# International Journal of Human-Computer Interaction Integrating Transparency, Trust, and Acceptance: The Intelligent Systems Technology Model (ISTAM) Integrating Transparency, Trus...

**Article** *in* International Journal of Human-Computer Interaction · April 2022

DOI: 10.1080/10447318.2022.2070107



**2 authors:**

E.s. Vorm
Naval Air Warfare Center Aircraft Division


David Combs
United States Naval Research Laboratory


Taylor & Francis
Taylor & Francis Group

International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hihc20

# Integrating Transparency, Trust, and Acceptance: The Intelligent Systems Technology Model (ISTAM)

E. S. Vorm & David J. Y. Combs

INTERNATIONAL JOURNAL OF HUMAN-COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2022.2070107

Taylor & Francis
Taylor & Francis Group

Check for updates

# Integrating Transparency, Trust, and Acceptance: The Intelligent Systems Technology Model (ISTAM)

E. S. Vorm[a] and David J. Y. Combs[b]

[a]Laboratory for Autonomous Systems Research, US Naval Research Laboratory, Washington, DC, USA; [b]Laboratory for Autonomous Systems Research, Knexus Research, Washington, DC, USA

**ABSTRACT**
Intelligent systems such as technologies related to artificial intelligence, robotics, machine learning, etc. open new insights into data and expand the concept of work in myriad domains. These technologies, while potentially useful, face high barriers to widespread adoption and acceptance by industries and citizens alike. The complexity and multi-dimensionality inherent in intelligent systems often renders traditional validation efforts (e.g., traceability analysis) impossible. In addition, contexts where predictions or computer-generated recommendations have real-world consequences, such as in medical prognosis, financial investing, or military applications introduce new risks and a host of moral and ethical concerns that can further hinder the widespread adoption of intelligent systems. Naturally, such reluctance by would-be users limits the potential of intelligent systems to solve real-world problems. This article reviews the challenges to technology acceptance through the lens of system transparency and user trust, and extends the Technology Acceptance Model (TAM) structure with issues germane to intelligent systems. We examine several prospective transparency frameworks that could be adopted and used by Human-Computer Interaction (HCI) practitioners involved in systems development. Our intention is to assist practitioners in the design of more transparent systems with a specific eye towards enhancing trust and acceptance in intelligent systems. Further, as a result of our review, we suggest that the well-known TAM should be expanded in the context of intelligent systems to include trust and transparency as key elements of the model. Finally, we conclude with a research agenda that might offer empirical evidence showing how transparency might enhance acceptance and use of intelligent systems.

## 1. Introduction

The field of Human-Computer Interaction (HCI) has historically focused on enhancing user experience and interactions with technology through improved interfaces, modalities, and design (Rogers et al., 2015) In addition to areas such as usability and human factors engineering, HCI is also largely focused on positively affecting the adoption of new emerging technologies. This concept is often referred to as "technology acceptance," and refers to a combination of attitudes and opinions formed by users that result in behavioral intentions towards given technologies (Venkatesh, 1999).

Technology acceptance was widely operationalized and became a validated concept in the late 1990s as a result of the work by individuals such as Davis (1989) as well as (Venkatesh & Davis, 2000) on the Technology Acceptance Model (TAM). The TAM suggests that technology acceptance is a function of users' assessment of (1) a technology's perceived usefulness with regard to mission accomplishment and (2) a technology's perceived ease of use.

According to Venkatesh and Bala (2008), illustrated in Figure 1, perceived ease of use is driven by factors such as a user's computer self-efficacy, perceptions of control/ability to learn the system, general computer anxiety (negative association), potential enjoyment, and intrinsic motivation to use a system. Perceived usefulness, on the other hand, is driven by factors such as actual ease of use (i.e., from the factors noted above), norms of use within the organization, overall job relevance, and enhanced output quality. While these overall hypothesized relationships have been tested and supported empirically in a variety of technologies, to our knowledge, intelligent systems have not been thoroughly examined in the context of the TAM (though, recently such research has occurred with more frequency, see below for additional details).

Recently, however, developments in machine learning and associated data processing techniques have given rise to a vast new slate of new intelligent system types. For example, autonomous systems, or systems that are capable of autonomously perceiving and acting in limited ways, are on the rise in domains from self-driving cars to military defensive weapons systems. In this context, machine learning-based prediction algorithms are quickly taking the place of older rule-based systems, and have proven themselves orders of magnitude more accurate than some traditional technologies (Avati et al., 2017; Rahwan et al., 2019; Starke

CONTACT E. S. Vorm eric.s.vorm.mil@us.navy.mil Laboratory for Autonomous Systems Research, US Naval Research Laboratory, Washington, DC, USA

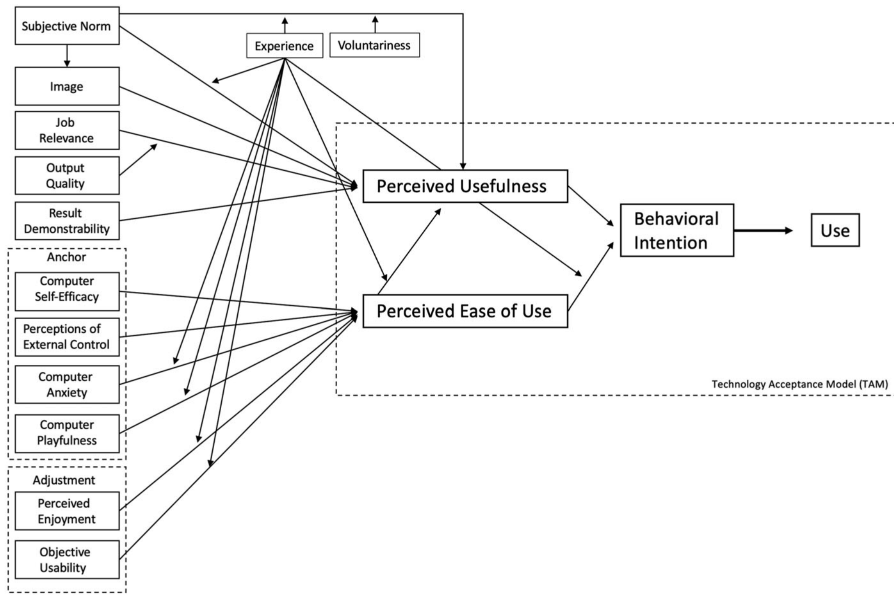


**Figure 1.** The technology acceptance model 3 (Venkatesh & Bala, 2008).

& Baber, 2020). But with these emerging technologies also come new concerns that may affect user technology acceptance.

For example, machine learning techniques such as deep learning are immensely complex and inherently opaque. When their outputs are inaccurate, it is nearly impossible to establish a direct causal understanding because the data and processes involved are vast beyond human comprehension. Autonomous systems built on advanced techniques such as deep reinforcement learning are capable of surpassing human performance at a variety of complex tasks such as playing chess, but their performance is imbalanced and often vulnerable to even small perturbations in their environment. This brittleness often results in system failures that are completely unexpected, and often bizarre, which can affect user experience and may lead some users to discard or avoid these technologies.

Managing complexity, of the sort noted above, in intelligent systems is nothing new for the field of HCI. Decades of research have resulted in a variety of successful design strategies and interface techniques that have stood the test of time and continue to afford users interactions that are safe and pleasant (Bowen et al., 2020; Ren & Bao, 2020). This newest generation of intelligent systems, however, presents a new hurdle that is forcing HCI practitioners to revisit their understanding of the determinants of technology acceptance: trust.

Trust is emerging as a main barrier to the successful rollout of many intelligent systems initiatives. Trust in the context of such systems is a frequent topic in both the popular press as well as academic venues (EU, 2019; Holzinger et al., 2020; IBM, 2020; Meske & Bunde, 2020; Vorm & Miller, 2020). Both policy and decision makers share concerns over the lack of understandability, interpretability, and transparency of these intelligent systems, and see these issues as central to affecting whether or not users will accept and use intelligent systems as they are intended.

As a result, many have attempted to validate trust as a determinant of technology acceptance in intelligent systems (Harrigan et al., 2021; Ma et al., 2021; Patterson-Hann & Watson, 2022; Poon & Sung, 2021; Siegrist, 2021; Van et al., 2021). Although limited results suggest that trust may be a contributor to technology acceptance, these studies commonly suffer from similar limitations: they are frequently small, underpowered studies based largely on online crowd-sourced research participants of questionable quality and often lack grounding in formal theory. This is somewhat understandable. The concept of trust is notoriously difficult to operationalize, and as a result is often difficult to empirically measure (Malle & Ullman, 2021). Additionally, focusing on trust as a determinant for technology acceptance presents HCI practitioners with a fuzzy and ambiguous target: how exactly should a system interface be designed to achieve a goal such as trust?

While user trust may be a desired end state of system developers, it is HCI practitioners who have the lion's share of responsibility for designing and creating the kinds of interactions that lead to appropriate and calibrated user trust. Thus, to gain widespread and appropriate trust in intelligent systems, the field of HCI will need to continue to focus on developing and testing techniques and strategies that can practically and reliably improve user trust through interactions and ultimately lead to enhanced technology acceptance.

To contribute towards this goal, this paper outlines several theories of human trust in the context of interactions with intelligent systems. Specifically, we describe three leading frameworks of trust, and explore the factors that underlie the construct in order to best understand how trust is developed and supported. In section 2 we discuss the concept of transparency and its foundational

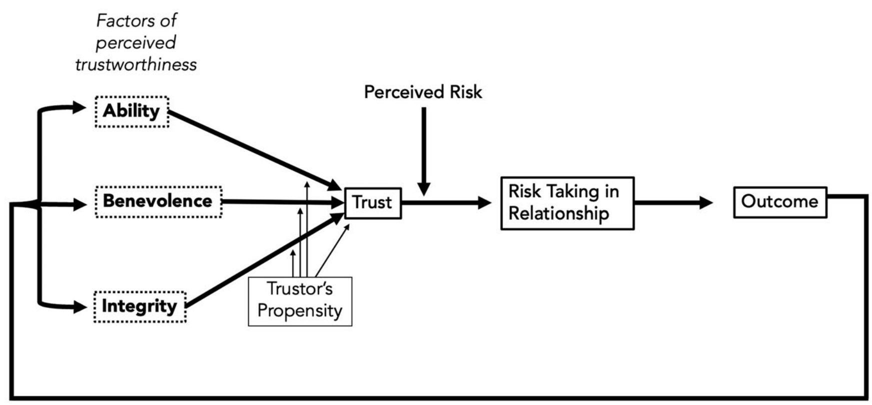


**Figure 2.** Mayer's integrative model of organizational trust (Mayer et al., 1995).

role in developing user trust. In section 3 we outline several useful frameworks to improve system transparency, and discuss them as potential solutions to improving user trust in AI and other emerging technologies. We then suggest that, as a result of our review, trust and transparency are so essential to technology acceptance that these concepts should be integrated into the Technology Acceptance Model structure (especially in the context of intelligent systems). In our concluding section, we offer recommendations for future research designed to (A) empirically demonstrate the importance of trust and transparency within the TAM framework and (B) provide HCI practitioners with practical approaches to designing for transparency, trust, and acceptance.

## 2. The role of trust in technology acceptance

Over the last several decades, research regarding trust in intelligent systems has increased tremendously (see Lee & See, 2004 for a classic perspective) and numerous researchers have come to the conclusion that trust in such systems is critical to attain if those systems are ever to be truly adopted by the public. Lee and See noted that "people tend to rely on automation they trust and tend to reject automation they do not" (Lee & See, 2004, p. 51). To this end, multiple models of trust in technology have been both adapted from other fields (e.g., Mayer et al., 1995), and some have been developed specifically for technological systems (e.g., Hoffman, 2017; Lee & See, 2004, etc.) with the intent of driving research on the matter.

Here we examine three notable models related to trust and advanced systems: Mayer's Integrated Model of Organizational Trust, Lee and See's model of Trust in Automated Systems, and Hoff and Bashir's Trust in Advanced Systems model. There are, unquestionably, a number of other perspectives and models that seek to provide perspective on this matter. We select these three models because of their foundational nature and their notable citation records (Mayer's model has been cited over 20,000 times, Lee and See over 3000 times, and Hoff and Bashir nearly 1000 times). While other models could be selected, we believe these three provide a diversity of perspectives and provide adequate coverage of the topic.

### 2.1. Integrated model of organizational trust

Perhaps the most celebrated model of trust is Mayer's integrated model of organizational trust (IMOT). This model was developed by Mayer and colleagues in an attempt to understand trust in a business context (Mayer et al., 1995). Mayer suggested that four primary factors drive trust between a trustor and a trustee: (1) ability; (2) benevolence; (3) integrity; and (4) personality factors. *Ability* refers to the degree to which a person believes another agent (e.g., another person, an organization, etc.) has the *capacity* to be successful in some domain. *Benevolence* refers to the degree to which a person believes an agent cares about the person. *Integrity* refers to the degree to which a person believes an agent has an acceptable internal moral compass, or holds similar values to one self. *Personality factors* refers to the presence of individual differences (i.e., personality traits, attitudes towards relationships, potential sources of bias, etc.) that have the capacity to affect a person's willingness to enter into a trusting relationship (see Figure 2 for overall IMOT structure).

From Mayer's perspective, as a person's perception of an agent's ability, benevolence, and integrity develops, eventually the person should begin to trust the agent. Importantly, Mayer defines trust as: the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action importance to the trustor, irrespective of the ability to monitor or control that other party.

Mayer's model has been tested hundreds of times and has been cited over 20,000 times (according to Google Scholar) across multiple academic fields. While Mayer's thinking was developed and subsequently tested in regards to human-to-human trust, of most importance for this paper, the model has been heavily leveraged in the context of trust in advanced robotic systems. For example, Lyons and colleagues (Lyons et al., 2020) examined the IMOT in the context of trust in autonomous security robots. In their study, participants watched a video of a security robot

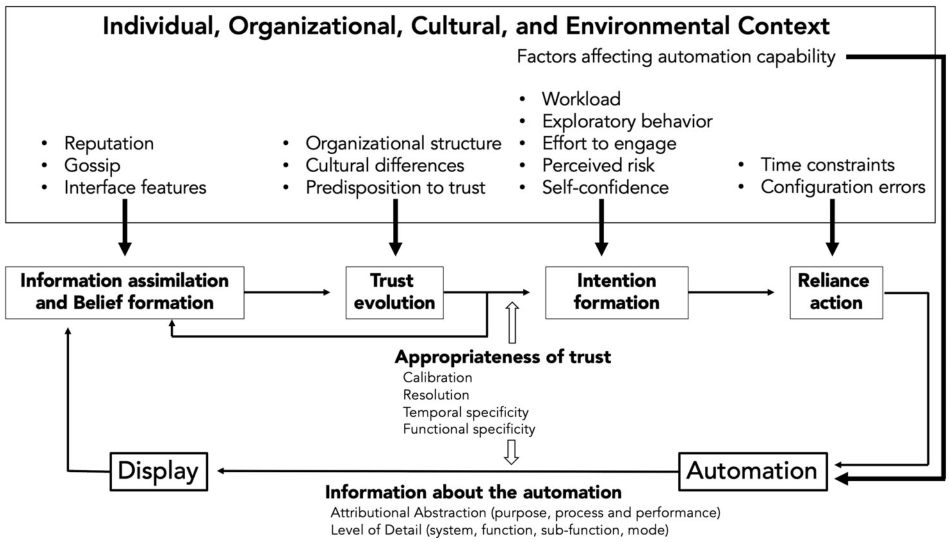


**Figure 3.** Lee and See's conceptual model of trust in automation (Lee & See, 2004).

checking ID badges of individuals attempting to enter a secure facility. The robot was described in different conditions as being programmed with various social intent (e.g., benevolence for people in the facility, benevolence for people waiting in line, self-sacrificial). In each case, the robot denied access to an individual under ambiguous circumstances. After watching the video of the robot, participants reported their trust in the robot as well as Mayer's ability, benevolence, and integrity concepts with regard to the robot. As would be expected, the robot was trusted more when participants were informed after watching the video that the robot correctly denied access as compared to when participants were informed that the robot had made a mistake. Importantly, though, the description of the intent behind the robot (e.g., the robot's programming described as being self-sacrificial) enhanced people's perceptions of the robot's benevolence and integrity. In another project, Kim and colleagues (Kim et al., 2020) leveraged the same automated security robot video, but this time examined the power of ability, benevolence, and integrity to predict trust in the robot. Their results found that perceived robot *ability* was the strongest driver of the three Mayer factors. Benevolence had a small, but still significant impact on trust, while integrity was not a significant predictor.

While Mayer's IMOT has utility for understanding trust in robotic systems, its utility for providing a solid predictive understanding of trust in other intelligent systems remains an open question. The concepts of benevolence, integrity and ability may require additional interpretation in order to be used in the assessment of inanimate software and computer code. For example, a person's perception of a technology's *ability* might include overall performance and capabilities of the whole system (a financial payroll system for a major corporation, for example), but the concept of ability may also be related more to factors like a system's overall reliability. *Integrity*, in the context of intelligent systems, might readily translate to expectations that a system holds values that are similar to the user's values. For example, how a smart investment system evaluates and mitigates risk in volatile markets, or how an autonomous drone chooses and prosecutes targets are all predicated on some sort of value system. Intelligent systems that engender a sense of *benevolence* might be those that demonstrate how the user is modelled and considered in the system's world model; or how a computer system attunes its actions to the user's needs, priorities, or desires. Ultimately, as such systems are developed, Mayer's model should be tested to see if its utility extends to such contexts. For now, the model is clearly a useful foundation for understanding trust in general and Mayer's definition of trust remains a solid foundation regardless of context.

## 2.2. Lee and see's model of trust in systems

Mayer's IMOT remains extremely influential in our scientific understanding of trust, whether that relates to interpersonal relationships, business or organizational settings, or relationships of humans interacting with advanced autonomy and other intelligent systems. While Mayer's IMOT is probably the most celebrated model of trust in the social sciences, Lee and See's model of trust in systems is probably the most foundational perspective on the specific topic of trust in technological systems (Lee & See, 2004). Lee and See's model in Figure 3, like Mayer, suggests that trust and reliance on automated systems is a feedback loop. From Lee and See's perspective, trust formation requires information and beliefs related to an automated system (for example, an autopilot system in a commercial aircraft). In many ways, Lee and See's perspective is analogous to Mayer's perspective about the impact of ability, benevolence and integrity on trust. In Lee and See's model, beliefs about the system are influential in producing trust and a subsequent intention to rely on a system (also see (Hoff & Bashir, 2015). The performance of the system and its displays ultimately produce information that feeds back and updates an individual's beliefs and knowledge about the system. In addition to the feedback loop similarities, Lee and See's definition of trust is also relatively similar to Mayer's. Specifically, they define

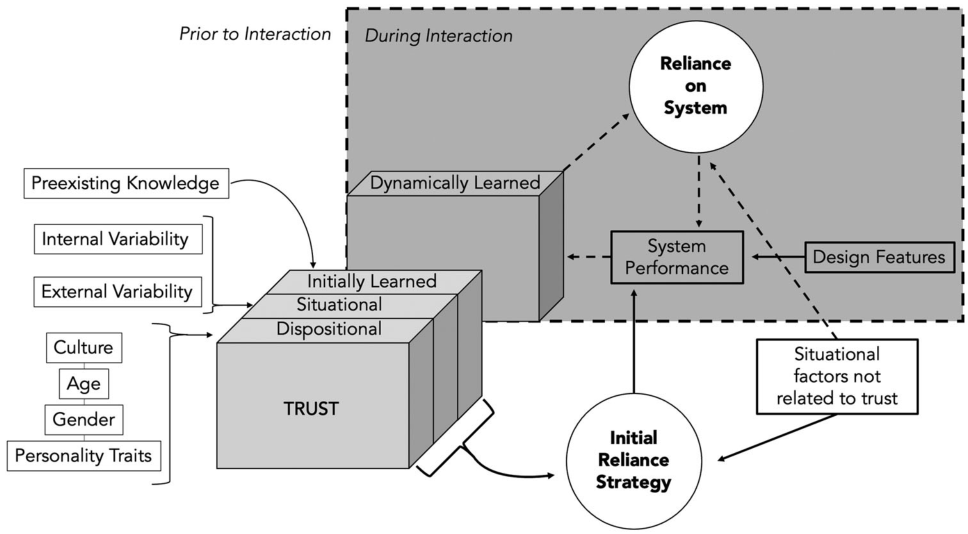


**Figure 4.** Hoff and Bashir's model of trust in automation (Hoff & Bashir, 2015).

trust as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability." While the definitions are not identical, they both point to a sense of vulnerability that an individual comes to accept in a potentially risky situation.

Elements of this model have been empirically tested and provide strong support for Lee and See's perspectives. For example, Lee and Kolodge (2020) note that multiple studies examining trust in automated vehicles have found that intentions to use an automated system are strongly driven by trust in a system. Tests of the impact of display systems on trust and reliance have also supported Lee and See's model. Koo and colleagues (Koo et al., 2014) found that individuals preferred systems (in the context of a self-driving car) that provided explanations for both how and why a system operated the way it did. For example, in Koo's study, when the self-driving system informed a driver of "what" it was doing (e.g., "the car is braking") users performed less capably using the system than when the system provided an explanation of "why" it was doing what it was doing (e.g., informing the driver that there is something in the road.) In a text analysis of potential determinants of trust in systems, Lee and Kolodge (2020) examined open-ended comments from the J.D. Power and Associates Tech Choice Study. Their approach discovered 13 distinct topics in the open-ended comments that were associated with trust in self-driving cars. One especially critical finding from this study was that "the basis of trust differs when the automation has not been directly experienced, leading to a focus on *societal* and *relational* bases rather than the more typically studied experiential basis" (p. 16, emphasis added). This point is especially congruent with Lee and See's perspectives on the importance of using a system in order to develop trust in that system. Lee and See (2004) provide an additional dimension to the concept of trust in intelligent systems–that trust is not a static state or attitude, but rather an evolving, adaptive state that can ebb and flow in response to experienced interactions with intelligent systems. This has special importance for new, emerging intelligent systems that introduce new modalities for interaction and new levels of complexity that users must become familiar with before they are likely to feel comfortable with and subsequently trust them. The additional dimension of *experience with systems* links well to the technology acceptance model's dimension of perceived ease of use, and suggests that, like first impressions, the way users experience interactions with intelligent systems has a broad impact on their likelihood to experience repeat interactions in the future.

### 2.3. Hoff and Bashir model of trust in automation

In addition to Lee and See, Hoff and Bashir's (2015) trust in automation model provides additional perspective on how trust in systems functions. Their model in Figure 4 suggests three overarching factors that drive trust in systems: dispositional factors, context factors, and learned factors. Dispositional factors, similar to Mayer's model (Mayer et al., 1995), tend to encompass an individual's personality traits, gender, age, and culture (among others). For the most part, these factors are thought to be relatively unchangeable, but designers should keep them in mind all the same. Context factors, according to Hoff and Bashir, are factors that impact the user of a system and come in two forms: external and internal. External contextual factors include concepts such as user workload, known capacity of the system, and the level of risk inherent in the potential use case situation. Internal factors include concepts such as user experience with other (similar) systems, user self-confidence in the use case situation, and user subject matter expertise. Finally, learned factors include the user's experience with the system after use. This is very similar to the feedback loop described in Mayer's IMOT, and also to Lee and See's concept of experience-based trust formation. The concept that a system's performance and reliability, naturally, will inform the user's understanding of the system and subsequent trust in that system is shared across all three trust models.

## 3. The potential impact of transparency on trust

The models described above make it clear that a number of factors drive trust, both between other humans as well as in technological systems. These models help conceptualize how users perceive and come to trust (or distrust) such systems. While there are clear distinctions in the factors that drive trust, they also have a number of common themes as well.

One common theme running throughout the major theories of trust described above is the importance of users being afforded *access* to information about the system's functions, components, and operations. Essentially, each model has some emphasis on allowing users to look "under-the hood" to obtain needed information. Systems that afford users with "under-the-hood" access to system operations are commonly referred to as "transparent" (see Bitzer et al., 2021 for a useful discussion).

The term transparency is a relatively common term in the scientific literature, especially in social and political sciences, but becoming more so in computer sciences as well (Lipton, 2018). Despite this, there are surprisingly few agreed upon definitions of the term in literature, and what definitions do exist vary widely. For example, HCI textbooks refer to transparency as providing "the necessary knowledge within the environment… to support the user in building an appropriate mental model of what is going on" (Dix et al., 2004, p. 283), and "easy-to-understand and intuitive ways of interacting with the system" (Rogers et al., 2015, p. 94). Literature from recommender systems refers to transparency as "exposing the reasoning and data behind a recommendation" (Herlocker et al., 2000, p. 241). Literature discussing intelligent agents describe transparency more broadly as "the descriptive quality of an interface pertaining to its abilities to afford an operator's comprehension about an intelligent agent's intent, performance, future plans, and reasoning process" (Chen et al., 2014). Studies in information systems have defined it as "…explaining to their human users both the knowledge they contain and the reasoning processes they go through" (Gregor & Benbasat, 1999, p. 498).

Common HCI design principles emphasize the importance of making users aware of the current system state, e.g., "…when there is nothing in the state of the system that cannot be inferred from the display." (Dix et al., 2004, p. 612). Another way that the concept of transparency is discussed is as a function of good design that informs users of what the system can do for them, or making users aware of affordances, e.g.,"…when it evokes an easy-to-understand system image in users" (Rogers et al., 2015, p. 94). Transparency is also concerned with aiding in the predicting of future state, or the consequences of an action, e.g.,"…a description of the potential effects that taking a course of action will have on the pre-planned mission" (Pharmer, 2004). This is tightly coupled with providing information about a system's intent or goal, e.g., transparency is "…the degree to which a system's action, or the intention of an action, is apparent to human operators and/or observers" (Orsosky et al., 2014, p. 1).

Regardless of which definition is chosen, system transparency has come to the fore in the discussion around trust in intelligent systems because, as noted above, users often struggle to understand how intelligent systems operate (e.g., how they produce their outputs), and this opacity likely plays a role in how users develop trust in systems. As a design strategy, therefore, improving the transparency of a system is a grounded and discrete goal that, according to leading models of trust discussed earlier, will likely improve user trust.

While the definition of transparency does indeed remain fuzzy, Bitzer et al. (2021) pointed out that perspectives from the business management literature might provide some helpful unifying guidance for characterizing transparency. Business leaders appear to struggle with similar issues regarding the various characterizations of transparency (often with regard to topics like transparency of organizational processes) as researchers in the intelligent systems space do. Specifically, Bernstein (Bernstein, 2017) noted that transparency is probably best thought of as consisting of four components that he suggests might be useful for management practitioners to consider: transparency for monitoring, transparency for process visibility, transparency for surveillance, and transparency for disclosure. Here we briefly describe these four components of transparency to explore their applicability to intelligent systems research. (see Bitzer et al., 2021 for additional details on merging Bernstein's perspective with intelligent systems research)

#### 3.1.1. Transparency for monitoring

In the business literature, Bernstein suggests that transparency for monitoring should allow business leaders to "monitor information about an activity or task and makes it more widely available" (p. 216). Such monitoring, from this perspective, is designed to be somewhat objective and would allow leaders to better understand if business objectives are being met. In the intelligent systems space, such monitoring might come in forms such as the feature in a navigation app that provides drivers with an ongoing sense of time remaining until the destination, overall traffic, speed traps etc.

#### 3.1.2. Transparency for process visibility

From Bernstein's business perspective, transparency for process visibility should allow business leaders to understand workflows, efficiencies, how business policies are engaged and carried out. Bernstein's business concept would seem to map onto intelligent systems perspectives that suggest transparency should allow users to understand the underlying algorithms within a system. As an example, from the intelligence systems space, such process visibility might come in the form of a feature for a navigation app that would provide users with a sense of why a navigation system is providing the logic it is providing, why it might suggest an alternate route, etc.

#### *3.1.3. Transparency for surveillance*

Bernstein's take on transparency for surveillance refers to a more tactical moment to moment surveillance of employees. He provides examples of organizations that have employees wear monitoring devices to watch them moment to moment on their computers. In the business context, such surveillance is likely onerous and suggests a lack of trust in employees, however, in the intelligent systems context, it may well be a helpful and natural form of transparency. Such an approach for an intelligent navigation system might look like an alerting system that constantly provides updates to drivers and allows for near real time partnership with the system for the driver to feel comfortable with moment-to-moment decisions.

#### *3.1.4. Transparency for disclosure*

Disclosure based transparency, in a business context, is designed to make previously hidden or secret information open to others for inspection, review, or criticism ( Bitzer et al., 2021; Bernstein, 2017; see also Diakopoulos & Koliska, 2017). For intelligent systems, this form of transparency might provide users of a system with increased understanding of the types of data a system uses and where it comes from (especially personal data). In the navigation system app example, this might include the kinds of location data a system collects on a user, how it is processed, stored, transformed, and leveraged.

The following section examines several empirical studies that attempt to link transparency to trust (and in some cases, technology acceptance). Each of the studies reviewed examined transparency in a slightly different way that, at a glance, might cause readers to believe that transparency is the muddled, ill-defined, concept we noted above. However, placed within the context of Berstein's transparency framework, it becomes clear that each is likely examining a different component of transparency.

## 4. Transparency improves trust and user acceptance

### *4.1. Transparency for monitoring*

Yang and colleagues (Yang et al., 2017) examined transparency in a way that appears to fit well within Berstein's monitoring concept. Participants were given a task in which they were to A) control four simulated drones and B) monitor incoming visualizations provided by the drones for threats. The control task was presented in one computer window and participants needed to toggle to another window to see the drone's incoming visualizations. The simulated drone had an alert system that would monitor the incoming visualizations for potential threats. In this study, transparency was manipulated via a threat alert system for users. Participants were assigned to receive either a low transparency binary warning (i.e., "Danger" or "Clear") or a higher-transparency "likelihood alarm" that had a more nuanced alert set (i.e., "Clear," "Possibly Clear," "Warning," or "Danger").

Participants were to press a "report" button if drones spotted a threat. If participants wished, they could toggle away from the drone control screen and personally examine the drone visualization images if they wished, or, they could rely on the monitoring detector system. Participants completed 100 trials of the activity.

As would be expected, correct information from the monitoring system (e.g., correct detection of threat, correct indication of absence of threat) generally increased trust for both the high transparency and low transparency systems with some relatively small differences. Of note, when the system was wrong (i.e., false alarm or false clear) trust was harmed in more extreme ways than it was improved when the system was correct.

Yang and colleagues called special attention to the notion that higher transparency had a stronger impact on participant trust (both when the higher transparency system failed and when it succeeded) than the low transparency condition. From their perspective higher transparency appeared to produce more properly calibrated reactions to a system than a lower transparency condition.

### *4.2. Transparency for process visibility*

Lyons and colleagues (Lyons et al., 2016) also provided evidence that affording users a sense of process visibility and underlying logic can improve trust in systems. His research team recruited commercial pilots and asked them to complete a simulated aviation-based experimental task (i.e., landing aircraft). Participants had access to an emergency landing assistant to help them with their landing. Participants completed the task using a system that had one of several levels of transparent information for the pilots: (A) a control baseline of landing related information, (B) the baseline information and a probability indicator which informed the pilots of the probability of their being able to land the plane successfully or have to circle back to make another approach, or (C) a system that had both types of information in addition to information about the logic behind the system's perspectives. Through multiple rounds of experimentation, trust was significantly higher in the logic-based transparency condition than in the baseline condition.

In a similar study, Mercado et al. (2016) had participants complete a task in which they controlled several simulated unmanned systems to the completion of a mission. Participants in this study had access to a recommender system that offered possible solutions to the problems they faced. As with Lyons, the system provided users with varying levels of transparency regarding the recommendation. In a control baseline condition, participants were only provided with a basic plan of action. Conditions with increased transparency provided participants with additional information such as the logic behind the recommender's reasoning, rational, as well as uncertainty information. Mercado's team found that participants were better able to complete their tasks in the conditions with greater transparency and participants reported greater trust in the systems as well.

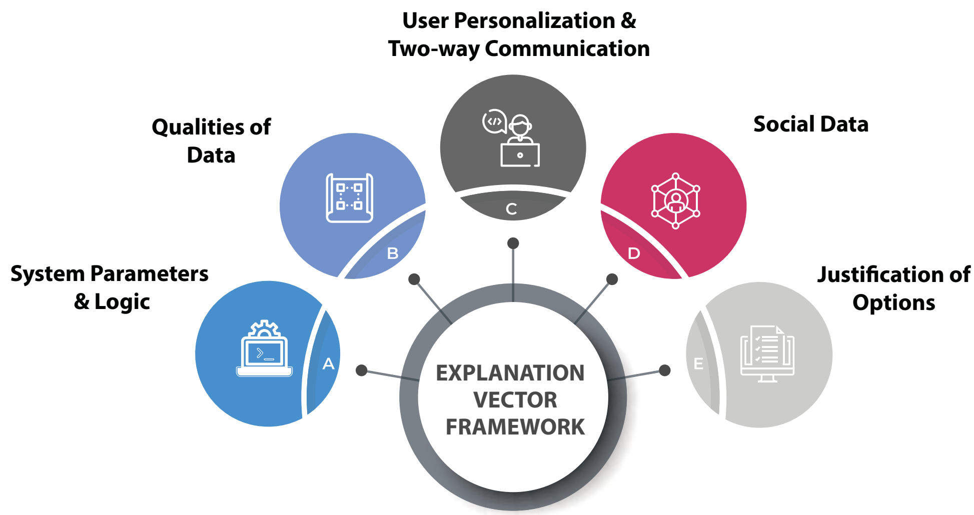


**Figure 5.** Vorm and Miller's Explanation Vector Framework (Vorm & Miller, 2020).

### 4.3. Surveillance-based transparency

Jung et al. (2021) tested the role of "interactivity" on trust within the technology acceptance model (TAM). Specifically, this study aimed to examine how a number of variables operate within the TAM framework to explain usage and reliance on the AirBnB platform for travel accommodation. In this study, researchers surveyed previous users of AirBnB and tested how the standard TAM factors impacted participants' intentions to use AirBnB again to book future accommodations. As would be expected, the typical TAM variables played a role in predicting re-use of AirBnB. However, for our purposes, Jung et al's "interactivity" concept is especially helpful. Interactivity, in this project, was characterized by variables such as ease of communication with the platform, active control, and synchronicity–concepts that are common design goals sought by HCI practitioners. Jung's research also notes that other projects have included items like "real time information" in the interactivity concept. This interactivity concept appears to bear a number of the hallmarks of Bernstein's transparency for surveillance noted above (e.g., real time information flow between platform and user, ability to have active control if needed, etc.). Of special relevance, interactivity (which we see as a potential example of surveillance-based transparency) was highly correlated with trust in AirBnB's technology.

### 4.4. Disclosure-Based transparency

Bitzer and colleagues (2021) conducted a study that was specifically designed to test the role of disclosure-based transparency on both trust and technology acceptance in the context of COVID-19 contact tracing apps. Their research was conducted at the height of the COVID-19 pandemic and, importantly, was conducted before governments and private industry had made substantial progress on a COVID-19 vaccine. Bitzer and colleagues had participants review a mock-up of an app store that presented them with multiple COVID-19 contact tracing app options, each with varying levels of disclosure-based transparency (e.g., levels of access to the systems inner workings, methodologies, overall functionality, analytic rigor, etc.). As hypothesized by Bitzer and colleagues, participants were significantly more likely to trust an app that had more disclosure-based transparency. In addition, participants were also more likely to select an app that was more transparent.

### 4.5. Summary

These empirical findings help demonstrate not only the relationship between system transparency and its effect on user trust in the context of intelligent systems, but also the potential utility of Berstein's transparency types as an organizing framework.

So far, we have explored theoretical models of how humans develop trust, and have attempted to outline the role that transparency (especially Bernstein's four transparency types) plays in driving trust. In this next section, we describe several practical frameworks that HCI practitioners have developed to specifically make technology more transparent. These frameworks were designed by their authors to solve HCI-related challenge sets faced by the authors in their respective fields. At a glance, the frameworks might appear to be substantially different in content and purpose. However, when viewed through the lens of Berstein's transparency types, it becomes apparent that these frameworks might have more in common with each other than might initially appear.

## 5. Practical transparency frameworks

The TAM, noted above, suggests that the acceptance of technology is heavily reliant on users' perceptions of a technology system's ease of use and the systems' usefulness. In addition, we have suggested that the TAM, when applied to intelligent systems, may be enhanced by considering trust (with a special emphasis on transparency and its various components) as a driver of acceptance.

We depict this potential relationship in Figure 5 below. In addition, it is also reasonable to suggest that enhanced trust might improve perceptions of ease of use and overall usefulness. As discussed in earlier sections, while transparency may be an appropriate priority for HCI practitioners, the term itself does not immediately lead designers and engineers towards specific and practical design approaches. What is needed, therefore, are practical and discreet design techniques that improve transparency as a function of their use.

Several useful frameworks for how designers can improve or adjust system transparency have been proposed and tested in the HCI space. We will review several of these frameworks and discuss their potential utility to create interface designs and communication strategies that directly improve system transparency in ways that are likely to improve user trust and subsequent acceptance. We review these practical transparency frameworks below and consider how each element of the respective frameworks might fit within Bernstein's structure. Naturally, as with the trust models noted above, there are additional transparency frameworks[1] that could be included in this discussion. We have selected three that offer a diversity of perspectives and are from varying academic/research traditions.

## 5.1. Lyons' transparency framework for human-robot interaction

Lyons (2013) and his team at the Air Force Research Laboratory commonly conduct research with regard to human trust in advanced robotic systems. As such, he and his team have developed a transparency framework that they believe designers of advanced robotic systems should consider when designing robots. In their thinking, advanced robotics designers should consider "robot-to-human transparency" and "robot-of-human transparency."

### 5.1.1. Robot-to-human transparency

Within Lyons' robot-to-human transparency type, he suggests that systems should produce four types of information for human operators to consume. Specifically, he states that human operators should be able to quickly assess:

1. *Intention*: Lyons suggests that users should have a clear understanding of a system's design purpose and when it has power to override a human. Lyon's intention concept allows a human operator increased visibility into system operations. As such, it seems natural to suggest that this element of Lyons' framework might fit well within Bernstein's transparency for process monitoring concept.
2. *Current task*: Lyons suggests that systems should allow a user to know that the system understands its tasks, goals, progression toward goals, and awareness of its own errors. Lyons' current task concept is clearly designed to function much like the transparency for monitoring concept within Berstein's framework.
3. *Analytics and decision-making approach*: Lyons suggests that users should be able to quickly assess the underlying analytical principles, understanding of how systems make decisions. Lyons' concept of allowing users access to underlying decision-making processes seems a natural fit within Bernstein's transparency for process concept.
4. *System environmental factors*: A user should know things like the environment the system is currently operating in, geography, weather, and the impact of the environment on its sensors etc. Lyons' environmental factors seem clearly designed to provide users with the ability for strong surveillance of the system, making this element of Lyons' model a natural fit within Berstein's transparency for surveillance bin.

### 5.1.2. Robot-of-human transparency

In addition to robot-to-human transparency, Lyons also suggests that researchers should consider what he calls "robot-of-human" transparency. That is, Lyons suggests that robots should be aware of their operator in two key ways.

1. *Teamwork*: Designers should endeavor to make robots aware of the teamwork nature of their work with an operator. The robot should understand things like the division of labor between an operator and itself, and the operator should have some interface that provides this information. The teamwork concept from Lyons probably fits within Bernstein's surveillance concept. In a way, this ability of a robot to understand which element of the human/machine team has what tasks, should allow a robot to know if it, or a human, has overstepped bounds and to keep in respective "swim lanes."
2. *Human Status*: Robots should have awareness of the human operator's physiological and psychological status. Designers should allow the system to understand the human operator's stress levels, exhaustion, etc. Like the teamwork concept, Lyons' human status concept appears designed to provide a robot with a near-real-time ability to surveil its human operator and detect any problems the human might be encountering. As such, the human status concept aligns well with Bernstein's surveillance-based transparency concept.

## 5.2. Multi-source AI scorecard table framework

One particularly unique context in which transparency has been raised as a critical issue is within the United States' intelligence community (IC). Because of the nature of the IC's activities, much information relating to how information is gathered, treated, and processed before being submitted into a report, whether classified or not, is hidden. Subsequently, many decision makers who are the intended recipients of intelligence reports sometimes express frustration and doubt over reports whose origins and details cannot be made available. In some cases, the actionable and credible intelligence found within these reports might actually be rejected outright because of this opacity. The ramifications of this reluctance to use intelligence are profound–intelligence reports inform a vast array of strategic decision making across all strata of government. Failing to act on credible intelligence constitutes a very serious failure of trust along the pipeline of intelligence activities.

To address this apparent lack of trust in the IC's products, the US congress implemented the Intelligence Reform and Terrorism Prevention Act (Title I of Public Law 108-458; 118 Stat. 3688). The result of this Act was, among other improvements to existing policies and practices, the Intelligence Community Directive (ICD) 203 (DNI, 2015). ICD 203 sought to create Analytic Tradecraft Standards that would directly address the apparent lack of trust in and subsequent unwillingness to act on credible and actionable intelligence information. ICD 203 created a set of nine standards (see below) that, when applied appropriately, would improve perceptions of analytic integrity and objectivity of intelligence reports. While greater transparency into how data was collected and sourced remains unavailable for obvious reasons, the standards outlined in ICD 203 provide practical steps that can have a measurable effect on users' perceptions of overall transparency. In other words, while the inputs and processes may remain somewhat intractable, ICD 203 provides meaningful and valuable techniques that improve the transparency, and subsequent user trust in, the outputs of said processes.

Blasch and colleagues (2019) recognized the similarities between the challenges of information opacity in the intelligence community and the intelligent systems communities. Perceiving ICD 203's relevance for use in making AI development processes more transparent, understandable, and trustworthy to end users, they adapted the ICD 203 into a multi-source AI scorecard table (MAST) which is designed to score the transparency of AI (or similar) systems.

The ICD 203 framework is below, accompanied by Blasch et al's (2019) perspective on their utility for AI system design. We also indicate where each of these standards might fall within Bernstein's overarching transparency concept.

*Standard 1—Sourcing*: ICD 203 states that reports must properly describe the quality and credibility of underlying sources, data, and methodologies used to collect data. From an AI transparency perspective, Blasch et al's MAST (2019) suggests that AI systems designers should take this standard for written reports and consider ways their systems can describe the quality and credibility of underlying sources of data, the qualities of data (age, noise, etc.), and the methodologies used in transforming, preparing and cleaning that data. Within the Bernstein framework, this standard matches to the transparency for disclosure concept as it would allow users access to data that might otherwise be hidden or unavailable to them.

*Standard 2—Uncertainty*: ICD 203 states that written reports must properly express and explain uncertainties associated with major analytic judgements. Such transparency is commonly noted as a critical element of intelligent systems (see above). The MAST suggests that intelligent systems designers should ensure that systems properly convey and explain uncertainties associated with major analytic judgments; accompanying any prediction or recommendation should include information about the relative confidence, risk, and level of uncertainty embedded in the system's reasoning. Within the Bernstein framework, providing users with a sense of uncertainty probably best fits within the surveillance transparency type as it should allow users the ability to rapidly make adjustments should too much uncertainty arise etc.

*Standard 3—Distinguishing*: ICD 203 states that written reports must properly distinguish between underlying intelligence information and analyst's assumptions and judgements. Like written reports, the MAST suggests that intelligent systems should convey to users where assumptions and judgements are incorporated into system outputs. Within the Bernstein framework, this distinguishing concept probably best fits within the transparency for disclosure concept as it should allow users access to information that is probably hidden or withheld in many situations.

*Standard 4—Analysis of Alternatives*: ICD 203 states that reports must incorporate an analysis of alternatives. The MAST suggests that intelligence systems should do the same for users. Specifically, intelligent systems, especially recommender systems, should provide users with various courses of action whenever possible ( not unlike a navigation system that provides multiple options for routes to a destination. Within the Bernstein framework, analysis of alternatives probably best fits within the transparency for disclosure concept as it should allow users access to information that they otherwise would not have.

*Standard 5—Relevance*: ICD 203 states that reports must demonstrate the relevance to customers and address the implications of findings. This is probably a more challenging concept for intelligent systems. Specifically, the MAST suggests that systems should communicate how outputs pertain to and affect user intentions, desires, and goals whenever possible. Within the Bernstein framework, relevance probably best fits within the transparency for process concept as it should allow users to understand how results fit into their workflow and needs. Admittedly, this element of the MAST may not have a direct fit within Bernstein's framework.

*Standard 6—Logic of Argumentation*: ICD 203 states that reports must use clear and logical argumentation. Likewise, MAST suggests that systems should convey predictions or recommendations using clear and logical argumentation that users can understand. This recommendation neatly maps onto Bernstein's process logic transparency concept.

*Standard 7—Consistency*: ICD 203 states that reports must explain any changes that have been made and should establish consistency of analytic judgements. MAST suggests that systems should, whenever possible, convey where major analytic judgements have been changed compared to previous versions; systems should explain current reasoning that supports these changes. This recommendation neatly maps onto Bernstein's process logic transparency concept as it would provide users a sense of the logic behind why recommendations etc. have changed.

*Standard 8—Accuracy*: ICD 203 states that reports should apply expertise and logic to make the most accurate judgements and assessments as possible. The MAST suggests that intelligent systems should do the same based on the information available and known information gaps. Outputs should express judgements clearly and precisely as possible, reducing ambiguity by addressing the likelihood, timing, and nature of the outcome or development. Accuracy most likely maps onto Bernstein's surveillance concept as it would allow users to quickly make decisions as fast as information is available.

*Standard 9—Visualization*: Reports must incorporate effective visualizations where appropriate. Naturally, MAST suggests that intelligent systems should afford users with visuals that are intuitive and assist in understanding. The visualization concept probably best fits within Bernstein's monitoring concept as it would allow users to have a quick ability to understand outputs as needed and then return to other tasks.

The MAST framework, while relatively new and yet to be formally tested in the intelligent systems domain, nevertheless has strong potential as a design guide for use in developing transparent and explainable intelligent systems. These standards, when appropriately and thoughtfully applied, provide excellent benchmarks against which future intelligent systems will be measured in order that users better understand their outputs.

### 5.3. Explanation vector framework

Human Computer Interaction designers must wrestle with the practicalities of developing interfaces often with very limited screen real estate. Cluttered, busy, visual environments, such as those found in many commercial airliners and other industrial interfaces, have long been the focus of human factors engineering research because of the potential for these environments to confuse and distract users in high-risk domains. Affording greater transparency, therefore, often involves a trade-off with maintaining a clean and uncluttered screen environment. But how exactly does one strike the right balance, especially when human performance and safety can be adversely affected by too much or little of transparency and visual information.

One potential solution to this challenge is to develop systems that can prioritize information according to the cognitive needs and expectations of end users. Adaptive systems that can infer user information needs and afford them the right information at the right time would serve multiple purposes in terms of improving user interaction, as well as improving system transparency. Research in human perception and decision making has made clear that information is not homogeneous, and that in the context of decision making, some information is more influential than others (Mumaw, 2017; Mumaw et al., 2000; Parasuraman et al., 2000; Riveiro et al., 2014 ). To develop an adaptive system, therefore, engineers must first understand, organize and characterize the many different *types of questions* that users of intelligent systems are likely to ask in order that they can develop effective explanation strategies to meet user needs.

Vorm and Miller (2020) developed a framework of explanation types that can be used to provide users with greater transparency in an adaptive format. Their "explanation vector" framework, in Figure 5, outlines five types of transparency information that users are likely to seek in order to understand how system inputs map to system outputs. A brief description of the explanation vector framework is below.

#### 5.3.1. System parameters and logic

From Vorm and Miller's perspective (and the perspective of other researchers, see above), providing information with regard to system parameters and logic, (i.e., how a system works, including its policies, logic, and limitations) can help users build appropriate mental models of systems and help users navigate or explain unexpected events. A mental model is a person's mental representation of what something is, what it is for, and how it works (Rouse & Morris, 1986). Users build mental models of systems through their experiences and interactions with them, which in turn determines subsequent interactions. Systems that restrict or hide information, therefore, can dramatically skew users' understanding of those systems (Marwick & Boyd, 2011; Viégas et al., 2006), which in turn influences how users use and interact with those systems. Mental models need to be accurate and appropriate in order to help users interact with a system and understand how to use it. In addition, mental models need to help users understand the reasoning of what lies beneath computations and processes that make the system function. As such, Vorm and Miller's explanation vector of system parameters and logic suggests that good transparency design should permit users to have strong and clear access to underlying system parameters and logic to help them build such models. Their perspective on the need for transparency with regard to system parameters and logic maps nearly identically onto Bernstein's transparency for system process concept.

#### 5.3.2. Qualities of data

In many instances, understanding the relationship of dependencies present in a system can provide meaningful insights into that system's functionality. A computer program may be functioning perfectly, but if the data on which it is operating is exceedingly noisy or corrupt, the system's outputs may still be incorrect or inappropriate. Providing users information on the qualities of data in intelligent systems–where the data came from; how old it is, etc.–has been shown to improve user ratings of ease of understanding, meaningfulness, and the convincingness of system outputs (Arrieta et al., 2020). Advances in visual analytic approaches have also been shown to improve the comprehensibility and intelligibility of data to users by presenting it in a manner that is more readily understood (Mühlbacher et al., 2014), and to improve user's understanding of cause-and-effect relationships between variables, even among users with little to no data analytical background (Bae et al., 2017). Displaying qualities of data can be achieved through a number of techniques–from listing data sources that users can explore, to color coding outputs with a color scheme that indicates the age or fuzziness of the underlying data. Any practical step to affording users this information is likely to help users determine when the system may be operating out of limits, and may help users determine when system recommendations should not be used. Within Bernstein's framework, Vorm and Miller's quality of data concept fits within the transparency for disclosure concept.

#### *5.3.3. User personalization and two-communication*

Many intelligent systems provide outputs in the form of recommendations or predictions. These outputs are often intended to be tailored to a specific user need–for instance, a vehicle navigation system may generate a route recommendation according to the understanding that users want to prioritize *the fastest route*, or perhaps *the shortest distance*. There are dozens of approaches that can be used to generate predictions of user taste and preference without users needing to express their preference directly. For instance, the frequent and successful use of automated collaborative filtering in music and video recommender systems is a classic example of this inference. In such low-risk domains, these approaches are preferred over manually training a system to know what types of music users want to hear. What data is collected from user interactions with these systems, and how that data is processed to make predictions and recommendations, however, is typically not shared with the user.

When recommender systems work well, most users are unconcerned about the inner workings of how those recommendations were made (Sen et al., 2016). When recommendations appear out of place or inappropriate, however, users may want to understand why. Knowing how the user is modelled by a system, if at all, and to what extent system outputs are personalized for them could help resolve conflicts that arise from unexpected or inappropriate results. Users who are unsure about what interactions are recorded and used for predictions and recommendations may therefore "tread lightly" and feel less willing to explore and use a system. Conversely, research has demonstrated that users who are afforded an understanding of how their personal data is collected and used to make personalized recommendations demonstrate more active engagement and higher feelings of control (Eslami et al., 2015). In many cases, offering users a sense of two-way communication between the system and the user (Klein et al., 2004), which includes a representation of how the user is modelled and "known" by the system, may help users feel notably more comfortable with systems. Vorm and Miller's two-way communication concept might fit best within Bernstein's transparency for monitoring concept. Two-way communication might allow users to have a sense of the system without having to allocate constant attention to that system in the same way as transparency for surveillance.

#### *5.3.4. Social data transparency*

The central tenet of social computing is that computer systems that provide socially-related information better support everyday functionality (Wang et al., 2007). Digital realms are therefore structured in patterns that mimic structures of social life. The ways that users interact with computer systems are deeply informed by social signs and strategies, which affect how users perceive and shape expectations. The power of social media has been displayed in a variety of contexts over the past decade of its modern existence. Its role in daily life has morphed beyond a simple photo sharing tool to become a powerful tool for marketers and influencers as well. User data from social media has become highly lucrative and commodified. Systems that group users according to online behavior in order to predict preferences are abundant, and represent a new standard in modern marketing and sales. A user's understanding of how they are grouped by a system using social media information, (i.e., social data) can provide meaningful insights into why a system output, such as a targeted advertisement, was generated, and can help users resolve conflicts that may arise between a user and an inappropriate system output. Providing a user with information about how they are categorized and grouped socially may also affect decision making as well (Horrigan, 2017). Naturally, the social data concept Vorm and Miller propose fits within Bernstein's surveillance element of his framework.

#### *5.3.5. Justification of options*

People almost universally prefer to have choices in most decision-making contexts (Blume & Easley, 2016). Accordingly, many systems strive to offer choices to users (e.g., customization; control over workflow; turning on or off certain functionality, etc.) as a means of increasing engagement and satisfaction (Rogers et al., 2015). There are times, however, when providing multiple choices to a user may be undesirable. The use of a GPS navigation system, for example, may result in at most three route choices to the user, with one option typically highlighted by the system. There may be, of course, several hundreds or even thousands of potential route options available to the user, but displaying them all would likely confuse the user, and could in fact lead them to discard the technology due to its confusing and busy interface. Modern computer systems today seek to alleviate the 'tyranny of choices' (Schwartz, 2004) by limiting the number of options users can see, but when intelligent systems make recommendations, this design decision often forces a conflict for users who may prefer greater transparency into the "why" of system recommendations and options (Swearingen & Sinha, 2001). Providing information about the potential scope and scale of available choices or options in a decision space, therefore, may do well to increase a user's willingness to engage with system outputs, which may translate to gains in usability and acceptance. The justification of options notion that Vorm raises likely fits best within Bernstein's process transparency concept.

Of note, while there are many similarities between Vorm and Miller's explanation vector framework and other

Table 1. Features of practical transparency frameworks.

| Transparency for monitoring | Transparency for process | Transparency for surveillance | Transparency for disclosure |
|---|---|---|---|
| Current task (Lyons) | Analytics & Decisions (Lyons) | System Environment (Lyons) | Intention (Lyons) |
| Visualization (MAST) | Relevance | Teamwork (Lyons) | Sourcing (MAST) |
| Two-Way Communication (Vorm) | Logic of Argumentation (MAST) | Human Status (Lyons) | Distinguishing (MAST) |
| | Consistency (MAST) | Uncertainty (MAST) | Analysis of Alternatives (MAST) |
| | System Logic (Vorm) | Accuracy (MAST) | Quality of Data (Vorm) |
| | Option Justification (Vorm) | Social Data (Vorm) | |

frameworks mentioned above, the explanation vector framework is unique in that it seeks first to identify questions that users are likely to ask when interacting with intelligent systems. Understanding the question and its motivation is critical to providing appropriate and meaningful explanations of system processes that are aligned to users' needs. This is the principal method through which the explanation vector framework seeks to improve system transparency.

## 6. Transparency frameworks within the Bernstein perspective

The three transparency frameworks noted above were designed with the intent of enhancing transparency and trust in intelligent systems. Though each framework was designed with different challenge sets in mind, when viewed through the lens of Bernstein's four transparency components, it becomes clear that each framework addresses elements of Berstein's perspective. The table below attempts to place each element of the three frameworks noted above within the Bernstein perspective.

Undoubtedly, several of the elements of the three frameworks above could arguably fit in several of Bernstein's categories. Ultimately, the point we are raising is that individuals from multiple fields (e.g., business, the intelligence community, recommender systems, robotics, etc.) have proffered ways of thinking about transparency. At first glance, these approaches may seem a bit muddled and not necessarily pointing in the same direction. However, if one considers Bernstein's perspectives on four types of transparency as a unifying framework, then it becomes clearer that each of these separate frameworks appear to be trying to address transparency in a way similar to Bernstein's overarching transparency concept. In Table 1, below, we attempt to "bin" each one of the items noted above within Bernstein's four transparency components.

### 6.1. Summary and potential research directions

Thus far, we have attempted to make three key points in this paper.

1. Technology acceptance of intelligent systems is probably heavily dependent upon trust.
2. Trust in intelligent systems likely flows, in part, from the transparency of that system.
3. Transparency has often been a muddled concept across multiple academic literatures. However, Bernstein's perspective that transparency should actually be thought of as consisting of four distinct parts might well provide a unifying framework for both applied research into trust in intelligent systems and HCI transparency frameworks/guides. Importantly, Bernstein's perspective on transparency is well represented in practical HCI perspectives on transparency, even though the authors likely did not set out to support such an integrated perspective on transparency.

With these points in mind, we suggest a revision to the now famous Technology Acceptance Model when the TAM is applied to intelligent systems—the Intelligent Systems Technology Acceptance Model (ISTAM). The ISTAM introduces the concepts of transparency into the TAM as a key factor on the level of Ease of Use and Perceived Usefulness. We depict the revised TAM concept below in Figure 6.

Below, with this revised ISTAM in mind, we provide several potential lines of research effort to test the claims we made above with regard to transparency, trust, and intelligent systems technology acceptance. First, we suggest a market research-based project to provide initial evidence that the TAM should be updated to include transparency when the technology in question is an intelligent system. Second, we suggest a project designed to test which elements of Bernstein's overarching framework (as well as the frameworks by individuals such as Lyons etc.) are most desirable by users and if those elements vary by intelligent system type. Finally, we suggest an applied observational project designed to test the ISTAM concept in a field setting. Each project would build on one another and, if successful, provide strong evidence for the need for a revised TAM as well as enhanced nuance regarding the types of transparency needed/desired and under which circumstances they are needed/desired.

### 6.2. Testing for an updated TAM: a market research approach

Market research professionals have numerous methods designed to analyze the preferences of consumers and craft very detailed product specifications based on those analytics. In many cases product designers must make decisions about which features to include when designing a product. For example, a smart-phone manufacturer may need to prioritize features such as the size of a phone, the weight of the phone, material composition, operating system features, cost, and a host of others. How might such a manufacturer prioritize which features to include (and at which price point)? A restaurant owner might plan to revamp a menu—how might he or she prioritize the menu items most likely

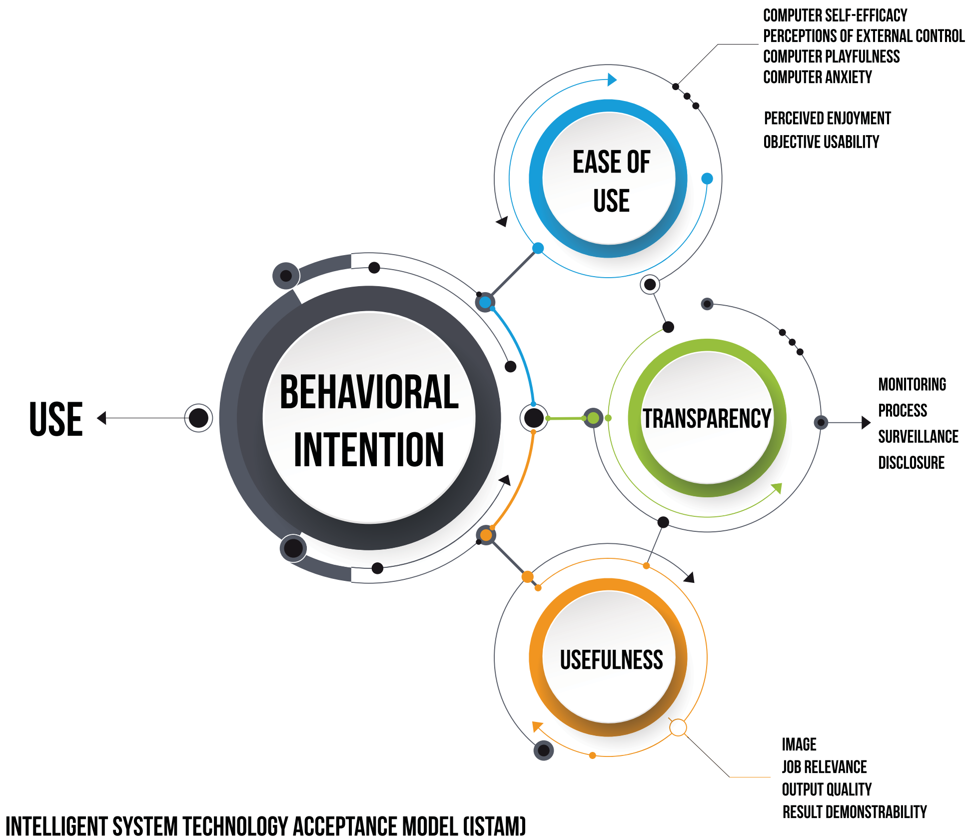


**Figure 6.** The Intelligent Systems Technology Acceptance Model (ISTAM). Transparency factors play moderating and supporting roles that combine to influence trust, and ultimately acceptance.

to be good sellers, with varying ingredients, varying calorie contents, etc., with limited menu space?

Market researchers have a number of methodologies such as MaxDiff analysis (i.e., "best-worst scaling") or conjoint analysis[2] that can help answer such questions (see What Can, 2018) for a full description of MaxDiff methods). Rather than, for example, asking survey respondents to rate preferences for features of a phone, which might well result in respondents rating all features as "very important," methods such as MaxDiff force respondents to make tradeoffs regarding features. In the case of MaxDiff, participants are shown a list of several features (e.g., 4 features out of a possible 50 of interest to the researchers) and asked to simply rate which feature listed is the best/most important and which is the worst/least important (see Figure 7).

After participants rate the items listed, they are prompted to rate another list of features in the same way. This procedure is repeated many times until a suitable number of ratings for all features is obtained.

This method of data collection places researchers in a helpful position as it allows them to make judgments about the importance of various features while only, in this example, obtaining ratings for cost and storage. Specifically, the rating in the example above allows a researcher to know that this participant views cost as more important than size, weight, and storage. In addition, the researcher also knows that storage is less important than all of the other features.

Methods such as MaxDiff are well suited for testing the kinds of features that might drive acceptance of intelligent systems. One approach to leveraging MaxDiff in the context of intelligent systems might be for researchers to conduct a randomized experiment in which participants are asked to rate a number of potential features of either an intelligent system or a non-intelligent system. For example, participants might be asked to consider features of a list of 50 features of a potential product. The features would be a mixture of items pertaining to the TAM ease of use and usefulness constructs as well as a number of features relating to transparency of systems (e.g., several of the features noted above from authors such as Lyons, Vorm, etc.). If transparency is an especially important driver of potential acceptance of intelligent systems, then participants in the intelligent system condition should rate transparency features as far more important features of the system compared to participants in a non-intelligent system condition.

Results of this sort, especially if replicated across multiple intelligent systems types (e.g., robotics, ML based, etc.) would potentially be a powerful commentary on the importance of targeted transparency for intelligent systems adoption. In particular, such an approach might provide good evidence that the TAM should be updated to include transparency (and by extension, trust) in the context of intelligent systems acceptance.

### 6.3. Assessing which transparency features, and when

The research teams who developed the transparency frameworks we reviewed (i.e., Bernstein, Lyons, Vorm, Blasch

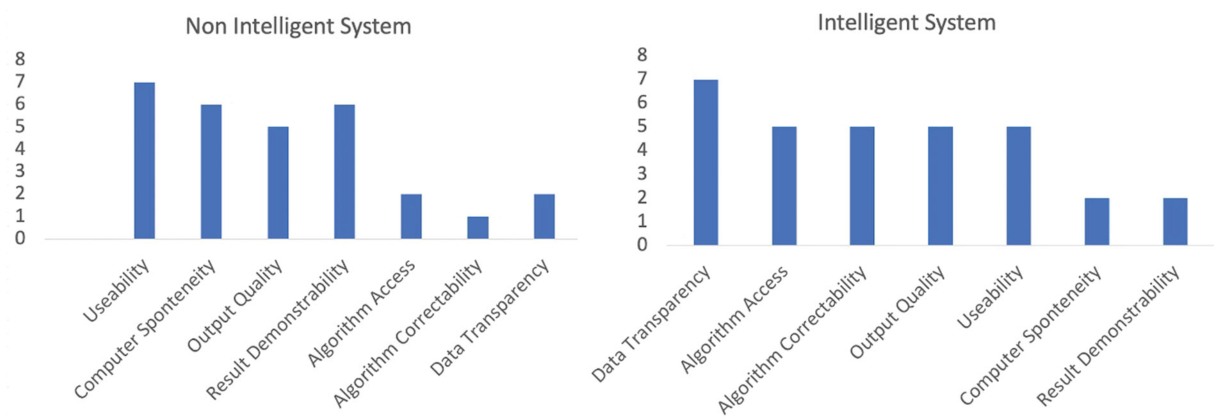


**Figure 7.** Example of evaluations of transparency using MaxDiff methodology.

et al.) each designed their respective transparency approaches to solve their respective challenge sets (e.g., transparency for robotics, transparency for recommender systems, etc.). As noted above, the transparency frameworks developed by Lyons, Vorm, and Blasch et al have some overlapping features and some distinct features, but ultimately all tend to fit within Bernstein's overarching multi-part transparency concept.

What remains an open question is whether any of the transparency concepts generated are more important to users than others and whether that utility varies by intelligent system type. For example, does Bernstein's transparency for process visibility (and the specific items developed by Lyons, Vorm, or Blasch) matter to users of recommender systems more than other types of intelligent systems? Does transparency for surveillance (and its items) matter more to individuals engaged with an intelligent robotic system than does transparency for monitoring (and its items)?

These issues of prioritization of transparency type remain an open question. One way to test the types of preferences and needs users might have with regard to the intersection of transparency type and intelligent system type is to leverage classic Q Methodology (see Vorm & Miller, 2020). Q Methodology is designed to find statistically relevant patterns in subjective unstructured opinions or data. The methodology has been referred to as the scientific study of subjectivity (Watts & Stenner, 2012). Q-methodology presents participants with some bank of stimuli (e.g., statements, questions, images, etc.) and has them rank their relative importance into a forced-choice matrix in the form of a pseudo-standard distribution (e.g., a bell curve).

Like the MaxDiff method noted above, participants in a Q study are *forced* to consider each item's relative importance and thus are unable to rank all items as important or unimportant. This format has significant relevance to expressing one's preference and value system, and models real-world decision making where trade-offs are most common. For this reason, Q methodology is far superior to traditional Likert-scale survey methods for issues that involve subjective opinions, values, or priorities from human beings. Ultimately, once all participants have completed sorted items in the study, researchers conduct a form of factor analysis to examine whether or not there are consistent patterns of importance ratings (see Vorm & Miller, 2020 for a helpful review of Q methodology).

With the current transparency and technology acceptance context in mind, Q methodology might be leveraged to examine how research participants prioritize certain transparency factors depending on which type of intelligent system is in question. For example, participants might be provided with a set of 4–5 scenarios describing different system types (e.g., a robotic system, a recommender system, a non-intelligent system etc.). Then, participants might be presented with a list of transparency items (likely based on the work of the authors above) and asked to sort the items based on importance to completing some tasks with the system in the scenario. Naturally, the analysis would be conducted to explore which patterns of transparency type might emerge and if those pattern types might differ depending on the system. Should the transparency type preferences differ by intelligent system type, it might provide researchers or HCI professionals with specific guidance for design based on system type. Further, such research might further solidify, and unify, the transparency frameworks noted above. For example, if some elements of the transparency frameworks are always highly prioritized regardless of system type, it would provide evidence that such a transparency item probably should be included in any type of intelligent system design. Likewise, should some elements of the transparency frameworks noted above always have low relevance to participants, then such evidence might indicate that such an item should be dropped from transparency frameworks.

### *6.4. Observational approaches to intelligent systems transparency*

A final approach to testing the utility of transparency and trust within the TAM is to leverage the field research approach used by Venkatesh and colleagues to obtain empirical evidence for the TAM. In their research, they partnered with four organizations who were each about to introduce a new technology system within their organization. The organizations had various business activities (e.g., financial services, manufacturing, accounting, investment banking) and were introducing different types of technologies to their employees. Venkatesh and his team developed scales to measure the elements of the TAM described above including ease of use, usefulness, intentions to use technology, and actual use of the technology (see Venkatesh & Bala, 2008 for full details and procedures). The researchers collected data at four different times over the course of several months. Overall, the model was very successful in predicting both intentions to use the new technology and actual use of the technology.

We recommend taking a similar approach to testing the potential integration of transparency and trust within the TAM– to test the ISTAM model. Such an approach would see a research team (or teams) partner with an organization(s) who is implementing the usage of an intelligent system(s). The research team would measure constructs such as perceived ease of use of the system, perceived usefulness of the system, and perceived transparency/trust in the system. For the TAM related constructs (ease of use and usefulness) the items developed by Venkatesh (Venkatesh & Bala, 2008) in his field assessment of the TAM would be useful to leverage. For the items related to transparency and trust, new items would need to be generated based on the results of the studies described above as well as the Table 1 which depicted the proposed facets of transparency detailed by the authors cited above.

As with Venkatesh, testing the relationships between ease of use, usefulness, trust and transparency as well as their ultimate impact on use of a system should probably be conducted at multiple points across some reasonable amount of time. The proposed relationship between these factors is presented in the simplified[3] figure below. A number of analytic procedures might be leveraged to test any statistical relationships obtained in this research effort, but structural equation modelling would likely be especially useful for this kind of project.

## 7. Discussion

This paper has examined the factors that drive acceptance of technology and made the case that intelligent systems (such as those based on machine learning, AI, etc.) obviously have additional hurdles to cross on the road to broad consumer acceptance. From our perspective, although trust in intelligent systems is a frequently discussed topic and is seen as a major hurdle, the practical and foundational construct of *system transparency* is the most appropriate goal for HCI practitioners because transparency is directly linked to affecting user trust, and has the added benefit of being an easily quantifiable design goal via multiple existing frameworks and design techniques.

We reviewed three transparency frameworks—Lyon et al's human-to-robot and robot-of-human transparency; Blasch et al's Multi Source AI Score Table (MAST), and Vorm and Miller's Explanation Vector framework. We synthesized those frameworks into one overarching approach. Further, we attempted to link these frameworks, transparency, and trust into the TAM. From our perspective, to drive acceptance of intelligent systems, designers should aim to build transparency for monitoring, transparency for process visibility, transparency for surveillance, and transparency for disclosure into their systems. Although, at present, there is relatively little direct empirical support for our approach, we laid out several initial study designs that might provide scientific backing for this line of argumentation.

## Notes

1. e.g., (Orsosky et al., 2014).
2. These procedures are often thought of as cousins of one another, see: https://www.qualtrics.com/blog/an-introduction-to-maxdiff/
3. For ease of depiction, we do not present each potential item to test nor the proposed moderating variables noted in Venkatesh & Bala (2008).

## Acknowledgements

We would like to acknowledge the assistance of Dr. Terry Fong, Chief Roboticist at NASA Ames for his valuable insights and assistance in theoreticizing these ideas. We would like to acknowledge Dr. Glen Henshaw of the US Naval Research Laboratory's Space Sciences Division for his useful and thoughtful feedback. We would like to acknowledge Dr. Chris Wickens, Professor Emeritus at the University of Illinois for his invaluable insight and guidance.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## ORCID

E. S. Vorm http://orcid.org/0000-0003-2844-1994

## References

Arrieta, A. B., Díaz-Rodríguez, N., Ser, J. D., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Avati, A., Jung, K., Harman, S., Downing, L., Ng, A., & Shah, N. H. (2017). *Improving palliative care with deep learning* (pp. 311–316). IEEE. https://doi.org/10.1109/bibm.2017.8217669

Bae, J., Ventocilla, E., Riveiro, M., Helldin, T., & Falkman, G. (2017). *Evaluating multi-attributes on cause and effect relationship visualization* [Paper presentation]. Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications – IVAPP (pp. 64–74). https://doi.org/10.5220/0006102300640074

Bernstein, E. (2017). Making transparency transparent: The evolution of observation in management theory. *Academy of Management Annals*, *11*(1), 217–266. https://doi.org/10.5465/annals.2014.0076

Bitzer, T., Wiener, M., & Morana, S. (2021). *Algorithmic transparency and contact-tracing apps: An empirical investigation*. Presented at the twenty-seventh Americas Conference on Information Systems, Montreal.

Blasch, E., Sung, J., Nguyen, T., Daniel, C. P., & Mason, A. P. (2019). *Artificial intelligence strategies for national security and safety standards*. Presented at the AAAI Fall Symposium Series, Arlington, VA.

Blume, L. E., & Easley, D. (2016). Rationality. In *The new Palgrave dictionary of economics* (pp. 1–13). Palgrave Macmillan. https://doi.org/10.1057/978-1-349-95121-5_2138-1

Bowen, J., Winckler, M., & Vanderdonckt, J. (2020). A glimpse into the past, present, and future of engineering interactive computing systems. *Proceedings of the ACM on Human-Computer Interaction*, *4*(EICS), 1–32. https://doi.org/10.1145/3394973

Chen, J., Procci, K., Boyce, M., Wright, J., Garcia, A., & Barnes, M. (2014). *Situation awareness-based agent transparency*. Army Research Laboratory Report ARL-TR-6905.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. *Digital Journalism*, *5*(7), 809–820. https://doi.org/10.1080/21670811.2016.1208053

Dix, A., Finlay, J., Abowd, G. D., & Beale, R. (2004). *Human-computer interaction* (3rd ed.). Pearson Prentice Hall.

DNI (2015). *Intelligence community directive 203 (ICD 203)*. https://www.dni.gov/files/documents/ICD/ICD%20203%20Analytic%20Standards.pdf

Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K., & Sandvig, C. (2015, April). *"I always assumed that I wasn't really that close to [her]": Reasoning about invisible algorithms in news feeds*. CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 153–162). https://doi.org/10.1145/2702123.2702556

EU (2019). *Ethics guidelines for trustworthy AI*. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Gregor, S., & Benbasat, I. (1999). Explanations from intelligent systems: Theoretical foundations and implications for practice. *MIS Quarterly*, *23*(4), 497. https://doi.org/10.2307/249487

Harrigan, M., Feddema, K., Wang, S., Harrigan, P., & Diot, E. (2021). How trust leads to online purchase intention founded in perceived usefulness and peer communication. *Journal of Consumer Behaviour*, *20*(5), 1297–1312. https://doi.org/10.1002/cb.1936

Herlocker, J. L., Konstan, J. A., & Riedl, J. (2000). *Explaining collaborative filtering recommendations* [Paper presentation]. ACM conference on computer-supported cooperative work, CSCW'00 (pp. 241–250). https://doi.org/10.1145/358916.358995

Hoff, K., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Hoffman, R. (2017). A taxonomy of emergent trusting in the human–machine relationship. In *Cognitive systems engineering: The future for a changing world* (1st ed., pp. 137–164). CRC Press. https://doi.org/10.1201/9781315572529-8

Holzinger, A., Carrington, A., & Müller, H. (2020). Measuring the quality of explanations: The System Causability Scale (SCS): Comparing human and machine explanations. *Kunstliche intelligenz*, *34*(2), 193–198. https://doi.org/10.1007/s13218-020-00636-z

Horrigan, J. B. (2017). How people approach facts and information. Pew Research Center.

IBM (2020). *What's next for AI – building trust*. https://www.ibm.com/watson/advantage-reports/future-of-artificial-intelligence/building-trust-in-ai.html#section2

Jung, J., Park, E., Moon, J., & Lee, W. S. (2021). Exploration of sharing accommodation platform airbnb using an extended technology acceptance model. *Sustainability*, *13*(3), 1185. https://doi.org/10.3390/su13031185

Kim, W., Kim, N., Lyons, J. B., & Nam, C. S. (2020). Factors affecting trust in high-vulnerability human-robot interaction contexts: A structural equation modeling approach. *Applied Ergonomics*, *85*, 103056. https://doi.org/10.1016/j.apergo.2020.103056

Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems*, *19*(06), 91–95. https://doi.org/10.1109/MIS.2004.74

Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., & Nass, C. (2014). Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *9*, 269–275. https://doi.org/10.1007/s12008-014-0227-2

Lee, J. D., & Kolodge, K. (2020). Exploring trust in self-driving vehicles through text analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *62*(2), 260–277. https://doi.org/10.1177/0018720819872672

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lipton, Z. (2018). The mythos of model interpretability. *Communications of the ACM*, *61*(10), 36–43. https://doi.org/10.1145/3233231

Lyons, J. (2013). Being transparent about transparency: A model for human-robot interaction. In *Trust and autonomous systems: Papers from the 2013 AAAI Spring Symposium*.

Lyons, J., Koltai, K., Ho, N., Johnson, W., Smith, D., & Shively, R. J. (2016). Engineering trust in complex automated systems. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *24*(1), 13–17. https://doi.org/10.1177/1064804615611272

Lyons, J. B., Vo, T., Wynne, K. T., Mahoney, S., Nam, C. S., & Gallimore, D. (2020). Trusting autonomous security robots: The role of reliability and stated social intent. *Human Factors*, *63*(4), 603–618. https://doi.org/10.1177/0018720820901629

Ma, R. H. Y., Morris, A., Herriotts, P., & Birrell, S. (2021). Investigating what level of visual information inspires trust in a user of a highly automated vehicle. *Applied Ergonomics*, *90*, 103272. https://doi.org/10.1016/j.apergo.2020.103272

Malle, B. F., & Ullman, D. (2021). *Trust in human-robot interaction* (pp. 3–25). Elsevier. https://doi.org/10.1016/b978-0-12-819472-0.00001-0

Marwick, A., & Boyd, D. (2011). To see and be seen: Celebrity practice on Twitter. *Convergence: The International Journal of Research into New Media Technologies*, *17*(2), 139–158. https://doi.org/10.1177/1354856510394539

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*(3), 709. https://doi.org/10.2307/258792

Mercado, J. E., Rupp, M. A., Chen, J. Y. C., Barnes, M. J., Barber, D., & Procci, K. (2016). Intelligent agent transparency in human-agent teaming for multi-UxV management. *Human Factors*, *58*(3), 401–415. https://doi.org/10.1177/0018720815621206

Meske, C., & Bunde, E. (2020). Transparency and trust in Human-AI-Interaction: The role of model-agnostic explanations in computer vision-based decision support. In H. Degen, & L. Reinerman-Jones (Eds.), *Artificial intelligence in HCI. HCII 2020. Lecture notes in computer science (LNISA, Vol. 12217)*. Springer. https://doi.org/10.1007/978-3-030-50334-5_4

Mühlbacher, T., Piringer, H., Gratzl, S., Sedlmair, M., & Streit, M. (2014). Opening the black box: Strategies for increased user involvement in existing algorithm implementations. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 1643–1652. https://doi.org/10.1109/TVCG.2014.2346578

Mumaw, R. J. (2017). Analysis of alerting system failures in commercial aviation accidents. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 110–114. https://doi.org/10.1177/1541931213601493

Mumaw, R. J., Roth, E. M., Vicente, K. J., & Burns, C. M. (2000). There is more to monitoring a nuclear power plant than meets the eye. *Human Factors*, *42*(1), 36–55. https://doi.org/10.1518/001872000779656651

Orsosky, D., Sander, T., Jentsch, F., Hancock, P., & Chen, J. (2014). *Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems*. Presented at the SPIE Defense + Security, SPIE. https://doi.org/10.1117/12.2050622

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics. Part A, Systems and Humans: A Publication of the IEEE Systems, Man, and Cybernetics Society*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Patterson-Hann, V., & Watson, P. (2022). The precursors of acceptance for a prosumer-led transition to a future smart grid. *Technology*

*Analysis & Strategic Management*, *34*(3), 307–315. https://doi.org/10.1080/09537325.2021.1896698
Pharmer, J. (2004). *An investigation into providing feedback to users of decision support* [Doctoral dissertation]. Electronic Theses and Dissertations (p. 224). https://stars.library.ucf.edu/etd/22
Poon, A. I. F., & Sung, J. J. Y. (2021). Opening the black box of AI-Medicine. *Journal of Gastroenterology and Hepatology*, *36*(3), 581–584. https://doi.org/10.1111/jgh.15384
Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., Jennings, N. R., Kamar, E., Kloumann, I. M., Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D. C., Pentland, A. … Wellman, M. (2019). Machine behaviour. *Nature*, *568*(7753), 477–410. https://doi.org/10.1038/s41586-019-1138-y
Ren, F., & Bao, Y. (2020). A review on human-computer interaction and intelligent robots. *International Journal of Information Technology & Decision Making*, *19*(01), 5–47. https://doi.org/10.1142/S0219622019300052
Riveiro, M., Helldin, T., Falkman, G., & Lebram, M. (2014). Effects of visualizing uncertainty on decision-making in a target identification scenario. *Computers & Graphics*, *41*, 84–98. https://doi.org/10.1016/j.cag.2014.02.006
Rogers, Y., Sharp, H., & Preece, J. (2015). *Interaction design: Beyond human - computer interaction* (4th ed.). Wiley Publishing. https://doi.org/10.5555/2031622
Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, *100*(3), 349–363. https://doi.org/10.1037/0033-2909.100.3.349
Schwartz, B. (2004). *The paradox of choice: Why more is less*. HarperCollins Publishers.
Sen, S., Geyer, W., Freyne, J., Castells, P., Amatriain, X., & Basilico, J. (2016*). Past, present, and future of recommender systems: An industry perspective* [Paper presentation]. Proceedings of the 10th ACM Conference on Recommender Systems (pp. 211–214). https://doi.org/10.1145/2959100.2959144
Siegrist, M. (2021). Trust and risk perception: A critical review of the literature. *Risk Analysis: An Official Publication of the Society for Risk Analysis*, *41*(3), 480–490. https://doi.org/10.1111/risa.13325
Starke, S. D., & Baber, C. (2020). The effect of known decision support reliability on outcome quality and visual information foraging in joint decision making. *Applied Ergonomics*, *86*, 103102. https://doi.org/10.1016/j.apergo.2020.103102
Swearingen, K., & Sinha, R. (2001). Beyond algorithms: An HCI perspective on recommender systems. In *ACM SIGIR 2001 Workshop on Recommender Systems*.
Van, H. N., Pham, L., Williamson, S., Chan, C.-Y., Thang, T. D., & Nam, V. X. (2021). Explaining intention to use mobile banking: Integrating perceived risk and trust into the technology acceptance model. *International Journal of Applied Decision Sciences*, *14*(1), 55–80. https://doi.org/10.1504/IJADS.2021.112933
Venkatesh, V. (1999). Creation of favorable user perceptions: Exploring the role of intrinsic motivation. *MIS Quarterly*, *23*(2), 239–160. https://doi.org/10.2307/249753
Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, *46*(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926.
Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences*, *39*(2), 39. https://doi.org/10.1111/j.1540-5915.2008.00192.x
Viégas, F. B., Golder, S., & Donath, J. (2006). *Visualizing email content: portraying relationships from conversational histories* [Paper presentation]. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '06 (pp. 979–988). https://doi.org/10.1145/1124772.1124919
Vorm, E. S., & Miller, A. D. (2020). Modeling user information needs to enable successful human-machine teams: Designing transparency for autonomous systems. In D. Schmorrow & C. Fidopiastis (Eds.), *Augmented cognition. Human cognition and behavior. HCII 2020. Lecture notes in computer science (LNCS, vol. 12197)*. Springer. Cham. https://doi.org/10.1007/978-3-030-50439-7_31
Wang, F.-Y., Carley, K. M., Zeng, D., & Mao, W. (2007). Social computing: From social informatics to social intelligence. *IEEE Intelligent Systems*, *22*(2), 79–83. https://doi.org/10.1109/MIS.2007.41
Watts, S., & Stenner, P. (2012). *Doing Q methodological research*. Sage Publishing.
Yang, X. J., Unhelkar, V. V., Li, K., & Shah, J. A. (2017). Evaluating effects of user experience and system transparency on trust in automation. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 408–416). https://doi.org/10.1145/2909824.3020230

## About the Authors

**E. S. Vorm** is a cognitive systems engineer with a PhD in Human-Computer Interaction from Indiana University. As the deputy director for the Laboratory for Autonomous Systems Research at the US Naval Research Laboratory in Washington, DC, he works to improve human-machine collaboration through the deliberate design of intelligent systems.

**David Combs** is an experimental social psychologist with a PhD from the University of Kentucky. He has over 10 years of experience across the government, non-profit, and private sectors. He specializes in leveraging social and behavioral science models to make the world a better place.