An Empirical Investigation Into Three Underlying Factors Affecting Automation Acceptance

Carl Westin and Clark Borst
Faculty of Aerospace Engineering
Delft University of Technology
Delft, The Netherlands
C.A.L.Westin@tudelft.nl, C.Borst@tudelft.nl

Brian Hilburn
Center for Human Performance Research
Voorburg, The Netherlands
Brian@chpr.nl

***Abstract*— The MUFASA project showed benefits to controller acceptance and performance of a conflict detection and resolution decision-aiding system, when suggested resolution advisories were conformal with the controller's own way of solving the conflict. Building on these results, this study investigated why controllers sometimes rejected their own previous solutions, when they (mistakenly) believed these came from automation. Three factors were independently investigated together with strategic conformance: problem-solving consistency, source bias, and interface representation. Fourteen controllers participated in a series of realtime simulations. While the impact of conformance and representation effects were small in simulations, questionnaire responses indicated that controllers perceived a human source favorably over automation, and thought the information richer triangle representation facilitated a better understanding of why the automation suggested a certain conflict solution. The degree of consistency varied among participants, and four different patterns of problem-solving consistency were observed.**

Keywords*- *air traffic control, automation, human-machine interaction, decision-making, strategic conformance, consistency

I. INTRODUCTION

Research has shown that there are benefits to both performance and acceptance of automation conformal to human-human interaction. Drawn from this research [1] argued for strategic conformance as a key concept facilitating human-automation teamwork and the acceptance and use of advanced decision aids. They define strategic conformance as the degree to which automation's problem-solving style matches that of the individual human, with problem-solving style referring to both the solution (product) and associated underlying reasoning (process). In experiments exploring the effects of strategic conformance, acceptance, agreement, and response time to advisories improved when the advisories were strategic conformal with the air traffic controller's. However, in 25% of cases, controllers rejected a conformal advisory. Since conformal advisories were based on the individual's own performance, controllers were actually rejecting their own solutions [2].

This paper follows up this study and attempts to answer why controllers sometimes rejected their own solution, when they (mistakenly) believed it came from automation. In addition, the study attempts to replicate the effects of strategic conformance. The three factors investigated, and the associated research questions were:

- Consistency – how consistent are controllers in their conflict solving over time?
- Source bias – to what extent are controllers biased against automation, or against any external source of advice?
- Representation – how does information transparency of the interface representation impact acceptance?

The paper is organized as follows: Section 2 provides a brief state-of-the-art literature review of each research areas. It follows with a methods section detailing the experimental design, simulator and stimuli, and how two experiments covered three research areas (Section 3). In Section 4 the results are discussed separately for each research area. Section 5 discusses the relevance and broader implications of the results. Finally, conclusions (Section 6) to this study and future research prospects are addressed (Section 7).

II. THREE ACCEPTANCE DRIVERS

A. *Consistency*

While controllers generally are considered homogeneous in conflict detection and resolution (CD&R) [3] and judging complexity [4], it is also widely acknowledged that performance and work practices differ significantly between countries and control centers [6]. Furthermore, studies that have established consensus between controllers need to be viewed with caution given the adversaries of groupthink and other cognitive biases that play part when using methods involving group responses.

Researchers have pointed out individual differences in conflict judgments and resolution strategies and it is acknowledged that controllers develop their own individual work styles [2]. Mogford and colleagues [5] suggested that respective work style could be traced back to controller training and the style of the instructor. The limited research conducted on intra-controller consistency provides an incoherent picture. There is contradictory evidence that controllers are consistent [7] and inconsistent [8] in traffic complexity judgments. Controller performance inconsistencies have also been shown to increase with complexity [9]. Some variability, however, can be expected given fluctuations in, for example, motivation, tiredness, and learning. In addition, it is important to acknowledge the advantage of decision-making flexibility and adaptability, which allows humans to successfully handle and solve new situations and problems.

B. *Source Bias*

The rejection of conformal advisories in [2] led us to question whether this had represented rejections of presumed automation output, or rejections of advisories per se. To explore this notion, we defined the term “source bias” to refer to a potential difference in operator acceptance of advisories, based on the presumed source of those advisories.

What little empirical evidence there is in this area suggests differences in how human and automated advisors are judged, and that this difference is intertwined with the concepts of expertise and pedigree. Arkes and colleagues [10] found that patients rated physicians lower when those physicians seemed to use an automated decision support system. Whereas humans are judged based on their dispositional (i.e., immutable, long standing) traits, automation is judged more by its performance [11]. Moreover, it seems that initial (a priori) trust tends to be higher for automation than for human advisors [12], though such trust is vulnerable to misperformance.

Drawing on research exploring the interaction between source (human vs machine) and pedigree (novice vs expert), Madhavan & Wiegmann [13] concluded that operators formulate decision criteria depending on the perceived pedigree and source of a system. The decision criterion is described as a dependence strategy based on the perceived accuracy (i.e. probability of generating a hit and correct rejection) of the advisor. When the operator starts using a system, the decision criterion equates to how well they think the system will perform. How well the system matches those expectations impacts factors such as trust, reliability, and acceptance.

In summary, it seems that humans are more likely to show a positive bias toward automation and higher (expert) pedigree. In the absence of pedigree information, a priori perceived reliability is higher for automation than for human. For novice and expert pedigree, a priori perceived reliability is higher for automation and human, respectively. Notice how important framing [14] of the pedigree is in setting a priori trust. Finally, operators appear to be more critical, and less forgiving of errors in automation. Though it is hard to draw a clear hypothesis in this case, it is reasonable to speculate that controllers will become more critical of non-conformal automation over time.

C. *Representation*

Representation refers to properties of the interface that impact the *transparency/opacity* of that interface, including the choice of display parameters and richness of information. The information provided by the interface can facilitate understanding of the problem-solving rationale underlying the advice given. Ironically, advanced automation is more often found to feature high levels of opacity, as system designers (sometimes intentionally) hide the system’s complexity from the operator. Hilburn and colleagues [2] maintained a fixed level of interface representation, in which the amount, structure and organization of information provided was kept constant. As such, it is possible that controllers rejected advice because they could not adequately understand what the automation was suggesting (high opacity), or alternatively that the representation facilitated alternative “better” solutions.

Generally, transparency addresses an interface’s ability to afford understanding of a system’s complex reality. The level of transparency afforded by automation is believed to play an important role in constructive communication and team building between humans and machines [15]. Several transparency studies in the context of e-commerce and the semantic web have shown that increased transparency, in terms of offering explanations underlying the behavior of the system, positively influences trust and acceptance [16].

Poorly designed explanations can obstruct understanding and counteract acceptance of recommendations [17]. While well-designed explanations can foster acceptance and trust, as an unwanted side effect, it can conceal automation errors [18]. There is a balance between the decision-making quality of an advisory, and the appeal of accepting it. Furthermore, transparency should be attuned to perception and information processing abilities to avoid overloading the user, as higher transparency can increase complexity and amount of information to be considered.

III. METHODS

The experimental design consisted of two independent experiments that overlapped the three research areas. Each experiment consisted of two human-in-the-loop simulations based on the experimental design in [2]. In the baseline ‘prequel’ simulation, participants’ manual CD&R performance was captured in a series of en route traffic scenario vignettes. Data collected in the prequel simulations (one for each group of participants) was used to study problem-solving consistency and consensus (of CD&R task).

In addition, prequel simulation data was analyzed and used to configure a personalized decision-aid. In the subsequent second ‘conformance’ simulation, participants encountered the same scenarios, assisted by an automated decision-aid that would suggest conflict resolutions. Controllers were free to either accept a given advisory, or to reject it and implement an alternative solution. Two separate conformance simulations were conducted, with different participants, to study the interaction effects of automated conformal/non-conformal resolution advisories with source (experienced controllers) and interface representation (controller trainees) respectively.

A. *Participants*

Fourteen Swedish air traffic controllers from three different control centers voluntarily participated. While the consistency study could utilize all participants, groups were assigned for the source and interface representation studies. A larger group of nine trainees (three females and six males) trained in terminal maneuvering area and approach (one year basic training completed and just starting on-the-job training) participated in the representation study. Age varied between 24 and 29 (*mean*=26 years). For the source study, a group consisting of five experienced controllers (four males and one female) from two different control centers participated. Age varied between 26 and 47 (*mean*=32,8 years) and experience between 15 months and 24 years (*mean*=8.7 years). In the source study, data was partially lost for one of the five participants due to technical reasons. As a consequence, all data reported here, except accept/reject count which redundantly was recoded manually, is based on four participants.

B. Apparatus

The Java-based ATC simulation ran on a laptop connected to an external monitor with a resolution of 1600x1200 pixels. The interface was based on a modified prototype of the Solution Space Diagram (SSD) currently under development at the Delft University of Technology. Building on an Ecological Interface Design (EID) approach, the SSD is a tactical decision support tool that displays the constraints of maneuverability of a selected aircraft based on the relative position of other aircraft [19]. The interface represents the outer ring surrounding the aircraft in the upper right corner in Fig. 1. Vectors were implemented by mouse clicking an aircraft of interest, dragging the velocity trend vector to a desired area, and executing the command by pressing the ENTER key. Speed was controlled by scrolling the mouse scroll wheel up for an increase, and down for a decrease. It was also possible to combine vector and speed commands.

C. Traffic scenarios

The four traffic scenarios consisted of one measurement scenario and three dummy scenarios. The latter were used to prevent recognition of the repeated measurement scenario. All scenarios consisted of a hypothetical en route sector in a squared format, 80 x 80 NM in size. The simulator ran at 2x real speed and aircraft plots on the display were updated every second to simulate a 1 Hz radar update frequency. Certain simplifying assumptions were made. For example, all traffic was restricted to flight level 270, level changes were not possible, and there were no environmental fluctuations.

The measurement scenario contained a carefully designed conflict consisting of two aircraft with perpendicular tracks (Fig. 1). A right angle conflict was selected to mitigate biased solutions and any "obvious" conflict solutions. Generally right angle conflicts are considered relative easy to detect, but more difficult to solve than other conflict geometries [20]. Context aircraft were used to increase complexity, prevent early conflict detection, and make scenarios more realistic. They were placed and configured so that their presence would not interfere with the designed conflict or restrict conflict solving.

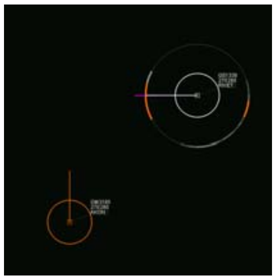


Figure 1. Designed conflict in measurement scenario.

D. Experimental Design

Table 1 presents an overview of the experimental design applied in the three different research areas. The consistency study investigated the variability in participants' conflict solving performance and identifying similarities in the four recorded solutions of the designed conflict. In both the representation and source study we used a 2x2 repeated measures design varying advisory conformance with interface representation and advisory source respectively. Since conformance and source concerned only the designed conflict, dependent measures were restricted to measures directly related to the resolution advisory. Since the representation manipulation applied to all traffic interaction, more scenario generic measures could be collected and analyzed (albeit only to investigate a potential main effect of representation). Presentation order was balanced between participants and traffic scenarios using a Latin Square design.

Strategic conformance of resolution advisories was varied, with advisories being either conformal with controllers' individual conflict solution style, as based on their own previous performance ("conformal" solutions), or a colleague's problem-solving style that was different but acceptable ("non-conformal"). Participants were not informed about the strategic conformance manipulation.

Source was manipulated by presenting resolution advisories as either human or automation generated. Prior to each run, participants received specific oral instructions about the underlying source (depending on human or automation source condition). Human: *"All resolution advisories suggested in this session are made by an air traffic controller."* Automation: *"All resolution advisories suggested in this session are generated by automation."* In addition, the source was stated in the dialogue box containing the resolution advisory. Information about respective sources was intentionally minimized and limited in order to avoid descriptive information that could influence participants' attitudes or the attribution of any pedigree.

Interface representation varied between a baseline heading band (HB) representation (as used in [2]) and the triangle representation (TRI, Fig. 2). Participants were first verbally instructed on what type of information each representation afforded. Secondly, they got to familiarize themselves and learn how to use the interface in the training runs preceding the measurement run. While both representations provided a

TABLE I. EXPERIMENTAL DESIGN (ACROSS RESEARCH AREAS).

Detail	**Study**		
	Consistency	***Representation***	***Source***
Design	Single-subject repeated measures	2x2 repeated measures	2x2 repeated measures
Sample	14	9	4
Variables	Designed conflict repeated four times	Interface representation & conformance	Source & conformance
Measures	Qualitative analysis of conflict solutions and the identification of consistent problem-solving patterns across repetitions.	· Acceptance of advisory (yes/no); · Agreement with advisory (1-100 scale); · Response time (to accept/reject), · Scenario difficulty (1-100 scale). · Number of SSD inspections; · Total amount of interactions; and · Amount of resolutions (heading, speed, combined)	· Acceptance of advisory (yes/no); · Agreement with advisory (1-100 scale); · Response time (to accept/reject), and · Scenario difficulty (1-100 scale).

visualization of the solutions spaces available, the TRI representation provided more instant information about the relative position of intruder aircraft. With the selected aircraft's entire speed envelope instantly visible, the resulting 'no-go' zones of intruder aircraft were rendered triangle shaped (illustrated in Fig. 2 (b)). The relative position of an intruder aircraft can be inferred by determining in which direction the triangle shape is convex. In contrast, the solution space depicted in the heading band representation was limited to the current speed of the selected aircraft (in Fig. 2 (a) the solution spaces, as restricted by colored 'no-go' zones, are only shown for the current speed of 270 knots). In order to overview the solution space of the selected aircraft's entire speed envelope, the participant had to scroll through the entire speed envelope (in increments of 10 knots) and mentally interpolate the solution space information. As such, both the heading band and triangle representations afforded the same information,

E. Procedures

The two simulations in each experiment were run over a three week period. In the first week, the manual prequel simulations were conducted to record participant resolutions to the designed conflict. Participation lasted one and a half hours and consisted of a simulation session and questionnaire part. After consent and briefing procedures, participants played 14 training runs and 10 scenarios. The measurement scenario was repeated four times, meaning that four data points were obtained for each individual participant. The other six scenarios consisted of three 'dummy' scenarios repeated twice per participant. Participants were instructed to maintain separation between aircraft, while vectoring them to their assigned exit points. Short-term conflict warnings were provided in two stages, starting with involved aircraft turning amber 60 seconds prior to a separation loss, and elevating to an aural alert together with both aircraft being displayed in red

(a) Heading Band (b) Triangle (full solution space)

Figure 2. Interface representations.

when less than 30 seconds remained. During the middle week, the dataset was analyzed and processed for purposes of creating personalized conformal and non-conformal automated resolution advisories for each participant. To ensure reliability, three researchers accomplished this process in parallel.

In the third week, the same participants played the same simulator, containing the same scenarios and designed conflict. Only this time, they were supported by a decision-aid that provided resolution advisories to the designed conflict by plotting it in the SSD interface. The automatically displayed SSD advisory was accompanied by a beeping sound and a dialog window that enabled participants to either 'accept' or 'reject' the advisory. Participants were at this stage also required to indicate their agreement with the resolution advisory. Participants were instructed that an advisory would always solve the conflict, but not necessarily in the most optimal way. As such, controllers were encouraged to find more suitable alternatives according to their preferences.

Participants were divided into two separate groups that would encounter conditions (i.e. source or representation) in different orders. The simulation session consisted of two runs (one for each condition) and participants received a short briefing prior to each run. This briefing included information and instructions relevant to the condition applied (i.e., human or automation source in the source study, and heading band or triangle representation in the interface representation study). Conformance, however, was alternated between scenarios in each run. Again, participation lasted one and a half hours and consisted of a simulation and questionnaire part. After each scenario, participants were asked to rate the subjective difficulty.

Post-simulation questionnaires were administered in both the source and representation study. The online questionnaires consisted of multiple 7-point Likert scale statements. The source questionnaire was based on a previously developed instrument used to assess operator trust in automated systems [21]. The representation questionnaire was partly adapted from two different interface transparency questionnaires [16] [22]. The second questionnaire consisted of multiple statements to be answered on a Visual Analogue Scale (VAS). VAS is a subjective questionnaire instrument with which participants indicate their agreement with a statement along a continuous line (1-100mm) with two endpoints.

IV. Results

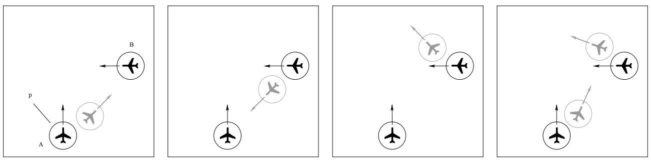
A. Consistency

We observed large variations in solution preferences across participants and in how consistently they solved the designed conflict. Fig. 3 illustrates four different and commonly observed solutions of the designed conflict. Four different groups of consistency patterns were identified (not to be confused with the solutions depicted in Fig. 3). Across the four patterns, a total of ten participants were found to have solved the designed conflict the same way in all four repetitions. All participants were found to consistently have solved the conflict, according to at least one of the patterns, in at least three out of four repetitions.

1) Solution parameters hierarchy analysis: Consistency determined by identifying similarities across specific solution parameters (e.g. choice of aircraft, type of resolution, resolution direction), and combinations thereof. Six participants consistently interacted with only one of the two aircraft, with four choosing aircraft A and two aircraft B. Irrespective of aircraft choice, seven participants preferred to solve the conflict by vectoring one aircraft to the right. Only one participant consistently vectored one of the aircraft to the left. In a combination of the above, only three participants were found to consistently vector one of the aircraft to the right.

2) Number of interactions undertaken: A limitation of the above analysis was that it only considered first interactions. Several participants, however, solved the designed conflict by interacting with both aircraft rather than only one (Fig. 3 (c)). This pattern was observed for four participants. In addition, post-simulation questionnaire responses confirmed the relevance of this solution strategy, with responses including comments "*I would change headings on both aircraft*" and "*Divide the delay across the two aircraft*."

3) Solution geometry: While the solution parameter hierarchy provided a logical method for analyzing conflict solutions, it did not adequately address the resulting relationship between aircraft irrespectively of aircraft choice. That is, it is possible to achieve the same relationship between aircraft A and B, by either vectoring aircraft A to the right behind B, or vectoring B in front of A. In this view, the solution in Fig. 3 (a) and (c) are considered identical. According to this pattern definition, conflict solutions from seven participants consistently generated a relationship where aircraft A went behind B, while only one participant produced a contrasting relationship with aircraft B going behind A. When asked how they would prefer to solve the designed conflict in post-simulation questionnaire responses, however, no participants stated that they would take aircraft A behind B. In contrast, four stated they would prefer to take aircraft B behind A.



(a) (b) (c) (d)

Figure 3. Patterns of consistent conflict solving styles.

4) Control problem analysis: In the fourth consistency pattern we looked at the solution from a control problem perspective. In this view, the aircraft interacted with first is considered the ownship, or *controlled* aircraft, and the other aircraft is an *intruder*. The analysis then investigated the consistency of interactions undertaken to avoid the intruder aircraft. In this simplified conflict situation, the only options were to either vector the controlled aircraft behind or in front of the intruder aircraft. In this view, the solution in Fig. 3 (a) and (b) are considered identical.

Seven participants consistently vectored the controlled aircraft behind the intruder, while four vectored it in front. Three participants were found inconsistent (vectoring the controlled aircraft twice behind and twice in front of the intruder aircraft). Overall, experienced controllers were found more homogenous than trainees. Furthermore, we found a significant positive Kendall Tau correlation between participant's actual consistency, as measured by the control problem analysis, and their self-rated consistency as measured in the post-simulation questionnaire (r_t=.49, p=.05).

B. Source bias

Acceptance rate was very high with only one out of twenty advisories rejected. This observation suggests that neither conformance nor source affected participants' acceptance of resolution advisories. The small sample size of 5 participants, a total of 20 data points, and the ceiling effect (with 95% advisories accepted) made it impractical to carry out any inferential statistics.

Results from one of the statements in the simulator questionnaire revealed reluctance among participants to reject resolution advisories. Two out of six participants agreed with the statement: "*I accepted resolution advisories even though I did not agree with them*" (*median*=2, *IQR*=6). This suggests that resolution advisories were accepted even though participants sometimes disagreed with them. Additional comments made by these participants confirmed this assessment. One participant stated that "*I followed/agreed with all suggestions. Had no reason to distrust them.*"

No source bias trends were apparent in the online questionnaire responses. Generally, responses were equally positive regardless of source condition. Irrespective of source, participants generally agreed with statements that the advisory system was reliable, trustworthy, dependable, and provided security. Participants generally disagreed with statements that the advisory system was harmful, deceptive, or underhanded.

Results from the VAS questionnaire, however, indicated that controllers did perceive the two sources differently. Participants felt that 'human-based' solutions were safer, more efficient, and more similar to the way they themselves would have solved a given conflict (respectively 6.6%, 8.6% and 19% in favor of human source). Overall, participants indicated a preference for working with the human source in the future (7.6%). In contrast, the automation source was perceived more risky and difficulty to work with (respectively 10.8% and 10% in favor of automation source).

C. Representation

Across all conditions, 26 out of 36 resolution advisories were accepted (72.2%). There was no difference between HB and TRI conditions and acceptance only varied with conformance, with conformal advisories (77.8%) accepted more often than non-conformal (66.7%). Cochran's Q test, however, did not yield any significant results (X^2=.600, *p*=.896). Z-scored difficulty rating data showed that scenarios with non-conformal scenario (*mean*=.085, *SD*=.166) were perceived slightly more difficult than conformal scenarios (*mean*=-.085, *SD*=.166). A 2x2 ANOVA did not indicate a significant main effect of conformance ($F(1,8)$=.266, *p*=.620) or representation ($F(1,8)$=1.137, *p*=.317) on the perceived scenario difficulty. Neither was there any interaction between conformance and representation ($F(1,8)$=.148, *p*=.711). Friedman's test revealed no significant differences in response time ($X^2(3)$=1.8, *p*=.615) or agreement rating ($X^2(3)$=.329, *p*=.954). The only notable difference was the faster response time to non-conformal advisories when using the heading band representation (*median*=10.2 seconds) as compared to the triangle representation (*median*=13.3 seconds)

Wilcoxon signed rank test indicated a trend (*T*=8, *p*=.086, *r*=.57) with number of SSD inspections lower with the TRI representation (*median*=33, *IQR*=13.3) than HB representation (*median*=37, *IQR*=11.5). The total number of interactions (combining heading, speed, and combined interactions) was not significantly affected by representation (*T*=20.5, *p*=.271, *r*=.37). Heading interactions were almost equally distributed between the HB and TRI representation (*T*=17.5, *p*=.944, *r*=.02). There was a trend for more use of speed interactions (*T*=19, *p*=.072, *r*=.60) with the TRI representation (*median*=0.5, *IQR*=3) than the HB representation (*median*=0.5, *IQR*=1). Combined interactions did not vary between representations (*T*=22 *p*=.172, *r*=.45).

Responses collected in the online questionnaire suggested that participants preferred the TRI over the HB representation. Especially, the TRI representation was perceived more understandable in terms of which aircraft caused which conflict, and better at facilitating the use of speed and combined solutions. The TRI representation was also perceived as being more cluttered than the HB representation.

The VAS questionnaire showed that participants preferred to work with the TRI representation. Participants indicated that, although it was more cluttered (46.4% in favor of TRI), the TRI representation reduced workload (14.2% in favor of TRI). Participants felt that it was easier to understand the rationale underlying a solution when using the TRI representation (50.2% in favor of TRI). Furthermore, the TRI representation was perceived as more helpful in conflict solving (31.6% in favor of TRI) and providing a better overview of the solution space available (44% in favor of TRI).

V. Discussion

First of all, we were able to replicate perhaps the most notable results in [2]: that acceptance of advisories increased with conformance. Acceptance rate in the trainee sample (representation study) was indeed higher for conformal than non-conformal advisories but the difference was much smaller than in [2]. For experienced controllers, acceptance was nearly complete (again, 95%). It is not clear what was driving this lack of conformance effect among the experienced controllers. Data suggest that the experienced controllers applied an "accept all" strategy. Questionnaire responses and simulation observations indicate that controllers accepted advisories even though they did not fully agree with the advisories.

Similar patterns have, however, been observed in other studies. When investigating participants' (undergraduate students') acceptance behavior with automated diagnostic aids, [23] observed two contrasting automation utilization strategies. One group agreed with the aid in the majority of all trials even when diagnosis was wrong (which it was in 20%). The author suggested that participants did so in order to assess aid reliability accurately without confusing it with their own decision-making reliability.

Possibly, the lack of a stronger conformance effect can be attributed to the definition of conformal and non-conformal resolution advisories. Although the underlying procedure relied on a similar process as that applied in [2], the allocation of conformal and non-conformal resolution advisories was partly less stringent, in that exact replays were not used, but an "averaged" solution as determined by the solution patterns observed in the four repetitions. We say 'partly' since in [2], conformal resolution advisories contained a mix of exact copies and averaged solution patterns as witnessed in repetitions. Furthermore, as the subsequent consistency study revealed, there were three more consistency patterns identified that better could explain the solution patterns for many participants. This was especially true for the experienced controllers who all were found to be consistent according to the control problem analysis. It is possible that resolution advisories based on some of these other consistency patterns would have resulted in different accept/reject patterns among some of the participants.

A. Consistency

The controllers who participated in our study did not all solve the designed conflict in the same way. Our qualitative analysis revealed individual conflict solution styles. This suggests that controllers differ in their preferences for solving conflicts. We were able to show that controllers were consistent in solving conflicts. The diversity in solutions, however, suggests that they could not be considered homogenous as a group. Overall these findings correspond with results obtained in [2].

All experienced controllers were found consistent according to the control problem analysis. The significant correlation between the control problem analysis and self-rated consistency supports the relevance of this consistency pattern. There was, however, not consensus among experienced controllers on whether to vector the controlled aircraft in front or behind the intruder aircraft. Trainees were less consistent and showed a larger spread in patterns.

Allocation of a participant's behavior to one consistency group did not preclude consistent behavior as defined by any of the other groups. In fact, several participants were found consistent in as many as three groups. A few were even found to be consistent in all groups. Depending on how the conflict solution was analyzed, participants could be determined to both match and diverge in their solutions. This finding was contradictory and surprising. It can, however, explain why some previous studies have concluded controllers to be homogenous. At certain levels of abstraction, controllers may be able to agree on how to solve conflicts, but when it comes to specific control actions, controllers disagree.

The definition of consistency is critical if we are to develop automation that acknowledges and is sensitive to controllers' problem-solving styles. Conflict solving is one of those tasks in which the underlying processes are intrinsically difficult to understand. Controllers find it difficult to explain why a certain solution was chosen - often it is "obvious" or "makes sense." As researchers we draw conclusions based on data we manage to extract from controllers. This makes it difficult to determine consistency since we cannot know, with full certainty, why solution X was chosen in situation Y. As a consequence, the relevance of the consistency patterns observed in this study can be questioned. We did, however, initially not want to limit the analysis by judging what may or may not be reasonable to a controller. Possibly, a person's behavior can better be explained in a way not readily apparent to that person (as argued in general naturalistic decision-making theories [24] and in ATC decision-making [5]).

Another discussion topic is that of personalized automation and exploration of heterogeneity in users problem-solving styles. Automation sensitive to the users preferences and abilities can benefit acceptance, enjoyment, and potentially teamwork with automated agent and performance, if not only from fact that automation is used. There are, however, aspects that may argue against personalizing automation, which instead advocates homogeneity. Personalized automation may work against proceduralized environments like ATC, where handovers from one controller to another requires understanding of what the other person is doing and why. There are benefits and drawbacks of personalized automation that require carefully consideration.

B. *Source*

Results could not establish the prevalence of any source bias as a result of varying the source of advisories. Because of the small sample size and observed ceiling effect, we were unable to draw any meaningful conclusions on the effects of source and conformance rate. We were surprised to learn that questionnaire responses and comments made by participants indicate that advisories were accepted even though participants sometimes did not agree with them. Unfortunately, the generally high ratings of agreement across conditions did not support this finding, despite agreement rating data displaying a larger spread compared to acceptance data.

The most relevant data were found in the VAS responses. Although effects were small, responses indicated that, even though solutions were identical between the source conditions, participants reported that the human source provided safer resolution advisories, and solutions more similar to how they would have solved the conflict. In contrast, automation was perceived as more risky and difficult to work with.

C. *Representation*

Simulation data did not reveal any main effects of interface representation, or interactions with advisory conformance, suggesting that interface representation cannot explain fluctuations in acceptance of conformal (or non-conformal) resolution advisories. Results suggest, however, that the two representations were perceived differently and influenced conflict solving. This was partly reflected in simulator performance data, with a trend evident for fewer interface inspections and increased use of speed interactions in the TRI condition. The influence of the TRI representation on the use of speed interactions was further supported by participants' questionnaire responses. It should be noted that participants seldom used speed clearances, and a closer look into participant data reveals that three participants who tended to use speed interactions more frequently primarily drove the observed difference.

Although questionnaire responses indicate that participants perceived the TRI representation as more transparent, in terms of facilitating understanding what the system suggested, these effects of transparency did not influence acceptance or agreement with resolution advisories measured in the simulations. It is possible that the transparency manipulation of the interface representation was too subtle. With a high level of transparency already built into the SSD interface (i.e., attributed to EID), the presentation of constraints in terms of limited heading bands or full triangles may have had a relatively small impact on participants' understanding of resolution advisories in an operational setting, albeit strong enough to make a difference in questionnaire responses.

Tintarev and Masthoff [25] argued that transparency has to be subjectively and contextually tailored. In our study, however, transparency was generic based in terms of how detailed the maneuverability constraints were depicted in the interface. As such, no consideration was explicitly given to the rationale driving a specific conflict solution, or varying the transparency of the decision-aids advisories. If so, the TRI representation only served to provide a better overview of the controlled aircraft's traffic situation and relationship to other aircraft (depicted as constrains in the interface). It did not necessarily explain why solution X was suggested in place of solution Y.

VI. CONCLUSIONS

Three research areas were identified and studied to determine whether they could explain why controllers, in the original MUFASA project, rejected personalized conformal resolution advisories 25% of the time. Although the impact of representation effects was small in simulations (the small sample size in the source study did not allow for inferential statistics), questionnaire responses revealed important differences in controllers' perception of advisory source and interface representation. The human source was perceived favorably over automation and controllers reported that the triangle representation better facilitated understanding of the underlying conflict solution rationale in automated resolution advisories. This is, however, not enough to conclude that either source or representation influenced the rate of rejected conformal advisories in [2].

Controllers were found to consistently solve conflicts, although the degree of consistency varied individually. The degree of variability supported the assessment that individual variability could have caused controllers to reject 25% of conformal advisories in [2]. Alternatively, it is possible that the conformance manipulation in [2] was based on an incorrect definition of consistency. In this study we identified four different groups of consistency patterns. Since a controller's conflict-solving style (underlying the development of conformal and non-conformal advisories) depends on the ability to determine a consistent pattern, it is important that a pattern relevant to that controller is used to define his/her conflict-solving consistency.

Over the course of the project, the team has refined its view of the strategic conformance construct, and the role that it can play in fostering operator acceptance of advanced automation systems, particularly at the initial stages of implementation. In situations, like ATM, that can lack "gold standard" criteria for optimal solution (shortest path is not always the "best" solution), the human can be the best judge. There is a critical potential paradox: given that automation is becoming increasingly capable of assuming strategic decision-making control, offered as advisories, automation is becoming more of an advisor and colleague. However, as with a human colleague, advice can be ignored or misused. The potential paradox lies in the implementation and familiarization phase, when trust must develop. A controller might not develop trust until he/she has adequate experience using the machine; but he/she might not use the machine before it is trusted. Although the concept of strategic conformance says nothing about the quality of decisions (indeed, some rightly argue that it will sometimes just reproduce human errors), it suggests that benefits can accrue in terms of acceptance and trust.

VII. GUIDELINES FOR INTRODUCING AUTOMATION

Hypothetically, the relationship between acceptance and strategic conformance can be used to harmonize operator and automation decision-making strategies. At the initial rollout of advanced decision-aiding automation, strategic conformance remains high for some period of time. As trust, usage and acceptance develop, automation can adaptively reduce strategic conformance (as appropriate), so as to increase the gap between the operator's baseline and current strategies. In this way, automation can begin to function not merely as an advisor but as a higher level trainer, ideally optimizing solutions (which might differ from the human's previous ones) while maintaining acceptance.

ACKNOWLEDGMENT

This work is co-financed by EUROCONTROL acting on behalf of the SESAR Joint Undertaking (the SJU) and the EUROPEAN UNION as part of Work Package E in the SESAR Programme. Opinions expressed in this work reflect the authors' views only and EUROCONTROL and/or the SJU shall not be considered liable for them or for any use that may be made of the information contained herein.

The authors would like to thank all staff and participating controllers from LFV for their support in hosting and facilitating simulations.

REFERENCES

[1] C. Westin, C. Borst, and B. Hilburn, "Strategic conformance: Overcoming acceptance issues of decision aiding automation?" *IEEE Trans. Human-Mach. Syst.*, in press.

[2] B. Hilburn, C. Westin, and C. Borst, "Will controllers accept a machine that thinks like they think? The role of strategic conformance in decision aiding automation," *ATC Quart.*, vol. 22, no. 2, pp. 115–136, 2014.

[3] B. Kirwan and M. Flynn, "Investigating air traffic controller conflict resolution strategies," EUROCONTROL, Brussels, Belgium, Rep. ASA.01.CORA.2.DEL04-B.RS, Mar. 2002.

[4] B. Hilburn, "Cognitive complexity in air traffic control: A literature review," EUROCONTROL, Brussels, Belgium, Rep. EEC Note No. 04/04, Mar. 2004.

[5] R. H. Mogford, E. D. Murphy, R. J. Roske-Hofstrand, and J. A. Guttman, "Application of research techniques for documenting cognitive processes in air traffic contro: Sector complexity and decision making," CTA Incorp., Atlantic City, NJ, Techn. Note DOT/FAA/CT-TN94/3, Jul. 1994.

[6] G. M. Flynn, C. Leleu, and B. Hilburn, "A complexity study of the Maastricht upper airspace centre," EUROCONTROL, Brussels, Belgium, Rep. EEC Report No. 403, Feb. 2006.

[7] S. Magyarits, and P. Kopardekar, "Dynamic density: The measure of air traffic control sector complexity for the en route environment," FAA,. Atlantic City, NJ, Rep. DD1-25-01, Apr. 2001

[8] C. Westin, *Investigating controller reliability in air traffic complexity judgments.* Master's thesis. Nottingham University, UK, 2002

[9] R. P. Thomas, B. Willem, J. Shanteau, J. Raacke, and B. M. Friel, "Measuring performance of expert controllers: An application to air traffic control," *Hum. Fac. Erg. Soc. Proc.*, vol. 45, no. 4, pp. 286-290, 2001.

[10] H. A. Arkes, V. A. Shaffer, and M. A. Medow, " Patients derogate physicians who uses a computer-assisted diagnostic aid," *Med. Decis. Making*, vol. 27, no. 2, pp. 189-202, 2007.

[11] F. J. Lerch, M. Prietula, and C. T. Kulik, "The turing effect: The nature of trust in expert systems advice," in *Expertise in Context*, J. F. Paul, M. F. Kenneth, and R. H. Robert, Eds. Cambridge, MA: MIT Press, 1997, pp. 417-448.

[12] J. J. Dijkstra, " User agreement with incorrect expert system advice," *Behav. Inform. Technol.*, vol. 18, no. 6, pp. 399-411, 1999.

[13] P. Madhavan and D. A. Wiegmann, "Effect of information source, pedigree, and reliability on operator interaction with decision support systems," *Hum. Factors*, vol. 49, no. 5, pp. 773-785, 2007.

[14] F. C. Lacson, D. A. Wiegmann, and P. Madhavan, "Effects of attribute and goal framing on automation reliance and compliance," *Hum. Fac. Erg. Soc. Proc*., vol. 49, no. 3, pp. 482-486, 2005.

[15] K. Christoffersen, and D. D. Woods, "How to make automated systems team players," in *Volume 2 - Advances in Human Performance and Cognitive Engineering Research*, E. Salas, Ed. St. Louis, MO: Elsevier, 2002, pp. 1-12

[16] H. Cramer, V. Evers, S. Ramlal, M. Someren, L. Rutledge, N. Stash, L. Aroyo, and B. Wielinga, "The effects of transparency on trust in and acceptance of a content-based art recommender," *User Model. User-Adap. Inter.*, vol. 18, no. 5, pp. 455–496, 2008.

[17] J. L. Herlocker, J. A. Konstan, L. G. Terveen, J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Trans. Inform. Sys*., vol. 22, no. 1, pp. 5-53, 2004.

[18] M. Dzindolet, S. A. Peterson, R. A. Pomranky, L. G. Pierce, H. P. Beck, "The role of trust in automation reliance," *Int. J. Human-Comput. Stud*. vol. 58, no. 6, pp. 697–718, 2003.

[19] G. A. Mercado Velasco, M. Mulder, and M. M. van Paassen, "Analysis of air traffic controller workload reduction based on the solution space for the merging task," in *AIAA GNC Conf.*, Toronto, Canada, Aug. 2-5. 2010, pp. 1–18.

[20] W. Knecht, "Requirements for automated collision-avoidance systems for Free Flight," *Hum. Fac. Erg. Soc. Proc.,* vol. 42, no. 1, pp. 42-46, 1998.

[21] J.-Y. Jian, A. M. Bisantz, and C. G. Drury, "Foundations for an empirically determined scale of trust in automated systems," *Int. J. Cogn. Ergon.*, vol. 4, no. 1, pp. 53-71, 2000

[22] R. Goritzlehner, C. Borst, J. Ellerbroek, C. Westin, M. van Paassen, and M. Mulder, "Effects of transparency on the acceptance of automated resolution advisories," in *SMC Conf.*, San Diego, CA, Oct. 5-8 2014, pp. 2965-2970.

[23] D. A. Wiegmann, "Agreeing with automated diagnostic aids: A study of users' concuurence strategies," *Hum. Fac. Erg. Soc. Proc.,* vol. 44, no. 1, pp. 44-50, 2002.

[24] C. E. Zsambok, and G. Klein, *Naturalistic decision making*. Mahwah, NJ: Erlbaum, 1997.

[25] N. Tintarev, and J. Masthoff, "Effective explanations of recommendations: User-centered design," in *ACM RECSYS*., Minneapolis, MN, Oct. 19-20. 2007, pp. 153-156.