Available online at www.sciencedirect.com

SciVerse ScienceDirect

Int. J. Human-Computer Studies 71 (2013) 410–424

International Journal of Human-Computer Studies

www.elsevier.com/locate/ijhcs

A framework for explaining reliance on decision aids ☆

Kees van Dongen[a], Peter-Paul van Maanen[a,b,*]

[a] Netherlands Organisation for Applied Scientific Research (TNO), P.O. Box 23, 3769 ZG Soesterberg, The Netherlands
[b] Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands

Received 25 October 2010; received in revised form 8 August 2012; accepted 18 October 2012
Communicated by S. Wiedenbeck
Available online 11 December 2012

Abstract

This study presents a framework for understanding task and psychological factors affecting reliance on advice from decision aids. The framework describes how informational asymmetries in combination with rational, motivational and heuristic factors explain human reliance behavior. To test hypotheses derived from the framework, 79 participants performed an uncertain pattern learning and prediction task. They received advice from a decision aid either before or after they expressed their own prediction, and received feedback about performance. When their prediction conflicted with that of the decision aid, participants had to choose to rely on their own prediction or on that of the decision aid. We measured reliance behavior, perceived and actual reliability of self and decision aid, responsibility felt for task outcomes, understandability of one's own reasoning and of the decision aid, and attribution of errors. We found evidence that (1) reliance decisions are based on relative trust, but only when advice is presented after people have formed their own prediction; (2) when people rely as much on themselves as on the decision aid, they still perceive the decision aid to be more reliable than themselves; (3) the less people perceive the decision aid's reasoning to be cognitively available and understandable, the less people rely on the decision aid; (4) the more people feel responsible for the task outcome, the more they rely on the decision aid; (5) when feedback about performance is provided, people underestimate both one's own reliability and that of the decision aid; (6) underestimation of the reliability of the decision aid is more prevalent and more persistent than underestimation of one's own reliability; and (7) unreliability of the decision aid is less attributed to temporary and uncontrollable (but not external) causes than one's own unreliability. These seven findings are potentially applicable for the improved design of decision aids and training procedures.

Keywords: Decision support systems; Automation trust; Automation reliance

1. Introduction

Information and communication technology is changing the nature of work. The use of decision aids in complex systems, such as aviation, nuclear power, health care or command and control systems is becoming increasingly common. The assumption behind the introduction of decision aids is that a team of a human and a decision aid will be more effective than either human or decision aid working alone. Performance improvement by introducing decision aids is difficult to predict, because decision aids are not always used appropriately. It is often found that users tend to rely too much or too little on decision aids (Parasuraman and Riley, 1997). For instance, Skitka et al. (1999) found that unaided participants made fewer errors than participants who worked with a decision aid. The last group relied too much on the decision aid and missed events that they could have discovered manually. Like human operators, in complex domains, it is not likely that decision aids are 100% reliable. A problem with decision aids is that these systems often have incomplete or unreliable data or knowledge and use simplifying assumptions that make them brittle (Guerlain et al., 1999). This means that users cannot blindly accept advice from a decision aid; sometimes they need to reject advice and rely

☆ Authors in alphabetical order and all have made a comparable contribution.

*Corresponding author at: Netherlands Organisation for Applied Scientific Research (TNO), P.O. Box 23, 3769 ZG Soesterberg, The Netherlands. Tel.: +31 8886 65952; fax: +31 346 353 977.

E-mail addresses: kees.vandongen@tno.nl (K. van Dongen), peter-paul.vanmaanen@tno.nl (P.-P. van Maanen).

1071-5819/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ijhcs.2012.10.018

on their own decision. The tendency to accept advice depends among others on the reliability of the decision aid.

The relationship between the reliability of the decision aid and the users' reliance on decision aids is complex and multifaceted (Thomas and Rantanen, 2006; Parasuraman and Riley, 1997; Dzindolet et al., 2003; Lee and See, 2004). It is difficult to determine a fixed threshold for an acceptable level of unreliability (Thomas and Rantanen, 2006; Wickens and Dixon, 2007). Reliance on advice is mediated by a range of cognitive variables of which trust in oneself (self-confidence) and trust in the decision aid are two central concepts. To increase the effectiveness of human–computer collaboration a body of frameworks has been developed to better understand how people use decision aids. This paper adds to this body with a new framework. To prove the soundness of this theoretical framework, seven hypotheses are derived from it and tested in a task and experimental environment specifically developed for this purpose.

2. Background

Several frameworks of trust have been developed to identify factors that affect reliance on decision aids (e.g., Dzindolet et al., 2001; Lee and See, 2004). The framework of Dzindolet et al. (2001) emphasizes cognitive, social and motivational factors that affect reliance on decision aids. Concerning cognitive factors, users for instance compare their perception of the reliability of the decision aid with their perception of their own reliability. Whether this leads to appropriate reliance on the advice from the decision aid, depends however on whether these perceptions of reliability correspond to reality. Dzindolet et al. (2001) have suggested that users have a perfect automation schema and often wrongly expect decision aids to perform nearly perfect due to their supposedly infallible calculation capabilities. Another example of a cognitive factor is that users sometimes rely on decision aids to save mental effort. Mosier and Skitka (1996) have used the term 'automation bias' to refer to the tendency to heuristically rely on a decision aid.

Concerning social factors, trust is an important factor. People rely more on decision aids when they trust the decision aid more than themselves (Mosier and Skitka, 1996). Other examples of social factors are: feelings of control and moral obligation to rely on oneself. Finally, diffusion of responsibility has been identified as a factor affecting the tendency to rely on decision aids (Mosier and Skitka, 1996). The idea is that decision makers feel less responsible for the performance of the human–computer system and thus invest less effort when working together with a decision aid.

The strength of the framework of Dzindolet et al. (2001) is that it identifies many psychological factors and processes that affect reliance on decision aids. A disadvantage of the framework is that it is not specific on the effects of experience and performance feedback on the dynamics of trust and not specific on the appropriateness of trust.

The framework of Lee and See (2004) and the model of Gao and Lee (2006) emphasize the dynamics of trust in, and reliance on, automation and take into account the role of feedback. Trust is not static, it changes over time as it is influenced by experience with performance. Like the framework of Dzindolet et al. (2001) this framework recognizes relative trust as a basic component of decisions about reliance: reliance is determined by the difference between a decision maker's trust in a decision aid and the confidence he has in his own performance. If this difference exceeds a particular threshold, i.e., when the trust in the decision aid is higher by some amount than the decision maker's self-confidence, then he will switch from relying on himself to relying on the decision aid and vice versa. Trust, in turn, depends upon one's own previous performance and that of the decision aid. This creates a feedback loop between the previous performance and trust which is an important element of the framework (Gao and Lee, 2006; Lee and See, 2004).

The framework of Lee and See (2004) also emphasizes how changes in trust are affected by system factors such as interface features; personality factors such as the tendency to trust; by indirect or organizational factors such as gossip and reputation; norms and expectations; and task and context factors such as workload and time constraints. In addition to a focus on the dynamics of trust this framework provides concepts to determine the appropriateness of trust and the relation between actual and perceived reliability. The advantage of this framework is that it not only emphasizes the effects of performance feedback on the dynamics of trust but also directs attention to concepts like trust calibration. Since reliance decisions are based on perceived performance reliability, rather than actual performance reliability, it is important to determine how well calibrated people's trust is. A disadvantage of this framework is that it does not specifically address reliance on recommendations of decision aids since it focuses on reliance on automation in general. For our purpose, explaining why people choose the recommendation from a decision aid over their own opinion, we need a focus on additional task and psychological factors.

3. Framework

3.1. Heuristic and systematic processing

We believe that integrating the strengths of the previous two frameworks and complementing it with additional task and psychological factors provides a useful view of factors affecting reliance on recommendations of decision aids. An important realization is that information on which people base their reliance decisions can be processed in different ways. According to Kahneman (2001) our thinking processes is directed by two systems: System 1 is fast, effortless and associative. System 2 operates slower, effortful and reasoned. System 2 can monitor the outcomes

of System 1: when one is not confident in its outcomes and willing to invest the effort required for System 2 processing, people switch from System 1 to System 2. From a System 2 perspective, reliance decision making can be seen as a rational choice. The decision maker will rely on a decision aid or on himself, and when doing so, he aims at maximizing correct outcomes and minimizing incorrect outcomes. In other words, a rational decision maker relies on the most reliable of the two. Within System 1 a distinction can be made between *skilled processing* and *heuristic processing*. In the first case judgments are based on 'valid cues' in the task environment and made by experienced decision makers with 'valid cue patterns in memory' (Kahneman and Klein, 2009). Skilled processing requires a more or less predictable environment in which individuals have the opportunity to learn the regularities of that environment. When there are no valid cues in the environment or when decision makers have no 'valid cue patterns in memory' people rely on heuristics and information that comes easy to mind (heuristic processing).

3.2. Order of advice, cognitive availability and reliance heuristics

Important factors in System 1 processing are the 'availability heuristic' and 'anchoring'. According to the *availability heuristic* people base their decisions on how easily information can be brought to mind (Tversky and Kahneman, 1973). *Anchoring* is the common human tendency to rely too heavily, or 'anchor' on one piece of information that is brought to mind when making decisions (Tversky and Kahneman, 1974). For reliance decisions this means that when one's own opinion is cognitively more available than that of the decision aid, people tend to rely more on themselves (i.e., self-reliance heuristic). Furthermore, when the opinion of the decision aid is cognitively more available, people tend to rely more on the decision aid (i.e., automation-reliance heuristic). These heuristics can lead to good performance, when differences in cognitive availability correspond to differences in reliability. When this is not the case (i.e., when people are repeatedly relying on the less reliable opinions), reliance is biased.

We expect that cognitive availability of one's own opinion, relative to that of the decision aid, is higher in a condition in which advice is provided after people have formed their own judgment, compared to a condition in which advice is provided before people have done this. In the first condition one needs to think for oneself, in second condition one does not need to. This is also in line with other work showing that people who see computer recommendations early on, tend to shorten their own analysis (e.g., Layton et al., 1994).

3.3. Transparency and cognitive availability of reasoning

A major finding of the advice-giving and advice-taking literature is that people tend to discount advice (Bonaccio and Dalal, 2006). People tend to weigh their own opinions more heavily than they weigh others' (e.g., Yaniv and Kleinberger, 2000). The discounting of advice has been attributed to three causes: differential information (Yaniv and Kleinberger, 2000), anchoring and availability (Tversky and Kahneman, 1973, 1974) and egocentric discounting (Harvey and Fischer, 1997). According to the differential information explanation, decision makers have privileged access to their internal reasons for holding their own opinions but not to the advisor's internal reasons (Yaniv and Kleinberger, 2000). Not only the cognitive availability of options seems to affect reliance decisions but also the cognitive availability of the reasons supporting these options. A common assumption in cognitive psychology is that the weight placed on a judgment depends on the evidence that is recruited to support that judgment. It has been argued that advice is often under-used because decision makers have no direct access to the reasons underlying the advice from an advisor (in this case a decision aid), while they do have direct access to the reasons supporting their own judgment as well as to the strength of those reasons (Yaniv and Kleinberger, 2000). To understand reliance decision making we believe that the effect of these cognitive factors needs to be studied.

3.4. Feeling of responsibility and invested effort

Kahneman (2001) suggests that when there are signs that you are in a cognitive minefield, you cannot be confident in System 1. In these situations System 2 is more effective, but as previously mentioned System 2 processing requires mental effort that people need to be willing to invest. The degree to which people are concerned with, and they feel responsible for, the overall task outcome, is expected to affect the effort people are willing to spend on a task. It has been argued that the availability of automated decision aids can feed into the general human tendency to travel the road of least cognitive effort and heuristic reliance on automation, replacing effortful information seeking and processing (Skitka et al., 1999). It should be noted that this assumes a task environment in which a tendency to heuristically rely on decision aids is induced. We expect that in task environments that induce a tendency to rely on oneself, and when decision makers feel more responsible, they invest more effort in changing their mind when they disagree with decision aids, assuming of course that they think this leads to a higher overall performance. Because human information processing depends on the effort they are willing to invest, it is important to study motivational effects, like the feeling of responsibility for task outcomes in relation with the other factors already mentioned.

3.5. Performance feedback, perceived reliability, attribution and relative trust

Where a perfect automation schema may lead to overestimation of a decision aid, another well-established bias, the overconfidence effect, suggests that people tend to

overestimate their own performance as well. Because in real-life situations decision aids are often used more than once, we think that studying the effects of performance feedback on performance expectancies and reliance decisions are important and need to be represented in a theoretical framework. When decision makers build up experience with the task, with themselves and with a decision aid, initial positive expectations about the reliability of decision aids, which may be based on a schema of perfect automation, are adjusted based on feedback about the reliability of the decision aid, of oneself and of the human–computer team as whole. This feedback may lead to underestimation rather than overestimation of performance reliability, as negative experiences have a greater influence than positive experiences. Since reliance decisions are based on relative trust and perceptions of reliability that not necessarily correspond to reality, understanding how people estimate and interpret performance and errors is important. People tend to attribute behavior to causes that capture their attention. When we observe other people for instance, the person is often the primary reference point and attributions for others' behavior are more likely to focus on the person we see, not the situational forces acting upon that person that we may not be aware of Gilbert and Malone (1995). This may also be the case when interpreting the performance of a decision aid or of oneself. People have privileged access to the causes of their own errors, but often not to those of the errors of a decision aid. It is more easy to find excuses for our own errors, then for those of the decision aid. As a result, your own errors may be differently weighed than those of a decision aid.

Based on the abovementioned aspects of reliance decision making the theoretical framework in Fig. 1 can be constructed. This framework takes the following aspects into account: (1) factors in the task environment; (2) factors affecting motivational, heuristic and systematic processes; and (3) behavior and a feedback loop indicating information about effectiveness of that behavior. The numbered triangles correspond to the hypotheses that are described in the following section.

4. Hypotheses

4.1. Order of advice, cognitive availability and reliance heuristics

Based on relative trust and System 2 processing, we expect that when people perceive the decision aid to be better than themselves, they will agree more often with it. Similarly, when people think they are better themselves, we expect them to disagree more often with advice from the decision aid. When the difference in the perceived reliabilities of oneself and the decision aid is large and stable, reliance decisions are easier compared to when the difference is small and unstable. Activation of System 2 processing assumes that people are not confident in whom to rely on and that people are willing to invest mental effort in seeking and processing information related to this decision. It also assumes that response options are cognitively available. Making a decision whom to rely on requires that both one's own opinion and that of the aid is cognitively available and under consideration. Whether this is the case depends on task and psychological factors like the order in which advice is provided.

Advice from a decision aid can either be presented before or after the decision maker has formed his own opinion. When advice is presented first, people can automatically follow that advice without thinking about the problem themselves. This causes their own knowledge about the decision problem to be mentally less available. Sniezek and Buckley (1995) found that the answers of participants who received advice first, matched more often with the answers of the advisor compared to participants who first formed their own opinion. When people automatically follow an advisor, their own opinion is not cognitively available and they do not have to face the dilemma of whom to rely on. As a result they do not need to reason about relative trust. We expect that relative trust is not a good predictor of reliance, when advice is presented before people have formed their own opinion. This tendency to rely on advice may be reduced by actively involving humans in decision making.

This can be done by providing advice after the decision maker has formed his own judgment, instead of before. This is also done in critiquing systems (Guerlain et al., 1999; Silverman, 1992). When advice is presented after people have formed their initial decision, they cannot automatically follow advice. They are required to think for themselves first, which makes their own opinion and reasons for it more cognitively available.

When provided with feedback about the correctness of their own answers and those of the decision aid, they also become more aware of relative reliabilities and the dilemma of whom to rely on. When people have to express their own opinion before receiving advice we expect them to be sensitive to relative trust. When people can automatically match the answers of the decision aid, these answers are not cognitively available and people are expected to be less sensitive to information about relative trust. This boils down to the following hypothesis:

Hypothesis 1. When people trust themselves more than the decision aid (i.e., relative trust higher), they disagree more with the decision aid than when people do not trust themselves more than the decision aid (i.e., relative trust lower), but only when they receive advice after they have formed their own opinion.

4.2. Understandability of underlying reasoning

In contrast to one's own reasoning, the reasoning of a decision aid is often not easily accessible or understandable to the user, especially when the decision aid is a computer rather than a human. At best, only part of the decision aid's reasoning can be made transparent. However, this is often not understandable and therefore cognitively not available. A common assumption in cognitive psychology is that the

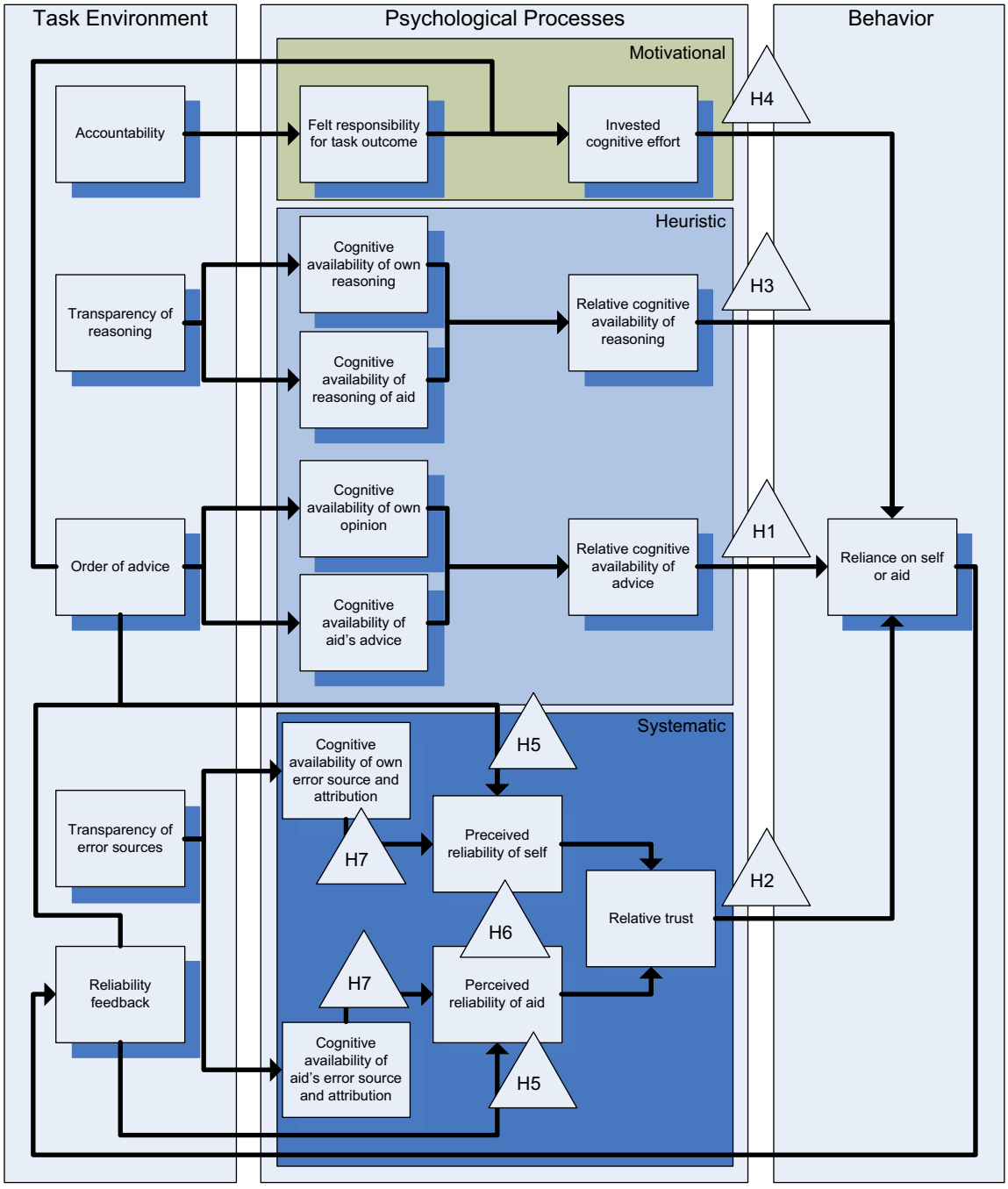

Fig. 1. Theoretical framework of reliance decision making.

weight placed on a judgment depends on the evidence that is recruited to support that judgment (Tversky and Koehler, 1994). Because the processes underlying the decision aid's advice are less available and understandable compared to those underlying one's own judgment, we expect that reliance on advice from the decision aid cannot be solely predicted by relative trust in performance reliability. This leads to the following two hypotheses:

Hypothesis 2. When people rely as much on themselves as on the decision aid, they still perceive the decision aid to be more reliable than themselves.

Hypothesis 3. The less people perceive the decision aid's reasoning to be cognitively available and understandable, the less people rely on the decision aid.

4.3. Feeling of responsibility

Working in a team has advantages and disadvantages. Although two may know more than one, two disadvantages of teamwork are that responsibility may be diffused between its members and that they do not feel accountable for the task outcome. Several researchers think of the human–computer system as a team in which one member is not

human (e.g., Bowers et al., 1996). The human may feel less responsible for the outcome when working with a (computer) decision aid than when working alone and may invest less mental effort in achieving the best team outcome. People feel the ultimate responsibility in a human–computer team to lie with the human user, rather than being shared between human and computer (Lewandowsky et al., 2000). It is expected that the less responsible the user feels, the less motivated the user is to invest mental effort in the task and in deciding whom to rely on, and the more likely he is to decide heuristically. In situations that induce the automation-reliance heuristic one expects users to heuristically rely on advice from the decision aid. In situations that induce the self-reliance heuristic, however, we expect the opposite effect.

When disagreement with the decision aid is salient, we expect people to invest the mental effort to change their mind and rely on the decision aid, but only when they feel responsible for the task outcome. When the felt responsibility for task outcome is low, we expect that people do not change their mind, tend to reject conflicting advice and rely on themselves. This results in the following hypothesis:

Hypothesis 4. The more people feel responsible for the task outcome, the more they rely on the decision aid.

4.4. Accuracy of perceived reliability

The above mentioned psychological concepts (i.e., perceived reliability, relative trust, cognitive availability of opinions and reasons, and felt responsibility for task outcome) may explain how reliance decisions are made, but do not explain whether reliance on advice is appropriate or not.

For appropriate reliance on advice one would expect a rational decision maker to rely on advice when this would increase the probability of goal achievement and to reject advice when it would decrease this probability. The decision to accept or reject advice is, however, not based on a comparison of the actual reliability of oneself or decision aid, but on how these reliabilities are perceived. Unfortunately these perceptions not necessarily correspond to reality and may be prone to random and systematic errors. Perceived reliability of oneself and perceived reliability of the decision aid may be underestimated or overestimated. When the direction or magnitude of estimation errors differs between oneself and decision aid this could lead to over-reliance or under-reliance on advice.

Concerning perception of one's own performance, studies of judgment under uncertainty have indicated that humans are often over-confident (e.g., Alba and Hutchinson, 2000). An explanation for this is that people tend to focus on supporting rather than contradictory evidence for a judgment, decision or prediction. Although pervasive in the literature, over-estimation of one's own performance is not universal (Brenner et al., 1996). May (1987, 1988)'s results for instance yielded 9% over-confidence when confidence in performance was estimated after each answer, whereas a 9% under-confidence was found when confidence in performance was estimated after each block. An explanation for this is that the estimated percentage correct is likely to be based on a general evaluation of the difficulty of the task or based on feedback about performance, rather than on a balance of arguments for and against each specific judgment (Brenner et al., 1996). Whether over- or underestimation of one's own performance is observed seems to depend on how and when people are asked to estimate their performance rate.

Concerning the perception of the decision aid's performance, Wiegmann et al. (2001) found that this is often underestimated. One reason for this may be that decision aids do not perform as expected. Dzindolet et al. (2001) argue that the perception of the reliability of an automated decision aid is filtered through the operator's 'perfect automation schema' (i.e., the expectation that automation will perform at near perfect rates). This sometimes unrealistic expectation may lead operators to pay too much attention to information that is in conflict with the schema: errors. Consequently, errors made by automation trigger a rapid decline in trust (Dzindolet et al., 2002). Whether the decision aid's performance is over- or under-estimated depends on what level of performance is expected in advance.

Providing users of decision aids with realistic information about the user's reliability and that of the decision aid results in more appropriately calibrated trust. Although performance feedback is expected to improve the accuracy of perceived reliability of oneself and decision aid, it is not expected to lead to a perfect correspondence. It has been argued that trust is a nonlinear function of performance and that it tends to be conditioned by negative experiences. Negative experiences have a greater influence on the perception of the reliability of the decision aid than positive experiences (Lee and See, 2004). Although performance feedback is expected to improve perceptions of reliability, underestimation of performance is expected because of this negativity effect. This leads to the following hypothesis:

Hypothesis 5. When feedback about performance is provided, people underestimate both one's own reliability and that of the decision aid.

4.5. Attribution bias

Factors and processes in the environment of a decision aid or a human decision maker can influence performance. The aid or human may perform well in some situations and not in others. Reliance on decision aids is thus not only affected by beliefs about performance reliability itself but also by beliefs about the factors and processes that affect this performance (Lee and See, 2004). When the processes underlying the decision aid's advice and the factors that affect the reliability of these processes are not transparent

and observable, causes of unreliability are *inferred* instead of observed. According to Weiner (1986)'s attribution theory, these causal attributions result in affective reactions, which may affect the level of trust in the decision aid or oneself. The attribution theory claims that how people assign success and failure can be divided into three categories. The first is internal or external attribution (locus). External attribution means that performance is perceived to be influenced by attributes outside the decision aid, such as the dynamics, complexity or unpredictability of the task. Internal attribution means that performance is perceived to be influenced by factors inside the decision aid or oneself, such as the competence or motivation to perform the task. The second is attribution to factors that are temporary or permanent (stability). When errors are thought to be caused by temporary factors, more optimism is expected than when errors are attributed to permanent attributes of the agent or task. The third category is attribution to factors one can or cannot control (controllability). When errors are assigned to causes that are not under control (e.g., unpredictability of situation) people are more forgiving than when errors are perceived to be under control (e.g., motivation). Relationships between people are qualitatively different from relationships between people and automation. Concepts like trust are however increasingly used to describe this relation. A challenge in extrapolating trust between people to trust between people and automated systems is that computers lack intentionality and motivational processes. Many attribution processes are heuristic and people may apply heuristics applied on other people also to systems.

Unfortunately, people are known to be biased in how causes are attributed to success and failure. Often, asymmetries are found in attribution of one's own performance to causes and that of others. One common bias in assigning causes is called the *fundamental attribution error* or *correspondence bias*. This is the tendency of people to under-emphasize situational causes for the behavior of others. Our own errors are more likely to be attributed to temporary, external or uncontrollable factors, while errors of others are more likely to be attributed to permanent, internal and controllable factors. In other words, we have excuses for our own errors, but not for those of others. Gilbert and Malone (1995) point out that for a correct attributional analysis that takes into account the role of situational causes for the behavior of others (i.e., decision aids in this case) one must not only have realistic expectations about their performance but also perceive and recognize situational constraints for others. The problem is, however, that factors constraining the reliability of a decision aid, such as the unreliability of the data it uses or the inherent unpredictability of the situation it operates in, are often not known or observable for the user. Because situational causes that constrain the user's task performance are more salient for the user than those that constrain the performance of the decision aid, these causes are also expected to be less cognitively available when causal attributions are made. As a result users are expected to be less forgiving and less optimistic about the performance of the decision aid than about their own performance.

Since the human–computer system can be regarded as a team in which one member is not human (e.g., Bowers et al., 1996), the theory on causal attribution in humans will also hold in the context of human–computer collaboration. This leads to the following two hypotheses:

Hypothesis 6. Underestimation of the reliability of the decision aid is more prevalent and more persistent than underestimation of one's own reliability.

Hypothesis 7. Unreliability of the decision aid is less attributed to temporary, external and uncontrollable causes than one's own unreliability.

5. Method

5.1. Participants

Seventy nine college students participated in the experiment. Their ages ranged from 18 to 38 yr ($M=23$). Participants were paid €35 for their participation.

5.2. Apparatus

5.2.1. Task and procedures

Before the training and experimental trials, participants read a cover story about a software company interested in evaluating the performance of their pattern learning software before applying it to more complex tasks on naval ships. To neutralize the effect of unrealistic expectations (i.e., perfect automation schema) the story pointed out that the level of reliability of both software and human performance was imperfect and was correct for 70% of the time, depending on the amount of training. This level was chosen because at this threshold humans tend to switch between relying on themselves and a decision aid (Wickens and Dixon, 2007). Prior pilots also showed that this was indeed the case.

Participants were asked to maximize the number of correct final predictions by relying on their own predictions as well as on the advice from the decision aid. The task required participants to predict what number (1, 2 or 3) would occur in the present trial. This prediction had to be based on the gradual discovery of a repeated pattern of numbers revealed in the previous trials. The pattern used (i.e., 2, 3, 1, 2, 3) was repeated until a sequence of 100 numbers was formed. These numbers were then partly randomized. That is, 10% of the numbers were altered to a different number, being either 1, 2 or 3. The reason for this randomization was to control for the difficulty of detecting the pattern and to make the participants think they still did not fully discover the correct pattern (otherwise performance would become 100% after a while; see also Section

5.2.3). At the end of each trial the correct number was revealed.

After the instructions participants performed 40 practice trials in which they had to discover a pattern in the data. The sequence of numbers for the practice trials was constructed in a similar way as described above. The participants could experience that their own performance and the advice from the decision aid was not perfect. For the first (practice) trials participants had no information about the correct sequence of numbers and could only guess. After a few trials, participants could form a more or less stable, but imperfect, mental model of the pattern of numbers based on the feedback they received. By building up, remembering, using and adjusting this model, participants could predict with some degree of success (i.e., aiming at an average success rate of around 70% under normal conditions) what number occurred next. After these training trials the actual experiment started.

To be able to observe possible learning effects each participant performed two experimental blocks, each consisting of 100 trials. After each experimental block participants had to fill in questionnaires. Between the two blocks participants had a short break.

5.2.2. Reliability of decision aid

The actual reliability of the decision aid was set to vary between 60% and 80% with an average reliability of 70% and an *SD* of 3% for each block (100 trials). Errors were defined as a deviation from the correct sequence of numbers as provided by the feedback. The causes of unreliability were not made transparent to the user and occurred at random intervals such that their occurrence could not be anticipated.

Because the average decision aid reliability was chosen to be equal to the expected human reliability (more on controlling for human reliability in Section 5.2.3), the number of situations in which participants were required to rely on themselves and the number of situations in which participants were required to rely on the decision aid, was also expected to be equal. If, for instance, the decision aid reliability was chosen to be much higher (e.g., 90%) or much lower (e.g., 50%) than the expected human reliability, participants were expected to be confronted with less challenging reliance decisions to be made. A similar argument can be made for the choice for a decision aid reliability of 70% as suggested by Wickens and Dixon (2007).

5.2.3. Task predictability

Since the reliability of individual participants was not under direct experimental control, we controlled the predictability of the task which influences the their reliability. About 10% of the sequence of numbers differed randomly from what would be expected by extrapolating the dominant and recurring pattern (2, 3, 1, 2, 3). Like the unreliability of the decision aid's advice, the unpredictability of the pattern occurred at random intervals. This made the decision to rely on oneself or the decision aid more difficult. Without the control of task predictability, floor or ceiling effects in the performance of reliance would occur, i.e., when the reliability of the participants' advice becomes predictable, their reliance decision also becomes predictable. The used sequence of numbers (of length 100) was determined beforehand and tested to have a Hamming distance of 10. Pilots (and later post-experimental questionnaires) showed that this partial randomization was enough to vary in a controlled manner the difficulty of the pattern. Participants did not suspect any randomization and did not find out that the pattern could never be discovered. This was due to the fact that humans tend to see patterns in noise and because of the convincing cover story told in the beginning of the experiment (which was also both confirmed in post-experimental questionnaires). So participants just suspected that their idea of what the pattern should be was imperfect and hence the confidence in themselves decreased (or increased when they were right). If the decision aid was correct, the confidence in the decision aid increased.[1] Eventually, calibration to this feedback leads to improved reliance decisions by the participants. Of course participants were unable to predict whether there was an instance of randomization, but the mentioned reliance decision could already be made independently of the advices given (i.e., the reliance decision could already be made before the predictions were made).

5.3. Design

Hypotheses were tested with a between subjects design, with the factor order of advice. The 79 participants were randomly assigned to a 'human first' (40) and a 'decision aid first' (39) condition. Each condition consisted of two blocks. During these blocks participants had to make 100 predictions and reliance decisions. See Table 1 for an overview of the gathered number of data points in each condition and block.

To test Hypothesis 1 we used the data set of the two blocks in the 'human first' and 'decision aid first' condition using the dependent variable 'agreement with the decision aid' ($N=158$). To test Hypotheses 2, 3 and 4, we used the data set of the two blocks in the 'human first' condition ($N=80$). This was done because only in the 'human first' condition the pre-advice prediction of participants is made explicit and available for analysis (more on this in Section 5.5). This data set was also used to test Hypotheses 5, 6 and 7.

5.4. Independent variables

5.4.1. Order of advice

In Figs. 2 and 3 the interfaces are shown of the task in the 'human first' and 'decision aid first' condition,

[1]Note that the reliability of the decision aid was not affected by randomization for the control of task predictability. Randomization for adapting the reliability of the decision aid was done using the already randomized sequence which was used to give feedback to the participant.

Table 1
Number of gathered data points in each condition and block.

Condition	Block 1 (100 trials)	Block 2 (100 trials)	Total
Human first	40	40	80
Decision aid first	39	39	78
Total	79	79	158

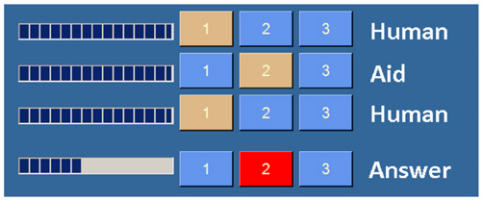

Fig. 2. Interface of the pattern learning task: human first.

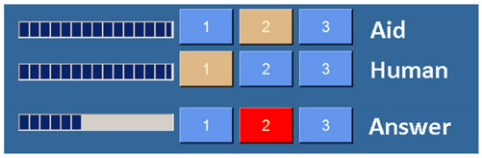

Fig. 3. Interface of the pattern learning task: decision aid first.

respectively. The interface was presented on a normal computer screen, without any other potentially distracting computer programs visible or running.

In the 'human first' condition the order of activities in each trial was: (1) predict; (2) receive advice from decision aid; (3) revise prediction (or re-select same prediction); and (4) receive feedback with the correct answer, corresponding to each successive row in Fig. 2. Participants first made their own independent prediction (initial prediction) by clicking on one of the three numbers in the first row. Then the decision aid communicated its advice by highlighting one of the three numbers on the second row. Neither the data nor the rules, on which this advice was based, were made transparent to the participant. On the third row participants had to formulate their answer again (final prediction) and were allowed to revise their initial prediction. When their initial prediction differed from the advice from the decision aid, they could either follow their own initial prediction or the advice from the decision aid.[2] On the fourth row the correct answer (highlighted button) and feedback about the success of the final prediction (red or green color) was provided. Participants had 3 s to push a button on each row and the amount of time left was indicated by the so-called progress bars (shown on the left side of each row in Fig. 2). By comparing the correct answer with the responses on the first three rows participants were able to calibrate their perceptions of the reliability of: (1) their own initial predictions; (2) the decision aid's advice; and (3) their final predictions. This calibration is expected to help the participant determine whether to rely on their own initial prediction or on the advice from the decision aid.

In the 'decision aid first' condition the order of activities in each trial is: (1) receive advice from decision aid; (2) predict; and (3) receive feedback with correct answer, corresponding to each successive row in Fig. 3. In each trial participants first received advice before they could express their own prediction. On the second row the participants expressed their own prediction. They could follow the advice from the decision aid or make their own prediction. On the third row, the correct answer (highlighted button) and feedback about the success of the decision (red or green color) was provided.

5.5. Dependent variables

5.5.1. Agreement with the decision aid

Percentage agreement with the decision aid is measured during experimental trials and is defined by the correspondence of the final prediction of the participant with the advice from the decision aid (Bonaccio and Dalal, 2006). Percentage agreement allows us to compare the degree to which participants rely on themselves rather than on the decision aid in both conditions (i.e., a self-reliance heuristic or not).

Agreement measures are insensitive to changes in pre-advice and post-advice decisions. They cannot distinguish between whether one agrees with a decision aid because one is holding to one's pre-advice decision or whether one adopts advice from a decision aid that conflicts with one's pre-advice decision. However, only the latter is a true decision to rely on the decision aid. To determine what factors affect decisions to rely on oneself or the decision aid we measured reliance when human and decision aid disagreed. Note that this could only be done in the 'human first' condition.

Since the decision aid advice was predetermined, it was not possible to 'dynamically' ensure the amount of disagreement was equal and substantial enough in each condition. But due to randomization of the decision aid advice (30%) and the large number of trials in each condition, it was expected to be similar and substantial enough for reliable data analyses.

5.5.2. Perceived reliability

After each experimental block participants estimated the reliability of both their own performance and that of the decision aid on a scale between 0% and 100% correct with steps of 10%. Relative trust in oneself is calculated by subtracting perceived reliability of the decision aid from perceived reliability of oneself. A positive value indicates that trust in oneself is higher than trust in the decision aid and a negative value that trust in oneself is lower.

[2]Participants could, of course, also decide to predict something other than their own initial prediction or the advice from the decision aid, but this almost never occurred.

To determine whether the effect of 'order of advice' on 'agreement with the decision aid' depends on 'relative trust' we created two additional groups after we collected the data: In the 'relative trust higher' group, participants trusted themselves (T_s) more than the decision aid (T_a). Hence for this group it held that $T_s - T_a > 0$. In the 'relative trust lower' group, participants trusted the decision aid more than themselves. Hence for this group it held that $T_s - T_a < 0$.

5.5.3. Actual reliability

Actual reliability of both the participant and the decision aid was measured during task execution. Reliability is defined as the percentage correct predictions (in rows 1–3 in Fig. 2 and rows 1 and 2 in Fig. 3).

5.5.4. Understandability

Participants indicated on a Likert-scale from −3 to 3 (in steps of one) whether they thought their own decision making process and that of the decision aid was understandable, where −3 indicated that it was completely not understandable and 3 meant that it was completely understandable. Relative understandability of oneself is calculated by subtracting understandability of the decision aid from understandability of oneself. A positive value indicates that understandability of oneself is higher than that of the decision aid and a negative value that understandability of oneself is lower.

5.5.5. Responsibility

Participants indicated on a Likert-scale from −3 to 3 (in steps of one) whether they felt responsible for the outcome of the task, where −3 meant they did not feel responsible at all and 3 meant they felt completely responsible.

5.5.6. Attribution of unreliability

Participants indicated on a Likert-scale from −3 to 3 (in steps of one) whether unreliability in one's own performance and performance of the decision aid is attributed to 'temporary factors', 'external factors' and 'uncontrollable factors', respectively (i.e., three scales), where −3 meant that they thought that performance could absolutely not be attributed to those factors and 3 meant that they thought those factors absolutely did play a role.

6. Results

There were three missing values in the 'human first' condition and one in the 'decision aid first' condition due to technical reasons during data retrieval. This means that the total of 80 in the 'human first' and the total of 78 in the 'decision aid first' condition mentioned in Table 1 both became 77, and the total in general became 154. The statistical analyses based on questionnaires could remain the same (i.e., $N=80$).

6.1. Order of advice, cognitive availability and reliance heuristics (Hypothesis 1)

Hypothesis 1 claimed that when people trust themselves more than the decision aid (i.e., relative trust higher), they disagree more with the decision aid than when people do not trust themselves more than the decision aid (i.e., relative trust lower), but only when they receive advice after they have formed their own opinion. The results related to this hypothesis are discussed below.

The percentage agreement with the decision aid in the 'decision aid first' condition ($M=72.2\%$) did not differ from that in the 'human first' condition ($M=72.5\%$), $t(153)=-0.19$, $p=0.85$. In both conditions, on average, trust in the decision aid was higher than trust in oneself. But participants in the 'human first' condition perceived the decision aid to be 30% better than themselves compared to only 6% in the 'decision aid first' condition, $t(153)=3.97$, $p<0.01$. Concerning relative trust, one would expect that when participants perceive the decision aid to be a lot better than themselves, they would agree more with the aid. This proved not the case. Despite the fact that the decision aid was perceived to be 30% better in the 'human first' condition compared to only 6% better in the 'decision aid first' condition, no differences in agreement between the 'human first' and 'decision aid first' conditions showed up. A further analysis showed that the perceived reliability of the decision aid's performance was only 7% higher in the 'decision aid first' condition ($M=67.2\%$, $SD=13.7$) than in the 'human first' condition ($M=63.1\%$, $SD=14.0$), $t(153)=1.93$, $p=0.055$. The perceived reliability of own performance was 30% lower in the 'human first' condition ($M=48.5\%$, $SD=17.4$) than in the 'decision aid first' condition ($M=63.1\%$, $SD=14.4$), $t(153)=-5.72$, $p<0.01$. It seems that feedback about one's own performance errors is much more salient in the 'human first' condition than in the 'decision aid first' condition.

As mentioned in the method section, to determine whether the effect of 'order of advice' on 'agreement with the decision aid' depends on 'relative trust' we created two additional groups: a 'relative trust higher' group ($N=66$) and a 'relative trust lower' group ($N=88$).

In agreement with our hypothesis we found that participants disagree more with the decision aid when trust in themselves was higher than trust in the decision aid (i.e., relative trust higher). This difference was significant in the 'human first' condition (left side of Fig. 4), $t(76)=2.15$, $p=0.03$, but not in the decision aid first condition (right side of Fig. 4), $t(76)=0.97$, $p=0.33$.

These results seem to indicate that in the 'human first' condition people use relative trust in making reliance decisions and that people do not use this information in the aid first condition. It seems that people are cognitively engaged in making reliance decisions when advice is presented after people have formed their own opinion and that people heuristically rely on the decision aid, when advice is provided beforehand.

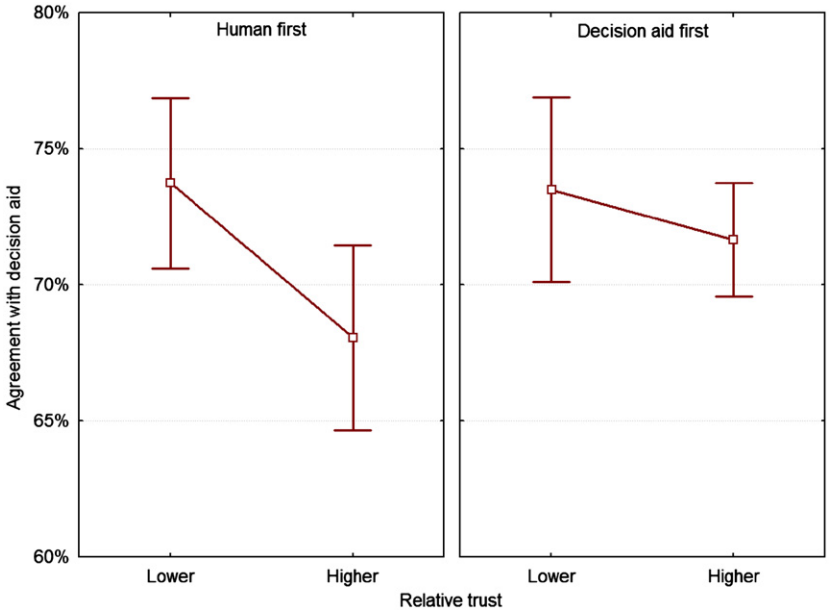

Fig. 4. Agreement with the decision aid is higher when trust in oneself is lower than trust in the decision aid (lower relative trust), but only in the 'human first' condition.

6.2. Understandability of underlying reasoning (Hypotheses 2 and 3)

Hypothesis 2 claimed that when people rely as much on themselves as on the decision aid, they still perceive the decision aid to be more reliable than themselves. The results related to this hypothesis are discussed below.

The data from the 'human first' condition were used to determine whether the factors relative understandability of underlying reasoning and feeling of responsibility in addition to relative trust in performance explain reliance on oneself or the decision aid when disagreement occurred.

On average, participants estimated their own reliability to be 48.5% and that of the decision aid 63%. In other words, they thought the decision aid was 14.5% more reliable than themselves (and 30% more reliable, relatively speaking), $t(79) = -5.79$, $p < 0.01$. When the initial prediction of the participant differed from the advice from the decision aid, participants relied for 52% on the decision aid and for 48% on themselves. This difference is not significant, $t(79) = -0.78$, $p = 0.44$. When participants based their decisions to rely on themselves or on the decision aid on relative trust alone, one would expect them to rely at least 30% more often on the decision aid than on themselves. The results seem to be in agreement with Hypothesis 2: Participants did not rely more often on the decision aid when in disagreement, although they perceived the decision aid to be 30% more reliable.

Hypothesis 3 claimed that the less people perceive the decision aid's reasoning to be cognitively available and understandable, the less people rely on the decision aid. The results related to this hypothesis are discussed below.

As expected, on average, participants found their own decision making process to be understandable ($M = 0.64$, $SD = 1.4$), in contrast to that of the decision aid ($M = -0.93$, $SD = 1.32$), $t(79) = 7.07$, $p < 0.01$. Despite a correlation between relative trust and relative understandability ($r = 0.27$, $p < 0.05$), which is caused by a correlation between perceived reliability of oneself and understandability of oneself ($r = 0.37$, $p < 0.05$), results of the regression analysis indicated that relative understandability also contributed to predicting reliance on oneself ($\beta = 0.28$). The more understandable participants thought their own decision making process was compared to that of the decision aid, the more they relied on their own initial prediction. These results are in agreement with our Hypothesis 3: people rely less on conflicting advice when they perceive the decision aid's reasoning to be cognitively less available and understandable than their own reasoning.

6.3. Feeling of responsibility (Hypothesis 4)

Hypothesis 4 claimed that the more people feel responsible for the task outcome, the more they rely on the decision aid. The results related to this hypothesis are discussed below.

On average participants felt responsible for the accuracy of the final decision ($M = 1.25$, $SD = 1.11$). Differences in responsibility between individuals ranged between negative (-2) and absolutely positive (3).

Results of the regression analysis indicate that responsibility also contributed to predicting reliance on oneself ($\beta = -0.29$). The more responsible the participants felt for the task outcome the more they relied on the conflicting advice from the decision aid rather than their own initial prediction. These results are in agreement with Hypothesis 4: decision makers are more likely to accept (more reliable

but) conflicting advice from the decision aid when they feel more responsible for the outcome of the decision.

Together relative trust (Section 6.1), relative understandability (Section 6.2) and responsibility (Section 6.3) explain 38% of the variance in reliance. Based on the magnitudes of the beta-coefficients, the squared partial and semi-partial correlations (see Table 2), the relative contribution of these factors seems to differ little.

6.4. Accuracy of perceived reliability (Hypothesis 5)

Hypothesis 5 claimed that when feedback about performance is provided, people underestimate both their own reliability and that of the decision aid. The results related to this hypothesis are discussed in the two sections below. The first section deals with the perceived reliability of oneself and the second with that of the decision aid.

6.4.1. Perceived reliability of oneself

Results showed that some participants underestimated their own reliability, while others overestimated their own reliability (Fig. 5), but averaged over two blocks, the perceived own reliability was 4% lower than their actual reliability, $t(79) = -2.53$, $p < 0.01$.

For the first block participants underestimated their own reliability with 5%, $t(39) = -2.16$, $p < 0.05$. But this underestimation was not statistically significant in the second block, $t(39) = -1.46$, $p > 0.05$ (Fig. 6). Correlations between perceived own reliability and their actual reliability increased from $r = 0.42$, $p < 0.05$ in the first to $r = 0.51$, $p < 0.05$ in the second block. We also found that estimations of own reliability improved over time and that this underestimation seems to disappear over time.

6.4.2. Perceived reliability of the decision aid

Most participants underestimated the reliability of the decision aid, but both pessimists who over-weighed errors as well as optimists who under-weighed errors were found (Fig. 7). Averaged over two blocks the perceived reliability of the decision aid was 7% lower than the actual reliability, $t(79) = -4.41$, $p < 0.01$.

For the first block participants underestimated the performance of the decision aid for 7%, $t(39) = -2.47$, $p < 0.05$ and for 8% in the second block, $t(39) = -3.79$, $p < 0.01$ (Fig. 6). In the first block the standard deviation

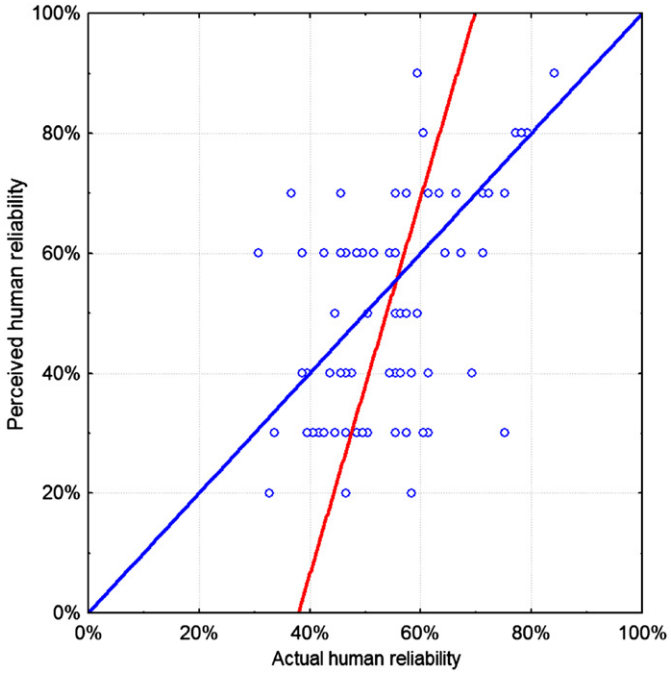

Fig. 5. Calibration human reliability.

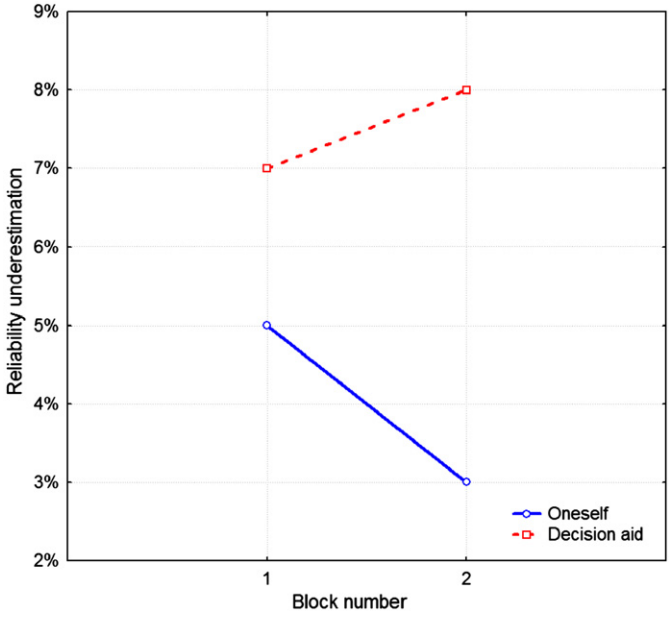

Fig. 6. Effects of learning on estimation of reliability of oneself and decision aid.

Table 2
Regressing reliance on relative trust, relative understandability and responsibility.

	Regression coefficients		Squared partial correlations	Squared semi-partial correlations
	Beta	Std error		
Relative trust	0.32*	0.098	0.12	0.09
Relative understandability	0.28*	0.097	0.10	0.07
Responsibility	−0.29*	0.094	0.11	0.08

*$p < 0.01$.

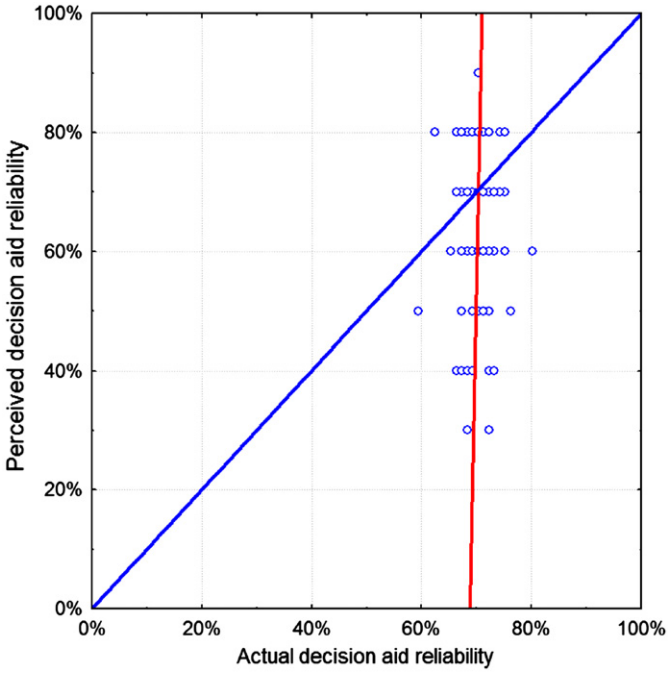

Fig. 7. Calibration decision aid reliability.

of perceived reliability was slightly higher ($SD=14.30$) than in the second block ($SD=13.53$).

These results are in agreement with Hypothesis 5: on average reliability of own performance and that of the decision aid is underestimated when people are provided with feedback about performance.

6.5. Attribution bias (Hypotheses 6 and 7)

The results from Section 6.4 are also in agreement with Hypothesis 6, which claimed that underestimation of the reliability of the decision aid is more prevalent and more persistent than underestimation of one's own reliability. In sum, we indeed found underestimation of own reliability and that of the decision aid, but underestimation was higher for the decision aid. Also underestimation of own reliability decreased after practice; that of the decision aid did not.

Hypothesis 7 claimed that unreliability of the decision aid is less attributed to temporary, external and uncontrollable causes than one's own unreliability. The results related to this hypothesis are discussed below.

On average, unreliability of the decision aid is less attributed to temporary factors ($M=0.05$) than own unreliability ($M=0.41$), $t(79)=2.02$, $p<0.05$. Unreliability of the decision aid is also less attributed to uncontrollable factors ($M=-0.85$), than own unreliability ($M=-0.26$), $t(79)=2.92$, $p<0.05$. However, both own unreliability ($M=-0.79$) and that of the decision aid ($M=-1.09$) was not attributed to external factors and no difference was found between them, $t(79)=1.66$, $p>0.05$. The above results are partly in agreement with Hypothesis 7: unreliability of the decision aid is less attributed to temporary and uncontrollable causes, but, like own unreliability, is not less attributed to external causes.

7. Conclusion

In this study we present a framework for understanding task and psychological factors affecting reliance on advice from decision aids. The theoretical framework describes the relations between (1) factors in the task environment; (2) factors affecting motivational, heuristic and systematic processes; and (3) behavior and a feedback loop indicating information about effectiveness of that behavior. We believe that both heuristic and systematic processes play a role in reliance decision making and that the degree to which these processes play a role depends on the amount of effort invested.

This study showed that people are sensitive to relative trust in a condition in which their own opinion is cognitively more available than that of the aid, that is, when advice is provided after people have formed their own opinion (Hypothesis 1).

Furthermore, we found that although participants think (i.e., via System 2) that the decision aid is more reliable than they are themselves (30% better), they do not rely more on it (Hypothesis 2). This inconsistency and deviation from rationality (a self-reliance bias) is best explained by factors affecting automatic and effortless processing (i.e., via System 1) like the 'availability heuristic' and 'anchoring'.

Results also suggest that decision makers rely less on conflicting advice because they perceive the decision aid's reasoning to be cognitively less available and understandable than their own reasoning (Hypothesis 3).

The invested effort provides a useful indication of whether a given mental process should be assigned to System 1 or System 2. We found that people who felt more responsible for the task outcome relied more on conflicting advice than people who felt less responsible (Hypothesis 4). It seems that when people feel more responsible they are more willing to invest the mental effort required to abandon their initial decision (anchor) and accept conflicting advice from the more reliable decision aid. Although it was not the subject of this study, we think that the felt responsibility and effort invested can be affected by holding the human decision maker accountable for the task performance of a human–computer team.

In our experiment, together with relative trust, relative understandability and felt responsibility for joint task performance explain 38% of the variance in reliance behavior.

When people wrongly think they perform better than the decision aid or vice versa, reliance decisions based on relative trust can result in undesirable outcomes. We found that, when provided with feedback, the perceived reliability of both oneself and decision aid is underestimated (Hypothesis 5) and it seems that negative experiences have a greater influence than positive experiences.

Since relative trust is based on the difference between perceived reliability of oneself and decision aid, as long as the degree of underestimation does not differ between oneself and decision aid, no problems with relative trust and the decision to rely on advice are expected. However, when the magnitude or direction of underestimation differs, this may result in inappropriate reliance decisions. Our results suggest however that underestimation of the reliability of the decision aid is more prevalent and more persistent than underestimation of the reliability of oneself (Hypothesis 6).

It seems users are less forgiving for errors made by the decision aid, probably because errors of the decision aid are less attributed to temporary and uncontrollable causes (Hypothesis 7, no evidence was found for attribution to external causes). This asymmetry can be explained partly in terms of information asymmetry and partly in terms of attributed biases. According to the differential information explanation, decision makers have privileged access to their factors affecting their own performance errors, but not to those affecting errors of the decision aid. Because factors affecting the reliability of the aid are not transparent and cognitively available for the decision maker, errors are more likely to be attributed to more dispositional (permanent, internal and controllable factors) than situational factors. Providing feedback seems to introduce its own set of heuristics and biases.

8. Discussion

Due to the fact that we 'merely' studied effects on participants executing a pattern learning task, one might argue that the generalizability of the above conclusions is an issue. The reason for using a pattern learning task in this study is that it can be controlled very well and hypotheses can be tested quite precisely. More realistic settings to which the results of this study are expected to generalize are for example tasks that incorporate decision making based on advice from different agents (man or machine). The reliance decisions studied in this paper can be seen as largely independent of the task at hand and therefore the drawn conclusions are expected to generalize to these more realistic and more ecologically relevant tasks.

Here, it is important to discuss the decision aid design implications of the presented framework and research findings. Appropriate reliance on decision aids is not guaranteed when only focusing on optimizing the reliability of decision aids. There are several things one could do in the design phase of a decision aid. First of all, a general strategy to improve the quality of reliance decision making would be to reduce information asymmetries. We expect reliance biases to be reduced when options for oneself and the decision aid, supporting reasons and sources of error and the accuracy of advice, are equally and easily accessible. Second, people should be given feedback about their own individual performance, that of the decision aid and the performance of the human–computer team. This feedback should be corrected for differences in saliency and the bias that negative information is given more weight. This feedback can improve the calibration of trust in oneself and the decision aid and therefore stimulate appropriate reliance. Third, by providing advice after, rather than before, people have formed their own opinion, they bring more knowledge to the task. The degree to which this is desirable depends on how skilled the decision maker is and whether 'cue patterns in memory' are valid for the task at hand. It can be achieved for instance by allowing the decision aid to only provide advice when requested by the human, or by allowing an adaptive autonomous system to estimate the precise level of support needed. Such a design is not focused on reducing workload by automation, but focused on human–machine collaboration with the goal of increasing decision accuracy. Receiving advice afterward may also increase confidence of the decision maker when human and system agree or make people think twice when they disagree. The designer should aim at reducing the effort to rely on oneself and the decision aid by making human–computer collaboration more flexible. Reliance heuristics, leading to non-desirable outcomes may be induced when reliance on oneself or the decision aid require a lot of effort. One should make the reasoning of the decision aid available and understandable in the human–computer interface. For instance by using simple algorithms and by revealing intermediate results in a comprehensible way. Also, people should be made to feel accountable for the performance of the human–computer team and they should be held responsible for the quality of the outcome. Finally, one should control for the attribution of errors. For instance by making sources of error transparent or by making operators aware of their biases in attribution. The idea is that providing information regarding why the automation might be mistaken increases trust (Dzindolet et al., 2003). Teach users the conditions in which the decision aid performs well and the conditions in which it does not. We believe that when these recommendations are taken into account in the design of decision aids, these decision aids will be used more appropriately, resulting in improved human–computer team performance.

Acknowledgments

This research was partly funded by the Royal Netherlands Navy under Program numbers V206 and V929. The authors are grateful to Lisette de Koning, Jan Maarten Schraagen, Jasper Lindenberg, Tibor Bosse, Anja Langefeld and Jan Willem Streefkerk for their valuable input, comments and suggestions.

References

Alba, J.W., Hutchinson, J.W., 2000. Knowledge calibration: what consumers know and what they think they know. Journal of Consumer Research 27, 123–156.

Bonaccio, S., Dalal, R.S., 2006. Advice taking and decision making: An integrative literature review, and implications for the organizational sciences. Organizational Behavior and Human Decision Processes 101, 127–151.

Bowers, C.A., Oser, R.L., Salas, E., Cannon-Bowers, J.A., 1996. Team performance in automated systems. In: Parasuraman, R., Mouloua,

M. (Eds.), Automation and human performance. Lawrence Erlbaum Associates Inc., Mahwah, NJ, pp. 243–263.

Brenner, L.A., Koehler, D.J., Liberman, V., Tversky, A., 1996. Overconfidence in probability and frequency judgments: a critical examination. Organizational Behavior and Human Decision Processes 65, 212–219.

Dzindolet, M.T., Beck, H.P., Pierce, L.G., Dawe, L.A., 2001. A Framework of Automation Use. Technical Report ARL-TR-2412. Army Research Laboratory. Aberdeen Proving Ground, MD.

Dzindolet, M.T., Peterson, S.A., Pomransky, R.A., Pierce, L.G., Beck, H.P., 2003. The role of trust in automation reliance. International Journal of Human Computer Studies 58, 697–718.

Dzindolet, M.T., Pierce, L.G., Beck, H.P., Dawe, L.A., 2002. The perceived utility of human and automated aids in a visual detection task. Human Factors 44, 79–94.

Gao, J., Lee, J.D., 2006. Extending decision field theory to model operator's reliance on automation in supervisory control situations. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans 36, 943–959.

Gilbert, D.T., Malone, P.S., 1995. The correspondence bias. Psychological Bulletin 117, 21–38.

Guerlain, S., Smith, P.J., Obradovich, J.H., Rudmann, S., Strohm, P., Smith, J.W., Svirbely, J., Sachs, L., 1999. Interactive critiquing as a form of decision support: an empirical evaluation. Human Factors 41, 72–89.

Harvey, N., Fischer, I., 1997. Taking advice: accepting help, improving judgment, and sharing responsibility. Organizational Behavior and Human Decision Processes 70, 117–133.

Kahneman, D., 2011. Thinking, Fast and Slow. Farrar, Straus and Giroux.

Kahneman, D., Klein, G., 2009. Conditions for intuitive expertise. A failure to disagree. American Psychologist 64, 515–526.

Layton, C., Smith, P., McCoy, C., 1994. Design of a cooperative problem-solving system for en-route flight planning: an empirical evaluation. Human Factors 36, 94–119.

Lee, J.D., See, K.A., 2004. Trust in automation: designing for appropriate reliance. Human Factors 46, 50–80.

Lewandowsky, S., Mundy, M., Tan, G., 2000. The dynamics of trust: comparing humans to automation. Journal of Experimental Psychology—Applied 6, 104–123.

May, R.S., 1987. Calibration of subjective probabilities: a cognitive analysis of inference processes in overconfidence. Peter Lang, Frankfurt (in German).

May, R.S., 1988. Overconfidence in overconfidence. In: Chaikan, A., Kindler, J., Kiss, I. (Eds.), Proceedings of the 4th FUR Conference. Kluwer, Dordrecht.

Mosier, K.L., Skitka, L.J., 1996. Human decision makers and automated decision aids: made for each other?. In: Parasuraman, R., Mouloua, M. (Eds.), Automation and human performance: theory and applications. Lawrence Erlbaum Associates Inc., Mahwah, NJ, pp. 201–220.

Parasuraman, R., Riley, V.A., 1997. Humans and automation: Use, misuse, disuse, abuse. Human Factors 39, 230–253.

Silverman, B., 1992. Survey of expert critiquing systems: practical and theoretical frontiers. CACM 35, 106–127.

Skitka, L.J., Mosier, K.L., Burdick, M., 1999. Does automation bias decision-making? International Journal of Human–Computer Studies 51, 991–1006.

Sniezek, J.A., Buckley, T., 1995. Cueing and cognitive conflict in judge–advisor decision making. Organizational Behavior and Human Decision Processes 62, 159–174.

Thomas, L.C., Rantanen, E.M., 2006. Human factor issues in implementation of advanced aviation technologies: a case of false alerts and cockpit displays of traffic information. Theoretical Issues in Ergonomics Science 7, 501–523.

Tversky, A., Kahneman, D., 1973. Availability: a heuristic for judging frequency and probability. Cognitive Psychology 5, 207–232.

Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. Science 185, 1124–1131.

Tversky, A., Koehler, D.J., 1994. Support theory: a nonextensional representation of subjective probability. Psychological Review 101, 547–567.

Weiner, B., 1986. An attributional theory of motivation and emotion. Springer-Verlag, New York.

Wickens, C.D., Dixon, S.R., 2007. The benefits of imperfect diagnostic automation: a synthesis of the literature. Theoretical Issues in Ergonomics Science 8, 201–212.

Wiegmann, D.A., Rich, A., Zhang, H., 2001. Automated diagnostic aids: the effects of aid reliability on user's trust and reliance. Theoretical Issues in Ergonomics Science 2, 352–367.

Yaniv, I., Kleinberger, E., 2000. Advice taking in decision making: egocentric discounting and reputation formation. Organizational Behavior and Human Decision Processes 83, 260–281.