11-13-2015

Attitudinal predictors of relative reliance on human vs. automated advisors

Stephanie M. Merritt
University of Missouri-St. Louis, merritts@umsl.edu

Ruchi Sinha
Indian School of Business

Paul G. Curran
Michigan State University

Daniel R. Ilgen

Recommended Citation

Merritt, Stephanie M.; Sinha, Ruchi; Curran, Paul G.; and Ilgen, Daniel R., "Attitudinal predictors of relative reliance on human vs. automated advisors" (2015). *College of Business Administration Faculty Works*. 18.
DOI: https://doi.org/10.1504/IJHFE.2015.072982
Available at: https://irl.umsl.edu/business-faculty/18

Attitudinal predictors of relative reliance on human vs. automated advisors

Stephanie M. Merritt*

Department of Psychological Sciences,
University of Missouri – St. Louis,
421 Stadler Hall, St. Louis, MO 63121, USA
Email: Merritts@umsl.edu
*Corresponding author

Ruchi Sinha

Indian School of Business,
Academic Centre, # 8117,
Gachibowli, Hyderabad, AP 500032, India
Email: Ruchi_Sinha@isb.edu

Paul G. Curran

Michigan State University,
178 Giltner Hall, East Lansing, MI 48824, USA
Email: Curranp1@msu.edu

Daniel R. Ilgen

Department of Psychology,
Psychology Building,
316 Physics-Room 340A,
East Lansing, MI 48824, USA
Email: ilgen@msu.edu

Abstract: Trust and liking are attitudes with important implications for automation reliance in single-advisor settings; however, the extent to which their relationships with reliance generalise to settings in which the user receives conflicting advice from a human and automation is unknown. Participants completed an X-ray screening task and received simultaneous advice from what they believed was another human and an automated aid. High disuse was found for both advisors. Among participants who relied on advice, those with greater relative liking for the automation than for the human significantly increased their reliance on the automation relative to the human during the first half of the task. No significant relationships were found between relative trust or relative liking with reliance in the later part of the task, suggesting that reliance processes in dual-advisor settings may differ from those in single-advisor settings.

Keywords: trust; screening; reliance; automation; human; advice; advisor; decision-making; dual advisor; X-ray; vigilance.

Reference to this paper should be made as follows: Merritt, S.M., Sinha, R., Curran, P.G. and Ilgen, D.R. (2015) 'Attitudinal predictors of relative reliance on human vs. automated advisors', *Int. J. Human Factors and Ergonomics*, Vol. 3, Nos. 3/4, pp.327–345.

Biographical notes: Stephanie M. Merritt is an Associate Professor of Psychology at the University of Missouri-St. Louis. She received a PhD from Michigan State University in Industrial/Organizational Psychology in 2007. Her research focuses on how attitudes and affect combine with more rational processes in user decisions about reliance on automated systems. With her colleagues, she has been a recipient of the Jerome H. Ely Article Award for the journal *Human Factors* and a finalist for the 2014 Human Factors and Ergonomics Society's Prize for Excellence in human factors/ergonomics research: human-automation interaction/autonomy.

Ruchi Sinha is an Assistant Professor of Business in the Organizational Behavior Area at the Indian School of Business. She received a PhD from Michigan State University in Industrial/Organizational Psychology in 2010. Her work is published in international academic journals and in several books. She teaches management courses on leadership development and negotiation skills. She is an active member of the Academy of Management, Society for Industrial and Organizational Psychology and International Academy of Conflict Management. She has been a recipient of a SHRM grant to conduct research on work-life balance. She is currently conducting research that explores the effects of power and status hierarchies on team performance. Prior to her academic position, she worked as an independent human resource consultant in India. She is a founding member of Community of Organizational Scholars in India and participates in several community engagement initiatives with small and medium size enterprises in India.

Paul G. Curran is a Visiting Assistant Professor of both Mathematics and Statistics and Psychology at Kenyon College. He received a PhD in Industrial/Organizational Psychology from Michigan State University in 2012.

Daniel R. Ilgen is the John A. Hannah Distinguished Professor Emeritus at Michigan State University. He received his PhD in Industrial/Organizational Psychology from the University of Illinois, Champaign-Urbana.

This paper is a revised and expanded version of a paper entitled 'Comparing trust in humans and machines: development of a measure' presented at the 2011 International Annual Meeting of the Human Factors and Ergonomics Society, Las Vegas, NV, 19–23 September 2011.

1 Introduction

Individuals often receive advice from various sources. Increasingly, these sources include automated decision aids. Although such aids generally improve safety and performance, improper reliance on automation can have disastrous consequences (e.g., Parasuraman and Riley, 1997). Sparaco (1994) described a fatal airplane crash that could have been prevented had the pilots taken action to override the autopilot only four seconds sooner. Poor human-computer interaction has also been implicated in medical patient overdoses (Neuman, 1986; Sollins, 1986) and the nuclear incident at Three Mile Island (Connors et al., 1994). Thus, identifying factors associated with appropriate reliance on automated aids may be a key to improving safety and performance in critical systems.

To date, few empirical studies have examined advice reliance in contexts in which advice is received from both a human and an automated system. However, these contexts also have great implications for safety and performance. Lyons and Stokes (2012) describe a case in which a passenger plane and cargo plane collided mid-air, killing 71. The accident occurred when one of the pilots disregarded an automated warning system in favour of conflicting advice from an air traffic controller. Situations such as these demonstrate the key importance of determining how individuals rely on conflicting advice from humans and automation.

We examine potential differences in reliance on advice from human and automated sources in a context in which advice is provided from both sources simultaneously. When these advice sources disagree, the individual must choose between them. The individual's relative attitudes toward the human and automated 'advisors' may be key factors in this decision-making process. Individual differences in attitudes such as trust have been shown to significantly associate with single-source advice reliance (e.g., De Vries et al., 2003; Lee and Moray, 1992; Merritt and Ilgen, 2008) and may also play a key role in dual-advisor situations where one advisor is human and the other is automated. The goal of the present study is to examine relative trust and relative liking as predictors of the choice to rely on either a human or an automated advisor.

1.1 Reliance in contexts in which one advisor is automated

Thus far, little research has been conducted in situations where a user received conflicting advice from human and automated advisors. In one key study, Lyons and Stokes (2012) presented participants with a task in which they were required to make route choice decisions based on information from a human advisor and an automated map. They were presented with decisions of varying risk. In the low risk condition, the route was characterised as having little hostile activity and the human and automation provided consistent advice. In the high risk condition, the human and automation presented conflicting advice and the route was characterised as highly hostile. They found that reliance on the human co-worker's advice was significantly lower – when conflicting advice was presented by an automated system. This suggests that reliance on advice from a human co-worker is affected by the presence of conflicting advice from automation in high-risk scenarios. However, their focus was on reliance on a human advisor in the context of conflicting automation advice, not *relative* reliance decisions, nor did the Lyons and Stokes study examine relationships between trust and reliance. Therefore, the extent to which the relationship of trust with reliance generalised from single-advisor to dual-advisor situations is unknown. In the present study, we examine the relationships of individual differences in trust and liking on advice reliance in a dual-advisor setting in which one advisor is human and one is automated.

1.2 Trust

Trust is "the attitude that an agent will help achieve an individual's goals in a situation characterised by uncertainty and vulnerability" [Lee and See, (2004), p.51]. When people trust automation too little, the result is behavioural 'disuse', or sub-optimal levels of reliance. When people trust it too much, the result is 'misuse', or over-reliance on automation (Parasuraman and Riley, 1997). Thus, understanding how to calibrate trust and how trust in a source of information associates with reliance on it, has been a key focus.

Antecedents of trust include characteristics of the automation such as reliability (Dzindolet et al., 2003; Moray et al., 2000; Merritt and Ilgen, 2008) and expertise (Madhavan and Wiegmann, 2007a); and the decision maker's individual differences including age (Wiegmann et al., 2006), personality (Merritt and Ilgen, 2008), risk-taking tendencies (Madhavan and Wiegmann, 2005), implicit attitudes (Merritt et al., 2013) and mood (Merritt, 2011). Actual and perceived reliability levels of the automation are also associated with trust (Dzindolet et al., 2003; Merritt and Ilgen, 2008). Several studies have found that trust is positively associated with behavioural reliance on automation (De Vries et al., 2003; Lee and Moray, 1992; Merritt and Ilgen, 2008).

Furthermore, research suggests that reliance on automation advice may actually reflect a comparative process. When the automation's advice disagrees with the user's initial opinion, users compare their level of trust in the automation with their level of trust in themselves (i.e., self-confidence) and the user relies on the party that is trusted more (Dzindolet et al., 2003; Lee and Moray, 1994; Parasuraman and Miller, 2004). We hypothesise that a similar comparative process may occur when users are presented with *conflicting* advice from two external advisors. Thus, users may compare their level of trust in the human and automated advisors and rely on the more-trusted source.

Some research suggests that when trust is compared between human and automated advisors, automation may have an initial advantage. Dijkstra et al. (1998) found that people rated automated aids as more objective and rational than human advisors and Dzindolet et al. (2002) found that participants rated an automated aid as more useful than a human advisor. Dijkstra (1999) found that participants were more likely to agree with advice from an automated system, even when it was incorrect, relative to correct advice from a human advisor.

However, after advisors make errors, users may trust human advisors more. Dzindolet et al. (2003) proposed that individuals tend to believe that automated aids perform perfectly. When aids violate these expectations by committing errors, user trust declines severely. However, individuals do not possess these types of beliefs about human advisors. It has therefore been suggested that user trust will decline more severely after automation has erred than after a human has erred (e.g., Madhavan and Wiegmann, 2007b).

Research is mixed concerning whether trust predicts reliance equally on humans and machines. In one study, Lewandowsky et al. (2000) found that while trust was associated with reliance on both human and automated advisors, the relationship was stronger for automation. Perhaps social motives such as sharing decision-making responsibility seem more salient for human versus automated advisors (Harvey and Fischer, 1997; Lewandowsky et al., 2000). Thus, users may rely on the advice of less-trusted co-workers in order to avoid social penalties inherent in disregarding a co-worker's advice. Therefore, the relationship between trust and reliance in these types of situations is unclear.

Based on the research discussed above, we suggest that trust in an information source will be a meaningful variable for both human and automated advisors. We further propose that reliance on information from a source involves a comparative process in which decision makers' relative levels of trust are associated with relative reliance on either a human or automated advisor. Thus, we hypothesise that:

- *Hypothesis 1:* relative levels of trust will be associated with relative reliance on human and automated advisors.

1.3 Liking

While trust is associated with the degree to which users perceive the automation as reliable and useful, recent work has identified a second attitudinal construct, *liking*, as another antecedent of reliance on an automated decision aid's advice (Merritt, 2011). Liking is "the degree to which the user feels positively toward the automated system" [Merritt, (2011), p.358]. Theoretically, this construct is expected to relate to factors such as the user's interest in technology, novelty of the automation, moods and the automation's interface. Research suggests that liking for technology can be influenced by social cues such as the gender of the technology's voice or personality match (dominance or submissive communication style) with the user (c.f., Nass and Moon, 2000). Thus, while trust may centre on the degree to which the advice source is perceived as useful in reaching the user's goals, liking theoretically centres on the user's affective reactions. Liking for automation is expected to be influenced by both characteristics of the automation and individual differences, such that different individuals will vary in their liking for a given aid.

One study has empirically examined liking for automation and found that trust and liking were correlated, but distinct constructs [Merritt, (2011), $\Phi = .68$]. Further, liking was significantly associated with reliance in early interactions beyond the effects of trust and aid reliability. In a five-block long task, liking uniquely and significantly predicted reliance in the earlier stages of the task (blocks 1 and 2), while trust did not. During the intermediate stage (block 3), neither was a significant predictor (although trust was marginally significant) and in end stages (blocks 4 and 5), only trust was a unique significant predictor of reliance. The findings suggested that liking predicted reliance early in the task, while trust predicted reliance later in the task. Here, we build upon that initial work by examining the relationship liking with reliance in a dual-advisor situation. Similar to trust, we expect that individuals will compare their levels of liking for the human co-worker and automated advisor and relative liking will be significantly associated with relative reliance on those two advisors. Based on past work, we expect that relative liking will have significant associations with reliance only early in the experimental task.

- *Hypothesis 2:* relative liking will be significantly associated with relative reliance early in the experimental task.

2 Method

2.1 Participants

236 participants from a large US university were recruited through the psychology subject pool. Two multivariate outliers were discarded (*ps* = .0001) due to the pattern of their responses across items. Of the remaining 234, the mean age was 19 years (SD = 1.45); 77% were female. Ethnically, 68.9% were White, 9.4% were

African-American, 6.8% were Asian-American and 14.9% were other ethnicities or declined to report ethnicity.

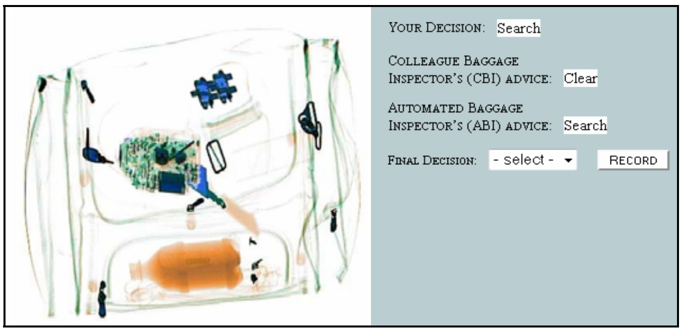
2.2 Overview

Participants completed a modified version of the X-ray screening task described by Merritt and Ilgen (2008). Participants viewed X-ray images of luggage, some of which contained weapons (guns and knives). Participants needed to determine whether each image contained a weapon. In doing so, they received simultaneous advice from two (fictitious) sources: one human and one automated. Participants were told that the automation screened each image using a software algorithm designed to identify guns and knives. The human advisor was portrayed as another research participant located in an adjacent room, as described further below. During the task, weapons appeared in 20% of images. No search time limit was imposed and the difficulty of identifying weapons varied across slides. On 'target trials', the human and automated advisors provided conflicting advice – thus, these target trials are our key focus. The target trial slides had been classified as medium-to-difficult based on participant success in past research (*citation removed for masking*).

2.3 Procedure

Participants were randomly assigned to one of two groups and each group was sent to separate rooms. Participants in both rooms were told that they had been assigned to the 'decision maker' condition, whereas the participants in the other room had been assigned to the 'co-worker' condition. Actually, both groups acted as the decision makers and received identical information. Debriefing indicated that participants believed they were receiving advice from an automated system and from a human participant in the other room. Note that at the time of check-in, participants may have seen other participants arriving and being sent to the other room, but at no point were participants made aware of which specific individual was purportedly their partner. Thus we do not anticipate any effects as a result of participants believing their partner was known/unknown, demographics, etc.

Figure 1 X-ray task screenshot (see online version for colours)



Participants viewed X-ray images of luggage containing various items. For each, they first recorded their initial decision (search/clear) based on whether they thought a weapon was present. Next, participants simultaneously viewed the advice from the automation and the human (see Figure 1). Lastly, they made a final decision and received immediate feedback. Participants made decisions regarding 100 images divided into two blocks of 50 images each. On 80% of these, the automation and human agreed. On the remaining 20% ('target trials'), they disagreed, allowing participants to develop differential levels of trust and liking for each advisor.

Participants' interactions with the human and automation were controlled in order to provide the strictest comparison of reactions to these sources. They received advice through a computerised interface in which each source's advice was provided simultaneously. Therefore, cues used to determine trust likely involved source reliability, perceived credibility and the participant's propensity to trust. Because people are expected to like sources that help them achieve their goals, the source's accuracy and perceived credibility were also expected to influence liking. Although we suspect that liking was affected by factors other than accuracy and perceived credibility, such effects were controlled across conditions by random assignment to conditions and standardisation of the interaction format.

2.4 Manipulations

Source reliability was manipulated in a three-condition, between-subjects design (automation more accurate, human more accurate, or both equally accurate). The primary goal of this manipulation was to ensure that adequate variance would be obtained on relative trust and liking. Differential reliability was established using the 'target trials' – the trials on which the two sources disagreed. When one source was better than the other, the more accurate one provided correct advice on 70% of the target trials. When equally accurate, each source was correct on 50% of the target trials. Because initial impressions can be important, each of these conditions was counterbalanced such that in half the cases the automation was the first to err, while in the other half the human was the first to err. This counterbalancing manipulation showed no significant effects; thus, counterbalanced halves were merged within condition.

On non-target trials (on which the human and machine agreed), the human and machine were correct on approximately 70% of trials. Thus, when target and non-target trials were combined, the accuracy rates were as follows: the more-accurate source was correct on 75% of total trials and the less-accurate source was correct on 65% of total trials. In the condition in which the human and machine's reliabilities were equal, they were both correct on 70% of total trials.

2.5 Measures

Trust. Trust was assessed using the seven-item scale developed by Merritt et al. (2011). Measure items have been designed to refer to either human co-workers or automated systems (i.e., human and automated referents). An example item is, 'I believe (my co-worker/the automated system) is a competent performer'. Each participant completed two versions of the scale (one referencing the automation, one referencing the human) at three times – prior to the start of the task, halfway through the task and after the task.

Liking. The liking scale was a five-item measure developed by Merritt (2011). Two versions were administered, one referencing automation and the other referencing the coworker. A sample item is 'I dislike working with the (automated systems/co-worker)'. This scale was also completed at the beginning, middle and end of the task.

Reliance. In computing reliance, we focused exclusively on target trials (the trials on which the human and automation disagreed). We conceptualised 'reliance' on a source as a 'switch' in which the user changed from his/her initial opinion to agree with a particular source. For example, if the user and human advisor initially said, 'clear' but the automation said 'search', reliance would be indicated if the user ultimately decided to 'search' in order to agree with the automation. This is a relatively conservative measure of reliance that was used to avoid the fallacy that simply siding with a 2–1 majority reflected reliance on one particular source (a failure to switch from one's initial opinion might not reflect a lack of trust so much as going with the majority opinion). Reliance was coded as follows: –1 = switch to rely on the human's advice; 1 = switch to rely on the automation's advice and 0 = no switch.

Relative trust and liking. We calculated difference scores for trust and liking at each time. This approach is consistent with Lee and Moray (1994), who calculated relative trust in automation versus user self-confidence in predicting reliance. At each time, trust and liking in the human advisor were subtracted from trust and liking in the automated advisor. Thus, negative scores represent relatively more trust/liking for the human, while positive scores represent relatively more trust/liking for the automation. We denote these difference scores as DiffTrust and DiffLiking.

3 Results

3.1 Descriptive statistics

Table 1 presents the sample sizes, means, standard deviations and alphas for the observed variables. All coefficient alphas (internal consistency reliabilities) were acceptable (.80 and above) and the standard deviations indicated that variance was observed on the measures. Mean values indicated that participants tended to rate the human higher than the automation in terms of both trust and liking. The difference in attitudes toward the co-worker and automation was significant in all cases except for time 1 trust, which was assessed prior to beginning the task. Mean levels of trust for both the human and the automated aid tended to decrease over time as participants encountered incorrect advice from both sources.

Table 1 Means, standard deviations, alphas and Cohen's d for observed variables

	Human advisor			*Automated aid*			*Cohen's d*
	M	*SD*	α	*M*	*SD*	α	
Trust time 1	3.41	.49	.86	3.42	.49	.85	–.02
Trust time 2	2.91	.63	.91	2.72	.53	.83	.33*
Trust time 3	2.79	.71	.94	2.65	.67	.93	.20*
Liking time 1	3.99	.48	.88	3.53	.58	.85	.86*
Liking time 2	3.70	.51	.87	3.30	.62	.86	.70*
Liking time 3	3.64	.57	.87	3.28	.64	.87	.59*

Notes: *Indicates a significant mean difference between human and automation, $N = 234$

3.2 Confirmatory factor analyses of trust and liking

In order to verify that trust and liking were two separate, but related, constructs, confirmatory factor analyses (CFAs) were performed separately for the human and machine advice sources at each time. We contrasted the fit of the hypothesised two-factor model (the scale items load on trust and liking, respectively) with a one-factor model in which all of the trust and liking items load onto a single factor. The results are displayed in Table 2. For human and machine advice sources at every time point, the two-factor model fit significantly better than the one-factor model according to chi square difference tests, indicating that trust and liking are two separate, but related, constructs.

Table 2 CFA for trust and liking in the human and automated advisors at each time point

		Human advisor				*Automated advisor*			
		χ^2 *and df*	*RMSEA*	*CFI*	*Φ*	χ^2 *and df*	*RMSEA*	*CFI*	*Φ*
Time 1	Two factors	130.82 (53)	.08	.96	.26	150.44 (53)	.09	.96	.45
	One factor	1,231.68 (54)	.31	.69		680.27 (54)	.22	.80	
Time 2	Two factors	129.25 (53)	.08	.97	.35	248.05 (53)	.13	.92	.56
	One factor	809.33 (54)	.25	.81		774.87 (54)	.24	.82	
Time 3	Two factors	268.13 (53)	.13	.94	.40	216.08 (53)	.12	.96	.68
	One factor	888.89 (54)	.26	.81		695.21 (54)	.23	.89	

Notes: All chi-square difference tests between the one-factor and two-factor models were significant at $p < .01$, indicating that trust and liking were best modelled as two separate, but correlated, constructs,
$N = 234$

3.3 Analytic concerns for dual-advisor settings

As previously discussed, past research using single-advisor scenarios with automated aids has found significant associations between both trust and liking with reliance (e.g., Merritt, 2011). Assessing these relationships for single advisor situations is relatively simple because whenever the automation and the user disagree, the user has only two options: stay with his or her initial opinion or accept the automation's advice. Thus, in such situations, greater degrees of trust and liking for the automation are expected to linearly relate to the likelihood of reliance on the automation's advice.

However, analysing associations with reliance in dual-advisor situations is significantly more complicated. In this case, three parties (the human user, the human co-worker and the automated system) each provide input on every decision. On target trials (in which the co-worker and automation were programmed to *disagree*) a 2–1 majority situation always occurred. Because there were only two decision choices available (search or clear) and three advisors, it follows that whenever the automation and human advisors disagreed with each other, one of them necessarily had to agree with the participant's initial choice. This scenario makes a simple linear association between trust or liking and reliance significantly less likely because in order for the user to change his or her initial opinion, trust in that advisor would need to exceed the *combined* level of trust in the self and the other advisor. For example, if the participant and the human advisor thought the correct decision was 'clear', then in order for the participant to change their final decision to 'search' in order to agree with the automation, they would need to trust the automation more than they trust themselves and the human advisor combined. Thus, we expected that the user would usually make a final decision consistent with the 2–1 majority opinion, perhaps regardless of trust in either of the two. In fact, this is what we found, as users provided a final decision that agreed with the majority opinion on an average of 93–94% of target trials in block 1 and 90–92% in block 2.

Because of this challenge, we adopted the following approach. First, we suspected that some individuals would display a strong tendency to rely on their own initial opinions, effectively disusing both the automated and human advice sources. To examine this possibility, we examined *non*-target trials, in which the co-worker and the automation always agreed. We identified cases in which the user's initial opinion conflicted with *both* the co-worker and the automation – these were cases in which the participant was on the minority side of a 2–1 majority. Next, we calculated the percentage of these instances in which the participant changed his/her initial decision to a final decision that agreed with the majority. Our logic was that if a given participant is unwilling to change his/her initial opinion even when outnumbered 2–1, it seemed very unlikely that he/she would change an initial opinion in target trials, when he/she was guaranteed to be in a 2–1 majority position. Thus, these participants were likely employing some other strategy, such as relying exclusively upon themselves [likely related to what Beck et al. (2007) termed 'intent errors']. These individuals would likely introduce error into the reliance analysis, decreasing the proportion of variance in final decisions that was possible to predict using trust and liking.

Our analysis revealed that a large percentage of participants disused/underutilised both the co-worker and automation even when outnumbered 2–1 by those sources (in other words, they failed to rely on the sources' advice). Even in situations where the participant was outnumbered, participants switched to agree with the majority only 18.72% of the time on average, with a median of 15%. Only 3.7% of our participants switched at a rate of 50% or more. This suggests a widespread tendency to ignore both sources of advice; to stick with one's initial opinion rather than siding with the majority. Furthermore, this tendency appeared to increase in block 2 (switch rate = 12.05%) relative to block 1 (switch rate = 27.86%).

To meaningfully discuss the extent and nature of reliance, we conducted more specific reliance analyses. For these, we chose to focus on only those participants who appeared willing to rely upon either the co-worker or automation. In selecting participants for these analyses, we needed to balance our desire to focus on these individuals with the fact that there were relatively few of them. We therefore elected to retain only participants who had a switch rate of at least 20% on non-target trials (N = 90). That is, we retained participants who, when outnumbered 2–1 by the coworker and automation, switched to agree with the majority at least 20% of the time. This choice was in line with our conservative conceptualisation of 'reliance' on a source as a 'switch' in which the user changed from his/her initial opinion to agree with a particular source.

The characteristics of this reduced sample were similar to those of the main sample (M_{age} = 19, 77% female), although the racial composition changed somewhat, with white participants comprising a larger percentage of the reduced sample (80% white). Each of the three reliability conditions were relatively evenly represented in the reduced sample (human better = 34%, machine better = 37%, human and machine equal = 29%).

3.4 Hypothesis testing

Because our data include within-subject observations that were nested within persons, it is multilevel. We therefore assessed the relationship between trust and liking with reliance using multilevel modelling in HLM 6.0 (Raudenbush et al., 2001). Separate models were used for task blocks 1 and 2. In this analysis, level-1 represents the within-person behaviours on the ten target trials in each block. Thus, the level-1 model reflects each individual participant's pattern of reliance throughout the target trials in the task block.

The level-2 variables included DiffTrust and DiffLiking. Because DiffTrust and DiffLiking have a meaningful zero (zero indicates equal trust or equal liking between the human and machine advisors), these variables were not centred. For block 1, the time 1 attitudes were used as predictors. For block 2, separate models were assessed using the time 1, time 2 and time 3 attitudinal predictors. A significant effect of a level-2 predictor on the intercept (β_0) would indicate a main effect of that predictor on reliance on trial 0, which was the first target trial presented. A significant effect of a level-2 predictor on the slope (β_1) would indicate an interaction between that predictor and trial, suggesting that the predictor affects the user's linear *pattern* of reliance over the course of the block. When time 1 predictors were used to predict reliance in block 1, chi-square difference tests of the deviance suggested that a fixed error structure be used ($\Delta\chi^2$ = 4.24, Δdf = 2, p = .12).

Table 3 Results of HLM Model predicting block 1 relative reliance with time 1 predictors

			Coefficient	*Robust standard error*	*T-ratio*	*d.f.*	*p-value*
Model			*Block 1 reliance; time 1 predictors*				
For	Intercept (π_0)						
	Intercept	β_{00}	–.02	.02	–.64	87	.52
	DiffTrust	β_{01}	.09	.05	2.02	87	.05
	DiffLiking	β_{02}	–.04	.03	–1.61	87	.11
For	Trial slope (π_1)						
	Intercept	β_{10}	.00	<.01	.73	892	.47
	DiffTrust	β_{11}	–.01	.01	–1.29	892	.20
	DiffLiking	β_{12}	.01	<.01	2.13	892	.03

Note: N = 90

Results are displayed in Table 3. Using robust standard errors, the effect of DiffTrust on the reliance intercept approached significance (β_{01} = .09, p = .05), suggesting a trend for those who trusted the automation more than the human to rely on the automation more than the human on the first target trial. In addition, a significant interaction was found between DiffLiking and trial (β_{12} = .01, p = .03). The interaction was graphed according to Preacher et al. (2006) and is displayed in Figure 2. Over the course of block 1, participants with greater liking for the human relative to the automation significantly increased their reliance on the human over the course of the block and vice versa.

Figure 2 Interaction of block 1 trial and time 1 DiffLiking predicting block 1 relative reliance

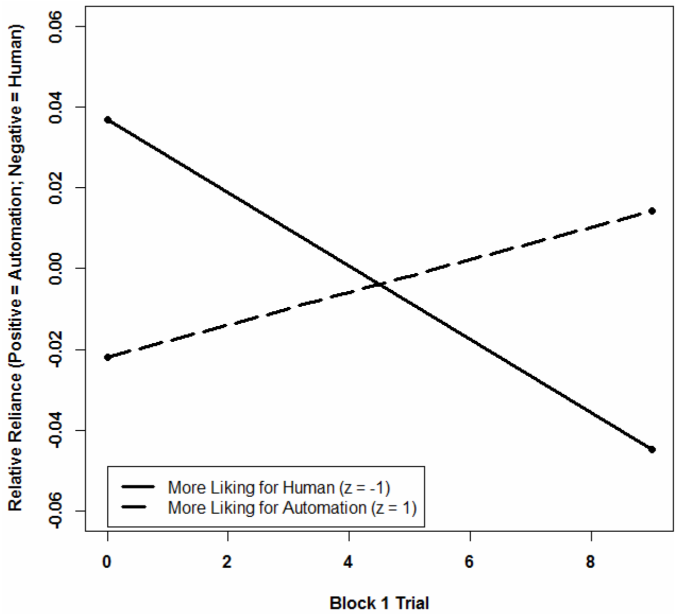


Table 4 Results of HLM model predicting block 1 relative reliance with time 2 predictors

			Coefficient	*Robust standard error*	*T-ratio*	*d.f.*	*p-value*
Model			*Block 1 reliance; time 2 predictors*				
For	Intercept (π_0)						
	Intercept	β_{00}	.02	.02	.60	87	.55
	DiffTrust	β_{01}	–.09	.04	–2.14	87	.04
	DiffLiking	β_{02}	.06	.04	1.41	87	.16
For	Trial slope (π_1)						
	Intercept	β_{10}	–.00	<.01	–.09	87	.93
	DiffTrust	β_{11}	.02	.01	2.76	87	.01
	DiffLiking	β_{12}	–.01	.01	–.94	87	.35

Note: $N = 90$

Next, time 2 predictors were used at level-2 (see Table 4). Note that because the time 2 attitudes were measured following block 1, this analysis can best be interpreted as the effects of reliance during block 1 on attitudes measured following block 1 (although the analyses cannot be run in this direction due to the necessity of a level-1 outcome variable). In this case, chi-square difference tests indicated a significant improvement in model fit when a random error structure was modelled ($\Delta\chi^2 = 8.39$, $\Delta df = 2$, $p = .02$).

A counterintuitive main effect emerged in which users who relied relatively more on the human's advice reported trusting the automation more than the human ($\beta_{02} = -.09$, $p = .02$). However, this effect was qualified by a significant interaction between DiffTrust and trial (β_{11} = .02, p = .01). The form of this interaction is displayed in Figure 3. As shown, individuals who increased their reliance on the human over the course of task block 1 reported having greater trust in the human at the conclusion of the block. In contrast, those who increased the extent to which they relied on the automation over the course of block 1 reported greater trust in the machine. Thus, it seems that participants' pattern of reliance during block 1 was significantly associated with relative trust following block 1.

Figure 3 Interaction of block 1 trial and time 2 DiffTrust predicting block 1 relative reliance

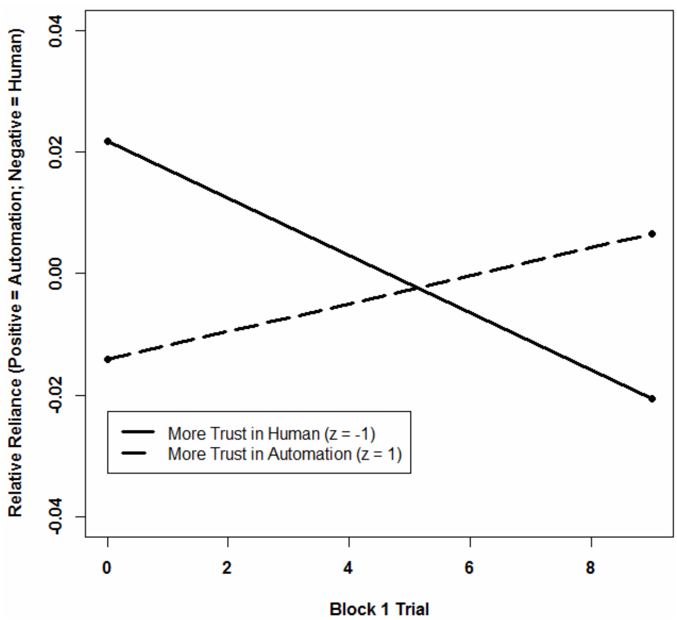


Table 5 Results of HLM models predicting block 2 relative reliance with time 1, time 2 and time 3 predictors

			Time 1 coefficient	*SE*	*d.f.*	*Time 2 coefficient*	*SE*	*d.f.*	*Time 3 coefficient*	*SE*	*d.f.*
For	Intercept (π_0)										
	Intercept	β_{00}	–.00	.02	87	.01	.03	87	.01	.02	87
	DiffTrust	β_{01}	.01	.04	87	.07	.05	87	.05	.04	87
	DiffLiking	β_{02}	–.01	.02	87	–.01	.04	87	.00	.02	87
For	Trial slope (π_1)										
	Intercept	β_{10}	–.00	.01	894	–.01	.01	894	–.01	.01	894
	DiffTrust	β_{11}	–.00	.01	894	–.01	.01	894	–.01	.01	894
	DiffLiking	β_{12}	–.00	< .01	894	.00	.01	894	.00	.01	894

Notes: All coefficients were non-significant,
$p > .05$,
$N = 90$

Behavioural reliance in block 2 was examined using time 1, time 2 and time 3 DiffTrust and DiffLiking. As shown in Table 5, none of these predictors showed significant relationships with block 2 reliance behaviour. Thus, while relative attitudes were significantly associated with relative reliance early in the task, these effects seemed to disappear in the second half of the task.

3.5 Relationship of source reliability with relative trust, liking and reliance

Ideally, users' relative levels of trust and liking for the advice sources should correspond to the advice sources' relative levels of reliability (accuracy). A close correspondence between relative reliability and relative trust and liking would suggest that trust and liking were correctly calibrated to actual reliability. To examine calibration, we correlated assigned reliability condition with DiffTrust and DiffLiking using both the full and reduced samples. We also calculated relative reliance by subtracting the percentage of the time the user switched his or her initial opinion to agree with the human from the percentage of the time the user switched to agree with the automation. These relative reliance variables were calculated using target trials only. The results of these correlations are displayed in Table 6.

Table 6 Correlations between relative reliability, DiffTrust, DiffLiking and relative reliance for the full and reduced samples

		Full sample (N = 234)		*Reduced sample (N = 90)*	
		Relative reliability (–1 = co-worker more reliable, 0 = equal reliability, 1 = machine more reliable)			
		r	*p*	*r*	*p*
DiffTrust (M-H)	Time 1	.07	*ns*	.11	*ns*
	Time 2	.21*	< .01	.24*	.02
	Time 3	.27*	< .01	.31*	<.01
DiffLiking (M-H)	Time 1	–.04	*ns*	.03	*ns*
	Time 2	.17*	.01	.30*	.01
	Time 3	.16*	.02	.27*	.01
Relative reliance (M-H)	Block 1	.06	*ns*	.07	*ns*
	Block 2	.39*	< .01	.36*	<.01

Note: $*p < .05$

As shown, time 1 DiffTrust and DiffLiking were not significantly associated with reliability. This lack of association is expected because these initial measurements were taken prior to the start of the task. In block 1, actual reliability was not significantly associated with relative reliance. This may indicate that users required some time to gain sufficient experience with the advice sources to perceive relative reliability.

However, at measurement Times 2 and 3, DiffTrust and DiffLiking were significantly associated with actual reliability (rs = .24–.31, ps < .05), indicating that users were able to calibrate their relative levels of trust and liking to some degree. In addition, relative reliability was significantly associated with relative reliance in block 2, such that when the machine was more reliable, participants tended to rely on its advice more heavily and vice versa (*rs* = .36 and .39, *ps* < .05).

4 Discussion

The first goal of the present study was to examine the relationship of trust and liking with reliance in a dual-advisor situation. Many of our results are consistent with those found in single-advisor situations, suggesting that some aspects of advice-taking seem to generalise to situations in which conflicting advice is provided by human and automated advisors. Consistent with the results reported by Merritt (2011) in a single-advisor context, we found that time 1 relative liking was significantly associated with users' pattern of reliance during the first task block. Time 1 relative trust had a marginally significant association with reliance on the first target trial; however, this association did not reach conventional levels of significance when relative liking was accounted for. Thus, it seems that the role that liking plays in early reliance decisions may generalise to from single-advisor to dual-advisor situations.

However, unlike past work in single-advisor settings, we found no significant effects of relative trust or relative liking during the second half of our task. Perhaps the decision criteria people use in dual-advisor settings differ, or change over time differently, from those in single-advisor settings. This finding is consistent, however, with work by Riley (1996) in a single-advisor setting, who found that learning about the task and failure states seemed to be associated with a shift in reliance strategies over the course of the task. As participants gained more experience and were better able to identify when the aid was failing, they seemed to move from a reliance strategy based on trust to one based on recognition of system states. In the present task, there were no clear indicates of whether the advisors were failing (e.g., the errors were not obvious to the participants in past studies using these stimuli); however, Riley (1996) also found individual differences in the reliance strategies used. It seems that as the task progressed, participants may have shifted from reliance strategies based on relative trust and liking to other types of reliance strategies.

Our examination of means for trust and liking indicated that over time, a general bias toward trusting and liking the human advisor more than the automated advisor appeared. This pattern is consistent with theory on the perfect automation schema (Dzindolet et al., 2003). Users with a stronger perfect automation schema are thought to find automation errors more surprising and thus weight automation errors heavily in their judgments of trust. Because no such schema is proposed to exist for humans, the perfect automation schema could explain the decreases in automation trust relative to human trust seen here. It is interesting that a similar pattern emerged for liking. Although little research has examined liking for automation, it seems reasonable that liking for an automated tool would depend more heavily on task performance than liking for a human might. While humans have a variety of personal characteristics that can make them likeable, automation may have relatively fewer and may be seen as existing to serve a sole purpose. Thus, errors in the task for which the automation was designed may result in severe decreases in liking relative to errors in human performance.

We also found it interesting to note that the actual relative reliability of the human and automated advisors was significantly associated with block 2 reliance. While relative trust and liking were significantly associated with relative reliability, they did not translate into reliance decisions. Instead, relative reliability was directly significantly correlated with relative reliance. Research suggests that in complex tasks, people may learn to recognise patterns at an unconscious level before they identify those patterns consciously (Bechara et al., 1997). It is possible that participants may have been able to behaviourally rely on the more accurate source even if conscious acknowledgement of differential trust was absent. However, because the association is low-to-moderate in magnitude, users may benefit from interventions designed to more effectively calibrate their relative levels of trust, liking and reliance.

We found it interesting to note the large percentage of participants who seemed to disuse both the human and automated advisors. Even when faced with a situation in which *both* advisors contradicted the participant, participants stuck with their initial decision 81.28% of the time. In contrast, in our past research using a single-advisor version of this task, participants stuck with their initial decision only 42.20% of the times when automation disagreed [data from (*citation removed for masked review*)]. The reason for this discrepancy between our single- and dual-advisor versions of the task is unclear. One possibility is that receiving advice from two sources is cognitively demanding and participants might seek to reduce cognitive load by dismissing the advice. Another is that disagreement from two sources may cause a greater degree of reactance in participants, motivating them to 'stick to their guns' and disuse both advice sources. Beck et al. (2007) discussed intent errors, in which participants accurately perceive an aid's capabilities but disuse it regardless. Such intent errors may reflect a sense of competition between the user and the aid – perhaps dual-advisor situations may increase such a sense of competition. Regardless of the reason, widespread rates of disuse can obscure the associations between attitudes and advice reliance. Identifying the mechanism behind disuse may help illuminate an important moderator of the attitude-reliance relationship.

4.1 Practical implications

The results of the present study provide further evidence that liking may be a key individual difference variable related to automation reliance, particularly in the early stages of interaction. Furthermore, relative liking for the automation significantly associated with the pattern of reliance found in the early stages of the task. These findings have implications for automation implementation. Disuse of new automation relative to human coworkers could be reduced by interventions designed to increase liking. Merritt (2011) found that inducing changes in mood can affect liking and several mood induction techniques have been identified (e.g., Stein et al., 2000; Velten, 1968).

We found significant, but relatively small, correlations between advisors' relative reliabilities and relative trust and liking. This suggests that while users can perceive differential reliabilities, their levels of trust and liking (and by extension, reliance) seem to be influenced by other factors as well. Thus, suggestions that organisations simply educate users on aid reliability may not have sufficiently large effects. Additional efforts may be required in order to correctly calibrate trust.

4.2 Limitations and suggestions for future research

Participants interacted with both the human and automation through a computer-mediated interface that was identical for both sources. This design feature was both a strength and a limitation. Standardisation allowed for the strictest comparison and provided confidence that the differences found were due to knowledge that one source was human and one was automated. However, it also limits generalisability to other settings.

Given the limited cues in these interactions, it is possible that liking could have been determined predominantly by propensity to trust[1]. Although propensity to trust was not a focal variable in this study, we had measured it via self-report for both humans and machines. In the full sample (N = 234), propensity to trust other humans was correlated moderately and significantly with liking for the human advisor at times 1–3 (*rs* = .39, .34, .24). For machines, propensity to trust had a higher correlation with liking for the automation at time 1, but the correlations decreased over the course of the task, becoming non-significant at time 3 (*rs* for times 1-3 = .55, .31, .22, respectively). These correlations indicate that propensity to trust was associated with liking, particularly earlier in the task. However, liking did not seem to be based completely on propensity to trust. Given that the interface controlled for the number of social cues provided from either the human or automation, the richness of interaction cues provided by the advisors may be an important moderator to examine in future research. Future studies might increase social cues for one of both sources (e.g., video conference and avatars could be used) in order to examine the potential moderating effects of interaction format richness.

In other contexts, additional factors may differ between humans and automation. Our participants had relatively few consequences associated with their performance. They were asked to imagine that they were acting as security screeners and responsible for the safety of others, but no real outcomes were associated with errors. Nor were participants asked to multi-task or balance multiple demands simultaneously. Therefore, future research should further investigate the effects of these and other potential moderators on relative reliance.

Finally, in the present study we did not assess user self-confidence, or trust in oneself. The high rate of disuse seen in this study could indicate that user self-confidence may be key in predicting reliance decisions. Thus, we recommend that future research directly measure self-confidence so that its interactions with trust in the other advisors can be examined empirically.

5 Conclusions

As more modern and sophisticated technology develops to aid task performance, human decision makers will have to compare, analyse and incorporate information/advice from a large number of sources. These sources will include both other humans as well as automated decision aids. Research on reliance on human and automated information aids in multiple advisor setting is still in its infancy. This paper provides some unique insights into how trust and liking may differentially influence reliance on humans and machine advisors – specifically in conditions when information sources provide conflicting advice. We hope this paper will encourage more researchers to explore this fascinating area.

References

Bechara, A., Damasio, H., Tranel, D. and Damasio, A.R. (1997) 'Deciding advantageously before knowing the advantageous strategy', *Science*, Vol. 275, No. 5304, pp.1293–1295.

Beck, H.P., Dzindolet, M.T. and Pierce, L.G. (2007) 'Automation usage decisions: controlling intent and appraisal errors in a target detection task', *Human Factors*, Vol. 49, No. 3, pp.429–437, Doi: 10.1518/001872007X200076.

Connors, M.M., Harrison, A.A. and Summit, J.S. (1994) 'Crew systems: integrating human and technical subsystems for the exploration of space', *Behavioral Science*, Vol. 39, No. 3, pp.183–212.

de Vries, P., Midden, C. and Bouwhuis, D. (2003) 'The effects of errors on system trust, self-confidence, and the allocation of control in route planning', *International Journal of Human-Computer Studies*, Vol. 58, No. 6, pp.719–735.

Dijkstra, J.J. (1999) 'User agreement with incorrect expert system advice', *Behaviour and Information Technology*, Vol. 18, No. 6, pp.399–411.

Dijkstra, J.J., Liebrand, W.B.G. and Timminga, E. (1998) 'Persuasiveness of expert systems', *Behaviour and Information Technology*, Vol. 17, No. 3, pp.155–163.

Dzindolet, M.T., Petersen, S.A., Pomranky, R.A., Pierce, L.G. and Beck, H.P. (2003) 'The role of trust in automation reliance', *International Journal of Human and Computer Studies*, Vol. 58, No. 6, pp.697–718.

Dzindolet, M.T., Pierce, L.G., Beck, H.P. and Dawe, L.A. (2002) 'The perceived utility of human and automated aids in a visual detection task', *Human Factors*, Vol. 44, No. 1, pp.79–94.

Harvey, N. and Fischer, I. (1997) 'Taking advice: accepting help, improving judgment, and sharing responsibility', *Organizational Behavior and Human Decision Processes*, Vol. 70, No. 2, pp.117–133.

Lee, J.D. and Moray, N. (1992) 'Trust, control strategies, and allocation of function in human-machine systems', *Ergonomics*, Vol. 35, No. 10, pp.1243–1270.

Lee, J.D. and See, K.A. (2004) 'Trust in automation: designing for appropriate reliance', *Human Factors*, Vol. 46, No. 1, pp.50–80.

Lee, J.D. and Moray, N. (1994) 'Trust, self-confidence, and operators' adaptation to automation', *International Journal of Human-Computer Studies*, Vol. 40, No. 1, pp.153–184.

Lewandowsky, S., Mundy, M. and Tan, G.P.A. (2000) 'The dynamics of trust: comparing humans to automation', *Journal of Experimental Psychology: Applied*, Vol. 6, No. 2, pp.104–123.

Lyons, J.B. and Stokes, C.K. (2012) 'Human-human reliance in the context of automation', *Human Factors*, Vol. 54, No. 1, pp.112–121.

Madhavan, P. and Wiegmann, D.A. (2005) 'Cognitive anchoring on self-generated decisions reduces operator reliance on automated diagnostic aids', *Human Factors*, Vol. 47, No. 2, pp.332–341, doi: 10.1518/0018720054679489.

Madhavan, P. and Wiegmann, D.A. (2007a) 'Effects of information source, pedigree, and reliability on operator interaction with decision support systems', *Human Factors*, Vol. 49, No. 5, pp.773–785, doi: 10.1518/001872007X230154.

Madhavan, P. and Wiegmann, D.A. (2007b) 'Similarities and differences between human-human and human-automation trust: an integrative review', *Theoretical Issues in Ergonomics Science*, Vol. 8, No. 4, pp.277–301.

Merritt, S.M. (2011) 'Affective processes in human-automation interactions', *Human Factors*, Vol. 53, No. 4, pp.356–370.

Merritt, S.M. and Ilgen, D.R. (2008) 'Not all trust is created equal: disposition and history-based trust in human-automation interactions', *Human Factors*, Vol. 50, No. 2, pp.194–210.

Merritt, S.M., Heimbaugh, H., LaChapell, J. and Lee, D. (2013) 'I trust it, but I don't know why: effects of implicit attitudes toward automation on trust in an automated system', *Human Factors*, Vol. 55, No. 3, pp.520–534, doi: 10.1177/0018720812465081.

Merritt, S.M., Sinha, R., Curran, P. and Ilgen, D.R. (2011) 'Comparing trust in humans and machines: development of a measure', Poster presented at *55th Annual Meeting of the Human Factors and Ergonomics Society*, Las Vegas, NV.

Moray, N., Inagaki, T. and Itoh, M. (2000) 'Adaptive automation, trust, and self-confidence in fault management of time-critical tasks', *Journal of Experimental Psychology: Applied*, Vol. 6, No. 1, pp.44–58.

Nass, C. and Moon, Y. (2000) 'Machines and mindlessness: social responses to computers', *Journal of Social Issues*, Vol. 56, No. 1, pp.81–103, doi: 10.1111/0022-4537.00153.

Neuman, P.G. (1986) 'AI in medical diagnosis and aviation', *ACM SIGSOFT, Software Engineering Notes*, Vol. 11, No. 2, p.5.

Parasuraman, R. and Miller, C.A. (2004) 'Trust and etiquette in high-criticality automated systems', *Communications of the ACM*, Vol. 47, No. 4, pp.51–55, doi: 10.1145/975817.975844.

Parasuraman, R. and Riley, V. (1997) 'Humans and automation: use, misuse, disuse, abuse', *Human Factors*, Vol. 39, No. 2, pp.230–253.

Preacher, K.J., Curran, P.J. and Bauer, D.J. (2006) 'Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis', *Journal of Educational and Behavioral Statistics*, Vol. 31, No. 4, pp.437–448.

Raudenbush, S.W., Bryk, A.S. and Congdon, R.T. (2001) *HLM for Windows*, Scientific Software International, Chicago.

Riley, V. (1996) 'Operator reliance on automation: theory and data', in R. Parasuraman and M. Mouloua (Eds.): *Automation and Human Performance: Theory and Applications*, pp.19–35, Erlbaum, Mahwah, NJ.

Sollins, K.R. (1986) 'A medical risk of computers', *AGM SIGSOFT, Software Engineering Notes*, Vol. 11, No. 3, p.11.

Sparaco, P. (1994) 'A330 crash to spur changes at airbus', *Aviation Week & Space Technology*, Vol. 141, No. 6, pp.20–22.

Stein, K.D., Goldman, M.S. and Del Boca, F.K. (2000) 'The influence of alcohol expectancy priming and mood manipulation on subsequent alcohol consumption', *Journal of Abnormal Psychology*, Vol. 109, No. 1, pp.106–115.

Velten, E. (1968) 'A laboratory task for induction of mood states', *Behaviour Research and Therapy*, Vol. 6, No. 4, pp.473–482.

Wiegmann, D., McCarley, J.S., Kramer, A.F. and Wickens, C.D. (2006) 'Age and automation interact to influence performance of a simulated luggage screening task', *Aviation, Space, and Environmental Medicine*, Vol. 77, No. 8, pp.825–831.

Notes

1 We thank an anonymous reviewer for this suggestion.