Objective: The aim of this study was to examine the human–automation interaction issues and the interacting factors in the context of conflict detection and resolution advisory (CRA) systems.

Background: The issues of imperfect automation in air traffic control (ATC) have been well documented in previous studies, particularly in conflict-alerting systems. The extent to which the prior findings can be applied to an integrated conflict detection and resolution system in future ATC remains unknown.

Method: Twenty-four participants were evenly divided into two groups corresponding to a medium- and a high-traffic density condition, respectively. In each traffic density condition, participants were instructed to perform simulated ATC tasks under four automation conditions, including reliable, unreliable with short time allowance to secondary conflict (TAS), unreliable with long TAS, and manual conditions. Dependent variables accounted for conflict resolution performance, workload, situation awareness, and trust in and dependence on the CRA aid, respectively.

Results: Imposing the CRA automation did increase performance and reduce workload as compared with manual performance. The CRA aid did not decrease situation awareness. The benefits of the CRA aid were manifest even when it was imperfectly reliable and were apparent across traffic loads. In the unreliable blocks, trust in the CRA aid was degraded but dependence was not influenced, yet the performance was not adversely affected.

Conclusion: The use of CRA aid would benefit ATC operations across traffic densities.

Application: CRA aid offers benefits across traffic densities, regardless of its imperfection, as long as its reliability level is set above the threshold of assistance, suggesting its application for future ATC.

Keywords: automation; human–automation interaction; air traffic control; conflict resolution; mental workload; situation awareness
Returning to the simpler ATC conflict alerting system, although the automation imperfections here are well documented (Wickens et al., 2009), it is apparent that even such imperfect automation can assist ATCOs in conflict detection and understanding relative to totally unaided performance (Metzger & Parasuraman, 2005). Indeed, a meta-analysis of studies of these and other automation-supported detection tasks reveals that the reliability of such automation can be as low as 70% to 75% and still assist human–system performance relative to human-only performance (Wickens & Dixon, 2007). This effect is particularly true when the workload imposed on humans’ cognition is high.

High workload influences the tendency to depend on, to agree with, and to accept more automation aid (Wickens & Dixon, 2007), particularly in a high-traffic environment (Prevot et al., 2012; Westin, Borst, & Hilburn, 2013) where ATCOs have limited time to perform ATC simultaneous tasks (Vossen, Hoffman, & Mukherjee, 2012). However, high task load does not necessarily affect trust. Trust is affected more critically by automation reliability (Wickens, Hollands, Banbury, & Parasuraman, 2013) and is generally rated lower with lower automation reliability (Metzger & Parasuraman, 2005).

In the current research, we use the CRA aid that integrates the conflict detection and resolution system as a tool for investigating the above issues in human–automation interaction (HAI). This choice allows us to examine the imperfections in the CRA system in a manner that has rarely been done within the very small body of research that has examined controller performance with the CRA aid (Cabrall et al., 2014; Prevot et al., 2012; Prevot, Homola, & Mercer, 2008; Trapsilawati et al., 2015). Only Trapsilawati et al. (2015) examined this issue of imperfect CRA aid, and those authors did indeed find that the imperfect CRA aid assisted ATCOs relative to manual performance. However they did not vary task load to examine if this factor amplified CRA aid benefits and imperfection costs.

In the current experiment we had a much higher level (2 times and 3 times) of traffic load, whereby dependence on automation might be expected to be considerably higher, thereby amplifying both its benefits and its costs when it errs (Wickens & Dixon, 2007). Thus the current study extends the general paradigm employed by Trapsilawati et al. (2015) to examine these issues. The issue of imperfect automation is thoroughly addressed in recent work by Onnasch, Wickens, et al. (2014) and by Wickens et al. (2015). Here the distinction is made between the overall level of performance supported by automation that is above around 75% correct on the one hand (aggregating both automation correct and automation failure trials) and the specific performance on the infrequent occasions (e.g., 25%) when automation fails on the other. The former value indicates overall automation assistance relative to unaided manual performance (the weighted average of correct and automation error trials). This is the value employed by Wickens and Dixon (2007) to define the threshold of automation assistance. However, the performance on the rare automation failure trial suffers. This degradation on the automation failure trial results, in part, because the operator increases dependence on automation during the frequent trials of its correct operation, becomes partially out of the loop (a complacency or automation bias effect), and hence loses SA, as revealed by the meta-analysis of Onnasch, Wickens, et al. (2014). As a consequence, there is a less fluid intervention when automation does fail, and this result is particularly prominent the first time automation fails (Wickens et al., 2015).

To examine these issues, we had participants with ATC experience control traffic manually and also use a CRA tool that was either fully or partially reliable under low and high levels of workload (traffic density). When the CRA aid was unreliable, we imposed either a short or long time allowance to secondary conflict (TAS). We examined a variety of performance and cognitive variables both when the CRA aid worked perfectly and imperfectly.

Based on prior research described earlier, seven hypotheses were offered (Table 1).

**METHOD**

**Participants**

Twenty-four participants ages 19 to 34 years ($M = 24$ years, $SD = 3.33$ years) were recruited. The participants consisted of 22 students and
two professional ATCOs in Singapore. The students majored in aerospace and aeronautical engineering from local institutions and had at least 2 weeks ATC simulator experience during their course. An interview was conducted before the study to make sure that all these participants were familiar with the procedures and terms used in the current ATC practice.

**Apparatus**

Two monitors were provided to show an ATC simulator and a CRA aid, in both ATCO and pseudopilot positions. A PC-based ATC simulation software, ATCSimulator2, was used. The ATC simulator consisted of three components shown in separate windows: a 60-NM range radar display and arrivals and departures flight progress strips (Figure 1). The traffic scenario in airspace was generated with an ATC Sector Design Kit (ATC-SDK). The flight plans, including fleet times, positions, and speeds of aircraft, were manipulated in order to generate conflicts.

A PC-based low-fidelity CRA aid was developed. The CRA aid showed the predicted conflicting pair of aircraft, the resolution maneuver advisory, and the option buttons to accept or reject the proposed resolution (see Figure 2). The list of possible abbreviations showed by the CRA aid is provided in Appendix A. The CRA aid worked based on the principle for Resolution Aircraft and Maneuver Selector (RAMS) proposed by Erzberger (2006). The CRA aid applied the altitude-first resolver principle that means a vertical maneuver would be suggested first over lateral and speed maneuvers due to its expediency (Rantanen & Wickens, 2012).

In the automated conditions, the CRA aid provided a resolution advice 2 min prior to a conflict to alert the participants and provide a resolution advice for the predicted conflict. Neither an alert nor resolution advice was provided in the manual condition. The conflict was defined when the vertical separation between two aircraft is less than 1,000 feet and the horizontal separation is less than 5 NM. When the conflict occurred, the ATC simulator activated a beeping sound and listed the pair of conflicting aircraft on the radar display, and it would remain activated until the conflict was resolved.

**Design**

A mixed-factorial design was adopted. The first factor, automation level, was a within-subjects factor with four levels: reliable, unreliable, ...
and short TAS (U-STAS), unreliable and long TAS (U-LTAS), and manual (Table 2). The four automation levels were presented across participants in a sequence specified by a balanced Latin square method. The second factor, traffic density, was a between-subjects factor with two levels: medium and high. Sector density for medium and high traffic density was 60 and 90 aircraft, respectively. The participants were randomly assigned to these two traffic density conditions, with each condition having 12 participants.

The conflict scenarios were similar across the four testing conditions. However, the aircraft call signs, waypoints, and occurrence times were changed. The traffic patterns were also rotated. This method produced generally similar scenarios across the testing conditions to ensure that other factors, such as conflict scenario (Thomas & Rantanen, 2006), would not influence the effects of manipulation in different testing conditions.

The dependent measures were the task performance measures, including percentage of resolved conflicts, conflict resolution time, and aid utilization rate. Percentage of resolved conflicts was operationally defined as the absence of loss of separation (LOS) relative to the number of designated conflicts. Conflict resolution time was defined only for the automation conditions and reflected the interval between the CRA aid onset and the pilots’ maneuvering response. Aid utilization rate assessed participants’ dependence on the CRA aid and was reflected by the ratio of the accepted advisories relative to the total number of advisories.

The NASA Task Load Index rating scale (total workload range from 1 to 100; Hart & Staveland, 1988) was used to obtain a subjective measure of mental workload. The objective measures of mental workload were derived from...
ready-response latency and percentage of time-outs in the situation present assessment method (SPAM; Loft, et al., 2015). Ready-response latency was defined as the interval between a ready prompt’s onset and participants’ response to the ready prompt. Percentage of time-outs was defined as nonresponded questions.

For the trust rating, the Likert-type rating scale (ranging from 0 to 7; Jian, Bisantz, & Drury, 2000) was administered after each testing condition.

SA was measured throughout the experiment using the SPAM (Durso et al., 1999, Durso & Dattel, 2004). This method incurs fewer burdens onto ATCOs’ memory and is less intrusive compared with other SA assessment tools (Bacon & Strybel, 2013), such as the situation awareness global assessment technique (Endsley, 1988). SPAM ready-response latency indicated the readiness of participants to answer SA questions and often correlates with subjective workload (Strybel, Vu, Kraft, & Minakata, 2008; Vu et al., 2012). Once participants indicated that they are ready, the accuracy and time taken to then answer the SPAM queries reflect SA (Durso & Dattel, 2004). Hence, two SA measures were recorded: (a) probe response latency, defined as the interval between the display of a question and participants’ answer, and (b) percentage of correct responses to SA probes.

Tasks

Each participant performed ATC tasks with the goals of maintaining separation and controlling the traffic flow. Participants used voice transmission to communicate with a pseudopilot. Participants had to handle all arriving and departing aircraft within their controlled area. For arriving aircraft, participants needed to (1) accept incoming aircraft to their area; (2) clear the aircraft altitude to the landing altitude; (3) assign an appropriate approach clearance and runway, either instrument landing system (ILS) or visual approach; and (4) hand over the aircraft to the tower controller. For departing aircraft, participants were responsible for (1) climbing the aircraft to the exit altitude and (2) handing over the aircraft to sector facilities.

Upon receiving voice instructions from the controllers, the pseudopilot inserted respective commands to the ATC simulator. The ATC simulator used in the study provides synthetic voices that met standard phraseology used in the real ATCO–pilot communication procedures when the pseudopilot inserted any instructed commands through keystrokes.

In the manual condition, participants were required to perform all the aforementioned tasks manually, without the CRA aid. In the automated conditions, the participants had available recommendations of the CRA aid. Six conflicts, including five preset primary conflicts and one secondary conflict, were imposed in the unreliable conditions. The detection of primary conflicts was 100% accurate and led to the automated resolution recommendation at 2 min prior to the LOS. Participants interacted with the onscreen CRA aid control using a computer mouse connected to the system. Participants could either accept or reject the resolution advisories provided by the CRA aid by clicking the accept or reject button, respectively.

If the ATCOs accepted the advice, the resolution advice would be automatically sent to the pseudopilot’s screen. The pseudopilot would directly apply the resolution by giving the preset macro commands to the simulator. The conflict resolution would always be effective to successfully resolve the primary conflict. If the ATCOs rejected the resolution advice, the CRA aid would stop processing the respective aircraft’s data and

<table>
<thead>
<tr>
<th>Condition</th>
<th>Reliable</th>
<th>U-STAS</th>
<th>U-LTAS</th>
<th>Manual</th>
</tr>
</thead>
<tbody>
<tr>
<td>High traffic</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Low traffic</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. U-STAS = unreliable and short time allowance to secondary conflict; U-LTAS = unreliable and long time allowance to secondary conflict.
would not be triggered again if the ATCOs implemented an ineffective resolution. The CRA aid would not have enough time to be triggered (i.e., less than 2 min) and generate a resolution for any ineffective resolution made by the ATCOs. This situation then was counted as an unresolved conflict for the conflict resolution performance.

Automation Implementation

Participants were provided with a resolution advisory for each potential conflict in the three automation conditions. In the reliable condition, all advisories provided by the CRA aid were correct (100% of aid reliability). In the unreliable conditions, the aid reliability was 83% (one failure out of six conflicts). The advice for the fourth conflict helped resolve the primary conflict but led to a secondary conflict with a traffic aircraft, and no advice was provided to avoid the secondary conflict. We put the failure in this fourth conflict trial because it would allow participants to develop their trust in the CRA aid during the prior three trials and for the primary conflict of the fourth trial before a failure occurred.

Our decision to impose unreliable automation by the CRA aid’s missing a secondary conflict allowed for the implementation of automation error in the automation advice while experimentally better controlling for the difference in maneuver preference between ATCOs. Furthermore, models for the development of on-ground CRA aid are complex and require a large number of rules to completely cover all possible encounter situations, and the CRA aid may fail in certain situations (Kuchar & Yang, 2000; Rantanen & Wickens, 2012). Hence, this study evaluated the failure situation that may likely happen in the future ATC environment.

In the U-STAS condition, if the ATCO accepted the CRA aid, then it would successfully resolve the first conflict as noted earlier, but this resolution would trigger a secondary conflict between one of the conflicting aircraft in the primary conflict and a traffic aircraft 100 s after implementing the resolution (Figure 3). The traffic aircraft was intentionally preset in the scenario. For the secondary conflict, the CRA aid would not be triggered again, since this was the meaning of CRA aid imperfection in the study. The radar system generated an alert for the ATCO if a secondary conflict (LOS) occurred by a beeping sound as the evidence to the ATCO of the error.

In the U-LTAS condition, the secondary conflict would again be triggered between one of the aircraft in primary conflict and a traffic aircraft, but now this conflict would be 4 min after implementing the resolution maneuver (Figure 4).

Procedure

All participants were provided with a training session. Participants completed two 30-min radar training (i.e., without and with the CRA aid). During the practice session, participants were told that some complex factors, such as flow constraints, area boundary, and iteration process, could be potential triggers for the CRA aid to err. All participants were able to successfully give clearance for all departing and arriving aircraft at the end of the training session.

During the experiment sessions, four 1-hr ATC scenarios were provided corresponding to the four experimental conditions, each on a different day. Participants were instructed to perform appropriate ATC activities to deal with the simulated air traffic situations. Participants were also instructed to respond to SA ready probes and resolve SA question probes that cover all SA levels if they had more attentional resources. Each probe appeared every 6 min; hence there

![Figure 3. The condition unreliable and short time allowance to secondary conflict.](hs.sagepub.com)
were nine probe questions in a 1-hr scenario. Seven probes appeared before the first CRA failure (i.e., the fourth conflict), one probe appeared during the CRA failure (in between primary conflict and secondary conflict, which was within 100 s and 4 min under U-STAS and U-LTAS, respectively), and one probe appeared after the postfailure trial. The ready and question probes would disappear after 1 min of no response. The list of the probes is provided in Appendix B.

Statistical Analysis

A 4 (automation level) × 2 (traffic density) mixed-design ANOVA was performed to analyze the percentage of resolved conflicts, mental workload, and SA. The trust ratings, conflict resolution time, and dependence were analyzed using a 3 × 2 mixed-design ANOVA since the data were collected only in the three CRA conditions.

In addition to the omnibus analyses, we performed a series of planned orthogonal contrasts to closely examine the effects of CRA automation and TAS. First, the Primary Conflicts 1 to 5 were analyzed by pooling the data of all the trials to examine the effects of automation versus manual condition with the medium and high traffic densities. Second, to investigate the effects of CRA reliability, a contrast analysis (i.e., reliable vs. unreliable) was performed exclusively on the fifth conflict in the automation conditions for the conflict resolution time, dependence, ready-response latency (i.e., workload indicator), and probe-response latency and accuracy (i.e., SA indicators). This fifth conflict was chosen because it was only here, in the two unreliable blocks, that those participants would be aware that the automation was imperfect, having experienced the automation failure on the previous conflict. Hence, on this conflict, we could measure the direct effects of unreliability on dependence. Finally, a targeted analysis on the secondary conflict (occurring in Trial 4) was performed to investigate the effects of TAS on the resolution performance.

RESULTS

Task Performance Measures

Percentage of resolved conflict. The main effect of automation level on the percentage of resolved conflicts was significant, $F(3, 66) = 40.30, p < .01, \eta^2 = .647$. All automation groups ($M = 87.50\%, SE = 4.24\%$) improved resolution performance relative to the manual condition, whereby participants were not equipped with the CRA aid ($M = 50.00\%, SE = 5.34\%$), $F(1, 66) = 103.98, p < .01$ (Figure 5). More conflicts were successfully resolved in the medium–traffic density ($M = 87.08\%, SE = 4.06\%$) than in high–traffic density ($M = 69.17\%, SE = 4.97\%$) conditions, $F(1, 22) = 17.55, p < .01, \eta^2 = .444$. A significant interaction effect between automation level and traffic density was observed, $F(3, 66) = 3.03, p = .04, \eta^2 = .121$, confirming the amplification of CRA aid benefits under high traffic. Thus both Hypothesis 1 and Hypothesis 2 were strongly confirmed.

For the fifth conflict, there was no performance difference between reliable ($M = 95.93\%, SE = 4.17\%$) and unreliable ($M = 89.58\%, SE = 4.46\%$) conditions, $F(1, 44) = 0.85, p = .36$, indicating the benefit of even CRA automation that had proven to be unreliable in the previous conflict.

For the targeted analysis of the secondary conflict, no significant difference between short TAS ($M = 54.17\%, SE = 10.39$) and long TAS ($M = 70.83\%, SE = 9.48\%$) was found, $F(1, 22) = 1.09$, $p = .32$.
Conflict resolution time. Conflicts were resolved significantly faster under medium ($M = 21.33$ s, $SE = 3.12$ s) than under high traffic density ($M = 38.33$ s, $SE = 4.95$ s), $F(1, 22) = 15.79, p < .01, \eta^2 = .418$. There were no other significant results on resolution time.

Aid utilization rate. For the fifth conflict, no difference in dependence between reliable ($M = 79.17\%$, $SE = 8.47\%$) and unreliable ($M = 75.55\%$, $SE = 7.37\%$) conditions was found, $F(1, 44) = 0.17, p = .68$. Higher dependence was observed under U-LTAS ($M = 87.5\%$, $SE = 6.89\%$) than under U-STAS ($M = 62.5\%$, $SE = 10.09\%$), $F(1, 44) = 4.57, p = .04$. There were no other significant results.

Mental Workload Versus Subjective Ratings of Mental Workload

Ready-response latency in the SPAM indexed objective workload. There was a significant difference for ready-response latency to suggest that automation ($M = 9.25$ s, $SE = 1.68$ s) did reduce workload compared with manually performing the tasks ($M = 13.58$ s, $SE = 2.44$ s) by shortening this latency, $F(1, 66) = 7.50, p = .01$. For the subjective workload measure, there were no significant results.

For the fifth conflict, higher workload as indicated by longer ready latency was found under U-LTAS ($M = 11.48$ s, $SE = 2.63$ s) compared with U-STAS ($M = 5.95$ s, $SE = 0.84$ s) condition, $F(1, 44) = 5.25, p = .03$.

SA

No differences between automated and manual conditions were found on probe accuracy ($M = 70.21\%$, $SE = 4.57\%$, vs. $M = 66.80\%$, $SE = 5.07\%$), $F(1, 66) = 1.01, p = .32$, as well as on probe latency ($M = 14.42$ s, $SE = 0.73$ s, vs. $M = 15.08$ s, $SE = 1.03$ s), $F(1, 66) = 0.80, p = .37$. There were no significant effects of traffic density on probe accuracy ($M_{\text{high}} = 67.39\%$, $SE = 7.85\%$, vs. $M_{\text{low}} = 71.69\%$, $SE = 5.25\%$), $F(1, 22) = 0.42, p = .52$, as well as on probe latency ($M_{\text{high}} = 14.23$ s, $SE = 0.97$ s, vs. $M_{\text{low}} = 14.94$ s, $SE = 1.29$ s), $F(1, 22) = 0.29, p = .59$.

For the fifth conflict, there were no differences between reliable and unreliable conditions in probe accuracy ($M = 88.89\%$, $SE = 5.68\%$, vs. $M = 87.09\%$, $SE = 5.86\%$), $F(1, 44) = 0.05, p = .82$, as well as in probe response latency.
Subjective Ratings of Trust

Trust was higher after the blocks in which participants were provided with reliable ($M = 5.54, SE = 0.21$) versus unreliable CRA aid ($M = 5.23, SE = 0.23$), $F(1, 44) = 3.99, p = .05$. There were no other significant results.

DISCUSSION

The current experiment addressed the extent to which an imperfect ATC CRA aid could assist controller performance and provided results relative to the more general theory of HAI. The hypotheses and the results are summarized in Table 3.

Our first hypothesis was confirmed regarding performance and objective workload, replicating the prior findings of Trapsilawati et al. (2015) but here at the much higher traffic load of 2 and 3 times the volume. Participants used (depended) on the automation, and their maneuver resolution performance with the aid was substantially better than unaided performance. Furthermore, workload was reduced by automation, since ATCOs’ ability to effectively and safely maintain aircraft separation depends upon the maintenance of their mental picture (Alexander, Wickens, & Merwin, 2005; Nunes & Mogford, 2003); the CRA aid could reduce the amount of mental computation and thus reduce their workload. We also observed increased CRA aid benefit in resolving conflicts under higher (47.8%) than under lower traffic load (27.2%), supporting Hypothesis 2.

Regarding Hypothesis 3, the analysis on Trial 5, at which time, from the participant’s perspective, automation was now seen to be 83% reliable in the two unreliable conditions, no decrement in performance was observed relative to the 100% reliable condition, hence providing additional data in support of the conclusions of Wickens and Dixon (2007) that reliability as low as 83% does not impose substantial costs and thereby generalizing their conclusions from automated detection tasks to those of automated decision aids.

Importantly, automation did not lower SA, in contrast to the prototypical model of automation dependency (Higham, Vu, Miles, Strybel, & Battiste, 2013; Onnasch, Ruff, et al., 2014; Onnasch, Wickens, et al., 2014). Our null finding here failed to confirm Hypothesis 4 but replicated the findings of Trapsilawati et al. (2015) regarding CRA aid benefits to overall SA and again is consistent with the assumption that our CRA aid-supported participants continued high engagement in the traffic situation. Thus, any possible costs of imposing automation to SA appeared to have been offset by its workload reduction benefits, which availed more resources for monitoring the raw traffic data. The absence of an interaction of automation with traffic load on SA also failed to confirm Hypothesis 5.

The effects of CRA aid on trust and dependence were in the expected direction. Trust was degraded in the unreliable automation blocks. However, dependence on the CRA aid was not degraded after it had failed, supporting Hypothesis 6. This result occurred due to high workload, such that participants opted to rely on the CRA aid to preserve cognitive resources for other ATC tasks although they did not trust it as much.

Regarding Hypothesis 7, the effects of TAS in this experiment were muted, just as they were in Trapsilawati et al. (2015). TAS did not affect SA. This might be because the short TAS (i.e., 100 s) reflected the average look-ahead time of the existing conflict alert in the current ATC operations. Since the CRA aid here represented the integration of conflict detection and resolution system, participants might be familiar with the situation and thus did not compromise their SA. In contrast to Hypothesis 7, shorter TAS did not increase dependence (utilization) but in fact decreased it. Correspondingly, shorter TAS did not increase workload but also decreased that variable when measured by ready-response time. Thus these two dependent variables were linked in an interpretable way: When workload was inferred to be higher, dependence increased; but the cause of the increase in workload was, unexpectedly, a longer, not a shorter, TAS. We may interpret this somewhat unexpected result by assuming that longer TAS allowed participants...
more time and resources to attend to the raw radar data and to use their cognitive resources to help choose the appropriate maneuver. Greater reliance on one’s own cognition is a very plausible source of higher workload.

**Limitations and Implications**

There exist some limitations in the present study. First, the ATC simulator used in the present study may not fully represent the real ATC environment. Real-world stressors, such as weather changes, were not taken into account. Second, most of the participants were not active ATCOs. However, our participants all had ATC training and knowledge and were familiar with ATC systems through prior training. Further, both ATCOs and students have been found in previous studies to have similar patterns of resolution maneuvers (Rantanen & Nunes, 2005). Given that the CRA aid has not been applied in the current ATC workplace, this research was conducted to gather preliminary information in the CRA context. Using students as participants fit well with the explanatory nature of the present study (Goritzlehner et al., 2014). Third, the conflict-alerting function integrated into the CRA aid was always reliable, whereas previous studies indicated different performance consequences due to misses and false alarms of conflict detection algorithms (Dixon, Wickens, & McCarley, 2007; Rovira & Parasuraman, 2010; Wickens & Colcombe, 2007). Thus, further research incorporating this issue is required. Last, the analysis of single–automation failure trials must of necessity be of considerably lower statistical power than the overall performance measures because, by

---

**TABLE 3: Summary of the Hypotheses Results**

<table>
<thead>
<tr>
<th>Hypothesis No.</th>
<th>Hypothesis</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Imposing the CRA automation will increase both performance and reduce workload (when compared with fully manual conflict resolution), given that controllers use (depend upon) the automation.</td>
<td>Confirmed</td>
</tr>
<tr>
<td>2</td>
<td>The benefits in Hypothesis 1 will be amplified under high traffic load, when cognitive resources are scarcer.</td>
<td>Confirmed</td>
</tr>
<tr>
<td>3</td>
<td>The benefits of the CRA automation will be manifest even when it is imperfectly reliable, given that its reliability is greater than the threshold of automation assistance between 0.70 and 0.75 (Wickens &amp; Dixon, 2007).</td>
<td>Confirmed</td>
</tr>
<tr>
<td>4</td>
<td>CRA automation, whether perfect or imperfect, will decrease situation awareness of the traffic situation.</td>
<td>Not confirmed</td>
</tr>
<tr>
<td>5</td>
<td>The effects in Hypothesis 4 will be amplified under high workload (high traffic load).</td>
<td>Not confirmed</td>
</tr>
<tr>
<td>6</td>
<td>After automation has failed, the trust in the CRA aid will decline but the dependence will not decline, due to the high traffic load.</td>
<td>Confirmed</td>
</tr>
<tr>
<td>7</td>
<td>Shorter TAS should decrease situation awareness, increase workload, and increase dependence on automation.</td>
<td>Not confirmed</td>
</tr>
</tbody>
</table>

Note. CRA = conflict resolution advisory; TAS = time allowance to secondary conflict.
definition, unexpected failures occur rarely (Wickens, 2009). Hence, we must interpret more cautiously those null effects that are observed on the single-conflict trial (Trial 4 or 5).

The present study has positive implications for CRA aid use. First, the results imply that system designers must put high effort in providing a reliable automation’s decision in the first conflict-resolution iteration to turn “totally unexpected” effects to “surprising” effects. For the air traffic management industry, this implication shifts the complexity of conflict resolution automation development to a feasible level. As of today, research is still ongoing for the development of conflict resolution automation; however, considering every single constraint in the airspace to provide a fully secondary conflict–free resolution is nearly impossible (Kuchar & Yang, 2000). Yet, ATCOs are waiting for the real support in facing the imminent traffic growth. Next, we found that CRA aid could improve ATCOs’ performance and reduce workload across traffic levels, verifying its applicability across traffic densities. Moreover, participants worked well with the CRA aid and preferred to share the responsibility of separation assurance (Cabrall et al., 2014) with the CRA aid. Thus, setting automation at the moderate level (Bekier, Molesworth, & Williamson, 2012; Li et al., 2014; Parasuraman et al., 2000) that supports rather than replaces ATCOs (Prevot et al., 2012) was recommended in developing the CRA aid. This level allows ATCOs to complement the CRA aid safely with their manual intervention should any automation failure occur.

APPENDIX A

List of Maneuvering Abbreviations and the Units

<table>
<thead>
<tr>
<th>Notation</th>
<th>Abbreviation</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Climb</td>
<td>Flight level</td>
</tr>
<tr>
<td>D</td>
<td>Descend</td>
<td>Flight level</td>
</tr>
<tr>
<td>FH</td>
<td>Fly heading</td>
<td>Degree</td>
</tr>
<tr>
<td>IS</td>
<td>Increase speed</td>
<td>Knots</td>
</tr>
<tr>
<td>RS</td>
<td>Reduce speed</td>
<td>Knots</td>
</tr>
</tbody>
</table>

APPENDIX B

List of Probe Questions

Level 1
- What is Aircraft A’s speed?
- What is the direction of departure for Aircraft B?
- What is the altitude clearance for Aircraft C?

Level 2
- How many aircraft are flying southbound?
- Which aircraft has lower altitude?
- Is the difference in heading between Aircraft D and Aircraft E more than 90°?

Level 3
- Which aircraft must be handed off to another sector within the next 2 min?
- Which pairs of aircraft will lose separation if they stay on their current courses?
- Which aircraft will need a new clearance to achieve landing requirements?

ACKNOWLEDGMENTS

Research was supported by Civil Aviation Authority of Singapore (CAAS) and Air Traffic Management Research Institute (ATMRI), Nanyang Technological University (NTU), Singapore, project reference ATMRI:2014-R5-CHEN.

KEY POINTS

- The human–automation interaction issues of imperfect automation and other underlying factors in future air traffic control were examined in the context of conflict detection and resolution system.
- Conflict resolution advisory (CRA) aid benefits air traffic controllers, notwithstanding its possible imperfection, in different traffic densities (27.22% and 47.78% improvement in lower and higher traffic load, respectively, as compared with manually performing the task).
- The benefits of imperfect CRA aid were associated with the examination of incorrect resolution with traffic aircraft, which turned a “totally unexpected” event into a “surprising” event.
- Time allowance to secondary conflict (TAS) did not affect situation awareness. Longer TAS led to...
higher workload and thus higher dependence on the CRA aid.

REFERENCES


Fitri Trapsilawati is a PhD candidate at the School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore. She earned her BEng in industrial engineering from the Universitas Gadjah Mada, Indonesia, in 2010.

Christopher D. Wickens is a professor emeritus of aviation and psychology at the University of Illinois and is currently a senior scientist at Alion Science and Technology, Boulder, Colorado, and a professor of psychology at Colorado State University.

Xingda Qu is a professor at Shenzhen University, Shenzhen, China. He is also serving as the director of the Institute of Human Factors and Ergonomics at Shenzhen University. He earned his PhD in industrial engineering from Virginia Tech, United States, in 2008.

Chun-Hsien Chen is an associate professor and the director of the design stream at the School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore. He earned his PhD in industrial engineering from the University of Missouri–Columbia, United States.

Date received: June 30, 2015
Date accepted: May 24, 2016