and Face cultures. (This is because Dignity cultures make the swift trust assumption).*

— *H2: If using the automation were encouraged by the user's organization, Face culture operators will have higher ratings of trust and reliance than those from Honor and Dignity cultures. (This is because of the high power distance present in Face cultures).*

Studies have shown that decrease in system reliability will decrease operators' trust in and reliance on automation [Visser and Parasuraman 2011; Rovira et al. 2007; Chien et al. 2013].

— *H3: Unreliable automation will lower ratings of trust of operators from all cultures (both low- and high-PD) operators, but Face culture operators will be more likely to continue relying on automation. (We hypothesize that the Western literature finding that low trust is positively correlated with unreliable automation is present in all cultures. Because Face culture has higher power distance, Face culture operators will be more likely to continue relying on the automation, if it is perceived as an authority).*

— *H4: Face culture operators will recover their trust in automation after failure more quickly than Honor and Dignity culture operators. Honor culture operators would be slowest in recovering trust. (This hypothesis on Face culture relies on the rationale for H3 above. Because the social context of honor cultures is unstable social hierarchies, members of those cultures would be the slowest to exhibit trust and slow to regain it once lost).*

— *H5: Operators from Dignity and Honor cultures will be more self-confident and therefore are less likely to rely on or ignore the automation than Face culture operators. (Dignity cultures are characterized by high individualism and self-reliance, hence* *their members will be more self-confident. In Honor cultures self-worth is derived both internally and externally. Therefore, members of honor culture will be more self-confident than those of Face culture whose self-worth is derived externally).*

—*H6: Honor and Face culture operators will exhibit more vigilance and more monitoring behavior than operators from Dignity cultures. (Honor cultures are more distrustful, hence they will engage in more vigilance than Dignity cultures. In Face cultures social interactions are governed by norms. Norm conformity is monitored and managed by institutional sanctioning).*

— *H7: Honor culture operators will take longer interaction times than operators from Dignity and Face cultures to develop equal degrees of trust. (Honor cultures are more distrustful than Face and Dignity due to need to protect their honor).*

—*H8: Honor operators will either disuse or take longer to regain trust after a failure occurs and may not recover trust to the original level (miscalibrate), as compared with Face and Dignity operators. The dynamic relation between use and trust may magnify these effects. (honor cultures have high uncertainty avoidance, are subject to the betrayal aversion effect and are more mistrustful than Face and Dignity members).*

Some effects associated with trust (such as complacency) have been found to occur only under multitasking or heavy workload conditions [Visser and Parasuraman 2011]. Because fewer resources are available for secondary tasks in high workload situations, participants may have a higher tendency to rely on automated assistance when experiencing heavy task loads.

— *H9: Operators will accept more automated recommendations or exhibit fewer checking behaviors on automation in high workload conditions. (We hypothesize this will be a general finding, valid in all cultures).*

—*H10: The trust of Face culture operators will be relatively more influenced by information about the purpose/benevolence of automation than Honor or Dignity culture operators. (Since the social context of Face cultures is stable hierarchies, information about the purpose of the automation will engender relatively higher trust).*

The above hypotheses have been evaluated through cross-cultural experimental studies, using RESCHU, a multiple unmanned aerial vehicle (UAV) testbed. In the experiments, participants' trust was measured using the trust instrument we developed in prior research [Chien et al. 2014; Chien et al. 2015, Chien et al. 2017].

5. EXPERIMENT

To measure cultural effects on trust in automation, experiments were conducted in the U.S., Taiwan, and Turkey. These countries were chosen based on Hofstede's Cultural Dimensions and Leung and Cohen's theory of Cultural Syndromes, which posit maximal cultural differences among the three countries.

5.1 Simulation

We modified RESCHU (Research environment for Supervisory Control of Heterogeneous Unmanned Vehicles) [Boussemart and Cummings 2008], a widely used UAV command and control simulation, to allow manipulation of factors previously found to influence trust and reliance/compliance with automation. RESCHU was selected because it provided both multitasking demands found to be a necessary condition for observing overtrust [Parasuraman et al. 1993] and subtasks alterable to span the range of levels of autonomy from target detection to fully automated control.

RESCHU simulates multiple UAVs conducting a search and attack task by performing navigation and target identification subtasks. UAVs travel along planned paths toward their target. Upon arrival, a payload task is spawned, requiring the operator to search for an assigned target in a separate window. After completing the payload task, the UAV is assigned a new target and the process repeats. The reliability of automation for the target identification task was manipulated through introduction of false alarms into the 'target finder'. Task load was manipulated by altering UAV speed resulting in more frequent payload requests for the target finder due to shortened travel times.

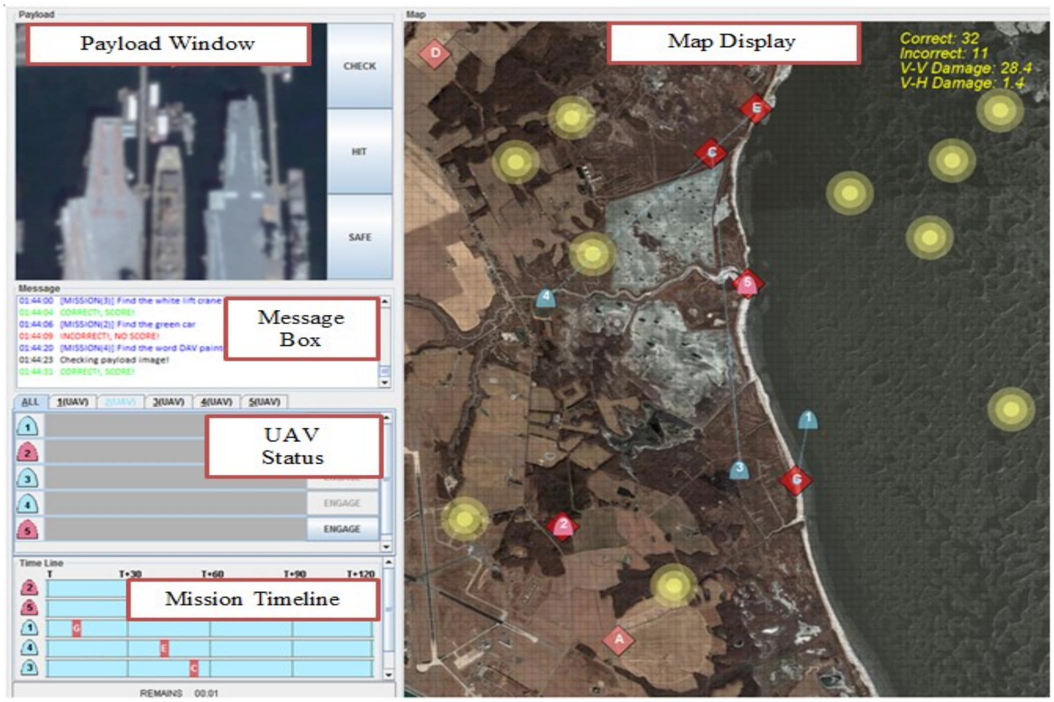


Figure 2. The RESCHU user interface. The map window shows the numbered UAVs (blue ovals) with paths to targets (red diamonds); threat areas are marked by the yellow circles. While an UAV reaches a target, the engage button will be switched on in the UAV status window and the UAV icon will begin flashing in the mission timeline.

In the experiments five UAVs were tasked to identify and attack hostile targets (payload task) while monitoring and rerouting the UAVs' paths if necessary (navigation tasks). RESCHU provided a payload window (shown at the top left in Fig. 1) for target identification tasks, map display (shown in the right window in Fig. 1) for UAVs planned routes, message box for assigned enemy targets for payload tasks, UAV status window for vehicles' current situations, and mission timeline for the ongoing and upcoming missions. To maintain necessary situation awareness and optimize the multi-tasking strategy, task results were included in the message box and real-time feedback panel (yellow text, shown at the top right in Fig. 2).

5.2 Payload Tasks

Upon reaching a target, (shown as red squares in Fig. 3), the operator was presented with a pannable view in the payload window and asked to search for a specific target shown in the message box. The target finder (details in the next section) mimics a familiar form of automation by drawing a rectangle around a suspected target much as image processing software is used to draw bounding boxes around faces in digital photography. The identifiability of an object, such as a red car, within the bounding box depends on perspective, background color, etc. providing a continuum of discriminability for potential targets. The operator first observed a low-resolution image in the payload window, along with three options: Check, Hit, and Safe. By clicking the “Check” button the system provided, after a three-second delay, a picture with better resolution for improving the identification of the assigned target.

If an operator believed the target was not present in the scene, “Safe” should be chosen to terminate the attack; otherwise, “Hit” should be chosen to attack the target. Following a payload submission (either Hit or Safe), the message box and real-time feedback panel informed the operator of whether the submitted decision was correct and the UAV was assigned to another target and the process was repeated.

(3a) Before checking (3b) After checking

Figure 3. Target identification tasks.

5.3 Target Finder

The target finder used a likelihood alarm system, which generated three levels of alarm explained to the participants as based on the estimated likelihood of a target [Wiczorek and Manzey 2014; Wickens and Colcombe 2007]. The target finder placed a bounding box on top of the suspected target and the payload window was highlighted in the appropriate color. High certainty (alarm) was represented by the red border (Fig. 4a), while a yellow border (warning) specified (Fig. 4b) a higher level of information uncertainty, A green border indicated a non-alert event (Fig. 4c), with a low possibility of a target. We will refer to these alarm levels as Red, Yellow, and Green.

In the alarm and warning conditions the operator needed to determine whether the bounding box indicated the assigned target or not making agreement with this choice a matter of compliance [Meyer 2004]. Hit was chosen, if the operator believed the box located the target correctly, otherwise Safe was selected. If the operator detects a target in the green non-alert condition a bounding box must be drawn on the suspected target and hit selected. As no decision is advanced by the automation, agreement in this condition is classified as reliance [Meyer 2004].

(4a) Alarm condition (4b) Warning condition (4c) non-alert condition

Figure 4. Conditions shown in the target finders.

5.4 Experimental Design

Research shows that perceived workload and system reliability are major factors mediating trust and reliance on automation [Lee and See 2004]. To investigate these effects, task load and automation reliability, were manipulated.

To simulate the effects of imperfect automation, reliability of the target finder for the alarm and non-alert events (the red and green cues, respectively) remained at 80% across all experimental conditions, whereas the warning condition (the yellow border) was 80% in the high-reliability condition and 20% in the low-reliability condition. The reliability comparison is between a “balanced” high reliability condition and a False Alarm (FA) prone low reliability condition.

Task load was manipulated through changes to the UAVs’ moving speed, in which the vehicles moved at 5.0 pixels/second in the high task load condition and 2.5 pixels/second in the low task load condition, a difference [Gartenberg et al. 2012] found sufficient to produce significant effects on their dependent variables. Change in UAV speed affects the payload task by shortening travel times leading to more frequent payload requests.

5.5 Participants and Procedures

American participants were recruited from the University of Pittsburgh (80 females and 40 males with average age of 19.57), Taiwanese participants were recruited from National Chengchi University (80 females and 40 males with average age of 21.60), and Turkish participants were recruited from Özyeğin University (95 females and 25 males with average age of 21.58). None had prior experience with UAV supervision or similar aviation tasks although most reported frequent computer use, defined as more than 8 hours in a typical day. To better capture cultural characteristics, a qualified participant was required to have attended K–12 schooling in the represented country.

After providing demographic data and completing an individualized measure of Hofstede’s dimensions, CVSCALE [Yoo et al. 2011], and a standard personality instrument, The Big 5 Inventory [John and Srivastava 1999], participants were asked to rate their initial trust in automation using the general scale of our cross-cultural trust instrument. Chinese and Turkish versions of the instruments were used in their respective countries. In the following 20-minute training session participants took an interactive training tutorial to learn control operations with the automated applications (target finder and/or conflict detector) and were informed that the automation was fairly (but not perfectly) reliable with the goal of the payload task being to identify and attack as many targets as possible.

After the training, participants began the first 10-minute experimental session in which they performed the target classification tasks controlling five UAVs. At the conclusion of the session, participants were asked to complete the specific trust instrument to evaluate their trust in the automated applications as well as the NASA-TLX. After a brief break, the other task load condition was run accompanied by a repeated specific trust questionnaire and NASA-TLX. Conditions were counterbalanced for reliability of the target finder.

6. RESULTS

Our full experimental design crossed low and high reliability conditions of the payload task with four levels of navigation automation and the three nationalities. Tests for interaction of the navigation task with payload measures found no interaction for the performance measure checks/engagement ($F_{3,672}$=.1.315, p=.398, η^2=.006) although the ratio of correct responses to engagements shows a small interaction ($F_{3,672}$=.2.991, p=.03, η^2=.013) accounting for approximately 1% of the variation in performance. There is also a marginally significant interaction for specific trust ($F_{3,672}$=.2.424, p=.065, η^2=.011). Because of the difficulty of interpreting 3 way interactions between the two tasks and nationalities and the small size of the interactions involved we have chosen to treat the tasks as independent. This allows analysis of the payload task as a simple 2x2 design with target finder reliability as between subject factor and task load the within subject factor.

Data were analyzed using a mixed-model ANOVA with reliability (high: 80% vs. low: 60%) of target finder, and countries (U.S., Taiwan, and Turkey) as the between-subject factors with task load (high vs. low) as the within-subject variable. These analyses adopt a significance level of p< .05, referring to alpha levels of .05-.07 as marginally significant. All post hoc comparisons use Bonneferoni correction. Taking NASA-TLX perceived workload ratings as a manipulation check, perceived workload was found to be higher under high task load conditions for the mental (p=.031), temporal (p<.001) and effort (p=.017) subscales as well as marginally higher (p=.061) for full score.

6.1 Survey data- general trust

The analyses revealed significant cultural effects on initial trust of automation for the performance ($F_{2,357}$=2.969, p=.053), process ($F_{2,357}$=66.225, p<.001), and task context ($F_{2,357}$=18.697, p<.001) scales, Fig. 5. Taiwanese rates differed from the other two countries with higher trust scores for performance (TW>TK, p=.066)—but the lowest trust values associated with task context (US>TW, p<.001; TK>TW, p=.005).

To measure the overall effect, the average score was computed by the mean value of the three constructs ($F_{2,357}$=16.225, p<.001). T-tests revealed significant differences between the U.S. and Turkey (p<.001), U.S. and Taiwan (p=.022), and Taiwan and Turkey (p=.009), in which the U.S. participants had the highest score in general trust and the Turkish participants had the lowest, with the Taiwanese rates falling in between.

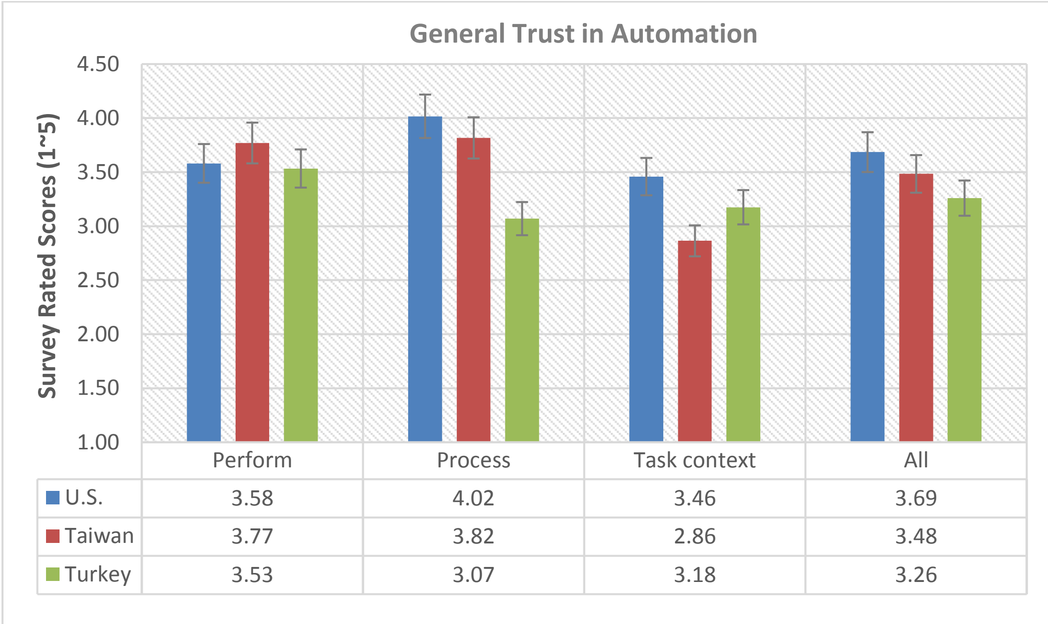


Figure 5. General trust in automation among three cultures.

6.2 Survey data- effect of task load on trust in target finder

Increasing task load (i.e., doubling the UAVs' travelling speed) affected trust in the target finder, in which trust was rated marginally higher in the high workload (HW) condition for the performance measure ($F_{1,672}$=3.831, *p*=.051); however, the rest of comparisons remained insignificant. In addition, no relation was found between workload as measured by NASA-TLX and trust for Taiwanese or Turkish participants, however, a slight positive correlation (r=.148, *p*=.022) was found in the U.S. sample.

6.2.1 Survey data- cultural effects on trust in target finder between task load conditions

To examine the relationship between culture and task load conditions, effects of task load were compared across cultures. The analysis (Fig. 6) found a cultural main effect for overall trust in the target finder under both LW ($F_{2,357}$=9.339, *p*<.001) and HW ($F_{2,357}$=3.668, *p*=.027). T-tests revealed similar trust ratings for U.S. and Taiwanese participants, which were higher than those of the Turkish participants in LW (US>TK, *p*<.001; TW>TK, *p*=.012) and marginally higher in HW (US>TK, *p*=.066; TW>TK, *p*=.052). Both Taiwanese (*p*=.045) and Turkish (*p*=.003) participants rated overall trust higher in the HW condition, however, in this experiment no task load difference was found for the U.S. sample.

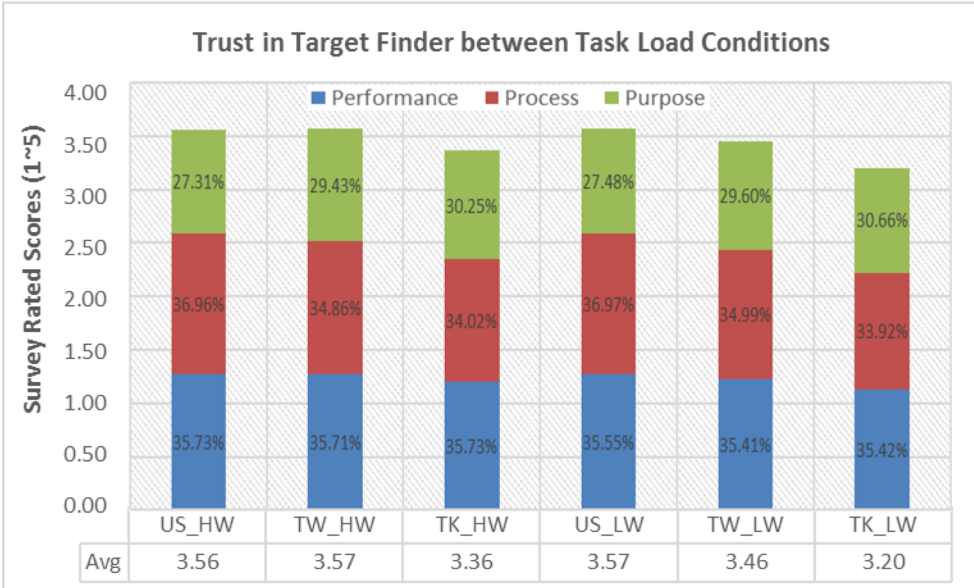


Figure 6. Average trust score in target finder between task load conditions. Performance, process, and purpose constructs are each represented by their proportion of each bar in the chart.

6.3 Survey data- trust in target finder between reliability conditions

To examine the effects of source reliability on trust in automation trust ratings for the full sample were compared between, two reliability levels, high (80%) and low (60%). The high reliability (HR) condition led to higher ratings of trust than the low reliability (LR) on each subscale as well as full scale as shown in Table I.

Table I. Trust in target finder between reliability conditions

Payload Tasks: Trust in *Target Finder* between *Reliability* conditions			
Measures	**$F_{1,672}$**	***p*-value**	**Post hoc**
S_Performance	16.413	<.001	HR>LR
S_Process	15.329	<.001	HR>LR
S_Purpose	9.368	.002	HR>LR
Overall (average value)	19.089	<.001	HR>LR

6.3.1 Survey data- cultural effects on trust in target finder between reliability conditions

To examine the relationship between culture and reliability, the effects of target finder reliability were compared across cultures (Fig. 7). No statistically significant differences were observed between the U.S. and Taiwanese participants. However, the analysis revealed significant differences between the U.S. and Turkish participants in both the HR (p=.001) and LR (p=.006) conditions, as well as a significant difference between Taiwanese and Turkish participants in HR (p=.001) but not in the LR condition.

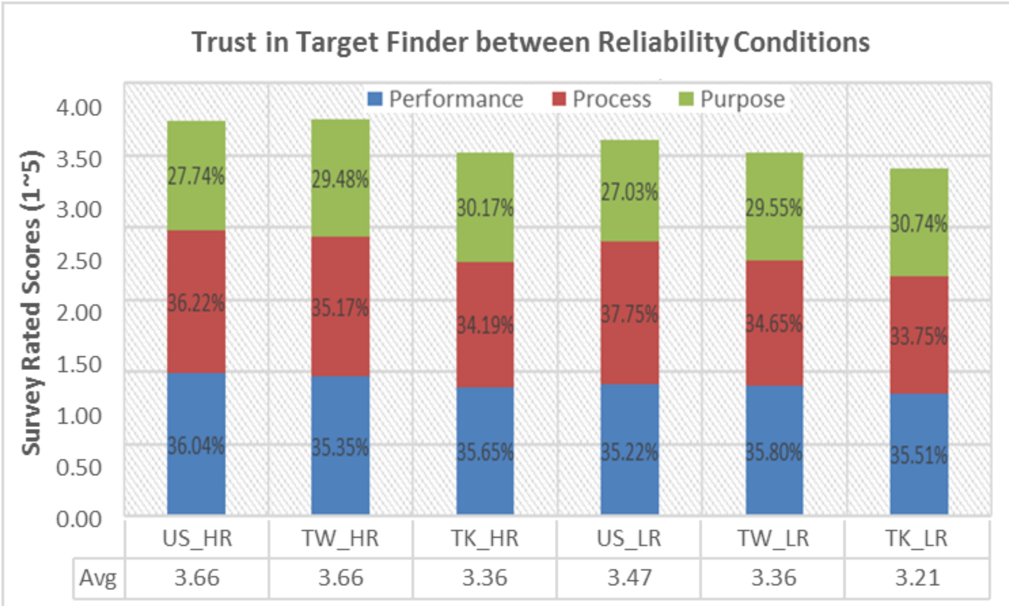


Figure 7. Average trust score in target finder between system reliability conditions. Performance, process, and purpose constructs are represented by their proportion in each bar of the chart.

These results confirm earlier findings that increases in system reliability contribute to higher trust in automation regardless of culture. In general, participants from the U.S. and Taiwanese cultures had similar levels of overall trust in the target finder, regardless of reliability conditions while Turkish participants again showed the least trust in the automated aid.

6.4 Performance data- correct target identification

The performance of payload tasks was examined for the ratio of correct target identifications per engaged payload tasks. ANOVA showed that the task load ($F_{1,672}$=6.084, p=.014, η^2=.009), reliability level ($F_{1,672}$=14.359, p<.001, η^2=.021), and culture ($F_{2,672}$=21.518, p<.001, η^2=.060) significantly affected the correct target identifications (Fig. 8). The analysis also found a significant interaction between task load and culture ($F_{2,672}$=7.128, p=.001, η^2=.021) and between task load and reliability ($F_{1,672}$=21.335, p<.001, η^2=.031). T-tests showed significantly higher engagement of correctly identified targets in HW than LW (p=.014) conditions as well as improved accuracy in HR rather than LR (p<.001) target finders. The comparisons also revealed that American participants were more accurate in finding targets than Turkish (p<.001), and Taiwanese operators had higher accuracy than Turkish operators (p<.001). No significant effect was found between American and Taiwanese participants.

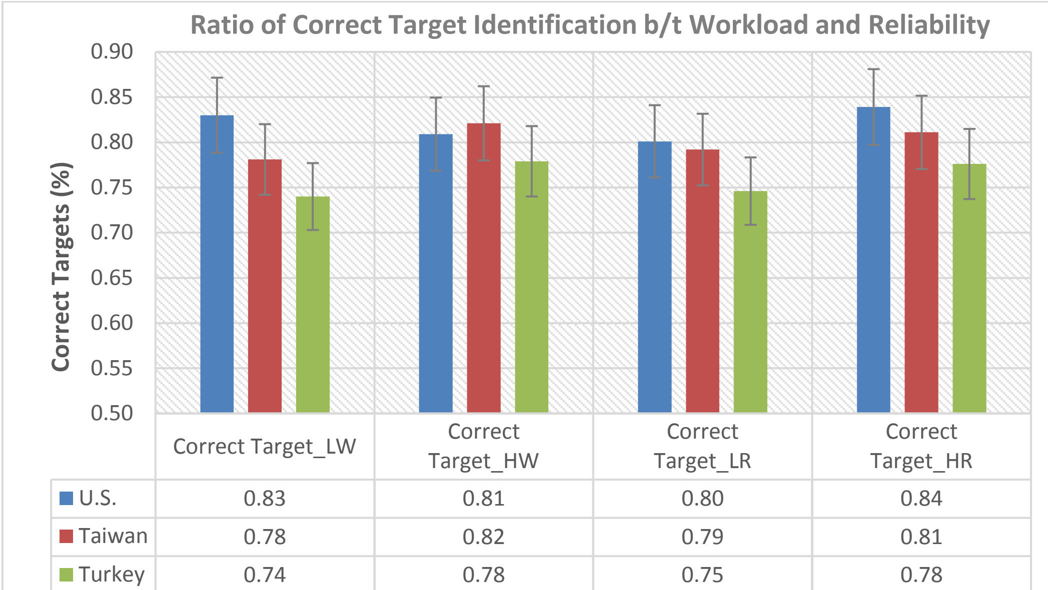


Figure 8. Ratio of correct target identification between workload and reliability conditions.

To examine the effects of unreliable automation on performance in target identification tasks, the results were compared by uncertainty levels (alarm-red, warning-yellow, non-alert-green), Fig. 9. The analysis revealed main effects for culture in Green ($F_{2,672}$=8.699, p<.001, η^2=.025), Yellow ($F_{2,672}$=16.380, p<.001, η^2=.046), and Red ($F_{2,672}$=8.005, p<.001, η^2=.023) conditions. Post-hoc tests showed American participants identified more targets than Taiwanese (p=.001) or Turkish (p=.002) participants in the Green condition; whereas similar proportions were correctly identified by American and Taiwanese participants in the other two conditions. Both U.S. and Taiwanese participants were significantly more accurate than Turkish participants in both Yellow (US>TK, p<.001; TW>TK, p<.001) and Red (US>TK, p<.001, TW>TK, p=.011) conditions.

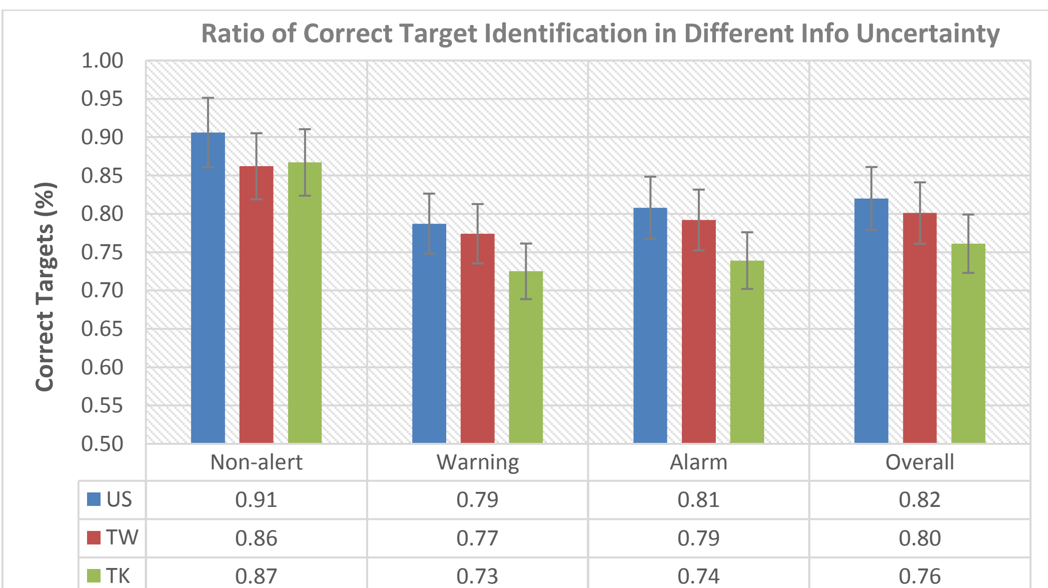


Figure 9. Ratio of correct target identification in different information uncertainty conditions.

6.5 Behavioral data- checking behaviors in payload tasks

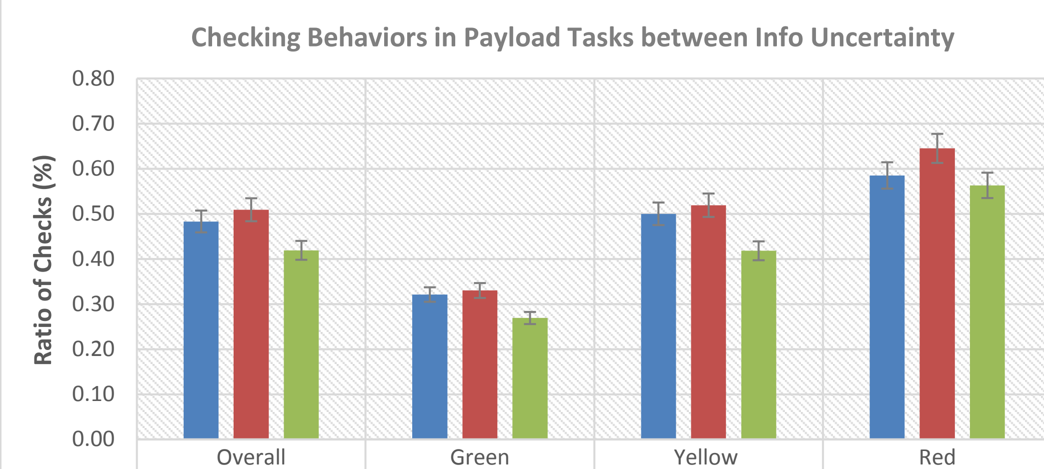
While participants were presented automated assistance (target finder) for payload tasks, before accepting or rejecting the automation's identification, the participants were allowed to verify the marked target by selecting "check" to view the picture with higher resolution to confirm the identification. The results for checking behavior (Table II) showed a main effect for culture ($F_{2,672}$=7.855, *p*<.001), and also found significant interactions between workload and reliability ($F_{1,672}$=7.512, *p*=.006) and between country and reliability ($F_{2,672}$=3.425, *p*=.033). Post-hoc analyses again found no significant differences in use of checks between the U.S. and Taiwan, while both were more likely to check identifications than Turkey (US>TK, *p*=.018, TW>TK, *p*<.001).

Table II. Checking behaviors in payload tasks across experimental conditions

Checking Behaviors	F-value	*p*-value	Post-hoc
Workload	$\mathbf{F_{1,672}}$= 2.262	.133	Not Significant
Reliability	$\mathbf{F_{1,672}}$= .296	.586	Not Significant
Country	$\mathbf{F_{2,672}}$= 7.855	<.001	US≈TW (*p*=.281) US>TK (*p*=.018) TW>TK (*p*<.001)
Workload x Reliability	$\mathbf{F_{1,672}}$= 7.512	.006	LW_HR (.516)>LW_LR (.453) HW_HR (.477)>HW_LR (.435)
Country x Reliability	$\mathbf{F_{2,672}}$= 3.425	.033	US_LR (.496)>US_HR (.470) TW_HR (.549)>TW_LR (.468) TK_LR (.431)>TK_HR (.408)

6.5.1 Behavioral data- checking behaviors in payload tasks by uncertainty level

The ratio of checks to engagements was calculated by dividing the number of checks in each color level by the number of engagements in that level. Cultural differences were found for the overall comparison ($F_{2,672}$=7.855, *p*<.001) as well as within each of the cue conditions (Red: $F_{2,672}$=4.205, *p*=.015; Green: $F_{2,672}$=3.099, *p*=.046; Yellow: $F_{2,672}$=9.141, *p*<.001), Fig. 10. The results also showed that the number of checks in the Red condition were significantly higher than Yellow (*p*<.001) as well as Green (*p*<.001) conditions, and checking in the Yellow state was higher than the Green (*p*<.001) state. Post-hoc analysis showed that the U.S. and Taiwanese participants had significantly higher checking rates than the Turkish participants in the Yellow (US>TK, *p*=.003; TW>TK, *p*<.001) and overall conditions (US>TK, *p*=.018; TW>TK, *p*<.001). In addition, the Taiwanese participants checked more frequently than Turkish participants in the alarm condition (p=.016) and marginally more often in the green non-alert (p=.065) condition.

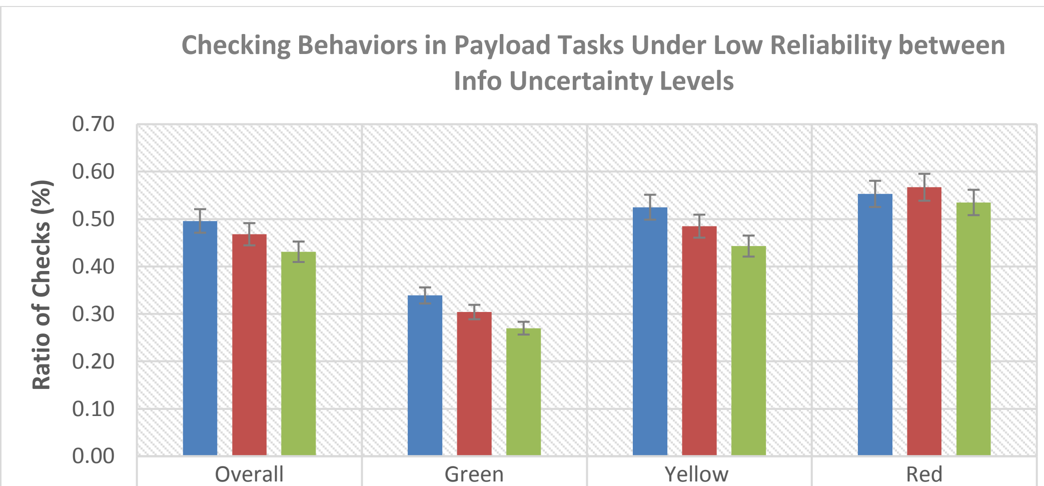


	Overall	Green	Yellow	Red
U.S.	0.48	0.32	0.50	0.59
Taiwan	0.51	0.33	0.52	0.65
Turkey	0.42	0.27	0.42	0.56

Figure 10. Checking behaviors in payload tasks between information uncertainty conditions.

6.5.2 Behavioral data- effects of reliability and uncertainty level on checking behavior

To examine the effects of reliability and uncertainty level on checking we again tested the ratio between checks and engagements at each uncertainty level. An ANOVA showed no significant difference in the low reliability condition (Fig. 11).



	Overall	Green	Yellow	Red
U.S.	0.50	0.34	0.53	0.55
Taiwan	0.47	0.30	0.49	0.57
Turkey	0.43	0.27	0.44	0.54

Figure 11. Checking behaviors in payload tasks under low reliability between information uncertainty levels.

By contrast an ANOVA showed a significant effect in HR condition Overall ($F_{2,336}$=9.730, p<.001) and for Red ($F_{2,336}$=6.224, p=.002) and Yellow ($F_{2,336}$=10.393, p<.001); Fig. 12. T-tests found that Taiwanese participants exhibited significantly higher checking patterns than those of the other two cultures with the fewest checking behaviors observed in the Turkish participants and the U.S. participants falling in between.

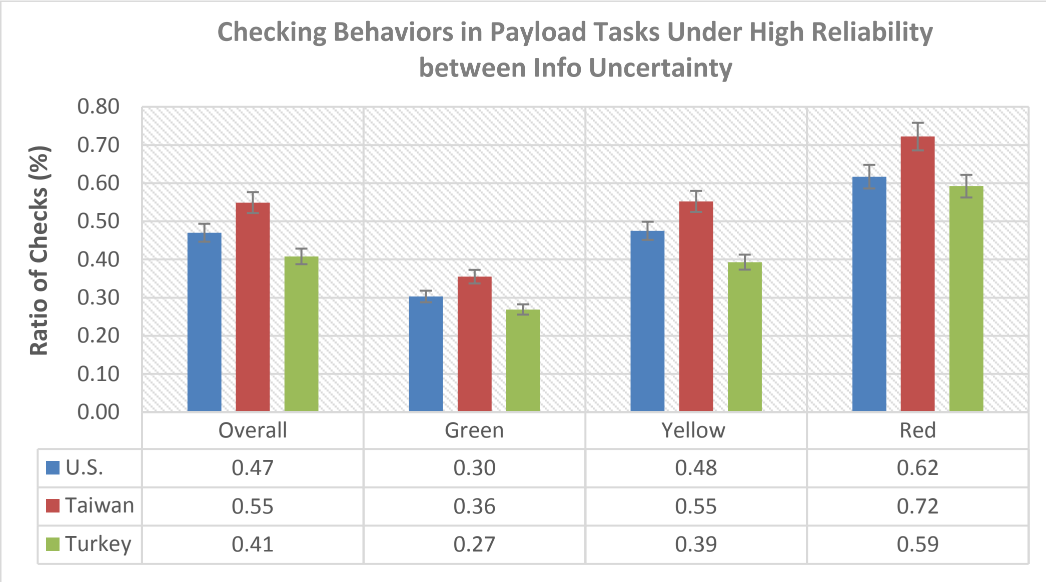


Figure 12. Checking behaviors in payload tasks between information uncertainty conditions.

6.6 Behavioral data- compliance/reliance in payload tasks

Operators' compliance behavior (Table III) was measured as the ratio of agreement to total number of engagements between operator and target finder in the Red and Yellow uncertainty conditions where targets were explicitly marked and alarmed. Compliance was higher in HW and HR conditions and varied by country with U.S. and Taiwanese groups complying more often with correct identifications and less often with incorrect ones than the Turkish group. In the Green uncertainty condition U.S. and Taiwanese groups relied more overall and agreed with more correct rejections while catching more automation suggested misses than the Turkish group. The effects of task load and reliability on reliance, however, were reversed from compliance with highest overall reliance and correct reliance occurring in low task load and low reliability conditions. Misses, however, were caught with greater frequency in HW condition.

Table III. Compliance and reliance

Overall compliance [Agree on Hits + False Alarm prone] Red and Yellow certainty levels				
Variable	*F-value*	*p-value*	*η²*	*Effect*
Country	$F_{2,672}$= 992	P=.371	0.003	Not Significant
Workload	$F_{1,672}$= 12.977	P<.001	0.019	HW>LW
Reliability	$F_{1,672}$= 2280.396	P<.001	0.772	HR>LR

Compliance [Agree on Hits- Red & Yellow]				
Variable	*F-value*	*p-value*	*η^2*	*Effect*
Country	$F_{2,672}$= 16.079	P<.001	0.046	US≈TW, N.S. US>TK, p<.001 TW>TK, p<.001
Workload	$F_{1,672}$= 529.757	P<.001	0.441	HW>LW
Reliability	$F_{1,672}$= 12609.985	P<.001	0.949	HR>LR
Non-compliance [Disagree on False Alarm-Red & Yellow]				
Variable	*F-value*	*p-value*	*η^2*	*Effect*
Country	$F_{2,672}$= 5.704	P=.003	0.017	US≈TW, N.S. US>TK, p=.002 TW>TK, N.S.
Workload	$F_{1,672}$= 44.614	P<.001	0.062	LW>HW
Reliability	$F_{1,672}$= 3036.500	P<.001	0.819	LR>HR
Overall reliance [Agree on Correct Rejection & Miss Green]				
Variable	*F-value*	*p-value*	*η^2*	*Effect*
Country	$F_{2,672}$= 4.413	p=.012	0.013	US≈TW, N.S. US>TK, p=.055 TW>TK, p=.019
Workload	$F_{1,672}$= 486.635	p<.001	0.420	LW>HW
Reliability	$F_{1,672}$= 10.552	p=.001	0.015	LR>HR
Correct reliance [Agree with Correct Rejection- Green]				
Variable	*F-value*	*p-value*	*η^2*	*Effect*
Country	$F_{2,672}$= 4.115	p=.017	0.012	US≈TW, N.S. US>TK, p=.027 TW>TK, p=.061
Workload	$F_{1,672}$= 562.973	p<.001	0.456	LW>HW
Reliability	$F_{1,672}$= 11.809	p=.001	0.017	LR>HR
Correct Non-reliance [Disagree with Miss- Green]				
Variable	*F-value*	*p-value*	*η^2*	*Effect*
Country	$F_{2,672}$= 4.019	p=.018	0.012	US≈TW, N.S. US>TK, p=.014 TK≈TW, N.S.
Workload	$F_{1,672}$= 3299.239	p<.001	0.831	HW>LW
Reliability	$F_{1,672}$= .907	p=.341	0.001	Not Significant

6.7 Behavioral data- behaviors after experiencing the first failure in payload tasks

Prior research has suggested that operators' trust and use of automation declines after a failure/error. To test this effect, we examined the following behaviors at payload tasks after an operator experiences the first automation failure (Fig. 13). Following behaviors are defined as decisions (hit/safe) consistent with the display's certainty level (i.e., 'hit' in Red or Yellow, and 'safe' in Green). Over-reliance occurs when this decision is incorrect (FA in Red or Yellow; Miss in Green) while under-reliance occurs when operator makes a decision both inconsistent with the certainty level and incorrect.

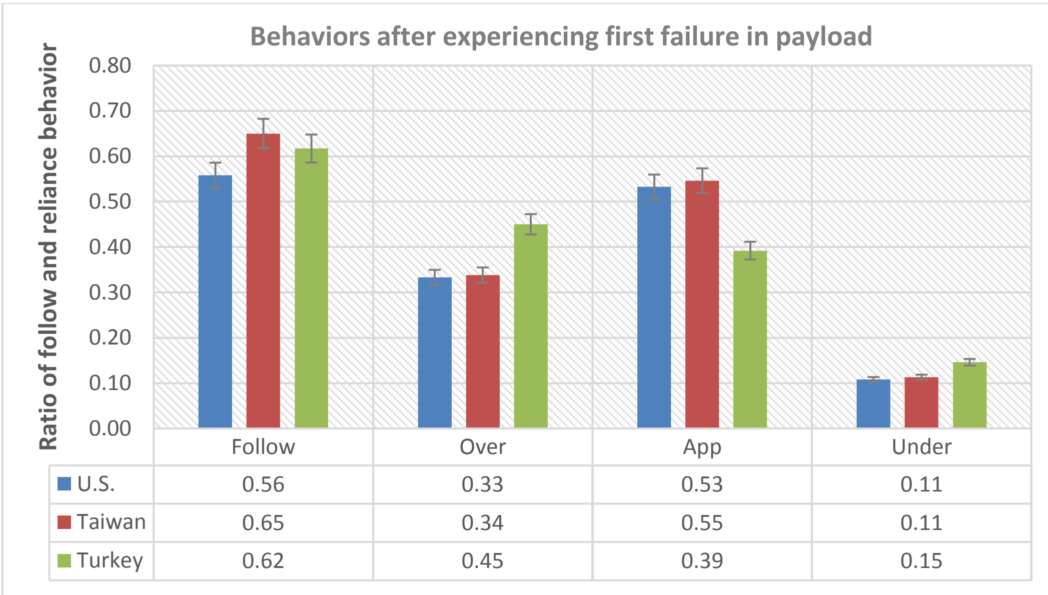


Figure 13. Following and reliance behaviors after experiencing the first failure in payload tasks.

No overall effect was found for following behaviors ($F_{2,672}$=2.509, p=.082). However, significant cultural differences were observed in over-reliance ($F_{2,672}$=7.113, p=.001) and appropriate-reliance ($F_{2,672}$=8.359, p<.001), in which Turkish participants had significantly higher over-reliance (TK>US, p=.003; TK>TW, p=.004) and lower appropriate reliance (US>TK, p=.002; TW>TK, p=.001) than the U.S. and Taiwanese participants. However, no difference was seen in the under-reliance results.

7. DISCUSSION

The present study investigated the impact of cultural factors on trust in automation, in theoretically guided experiments conducted in the U.S., Taiwan and Turkey, with 120 student participants from each of the countries (in total, 360 responses were collected). Task load and automation reliability of the target detector subtask of a modified RESCHU multi-UAV simulation were manipulated to investigate their effects on use of automation and trust measured using our previously validated cross-cultural trust questionnaires. US and Taiwanese participants with a few notable exceptions were highly similar in levels of trust in automation and their responses to changes in reliability and task load. Both differed substantially from the Turkish sample which exhibited lower initial trust and appeared less effective in calibrating trust and reliance on automation.

7.1 General trust

As hypothesized from the theory of cultural syndromes, an individual from a dignity culture (e.g., America) should have a higher level of general trust, than an individual from an honor culture (e.g., Turkey). Our first hypothesis assumed that *individuals from dignity cultures are more likely to have higher levels of initial trust in automation than those from both honor and face cultures*. These cultural effects were confirmed, with Turkish participants having the lowest scores in general trust in automation and American participants having the highest initial trust scores, with those participants from a face culture (e.g., Taiwan) falling in between. Contributions of constructs, however, were distinct with Turkish respondents emphasizing performance but not process, Taiwanese de-emphasizing task context and U.S. participants favoring both Performance and Process as bases for trust. The lack of Turkish concern for process may result from the lack of faith in lawfulness typifying Honor cultures while Taiwanese disregard for context may reflect limitations in the perceived scope of automation benefits consistent with Face society restrictions on roles. The U.S. sample, by contrast, valued Process most highly and relative to the other countries placed less emphasis on Performance and more on Context. This dual focus on Process and Context may reflect sensitivity to the appropriateness of automation, a feature of U.S. performance at the target-identification task.

7.2 Effect of task load

Raising task load by increasing UAV speed required operators to allocate more attention to the navigation subtask, leading to fewer resources available for payload tasks, at the very time UAVs are reaching their targets more rapidly creating new payload tasks at a higher rate. Increases in task load led to increases in compliance (Red and Yellow uncertainty levels) as often observed when operators must use automation to keep up with task demands [Daly 2002; McBride et al. 2011; Rovira et al. 2007; Rajaonah et al. 2008; Biros et al. 2004; Willems et al. 2002; Wang et al. 2011]. The slight performance advantage among Taiwanese and Turkish participants under high workload may result from this increase in compliance as their performance is in the vicinity or below that of the automation (see high reliability condition in Fig. 8 where automation is 80% correct). Overall reliance (Green uncertainty level), however, was higher in the low workload condition, accounted for primarily by higher rates of agreement for both trials without (correct) and with (incorrect) targets to be found. These results are consistent with Meyer's [2004] dual process theory of reliance and compliance and Rice's [2009] findings of relative independence between the processes.

A marginal increase in trust was found for the high task load condition on the performance subscale (η^2=.006, p=.051). This difference held for overall ratings of trust in the Taiwanese (p=.045) and Turkish (p=.003) groups but not the U.S. sample and contradicts a frequent finding in prior U.S. studies [Willems et al. 2002; Biros et al. 2004; Karpinsky et al. 2016; Rovira et al. 2007] of reduced trust under high workload conditions. A potential explanation for the higher degree of trust in the high task load condition might be that additional feedback, known to enhance trust [Hancock, Billings and Schaefer 2011], was provided by a larger sample of engagements. A relation between perceived workload and trust, again positive, was found but only for the U.S. sample and was very slight (r=.148, p=.022) accounting for only 2% of the variance mirroring Rajaonah's [2008] reports of no effect. No interaction as reported by [Daly 2002; Wetzel 2005] was found between task load and reliability in affecting trust. These findings confirm: *(Hypothesis-9) operators will accept more automated recommendations or exhibit fewer checking behaviors on automation under high* *workload conditions*. The American and Taiwanese participants had higher trust ratings in the target finder than Turkish participants with similar trust scores across most of the comparisons except purpose where the Taiwanese were higher. The slight increase in trust with task load found for the Turkish and Taiwanese but not U.S. participants suggests more accurate trust calibration in the U.S. sample but may represent a real cultural difference in the attribution of trust.

7.3 Effect of source reliability

With increased reliability, as expected, participants agreed more often with the automation, correctly identified more targets and provided higher ratings of specific trust in the automation (target finder) confirming: (*Hypothesis-3) Unreliable automation will lower ratings of trust of operators from all cultures (both low- and high-PD).* Particularly strong effects were found with correct compliance (η^2=.949, p<.001) as well as overall compliance (η^2=.772, p<.001) substantially exceeding the low reliability condition. Although false alarms were more likely to be corrected in the low reliability condition this is likely due to their higher probability and greater numbers. Overall (η^2=.015, p=.001) and correct reliance (η^2=.017, p=.001) by contrast were slightly higher in the low reliability condition but inconsequential relative to reliability effects on compliance.

Both American and Taiwanese participants reached similar levels of overall trust in the target finder in both high reliability (HR) and low reliability (LR) conditions, with ratings substantially higher than those of Turkish participants across reliability conditions. Therefore, our hypothesis, *(Hypothesis-7) honor culture operators will take longer interaction times than operators from dignity and face cultures to develop equal degrees of trust*, was confirmed. Calibration of trust was highest in the U.S. group and only slightly lower in the Taiwanese sample both of which were substantially higher than Turkish participants confirming hypothesis five. *(Hypothesis-5) Operators from Dignity and Honor cultures will be more self-confident and therefore are less likely to rely on or ignore the automation than Face culture operators.* In addition, As in *(Hypothesis-10) the trust of face culture operators will be relatively more influenced by information about the purpose-benevolence of automation than honor or dignity culture operators*, the differences in system reliability may greatly affect the operators' perceived purpose of automation. However, this hypothesis was not supported, in which increased source reliability failed to strengthen the Taiwanese participants' trust attitudes in the designed purpose of automation, rather than the other two cultures. In addition, our results partially supported *(Hypothesis-2) If using the automation were encouraged by the user's organization, Face culture operators will have higher ratings of trust and reliance than those from Honor and Dignity cultures.* Operators from the U.S. and Taiwanese cultures reported similar ratings of trust in the target finder regardless of reliability conditions, which were significantly higher than the Turkish participants.

7.4 Effect of checking behaviors

To clarify target identity, participants were allowed to check a higher resolution image before accepting or rejecting the target finder's identification. The most frequent checking was observed in the Red high certainty condition and the lowest in the Green low target probability condition with the Yellow warning condition falling in between. This suggests that participants may have responded to the compliance demands of alerting rectangles in the Red and to a lesser extent Yellow conditions to check for false alarms while checking in the Green reliance condition may have been limited to trials on which a possible target was observed. American and Taiwanese participants had higher checking rates than the Turkish participants for Yellow certainty and overall conditions. However, no difference was observed between American and Turkish participants in the Red alarm and Green non-alert conditions, whereas Taiwanese participants checked more often than the other two cultures.

Although the results showed little difference in the number of checks between reliability levels, cultural differences were found for the rate of checking, with American and Taiwanese participants checking more frequently than Turkish participants before making their decisions. Although American and Taiwanese operators identified similar numbers of targets the U.S. group used fewer checks to achieve that outcome. The results also revealed that American participants were significantly better in calibrating their trust than participants from the other two cultures having higher levels of appropriate reliance (agreeing with automation when correct and disagreeing when incorrect) and lower levels of over (agreeing with automation when incorrect) or under (disagreeing when correct) trust. The Taiwanese group was slightly less well calibrated while the Turkish sample was lower in appropriate reliance and higher in both over and under reliance.

As overall rates of agreement were comparable among the three samples the data suggest that Turkish operators were no less compliant and reliant than the others but merely less effective in their use of the automation. The resulting pattern in which the Turkish sample reported trusting less on both the initial measure of general trust and the later task specific measures yet checked less often to verify the (less trusted) target finder's identification, yet agreed with the automation at the same rate as the other cultures complicates the simple notion of trust supporting reliance. McBride et al. [2011] used a similar checking mechanism interpreting the absence of checks as an indication of trust and reliance. Our experiment showed exactly the opposite relation with groups with higher ratings of trust checking at higher rates while the group lowest in trust also checked less frequently. These findings disconfirm (*Hypothesis-6) Honor and face culture operators will exhibit more vigilance and more monitoring behavior than operators from dignity cultures* and suggest that our intuitions about trust and its influence on behavior may not always hold in generalizing to other cultures.

Human trust, compliance, and reliance typically drop after encountering an error and recover over subsequent error free trials [Lee andMoray 1994]. After experiencing an initial target finder failure, either from accepting an incorrect suggestion (over-reliance) or rejecting a correct non-identification (under-reliance), Taiwanese participants tended to agree with the next identification. American participants were least likely to accept a target finder identification immediately after the first error with the Turkish sample falling in between. These findings partially confirm (*Hypothesis-3) unreliable automation will lower ratings of trust of operators from all cultures (both low- and high-PD) operators, but face culture operators will be more likely to continue relying on automation*, and partially confirm *Hypothesis-4, face culture operators will recover their trust in automation after failure more quickly than honor and dignity culture operators. Honor culture operators would be slowest in recovering trust.* Once again the distinction seems to lie in appropriateness of reliance. Surprisingly operators from all three cultures were more likely to over-trust the automation on the trial following its first error than they were for the full session. U.S. participants, however, showed less over and under reliance than the other groups while matching the Taiwanese in appropriate reliance. The assumption of (*Hypothesis-8) honor operators will either disuse or take longer to regain trust after a failure occurs and may not recover trust to the original level (miscalibrate), as compared with face and dignity operator. appears* disconfirmed as Turkish participants relied on automation following the first failure at a rate statistically no different from the Taiwanese and higher than the U.S. participants.

8. CONCLUSION

Despite an array of cultural differences brought up in the discussion section substantial commonalities were found in the development of trust and its effects on use of automation across the three cultures. The trust instruments we developed found the same three constructs worked to successfully characterize attitudes of trust in automation in all three cultures [Chien et al. 2015]. The general relations between reliability, trust, and reliance were robustly supported in this study. Workload effects were more equivocal with a slight tendency toward higher ratings of trust at higher workloads found for non-U.S. participants, counter to earlier U.S. only findings. The relations between trust, reliance, and target checking found in the Turkish sample were unexpected and an illustration of why cross-cultural research of this sort is needed to avoid design assumptions that may not hold outside of our own cultures.

ACKNOWLEDGMENTS

This research has been sponsored by AFOSR FA9550-13-1-0129.

REFERENCES

Soroush Aslani, Jimena Ramirez-Marin, Zhaleh Semnani-Azad, Jeanne Brett, and Catherine Tinsley. 2013. *Dignity Face and Honor cultures implications for negotiation and conflict management*.

Soroush Aslani, Jimena Ramirez-Marin, Jeanne Brett, Jingjing Yao, Zhaleh Semnani-Azad, Zhi-Xue Zhang, Catherine Tinsley, Laurie Weingart, and Wendi Adair. 2016. Dignity, face, and honor cultures: A study of negotiation strategy and outcomes in three cultures. *J. Organ. Behav.* 37, 8, 1178–1201.

Nathaniel Bean, Stephen Rice, and David Keller. 2011. The Effect of Gestalt Psychology on the System-Wide Trust Strategy in Automation. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 55, 1 (2011), 1417–1421. DOI:https://doi.org/10.1177/1071181311551295

Charles E. Billings, John K. Lauber, Hallie Funkhouser, E. Gene Lyman, and Edward M. Huff. 1976. NASA aviation safety reporting system quarterly report.

David P. Biros, Mark Daly, and Gregg Gunsch. 2004. The influence of task load and automation trust on deception detection. *Gr. Decis. Negot.* 13, 2 (2004), 173–189.

Iris Bohnet, Benedikt Herrmann, and Richard Zeckhauser. 2009. The Requirements for Trust in Gulf and Western Countries. *Q. J. Econ.*

Iris Bohnet and Richard Zeckhauser. 2004. Trust, risk and betrayal. *J. Econ. Behav. Organ.* 55, 4 (December2004), 467–484. DOI:https://doi.org/10.1016/j.jebo.2003.11.004

Yves Boussemart and Missy Cummings. 2008. Behavioral recognition and prediction of an operator supervising multiple heterogeneous unmanned vehicles. *Humans Oper. unmanned*.

Joel Brockner, Tom Tyler, and Rochelle Cooper-Schneider. 1992. The influence of prior commitment to an institution on reactions to perceived unfairness: The higher they are, the harder they fall. *Adm. Sci.* 37, 2, 241–261.

Jennifer Carlson, Robin Murphy, and Andrew Nelson. 2004. Follow-up analysis of mobile robot failures. In *IEEE International Conference on Robotics and Automation 2004 Proceedings ICRA 04 2004*. Ieee, 4987–4994. DOI:https://doi.org/10.1109/ROBOT.2004.1302508

Jude Cassidy. 1988. Child-mother attachment and the self in six-year-olds. *Child Dev.* 59, 1 (1988), 121–134.

Shih-Yi Chien, Michael Leiws, Katia Sycara, Jyi-Shane Liu, and Asiye Kumru. 2017. The Influence of Cultural Factors on Trust in Automation [Data set]. Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E100532V2

Shih-Yi Chien, Michael Leiws, Katia Sycara, Jyi-Shane Liu, and Asiye Kumru. 2016. Relation between Trust Attitudes Toward Automation, Hofstedes Cultural Dimensions, and Big Five Personality Traits. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 60, 1 (September2016), 841–845.

Shih-Yi Chien, Siddharth Mehrotra, Michael Lewis, and Katia Sycara. 2013. Imperfect Automation in Scheduling Operator Attention on Control of Multi-Robots. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*.

Shih-Yi Chien, Michael Lewis, Sebastian Hergeth, Zhaleh Semnani-Azad, and Katia Sycara. 2015. Cross-Country Validation of a Cultural Scale in Measuring Trust in Automation. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*

Shih-Yi Chien, Michael Lewis, Katia Sycara, Jyi-Shane Liu, and Asiye Kumru. 2016. Influence of Cultural Factors in Dynamic Trust in Automation.

Shih-Yi Chien, Zhaleh Semnani-azad, Michael Lewis, and Katia Sycara. 2014. Towards the Development of an Inter-cultural Scale. In Springer International Publishing, 35–46.

Mark Daly. 2002. *Task load and automation use in an uncertain environment*. Air Force INST of tech wright-patterson afb oh school of engineering and management.

Kurt Dirks, Roy Lewicki, and Akbar Zaheer. 2009. Reparing Relationships Within and Between Organizations: Building A Conceptual Foundation. *Acad. Manag. Rev.* 34, 1, 68–84.

Ashley Fulmer and Michele Gelfand. 2010. Dynamic Trust Processes: Trust Dissolution and Restoration. 1–32.

Ken Funk , Beth Lyall, Jennifer Wilson, Rebekah Vint, Mary Niemczyk, Candy Suroteguh, and Griffith Owen. 1999. Flight Deck Automation Issues. *Int. J. Aviat. Psychol.* 109–123.

Daniel Gartenberg, Leonard Breslow, Malcolm J. McCurry, and Gregory Trafton. 2012. Time Pressure, Memory, and Task Knowledge Facilitate the Opportunism Heuristic in Dynamic Tasks. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 56, 2011 (2012), 1025–1029. DOI:https://doi.org/10.1177/1071181312561214

Peter Goillau, Vicki Woodward, Chris Kelly, and Gill Banks. 2003. Guidelines for Trust in Future ATM Systems : A Literature Review Guidelines for Trust in Future ATM Systems : A Literature Review Guidelines for Trust in Future ATM Systems : A Literature.

Brian C. Gunia, Jeanne M. Brett, Dishan Kamdar, and Amit Nandkeolyar. 2011. Paying a price: culture, trust, and negotiation consequences. *J. Appl. Psychol.* 96, 4 , 774–89. DOI:https://doi.org/10.1037/a0021986

Peter Hancock, Deborah Billings, and Kristine Schaefer. 2011. Can You Trust Your Robot? *Ergon. Des. Q. Hum. Factors Appl.* 19, 3, 24–29. DOI:https://doi.org/10.1177/1064804611415045

Sandra Hart. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human Mental Workload*. North-Holland , 139–183. DOI:https://doi.org/10.1016/S0166-4115(08)62386-9

Steven Heine. 2001. Self as cultural product: An examination of East Asian and North American selves. *J. Pers.* 69, 6, 881–906.

Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust . *Hum. Factors J. Hum. Factors Ergon. Soc.* 57, 3 (2015), 407–434. DOI:https://doi.org/10.1177/0018720814547570

Geert Hofstede. 1991. Cultures And Organizations - Software of the Mind. *Development* (1991), 1–29.

Geert Hofstede. 1973. National Culture. Retrieved May 22, 2014 from https://geert-hofstede.com/national-culture.html

Robert House, Paul Hanges, Mansour Javidan, Peter Dorfman, and Vipin Gupta. 2004. *Culture, leadership, and organizations*.

Jiun-Yin Jian, Ann M. Bisantz, and Colin G. Drury. 2000. Foundations for an empirically determined scale of trust in automated systems. *Int. J. Cogn.*.

Oliver John and Sanjay Srivastava. 1999. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Hand b. Personal. Theory*, 510.

Gareth R. Jones and Jennifer M.George. 1998. The Experience and Evolution of Trust: Implications for Cooperation and Teamwork. *Acad. Manag. Rev.* 23, 3 (July1998), 531. DOI:https://doi.org/10.2307/259293

Nicole D. Karpinsky, Eric T. Chancey, and Yusuke Yamani. 2016. Modeling Relationships among Workload , Trust , and Visual Scanning in an Automated Flight Task. In Human Factors and Ergonomics Society 2016 Annual Meeting, 1550–1554.

David Keller and Stephen Rice. 2009. System-Wide versus Component-Specific Trust Using Multiple Aids. *J. Gen. Psychol.* 137, 1 (2009), 114–128. DOI:https://doi.org/10.1080/00221300903266713

Arie Kruglanski, Erik Thompson, Tory Higgins, Nadir Atash, Antonio Pierro, and James Y. Shah. 2000. To“ do the right thing” or to“ just do it”: locomotion and assessment as distinct self-regulatory imperatives. *J. Personal.* 79, 5 (2000), 793–815.

John D. Lee and Neville Moray. 1994. Trust, self-confidence, and operators' adaptation to automation. *Int. J. Hum. Comput. Stud.* 40 (1994), 153–184. DOI:https://doi.org/10.1006/ijhc.1994.1007

John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Hum. Factors J. Hum. Factors Ergon. Soc.* 46, 1 (2004), 50–80.

John D. Lee and Neville Moray. 1992. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics* 35, 10 (1992), 1243–1270.

John D. Lee and Katrina A. See. 2004. Trust in automation: designing for appropriate reliance. *Hum. Factors* 46, 1 (January2004), 50–80.

Angela Leung and Dov Cohen. 2011. Within-and between-culture variation: Individual differences and the cultural logics of honor, face, and dignity cultures. *J. Pers. Soc. Psychol.* 100, 3 (2011), 507–526.
James Llinas, Ann Bisantz, Colin Drury, Younho Seong, and Jiun-Yan Jian. 1998. Studies and Analyses of Aided Adversarial Decision Making Phase 2: Research on Human Trust in Automation . *State Univ. New York Buffalo, Cent. Multisource Inf. Fusion,Department Ind. Eng.* , May 1998 (1998), 1.
Joseph Lyons, Charlene Stokes, Kevin Eschleman, Gene Alarcon, and Alex Barelka. 2011. Trustworthiness and IT Suspicion: An Evaluation of the Nomological Network. *Hum. Factors J. Hum. Factors Ergon. Soc.* 53, 3 (May2011), 219–229. DOI:https://doi.org/10.1177/0018720811406726
Joseph Lyons, Kolina S. Koltai, Nhut Ho, Walter B. Johnson, David E. Smith, and R. Jay Shively. 2016. Engineering Trust in Complex Automated Systems. *Ergon. Des.* , January (2016), 13–17. DOI:https://doi.org/10.1177/1064804615611272
Maria Madsen and Shirley Gregor. 2000. Measuring human-computer trust. *11th Australas. Conf. Inf. Syst.* 53 (2000), 6–8.
Nikolas Martelaro, Victoria C. Nneji, Wendy Ju, and Pamela Hinds. 2016. Tell me more: Designing HRI to encourage more trust, disclosure, and companionship. *ACM/IEEE Int. Conf. Human-Robot Interact.* 2016–April (2016), 181–188. DOI:https://doi.org/10.1109/HRI.2016.7451750
Roger C. Mayer, James H. Davis, and F. David Schoorman. 1995. An integrative model of organizational trust. *Acad. Manag. Rev.* 20, 3 (1995), 709–734.
Sara McBride, Wendy Rogers, and Arthur Fisk. 2011. Understand ing the Effect of Workload on Automation Use for Younger and Older Adults. *Hum. Factors J. Hum. Factors Ergon. Soc.* 53, 6 (2011), 672–686. DOI:https://doi.org/10.1177/0018720811421909
Joachim Meyer. 2001. Effects of warning validity and proximity on responses to warnings. *Hum. Factors* 43, 4 (2001), 563–572. DOI:https://doi.org/10.1518/001872001775870395
Joachim Meyer. 2004. Conceptual issues in the study of dynamic hazard warnings. *Hum. Factors* 46, 2 (2004), 196–204.
Robert Molloy and Raja Parasuraman. 1996. Monitoring an Automated System for a Single Failure: Vigilance and Task Complexity Effects. *Hum. Factors J. Hum. Factors Ergon. Soc.* 38, 2 (1996), 311–322. DOI:https://doi.org/10.1177/001872089606380211
Neville Moray and Toshiyuki Inagaki. 1999. Laboratory studies of trust between humans and machines in automated systems. *Trans. Inst. Meas. Control* 21, 4–5 (January1999), 203–211. DOI:https://doi.org/10.1177/014233129902100408
Neville Moray. 2000. Adaptive Automation , Trust , and Self-Confidence in Fault Management of Time-Critical Tasks. 6, 1 (2000), 44–58. DOI:https://doi.org/10.10371/0278-7393.6.1.44
Neville Moray. 2003. Monitoring, complacency, scepticism and eutactic behaviour. *Int. J. Ind. Ergon.* 31, 3 (2003), 175–178. DOI:https://doi.org/10.1016/S0169-8141(02)00194-4
Bonnie Muir. 1994. Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics* 37, 11 (1994), 1905–1922. DOI:https://doi.org/10.1080/00140139408964957
Bonnie Muir and Neville Moray. 1996. Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics* 39, 3 (1996), 429–460.
Richard Nisbett and Dov Cohen. 1997. Culture of Honor The Psychology of Violence in the South. *South. Cult.* 3, 1 (1997), 80–83. DOI:https://doi.org/10.1353/scu.1997.0033
Raja Parasuraman, Robert Molloy, and Indramani Singh. 1993. Performance consequences of automation induced "complacency." *Int. J. Aviat. Psychol.* 2 (1993), 1–23. DOI:https://doi.org/10.1207/s15327108ijap0301_1
Raja Parasuraman and Dietrich Manzey. 2010. Complacency and Bias in Human Use of Automation: An Attentional Integration. *Hum. Factors J. Hum. Factors Ergon. Soc.* 52, 3 (October2010), 381–410. DOI:https://doi.org/10.1177/0018720810376055
Raja Parasuraman, Mustapha Mouloua, and Robert Molloy. 1994. Monitoring automation failures in human-machine systems. *Hum. Perform. Autom. Syst. Curr. Res. Trends* (1994), 45–49.
Raja Parasuraman and Victor Riley. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. *Hum. Factors J. Hum. Factors Ergon. Soc.* 39, 2 (June1997), 230–253. DOI:https://doi.org/10.1518/001872097778543886
Raja Parasuraman, Thomas Sheridan, and Christopher D. Wickens. 2000. A model for types and levels of human interaction with automation. *IEEE Trans. Syst. Man. Cybern. A Syst. Hum.* 30, 3, 286–97.
Raja Parasuraman, Thomas B. Sheridan, and Christopher D. Wickens. 2008. Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs. *J. Cogn. Eng. Decis. Mak.* 2, 2, 140–160. DOI:https://doi.org/10.1518/155534308X284417.
Julian Pitt-Rivers. 1966. Honour and social status. *Honour shame values.*
Bako Rajaonah, Nicolas Tricot, Françoise Anceaux, and Patrick Millot. 2008. The role of intervening variables in driver-ACC cooperation. *Int. J. Hum. Comput. Stud.* 66, 3, 185–197. DOI:https://doi.org/10.1016/j.ijhcs.2007.09.002

Patrick Rau, Ye Li, and Dingjun Li. 2009. Effects of communication style and culture on ability to accept recommendations from robots. *Comput. Human Behav.* 25, 2 (2009), 587–595. DOI:https://doi.org/10.1016/j.chb.2008.12.025

Stephen Rice. 2009. Examining Single- and Multiple-Process Theories of Trust in Automation. *J. Gen. Psychol.* 136, 3 (2009), 303–319. DOI:https://doi.org/10.3200/GENP.136.3.303-322

Victor Riley. 1994. Human use of automation. *Unpubl. Dr. Diss.* University (1994), Department of Psychology.

Julian Rotter. 1967. A new scale for the measurement of interpersonal trust. *J. Pers.* 35, 4 (December1967), 651–65.

Julian Rotter. 1971. Generalized expectancies for interpersonal trust. *Am. Psychol.* (1971).

Ericka Rovira, Kathleen McGarry, and Raja Parasuraman. 2007. Effects of imperfect automation on decision making in a simulated command and control task. *Hum. Factors* 49, 1 (2007), 76–87.

Kristin Schaefer, Jessie Chen, James Szalma, and Peter Hancock. 2016. A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems. *Hum. Factors J. Hum. Factors Ergon. Soc.* 58, 3 (2016), 377–400. DOI:https://doi.org/10.1177/0018720816634228

Shalom Schwartz. 1992. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Adv. Exp. Soc. Psychol.* 3 (1992).

Thomas Sheridan and William Verpank. 1978. *Human and computer control of undersea teleoperators*,

Sim Sitkin and Nancy Roth. 1993. Explaining the limited effectiveness of legalistic “remedies” for trust/distrust. *Organ. Sci.* 4, 3 (1993), 367–393.

Randall Spain, Ernesto Bustamante, and James Bliss. 2008. Towards an Empirically Developed Scale for System Trust: Take Two. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 52, 19 (September2008), 1335–1339. DOI:https://doi.org/10.1177/154193120805201907

Randall Spain and James P. Bliss. 2008. The effect of sonification display pulse rate and reliability on operator trust and perceived workload during a simulated patient monitoring task. *Ergonomics* 51, 9 (2008), 1320–1337. DOI:https://doi.org/10.1080/00140130802120234

Richard Thackray and Mark Touchstone. 1989. A Comparison of Detection efficiency on an air traffic control monitoring task with and without computer aiding. *Aviat. Sp. Environ. Med.* 60, 8 (1989), 744–748.

Harry Triandis. 1994. *Culture and Social Behavior*, New York, New York, USA: McGraw-Hill.

Harry Triandis. 1995. *Individualism and collectivism*, Boulder, CO: Westview Press.

Ewar tDeVisser and Raja Parasuraman. 2011. Adaptive Aiding of Human-Robot Teaming : Effects of Imperfect Automation on Performance , Trust , and Workload. *Hum. Factors* 5, 2 (2011), 209–231. DOI:https://doi.org/10.1177/1555343411410160.

Lu Wang, Greg Jamieson, and Justin Hollands. 2011. The effects of design features on users' trust in and reliance on a combat identification system. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 55, 1 (2011), 375–379. DOI:https://doi.org/10.1177/1071181311551077

Lin Wang, Pei-Luen Rau, Vanessa Evers, Benjamin Krisper Robinson, and Pamela Hinds. 2010. When in Rome: The role of culture & context in adherence to robot recommendations. In *2010 5th ACMIEEE International Conference on HumanRobot Interaction HRI*. HRI '10. ACM, 359–366.

Ning Wang, David Pynadath, and Susan Hill. 2016. Trust calibration within a human-robot team: Comparing automatically generated explanations. *ACM/IEEE Int. Conf. Human-Robot Interact.* 2016–April (2016), 109–116. DOI:https://doi.org/10.1109/HRI.2016.7451741

Mark Weber, Deepak Malhotra, and Keith Murnighan. 2004. Normal acts of irrational trust: Motivated attributions and the trust development process. *Res. Organ.*

Jacob Wetzel. 2005. *Driver trust, annoyance, and compliance for an automated calendar system*.

Christopher Wickens and Angela Colcombe. 2007. Dual-task performance consequences of imperfect alerting associated with a cockpit display of traffic information. *Hum. Factors* 49, 5 (2007), 839–850.

Rebecca Wiczorek and Dietrich Manzey. 2014. Supporting Attention Allocation in Multitask Environments: Effects of Likelihood Alarm Systems on Trust, Behavior, and Performance. *Hum. Factors J. Hum. Factors Ergon. Soc.* 56 (2014), 1209–1221.

Ben Willems, Michele Heiney, Northrop Grumman, and Randy Sollenberger. 2002. *Study of an atc baseline for the evaluation of team configurations: Information requirements*,

Toshio Yamagishi, Karen Cook, and Motoki Watabe. 1998. Uncertainty, Trust, and Commitment Formation in the United States and Japan 1. *Am. J. Sociol.* 104, 1 (1998), 165–194.

Boonghee Yoo, Naveen Donthu, and Tomasz Lenartowicz. 2011. Journal of International Consumer Marketing Measuring Hofstede s Five Dimensions of Cultural Values at the Individual Level : Development and Validation of CVSCALE Measuring Hofstede ' s Five Dimensions of Cultural Values at the Individual Level : Devel. *J. Int. Consum. Mark.* 23, August 2011 (2011), 193–210.

APPENDIX A: CULTURE TRUST INSTRUMENT (ENGLISH VERSION)

Dimension	***Survey Items***	**Disagree strongly**	**Disagree**	**Neither agree nor disagree**	**Agree**	**Agree strongly**
General Automation Performance Expectancy	Using a smart phone increases my effectiveness on my jobs.	1	2	3	4	5
	Using a smart phone will improve my output quality.	1	2	3	4	5
	Using a smart phone will increase my chances of achieving a higher level of performance.	1	2	3	4	5
General Automation Process Transparency	The information that a smart phone provides is of high quality.	1	2	3	4	5
	A smart phone provides sufficient information.	1	2	3	4	5
	I am satisfied with the information that a smart phone provides.	1	2	3	4	5
General Automation Cultural-Technological Context	I prefer to use a smart phone to make decisions under high workload situations.	1	2	3	4	5
	Using a smart phone helps me to expend less effort to accomplish tasks.	1	2	3	4	5
	Using a smart phone helps me accomplish tasks with lower risk.	1	2	3	4	5
Specific Automation Performance Expectancy	GPS improves my performance.	1	2	3	4	5
	GPS enables me to accomplish tasks more quickly.	1	2	3	4	5
	GPS increases my productivity.	1	2	3	4	5
Specific Automation Process Transparency	My interaction with GPS is clearly understandable.	1	2	3	4	5
	GPS is user-friendly.	1	2	3	4	5
	GPS uses appropriate methods to reach decisions.	1	2	3	4	5
Specific Automation Purpose Influence	I am confident about the performance of GPS.	1	2	3	4	5
	When an emergent issue or problem arises, I would feel comfortable depending on the information provided by GPS.	1	2	3	4	5
	I can always rely on GPS to ensure my performance.	1	2	3	4	5