# How Important is Conflict Detection to the Conflict Resolution Task?

Joey Mercer[1], Cynthia Gabets[2], Ashley Gomez[2], Tamsyn Edwards[2], Nancy Bienert[2], Lauren Claudatos[2] and Jeffrey Homola[1],

[1] NASA Ames Research Center, Moffett Field, CA, USA, 94035
[2] San Jose State University / NASA Ames Research Center, Moffett Field, CA, USA, 94035
{joey.mercer, cynthia.gabets, ashley.n.gomez, tamsyn.e.edwards, nancy.bienert, jeffrey.r.homola}@nasa.gov

**Abstract.** To determine the capabilities and limitations of human operators and automation in separation assurance roles, the second of three Human-in-the-Loop (HITL) part-task studies investigates air traffic controller's ability to detect and resolve conflicts under varying task sets, traffic densities, and run lengths. Operations remained within a single sector, staffed by a single controller, and explored, among other things, the controller's conflict resolution performance in conditions with or without their involvement in the conflict detection task. Whereas comparisons of conflict resolution performance between these two conditions are available in a prior publication, this paper explores whether or not other subjective measures display a relationship to that data. Analyses of controller workload and situation awareness measures attempt to quantify their contribution to controllers' ability to resolve traffic conflicts.

**Keywords:** human factors · air traffic control · human-in-the-loop simulation · function allocation · human-automation interaction

## 1    Introduction

The transition to NextGen will likely include increasing levels of automation to help controllers perform their duties. A progression towards higher levels of automation could enable the controllers' working environment to move from tactical separation management to strategic decision-making. Such automation is envisioned to expand performance beyond today's limits by off-loading workload from controllers onto automated functions for the majority of routine operations [1]. However, the nature of this human-automation team is not well understood. It is still unknown exactly which tasks are best allocated to the human operator as opposed to the automation, and vice-versa. In considering this system as a whole, careful and thorough investigation is needed to better understand, not only how each team member performs in such environments, but also any associated human-automation cooperation issues.

## 2    Background

The motivation behind these investigations is to address a well-known problem: current-day air traffic control techniques are very labor intensive, and are limited to the amount of information controllers can process and keep in their working memory. Function allocation is but one approach to this problem, wherein automation can take responsibility for some tasks, theoretically easing the controller's workload.

The current series of studies fall under NASA's revised function-allocation research plan, which calls for advancing our understanding of the related air-ground and human-automation issues. In particular, the Airspace Operations Laboratory (AOL) focused on the following question: "Which separation assurance functions can air traffic controllers effectively perform in future air traffic management systems?" Understanding the strengths and weaknesses of individual team members is an important aspect in determining how to distribute tasks between team members. As a first step towards gaining such insights into human-automation teaming, our approach has been to conduct part-task HITL simulations that identify the capabilities and limitations of the controller in key separation assurance tasks.

### 2.1    Function Allocation Research

In March of 2015, the AOL at NASA's Ames Research Center [2] conducted the first in a series of studies that explored the capabilities and limitations of human operators with regard to the separation assurance element of air traffic control. Specifically, the research sought to better understand how best to allocate functions between controllers and automation. A second study conducted in May of 2015 continued that work, but with the conflict resolution task as its main focus. Although the first study differed in that it investigated the conflict detection task, both studies shared the same approach, in which they sought to tease apart the primary task from related secondary tasks. While looking across varying levels of automation, the studies measured the overall impact on the performance of the primary task. Of particular interest to the second study was discovering whether removing controllers' involvement in the detection task would impact their ability to resolve conflicts.

The first study, referred to as the Human-Automation Conflict Detection study (or HACD), and the second study, referred to as the Human-Automation Conflict Resolution study (or HACR), are reported in [3 - 5]. However, a brief summary of the HACR simulation environment is in order, to provide the appropriate context for the discussions of this paper.

**The HACR Simulation.** HACR examined controller performance on the conflict resolution task under different task sets, traffic density levels, and run lengths. The group of tasks under the controller's responsibility and those under the automation's responsibility defined a given task set. Traffic density and run length completed the study's set of independent variables. Although the full study featured a 5x2x2 within-subject repeated-measures design, the scope of this paper and its analyses are limited to two of the study's task sets (*Conflict Resolution* and *Conflict Detection & Resolu-*

*tion*.), both traffic densities (1x current-day traffic levels and 1.2x current-day traffic levels) and one run length (60 minutes).

Clearly, the key distinction between the two task sets of interest lies in whether or not the controller was responsible for the conflict detection task. The *Conflict Detection & Resolution* condition operated much like current-day air traffic control. The controller kept constant watch over their sector's radar display, observing the progress of air traffic in and around their sector, and issuing control instructions they deemed necessary. In contrast, the *Conflict Resolution* condition went to great lengths to isolate the conflict resolution task, and in doing so, removed the controller from the conflict detection task. The study accomplished such isolation by developing a clever display capability that suppressed all air traffic from the radar display unless the automation (i.e., a trajectory-aided conflict probe) detected a potential conflict. Once the automation detected a conflict, the system would turn off the 'blackout' mode, and displayed all traffic as it normally would, albeit with the aircraft in conflict highlighted (see Figure 1). At this point, the automation's task of detecting the conflict was complete, and it was then the controller's responsibility to, just as in the *Conflict Detection & Resolution* condition, issue whatever control instructions they deemed appropriate.
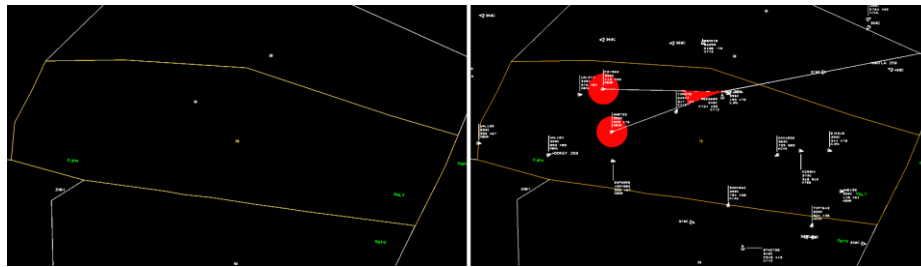


**Fig. 1.** Screen capture of the controller's radar display in the Conflict Resolution condition before the automation detects a conflict (left), and after the automation detects a conflict (right).

The airspace used during the simulation consisted of a single high-altitude sector, with a mix of overflights passing through at level altitudes, and transitioning aircraft descending to or climbing out from area airports. The scenarios progressed through a ramp-up, peak, and ramp-down phase, with each phase lasting approximately 20 minutes. Traffic levels reached 18 aircraft in the sector in the 1x traffic density, and 22 aircraft in the 1.2x density. The simulation's environment also included winds for the area, which were constant-at-altitude with a nominal forecast error. Eight retired FAA en route controllers (with an average of 24.9 years of experience among them) participated in the study, all of which worked the same conditions.

The primary simulation platform used for the study was the Multi Aircraft Control System (MACS) [2], which, for each controller workstation, hosted an En Route Automation Modernization (ERAM) emulation on a large-format monitor. The controller workstation also included a specialized keyboard and trackball, similar to those used in current air traffic control facilities, as well as a custom, stand-alone voice

application emulating the fielded communication system. Data recorded and collected at each workstation included aircraft flight states, operator task data and workload, automation states, voice communications, etc.

## 2.2 Previous Findings

The data presented in [4] compared the time at which the controllers issued a clearance to resolve a conflict, with the time of that conflict's detection. In the *Conflict Detection & Resolution* condition, the detection time was marked when the controller made a keyboard entry to signal they believed an aircraft pair to be in conflict. In the *Conflict Resolution* condition, the detection time was marked when the automation identified an aircraft pair to be in conflict (i.e., typically when the 'blackout' mode turned off). The difference between these two event times represents the Resolution Response Time measurement.

The findings showed that the controllers were able to issue resolution maneuvers within 30 seconds of conflict detection for 49% of cases in the *Conflict Resolution* condition, but did so for 59% of cases in the *Conflict Detection & Resolution* condition. Even after accounting for the traffic density variable, this trend held true: the proportion of resolution maneuvers issued within 30 seconds of conflict detection were 46% and 56% for the same conditions (respectively) at the 1x traffic density, and 51% and 64% at the 1.2x density. These results indicate that when removed from the conflict detection task, controllers more often needed more time in order to issue a resolution. Although measurements did not distinguish between solution identification time and solution execution time, when considering the fact that the solution execution methods available were constant across conditions, one can reasonably believe this data reflects an increase in the solution identification time (i.e., the controllers needed more time to determine how to solve the conflict).

Although the concentration of resolution response times showed more noticeable changes within the different comparisons, a repeated-measures analysis of variance (ANOVA) for resolution response time mean values did not provide significant results for the task set or traffic density variables ($p>0.05$). Table 1 lists the relatively similar descriptive statistics for the four combinations of task sets and traffic densities.

**Table 1.** Resolution response time mean and standard deviation values (in seconds) from the CR and CD&R task sets in both 1x and 1.2x densities.

| Resolution Response Time | CR | | CD&R | |
|---|---|---|---|---|
| | M | SD | M | SD |
| 1x | 48.00 | 17.47 | 56.88 | 28.33 |
| 1.2x | 45.63 | 14.78 | 49.25 | 31.28 |

## 3 Method

This paper explores whether or not other subjective measures display a relationship to the controllers' conflict resolution performance. The current analyses examine work-

load and situation awareness because prior research identified both as critical factors that frequently and negatively influence controller performance [6]. The results from [4] seem to support an obvious hypothesis: when controllers are not involved in the conflict detection process, they know less about the circumstances surrounding the conflict, and as a result, need more time to assemble a detailed enough picture in order to know what action(s) to take. This paper seeks to validate this idea using the available situation awareness data. Analyses will also compare the conflict resolution performance data against the controller workload data to seek out other hidden relationships between the objective and subjective data.

Although the full study included two treatments of run-length, the analyses in this paper are limited to only the 60-minute duration runs. However, the results from [4] collapse the run-length variable, combining data from the 60-minute runs with data from the 20-minute runs. In order to better align with the analyses presented here, new analyses of the performance data are also included.

### 3.1 Workload

Workload Assessment Keypads (WAKs) probed controller workload at three-minute intervals during the simulation trials. Controllers responded to the workload probes with Air Traffic Workload Input Technique (ATWIT) [7] ratings along a modified six-point scale (e.g., 1 as low workload, 6 as high workload).

### 3.2 Situation Awareness

The study collected situation awareness data using the Situation Present Assessment Method (SPAM) [8]. After responding to each of the workload prompts, a small window appeared on the display, presenting participants with a situation awareness question. Developed in collaboration with three retired air traffic controllers who were not participants in the study, the questions used a yes/no response format, implemented as separate response buttons within the question window. After answering the situation awareness question (i.e., after clicking either the 'yes' button or the 'no' button), the window automatically disappeared, allowing the participants to return to their air traffic control duties with minimal interruption. Results included in this paper benefit from two different measures of situation awareness: percentage of questions answered correctly (accuracy), and elapsed time between question presentation and correct answer (response time).

## 4 Results

The following describes the results from the current data analyses, all sourced exclusively from the 60-minute runs within the *Conflict Resolution* (CR) and *Conflict Detection & Resolution* (CD&R) task sets. The selected metrics are first considered individually, followed by multi-variate examinations that look to identify quantifiable relationships (via a series of Spearman's Correlation tests) between the objective conflict resolution performance data, and the subjective workload and situation awareness data. Other publications provide additional results from the HACR simulation [4, 5].

## 4.1 Resolution Response Time

Across task sets, the controllers were able to issue resolution maneuvers within 30 seconds of conflict detection for 51% of cases in the CR task set, and 53% of cases in CD&R. After further isolating the traffic density variable, data from the trials simulating 1x traffic density indicated that 52% of resolution maneuvers occurred within 30 seconds of detection in the CR condition, compared to 51% in CD&R. These numbers changed to 49% and 59% (respectively), at the 1.2x traffic density. This data differs from the findings reported in [4] that, at the highest level, associate controller involvement in the conflict detection task with more often needing less time to resolve a conflict. Such distinction is no longer present in this data, now characterized by largely similar distributions.

When looking at the mean values for resolution response time, ANOVA results approached significance for the comparison between task sets ($F(1,7) = 3.928$, $p = 0.088$), where CR (surprisingly) had faster resolution times (M=44.938, SD=4.74) than the CD&R condition (M=68.125, SD=10.853). Traffic density did not have a significant effect on resolution response time. Table 2 lists the relevant mean and standard deviation values.

## 4.2 Workload

A Kolmogorov-Smirnov test indicated that the workload data violated the assumptions of normality ($p<0.05$), thus requiring a Friedman's ANOVA for non-parametric data. This test revealed a significant difference between task sets and traffic densities, $\chi^2 (3) = 18.600$, $p = 0.01$. Post-hoc analyses applied a Bonferroni correction to Wilcoxon signed-rank tests and showed significant differences between task sets, in both the 1x ($Z = -2.100$, $p = 0.036$) and the 1.2x ($Z = -2.521$, $p = 0.012$) densities, with CR reporting lower workload ratings than the CD&R condition. Traffic density had a significant effect on workload, but only in the CD&R task set ($Z = -2.521$, $p = 0.012$), with lower workload ratings coming from the 1x density. There was no significant effect of traffic density within the CR condition ($Z = -1.120$, $p = 0.263$). Descriptive statistics reflect these trends (see Table 2). These workload results appear to support our expectation (and also align with the HACD data reported in [3]), that workload would increase under less automated working environments and during higher levels of traffic.

## 4.3 Situation Awareness

**Situation Awareness Accuracy.** Listed in Table 2, the average percentages of correctly-answered situation awareness questions remained fairly stable throughout the four combinations of task sets and traffic densities, with between 70.25% and 77.75% accuracy. A repeated-measures ANOVA confirmed this with no significant effect of condition or density ($F(1,7) = .747$, $p = 0.416$) and ($F(1,7) = 1.268$, $p = 0.297$), respectively. These results contest the expectation that removing controllers from the conflict detection task (i.e., the CR task set) would negatively impact their situation awareness.

**Situation Awareness Response Time.** A repeated-measures ANOVA showed a significant difference in situation awareness response time as a result of task set, $(F(1,7) = 7.555, p<0.05)$, where the CR condition had slower response times (M=7.114, SD=0.391) than the CD&R condition (M=6.057, SD=0.525). Tests also revealed that traffic density had no significant effect on situation awareness response time. In contrast to the situation awareness accuracy data, these results support the notion that the conflict detection task is an important contributor to the controller's understanding of the traffic in their sector. Admittedly, the statistical significance here represents a difference of less than two seconds; therefore such findings may have limited meaning. Descriptive statistics are listed in Table 2.

**Table 2.** Summary of means and standard deviations of the resolution response times (seconds), workload ratings, situation awareness accuracy, and situation awareness response times (seconds) from the CR and CD&R task sets in both 1x and 1.2x densities.

| Resolution Response Time | CR | | CD&R | |
|---|---|---|---|---|
| | M | SD | M | SD |
| 1x | 46.75 | 20.48 | 69.88 | 37.78 |
| 1.2x | 43.13 | 9.06 | 66.38 | 56.36 |
| **Workload** | CR | | CD&R | |
| | M | SD | M | SD |
| 1x | 1.69 | 0.42 | 2.48 | 0.72 |
| 1.2x | 2.11 | 0.60 | 3.12 | 0.55 |
| **Situation Awareness Accuracy** | CR | | CD&R | |
| | M | SD | M | SD |
| 1x | 74.50% | 10.20% | 77.75% | 11.34% |
| 1.2x | 70.25% | 16.15% | 74.25% | 7.34% |
| **Situation Awareness Response Time** | CR | | CD&R | |
| | M | SD | M | SD |
| 1x | 6.89 | 1.28 | 6.23 | 1.61 |
| 1.2x | 7.34 | 1.26 | 5.89 | 1.65 |

## 4.4    Resolution Response Time and Workload

Results from a Spearman's correlation test between resolution response time and workload approached significance in the CD&R-1.2x pairing ($r_s(18) = 0.452, p = 0.060$). This shows a positive relationship in that resolution response time and workload increased together. It is also interesting to note that while controller workload increased, the variability in resolution response time increased as well (see Figure 2). There were no significant or near-significant relationships found for any of the other combinations of task set and traffic density.
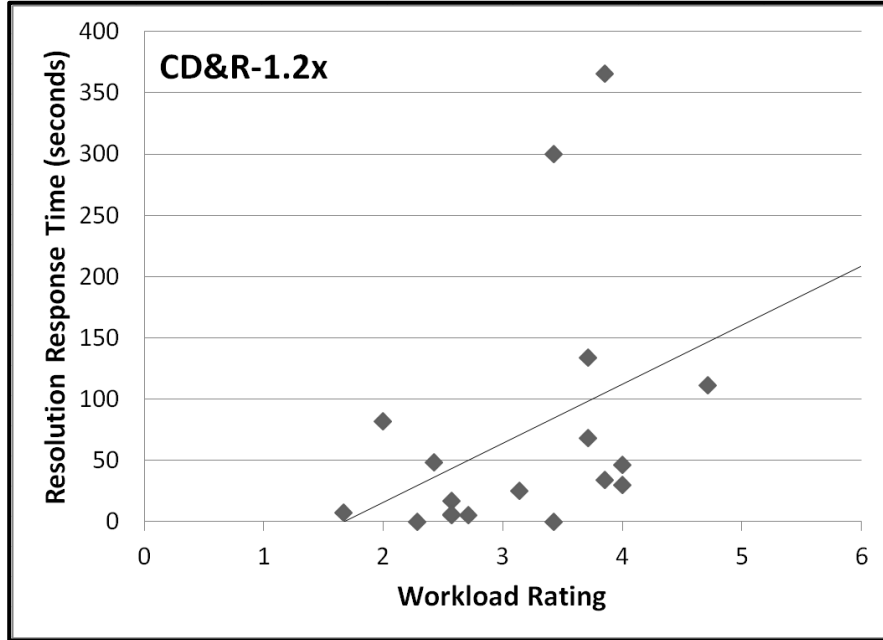
**Fig. 2.** Scatterplot of average resolution response time against average workload rating, including a linear trend-line, for the CD&R task set in the 1.2x traffic density.

### 4.5 Resolution Response Time and Situation Awareness

The correlation between resolution response time and situation awareness accuracy approached significance for two pairings: CR-1x and CD&R-1x (($r_s(24) = 0.374$, $p = 0.072$) and ($r_s(24) = -0.375$, $p = 0.071$), respectively). Although these two correlations are similar in strength, their directions are inverted relative to each other. The CR-1x correlation displayed a positive relationship, with resolution response time and situation awareness accuracy increasing together. Meanwhile, the CD&R-1x correlation revealed a negative relationship, where resolution response time decreased as situation awareness accuracy increased. For reference, these findings are reflected in Figures 3 and 4. Further tests were unable to find any correlations of significance or near-significance for any of the combinations of task set and traffic density ($p>0.1$).
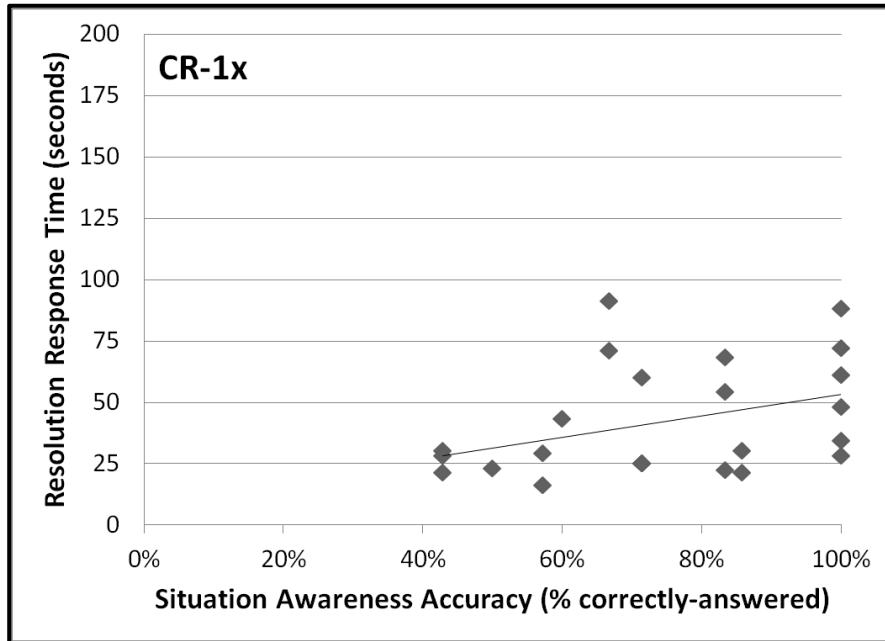
**Fig. 3.** Scatterplot of average resolution response time against average situation awareness accuracy, including a linear trend-line, for the CR task set in the 1x traffic density.
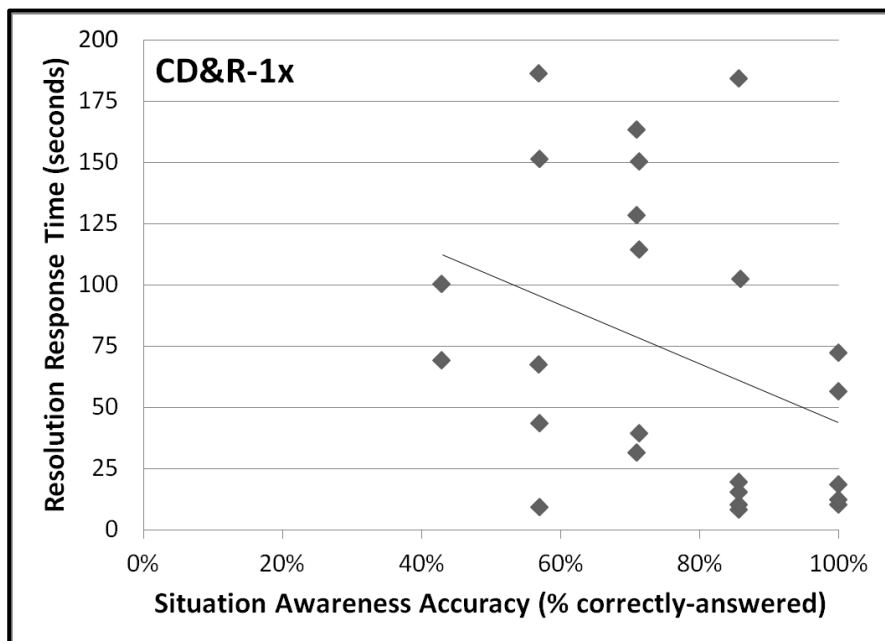


**Fig. 4.** Scatterplot of average resolution response time against average situation awareness accuracy, including a linear trend-line, for the CD&R task set in the 1x traffic density.

# 5    Discussion

When the confounds of run length and traffic density were present in the comparison of resolution response time data between the CR and CD&R task sets, the distribution of data indicated value in having controllers involved in the conflict detection task. However, the new analyses, which focused just on the 60-minute trials and split out the traffic density variable, no longer supported that argument. One possible explanation for this is that in the shorter, 20-minute trials controllers were more engaged and experienced less fatigue; perhaps supported by the larger standard deviations. Within the longer, perhaps more tiring trials, the demands of the more manual environment associated with the CD&R task set, seem to have led to the increased mean response times; enough to reach significance. Another element possibly contributing to these differences stems from a simulation artifact: in the CR condition, the trigger for the radar display's blackout mode was simply the absence of any detected conflicts. The following sequence captures an unintended consequence of this implementation: 1) conflict detected, blackout mode disengages; 2) controller issues a heading vector to maintain separation; 3) conflict resolved (i.e., no longer detected), blackout mode engages. The end result of this example is that an aircraft continues along an open vector, with the controller unable to see when to issue the follow-up heading instruction to put the aircraft back on course. During the simulation, controllers received training on how to manually disengage the blackout mode, facilitating the ability to follow-up on an open-ended instruction in order to 'close the loop', at which point they could manually re-engage the blackout mode. Comments from a few participants indicated this process was a bit cumbersome, and could explain a higher proportion of simpler, cruder resolution maneuvers (e.g., altitude instructions) that better supported the ability to more quickly complete the resolution process for a given encounter before moving on to the next thing.

The workload data describes very believable circumstances, where controllers felt less busy during conditions where they had (literally) nothing to look at for part of the time. Additionally, their workload ratings helped validate our traffic scenarios, reporting lower workload in the 1x traffic density. That the same difference between traffic densities was only observed in CD&R is likely because the effects of the CR task set's blackout mode outweighed the impact of traffic density.

A major concern about human-automation interactions is the possible reduction of operator situation awareness. While the situation awareness accuracy metric did not see any effect of task set, the situation awareness response time data did. The results not only highlight the need to examine situation awareness from multiple angles, but suggest that the controllers were able to perform equally well in correctly answering the situation awareness questions across both task sets, but only at the expense of response time. Given that such expenses amounted to less than two seconds of time, the likely implication is that both the level at which situation awareness was degraded and the amount of compensation needed to overcome it, were minimal.

Correlating the resolution response time and workload data uncovered a potential trend describing a positive relationship in which the resolution response times and workload ratings increased together. This relationship only appeared during the CD&R-1.2x condition, which generally speaking, was the most challenging of the analyzed conditions, since it paired the higher traffic density with the more manual

task set. Perhaps the more difficult nature of this condition explains why the statistical relationship did not appear anywhere else, suggesting that it was the only condition able to elicit a meaningful range of workload ratings from the participants. Also, the resolution response times appear to disperse more as the workload ratings increase, providing additional evidence of the relationship between the two measures: during the more complex situations likely associated with higher workload ratings, it would be reasonable for controllers to need more time to resolve a conflict.

The correlational analysis between resolution response time and situation awareness accuracy helped uncover a few key aspects of how they influenced each other. During the CR task set, any time spent by the controllers resolving a conflict directly corresponded to the amount of time that the blackout mode was disengaged, and consequently, the amount of time they were able to observe the traffic in their sector. Therefore, any increase in resolution response time brought with it more time for the controllers to observe traffic, and naturally led to better answers to the situation awareness questions. Whereas the resolution response time (and 'screen time', indirectly so) seemingly drives the situation awareness in the CR task set, that simple story may not hold true in the CD&R task set. Rather, it appears as if the situation awareness is driving the resolution response time. As controllers perform their conflict detection duties, they naturally need to observe more things and consider more things, and as a result, may need more time to resolve certain conflicts (note the larger spread in resolution response time data in Figure 4 vs. Figure 3). When we consider the situation awareness component, a controller with low situation awareness is likely to take even longer to resolve a conflict. Conversely, a controller with good situation awareness can more likely identify a resolution more quickly.

## 6    Conclusion

This paper examined the subjective measures of workload and situation awareness within the objective context of conflict resolution response time. Real-world service providers are considering future air traffic management systems that include more automation: automation that will likely work jointly with human operators. It is critical then, to understand the various impacts of human-automation interaction, in order to identify any costs or consequences that could inform good system design. In addition to showing that creating an environment which removes the controller from the detection task is more difficult than one might assume, the results here uncover not only the importance of analyses which co-examine multiple factors, but also offer evidence of the obvious relationship between conflict detection and conflict resolution. It was the situation awareness data that best identified how the detection and resolution tasks influence each other. Findings from the situation awareness accuracy data point to the idea that controllers can likely resolve conflicts with our without first detecting the conflict… but will do so in very different ways. Data here suggest removing the conflict detection task will limit situation awareness and may result in the consideration of only a few factors, producing resolutions of a more simplified nature; whereas detecting a conflict beforehand will add to the controller's situation awareness and may result in the considerations of several factors, producing resolutions of a more *optimized* nature.

# References

1. Joint Planning and Development Office. Concept of Operations for the Next Generation Air Transportation System. Version 3.2, JPDO, Washington, D.C., 2010.
2. Prevôt, T., Smith, N., Palmer, E., Callantine, T., Lee, P., Mercer, J., et al. (2014). *An Overview of Current Capabilities and Research Activities in the Airspace Operations Laboratory at NASA Ames Research Center*. AIAA-2014-2860, American Institute of Aeronautics and Astronautics, Reston, VA.
3. Edwards, T., Homola, J., Mercer, J., and Claudatos, L. (2016, in press). Multi-factor Interactions and the Air Traffic Controller: The Interaction of Situation Awareness and Workload in Association with Automation. *Proceedings of the 13th IFAC Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Kyoto, Japan*.
4. Mercer, J., Gomez, A., and Gabets, C. (2016, in press). Impact of Automation Support on the Conflict Resolution Task in a Human-in-the-Loop Air Traffic Control Simulation. *Proceedings of the 13th IFAC Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Kyoto, Japan*.
5. Mercer, J., Homola, J., Edwards, T., Gomez, A., Bienert, N., Martin, L., et. al. (2016, in press). Assessing Operator Capabilities in a Human-Automation Team Environment during Simulated Air Traffic Control Work. *Human Factors*.
6. Edwards, T., Sharples, S., Wilson, J. R., and Kirwan, B. (2012). The Need for a Multi-Factorial Approach to Safe Human Performance in Air Traffic Control. *Proceedings of the 4th AHFE International Conference, San Francisco: USA*.
7. Stein, E. *Air Traffic Controller Workload: An Examination of Workload Probe*. DOT/FAA/CT-TN84/24, Federal Aviation Administration, Washington, D.C., 1985.
8. Durso. F. T., Truitt, T. R., Hackworth, C., Crutchfield, J., Nikolic, D., Moertl, P., et. al. (1995). Expertise and Chess: A Pilot Study Comparing Situation Awareness Methodologies. In D. J. Garland & M. R. Endsley (Eds.), *Experimental Analysis and Measurement of Situation Awareness (pp. 295-304)*. Daytona Beach. FL: Embry-Riddle Aeronautical University Press.