Investigating the influence of autonomy controllability and observability on performance, trust, and risk perception

M.L. Cummings[1], Lixiao Huang[2,] Masahiro Ono[3]

[1]Duke University, Durham, NC, USA

[2]Arizona State University, Mesa, AZ, USA

[3]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Corresponding author: Mary Cummings, 304 Research Drive, Box 103957, Durham, NC 27708; email: m.cummings@duke.edu

Acknowledgments: The research was carried out in part at Duke University under a grant from NASA JPL, and in part at the Jet Propulsion Laboratory for a grant from the Office of Naval Research (N00014-17-IP00043).

ABSTRACT

Given the rise in autonomous vehicles in domains like space and underwater exploration, military applications, and surface transportation, human supervisors must often remotely assess risk in path navigation tasks under uncertainty. However, human risk assessment often diverges from an autonomous planner's quantitative calculation of risk, which can lead to degraded system performance. To further investigate this, an experiment was conducted to investigate the impact of controllability and observability on participants' performance, self-confidence, trust, and risk assessments in uncertain environments. Observability was expressed through a risk budget representation and controllability allowed participants to directly control the time horizon of autonomy-generated paths through different path leg lengths. Results demonstrated that observability of the risk budget did not impact performance but reduced self-confidence and trust in the autonomy when map complexity (i.e., number of obstacles) was high. Controllability of path planning algorithm leg lengths helped participants respect algorithm soft constraints, but they tended to take more risk as they approached the goal. This effort demonstrates that while risk-aware autonomy can generate optimized paths and quantify corresponding risks, designing interfaces to close the risk perception gap and enhance operator performance while promoting appropriate trust may not always have the intended effect.

Keywords: trust, risk perception, risk-aware autonomy, interface design, path planning, unmanned vehicles, decision making, risk budgeting, self-confidence, and risk evaluation

INTRODUCTION

With advances in artificial intelligence, there is increasing interest in using autonomous vehicles in domains like space and underwater exploration, military applications, and surface transportation. Such systems require human supervisors that must often remotely assess risk in the paths used to navigate the vehicles, often under uncertainty. One common source of risk for autonomous vehicle navigation is the uncertainty introduced by embedded path planning algorithms that generate paths of motion that, in theory, avoid collisions but in practice can lead to unsafe situations due to sensor error. In the context of autonomous vehicle path planning, we define the elements of risk as the source of the hazard or obstacle, an estimated likelihood of danger, and an estimated consequence of the danger (American Nuclear Society & Institute of Electrical and Electronics Engineers, 1983). Given that successfully navigating a clear path is a critical element of autonomous vehicle mission success, it is likely that how an autonomous planner computes such a path could directly influence trust and risk perception of operators overseeing such vehicles.

Risk perception and trust in an autonomous system influence an operator's reliance on that system (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003; Moray, Inagaki, & Itoh, 2000), and ultimately overall joint human-system performance. It has been proposed that risk-aware autonomy, which is an autonomous system that provides optimized solution options to human supervisors based on its risk probability calculations, could help humans better identify risks and develop mitigation strategies (Ono, Williams, & Blackmore, 2013).

To better understand what and how different variables influence a person's perception of risk and trust, especially in the presence of autonomy that is providing quantitative risk estimates to a remote supervisor, a notional influence diagram was developed to illustrate such relationships (Figure 1). The influence diagram in Figure 1 is grouped into four categories of variables (red = environment, yellow = autonomy, blue = operator, green = human-machine interaction), with overlap in the variables as illustrated, and is based on trust and risk research in supervisory settings (Clare, Cummings, & Repenning, 2015; Heitmeyer & Leonard, 2015; Hoff & Bashir, 2015; Kahneman & Tversky, 1979; Lee &

See, 2004; Mittu, Sofge, Wagner, & Lawless, 2016; Woods, 2016). The various elements feed into either the autonomy's ability to calculate risk or the human's assessment of risk. The influence diagram also includes the three layers of trust, including dispositional trust (marked as Human Trust in Figure 1), situational trust, and dynamically learned trust (Hoff & Bashir, 2015; Lee & See, 2004). Dispositional trust is independent of context and reflects a human's tendency to trust (or distrust) a system. Situational trust depends on the evaluation of the immediate situation and environment and is influenced by factors such as expertise and self-confidence. Lastly, we include learned trust, which is based on knowledge and experience of system past performance and can be dynamic, however all three types can evolve over time (Hoff & Bashir, 2015).

Figure 1 illustrates that human risk perception and trust have dramatically many more influences than for an autonomous system's risk computation, and a "risk interpretation gap" emerges when human risk perception and autonomy risk assessment do not align. This can be made worse when both do not accurately incorporate uncertainty into their computations, which means neither the human's nor the autonomy's risk assessment matches the real world. The diagram in Figure 1 illustrates that human development of risk perception, possibly mediated by trust, has very little, if any, overlap with how automation computes a quantitative risk assessment. Typically, human operators do not have any ability to influence how autonomy makes risk calculations in the operation of currently fielded systems. Thus, there is no way for operators to proactively align their risk perception with the risk assessment of the autonomy, and instead can only attempt to understand autonomy risk assessments reactively by remotely observing the performance of the vehicle in the presence of dynamic but uncontrollable environmental influences.

In this effort, we wanted to develop a more direct link between autonomy risk computation and human risk assessment, and we propose that this can be accomplished by allowing the human the ability to set, and thus control, the planning time horizon for an autonomous path planner, depicted by the dashed line in Figure 1. The time horizon, how far out in time the path planner considers in its computation of an optimal path, is directly linked to its solution quality (Murillo, Sánchez, Genzelis, & Giovanini, 2018).

Longer paths carry higher uncertainty, which human operators may find frustrating and at odds with their desire to plan a path from start to finish. However, the ability to control how far out in time an autonomous planner generates a path may help to close the risk perception gap since an operator would be both influencing the automation's computation or risk assessment as well as their own dynamic learned trust, per Figure 1, which influences human risk perception.

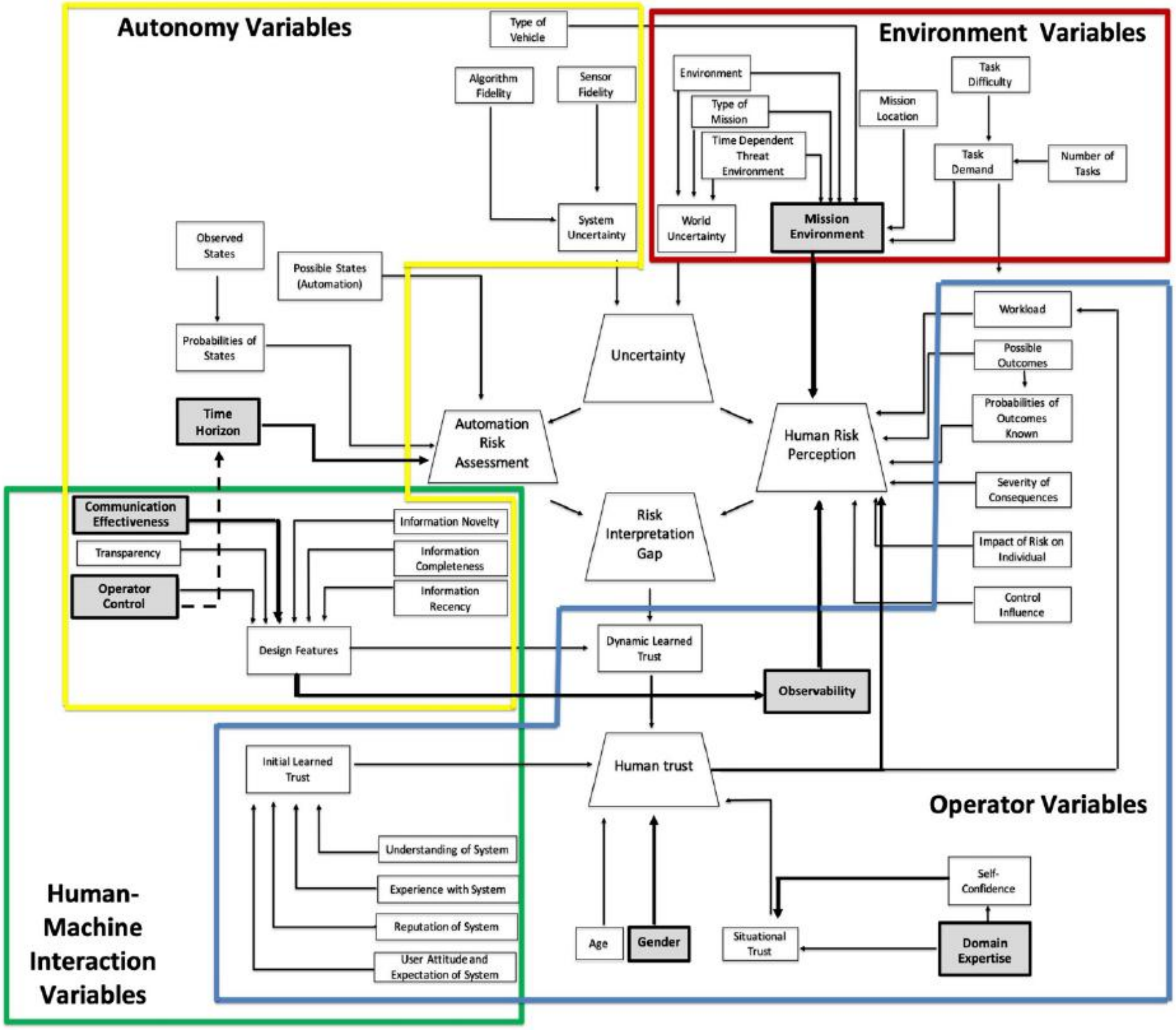


Figure 1. The influences of human and autonomy risk perception in a supervisory control system (red = environment, yellow = autonomy, blue = operator, green = human-machine interaction). Shaded boxes and bolded black lines indicate those variables under investigation in this study. The dashed line indicates a futuristic capability.

Indeed, this ability to control the time horizon could be seen as a way to bridge the gulf of evaluation, i.e., how a decision support system provides interpretable representations that can be assessed in terms of user expectations and intentions, in order to change the state of the world through bridging the gulf of execution, key elements of Norman's seven stages of actions (Norman, 1988). As shown in Figure 1, the ability to control the time horizon not only influences the autonomous planner, but since this is also a design feature in the human-machine interface, controlling the time horizon also could affect observability, the ability of the human operator to see risk information necessary information and relationships in the system they are supervising. Previous research has shown leveraging an effective control system helps bridge the gulf of execution which is the difference between a user's intent and whether and how a decision support system supports those actions (Leotti, Iyengar, & Ochsner, 2010). However, gambling research has shown a positive relationship between the perception of control over outcomes and risk-taking behavior (Nordgren, Van Der Pligt, & Van Harreveld, 2007), so such control could lead to risk-seeking behaviors. To this end, this study will examine whether the ability to control the autonomous planner in its actual risk computations leads to improved system performance, with a reduction in the risk interpretation gap without promoting risk seeking behavior.

In addition to examining the influence of human versus autonomous path planner time horizon control in this study, we also wanted to examine how the explicit representation of the amount of risk taken during path planning affected performance and operator trust. In a recent formulation of risk-aware autonomy, risk can be seen as a limited resource with a budget, with the operator potentially deciding how and when to take more or less risk (Ono, 2012). For example, an operator could be allowed to take a quantitative measure of risk in a system's operation, such as up to 10% risk of hitting an obstacle in a mission but decides when and where to take such risk that accumulates throughout a mission.

It is not clear how the concept of a risk budget could influence people's risk perception and regulation, especially in a path navigation scenario. While risk perception is a key component of trust (Hardin, 2006), humans often struggle to understand probabilistic estimates of risk (Tversky & Kahneman, 1974). Moreover, risk communication is often interpreted as vague (Spiegelhalter, 2017), and even how risk should be presented, numerically or verbally, is often debated (e.g., Erev & Cohen, 1990). Therefore, it is important to understand whether the concept of a risk budget is beneficial for human understanding and overall system performance, including the impact on trust and risk perception. In risk-aware autonomy, providing insight into automated risk computation through a risk budget could aid in the bridge of evaluation, and in Figure 1 this is represented as communication effectiveness, which is ultimately expressed as a design feature that promotes observability.

The impact of risk budget observability and controllability of the planning time horizon on risk perception and humans' trust in autonomy is not well understood. So, as illustrated in Figure 1, our focus for this study was to determine how two design features, operator control of the autonomous planner time horizon and a risk budget representation, influenced operators' risk perception, as well as trust, self-confidence, and ultimately system performance. Our hypothesis was that both design features would reduce the risk perception gap, which ultimately would lead to improved joint human-autonomous system performance, with improved trust and self-confidence. While all the variables in Figure 1 could not be tested in a single experiment, we elected to also explore other variables in this experiment, including two in the environment. These will be discussed further in the following sections.

1. THE HUMAN–AUTONOMY INTERFACE FOR EXPLORATION OF RISKS (HAIER)

This study leveraged the software platform *Human–Autonomy Interface for Exploration of Risks (HAIER)* to examine human interactions with a risk-aware, human-cooperative autonomous path planner. In HAIER, participants guide a single unmanned vehicle from start to finish across a field of obstacles. HAIER allows human supervisors to view autonomy-generated potential paths and adjust them according to their level of perceived risk. Although HAIER can be used to represent any generic unmanned vehicle, in this version, the unmanned vehicle was considered to be an unmanned underwater vehicle (UUV) that can occasionally surface to communicate with a hypothetical satellite to update its position, and thus locally resolve uncertainty about its position. The autonomous planner is considered "risk-aware" since it can propose the shortest path between its current position and the goal position while accounting for a user-specified corresponding risk level (low, medium, or high). High-risk paths result in paths that often pass very close to obstacles.

In HAIER, it is assumed that when a vehicle “surfaces,” it resolves all ambiguity about its position by communicating with satellites, which it cannot do underwater. However, the longer a vehicle remains underwater, the more uncertainty grows as to where the vehicle will surface (which is also true in the real world). Thus, while it is possible to go from the start to goal in a single leg, the uncertainty would be extremely high, and thus the vehicle needs to surface periodically to resolve any error. However, given that surfacing requires extra fuel and exposes UUVs to danger in this version of HAIER, the number of surfacing events should be minimized. Thus, participants have to balance the risk of hitting an obstacle against the risk of not completing a mission by surfacing too frequently. In this version of HAIER, operators are given a “surfacing budget” so they know how often they can surface for a specific mission. Each path between surfacing events is referred to as a leg. For this application of HAIER, the surfacing budget was always set to 6. This constraint is soft in that operators can surface more if needed, but they put the vehicle at much greater risk if they do so.

For this effort, two specific design features were tested in HAIER including observability in the form of a risk budget bar and controllability, which represents operator control of the underlying path planner time horizon. As discussed previously, allowing operators the ability to control the time horizon allows the human to directly influence the autonomy’s solution quality, while the risk budget representation gives operators a sense of how much risk is used in each leg of travel. These variables represent the intersection of autonomy and human-machine interaction features that can influence risk perception in Figure 1.

The basic HAIER interface (Figure 2a) contains a control panel that allows operators to select their proposed path risk levels (low, medium, or high) through the vertical buttons, as seen in Figures 2a and 2b. The term “observability” refers to the presence of a risk budget bar that explicitly shows how much of budgeted risk has been used at any given time (Figures 2b and 2d). Controllability of leg length (small, medium, or large) was manipulated in the control panel through the horizontal axis in Figures 2c and 2d. Both of these features are explained in more detail in the next sections.

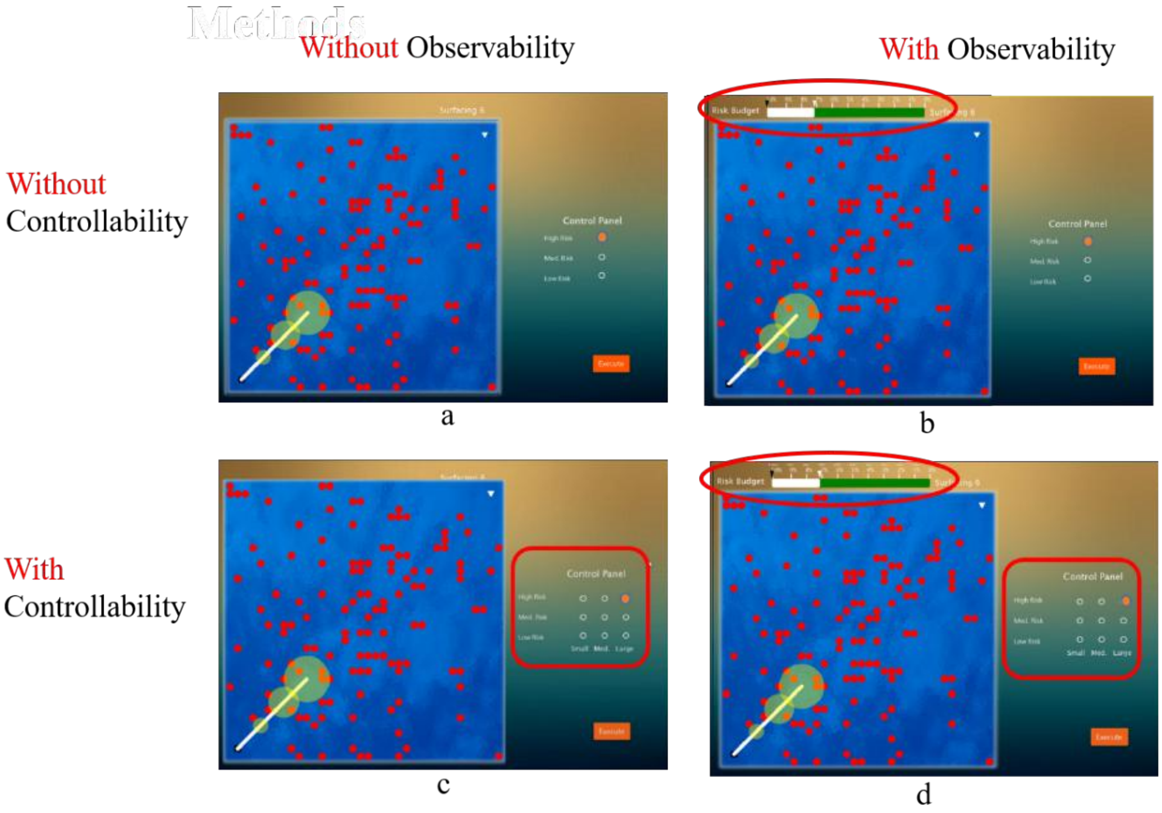


Figure 2. Four HAIER experimental interfaces. On the control panel, controllability is on the horizontal axis and risk acceptance is on the vertical axis. In 2a and 2b, operators only have access to risk acceptance levels and in 2c and 2d, the control panel allows for changes in the path leg length, known as controllability. The Observability experimental condition, i.e., the presence of a risk budget bar, is tested by comparing 2a and 2c to 2b and 2d.

The risk budget bar (upper left of Figures 2b and 2d) allowed participants to see exactly how much risk was associated with each autonomy-generated leg and how much risk would be left in a notional risk budget after executing one leg of planning. A mission was assigned an overall level of acceptable risk, i.e., the mission should not exceed more than 10% of risk from a collision. If, for example, one leg put the mission at a 2% risk of collision, then from that point forward, only 8% of overall mission risk remained. The percentage of risk computed for each leg by the autonomy was determined by the expected number of obstacles along the leg, the proximity of the obstacles to the path leg, the leg length, and the likelihood of deviation. The risk budget was also a soft constraint in that participants could exceed the allotted 10%, but the vehicle's risk of not completing the mission also grew commensurately.

Controllability allowed participants the ability to directly influence the planning time horizon through setting the leg length of the path planning algorithm, which in this version of HAIER was a Dijkstra planner. As seen in Figures 2c and 2d, participants could set the length of each leg to be small, medium or large, which represents growing uncertainty over time. The yellow circles represent the uncertainty of UUV's location due to the unpredictable ocean current, with the small, medium, large selections mirrored by the growing yellow circles. Without human control (Figures 2a and 2b), the underlying algorithm calculated the most optimal leg length, which was determined by mathematically optimizing local obstacle proximity, the final goal, and number of surfacings remaining.

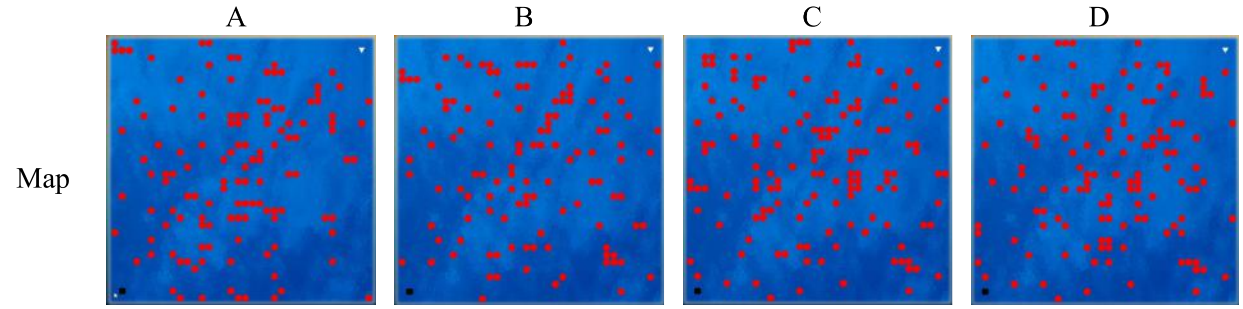
The primary research question for this effort was how the two design features of observability through the risk budget bar and controllability of leg length influenced participants' performance, trust in the planner's ability to carry out the operator's intent, self-confidence, and risk perception. As seen in Figure 2, there were four experimental conditions, which ranged from having neither observability nor controllability (Figure 2a) to those with both decision aids (Figure 2d).

2. METHOD

Seventy-five people, recruited from mailing lists and flyers from a southeastern US university, each received a $15 gift card for the one-hour experiment and an additional $100 gift card for the best performer. Each participant interacted with HAIER through a Dell desktop computer with a mouse. A 2x2 mixed factorial design was employed with Controllability as the between factor and Observability as the within factor. Participants completed two Observability test trials with two different maps (first without and then with the risk budget bar) for each controllability block. The order of Observability trials was not counterbalanced because once participants saw the With Observability interface, their impression might interfere with their understanding of the concept of risk budget.

Participants were told that their goal was to make the shortest path possible, try to surface no more than 6 times to reduce their exposure to enemy detection, and try to keep their risk budget under 10%. Training was provided prior to each testing trial by demonstrating the interface with a different training map, verbally introducing each element on the interface and performance criteria, and having participants practice path planning until finishing the training map. Participants' questions were answered by the experimenter during training. Testing trials started when participants verbally confirmed they felt ready for testing. Each training session took 5-10 minutes, and each testing trial took 5-15 minutes. Participants experienced two different maps (with or without the risk budget bar) in their testing trials.

The experiment used four maps of 36x36 grids (Figure 3) presented in a counterbalanced and randomized manner. Although the four maps were designed to be roughly equal in difficulty, given that the environment was not homogenous, map design was included as a secondary independent variable in the final statistical model, discussed in more detail in the results. In effect, these maps represent mission environmental variability, highlighted in Figure 1.



	A			B			C			D		
Obstacl e #	9	23	12	13	23	9	17	20	10	8	17	14
	14	24	7	11	22	13	19	31	14	12	26	15
	12	21	5	8	13	11	10	12	13	11	17	12
Total	127			123			146			132		
Length	35			34			36			34		

Figure 3. The four experiment maps, where Obstacle # = number of obstacles in a 3x3 matrix overlaid on a map; Total = total number of obstacles; Length = the shortest possible path length of the map in grid units.

Several dependent variables were measured, including the following:

Risk perception. Every time before participants planned a path leg, a pop-up window above the control panel asked “how risky the situation is” on a 3-point scale (1 = not risky, 2 = kind of risky, 3 = very risky). An overall risk perception mean rating was weighted by the number of legs participants elected to generate, which could range from 4-8 legs. Thus, the overall risk perception mean rating for a person who crossed a map in 4 legs would be the sum of their 4 legs risk ratings divided by 4, whereas another person who crossed in 6 legs would generate a mean based on the sum of risk ratings/6.

Self-confidence rating. While waiting for the planner to compute and display its solution based on input, a pop-up question asked participants to answer a question on a 5-point Likert scale (1 = not confident, 5 = very confident), “How confident do you feel about your [last path] planning?” Given that four to eight legs were required to reach the destination within the surfacing constraints, the overall self-confidence mean was based on the number of legs from start to the goal for each map in a similar fashion as the risk ratings variable.

Trust rating. During the planning wait time, a pop-up question asked participants to rate “How much do you trust that the autonomous planner generates the best solution?” on a 5-point scale (1 = do not trust, 5 = completely trust). Each leg of path planning involved a rating of trust, so the overall trust mean rating was also weighted by the leg number in the same manner as the overall risk perception mean rating.

Performance. Since participants were instructed to plan the shortest path possible, remain under the risk budget, and remain under the surfacing budget, three variables were used to measure a participant’s performance: *path length*, whether the *risk budget* was exceeded, and whether the *surface budget* was exceeded. Path length was the distance the UUV traveled from start to the goal, calculated through the number of grid units. Participants were told to plan the shortest paths possible without hitting any obstacles. In order to track whether people respected the risk and surface budgets, they were assessed on whether their total risk consumption and surfacing times stayed within or exceeded these respective budgets (10% and 6 times). Participants chose an acceptable risk level for each leg they planned, called leg risk.

Computer game experience. In the demographic survey, participants answered "How often do you play computer games normally?" (1 = rarely, 2 = a few times a month, 3 = once a week, 4 = a few times a week, 5 = daily). While not all of the operator-centric variables depicted in Figure 1 could be tested in a single experiment, we elected to use video game experience as a proxy for domain expertise because it has been shown to influence performance in similar experiments (Cummings, Clare, & Hart, 2010; Green & Bavelier, 2003; Lin, Wohleber, Matthews, & Funke, 2015). Thus, gaming experience was a covariate.

3. RESULTS

While the original pool of participants was 75, two participants' data were excluded due to problems with training. The remaining 73 participants (*male* = 38, *female* = 35; *age mean* = 22.8, *SD* = 7.1) resulted in 134 successful trials. Twelve trials resulted in collisions, which immediately ended a test session. These trials were analyzed separately. A MANOVA of continuous variables was conducted to assess the original research questions about controllability and observability. Then surfacing and risk budget adherence measures, also known as constraint violations, were assessed, followed by analyses of the temporal trends of the subjective measures of trust, risk perception, and self-confidence.

a. Controllability vs Observability Outcomes

Controllability (with or without) of leg length and observability (with or without the risk budget bar) were the primary independent variables. Gender was a secondary independent variable since the experiment was opportunistically roughly balanced. A 2x2x2x4 (observability x controllability x gender x map) MANOVA statistical model was used to investigate whether these factors influenced path length, as well as three subjective variables including mean subjective ratings for trust, risk perception, and self-confidence. Given the relatively high number of independent (4) and dependent variables (4), a reduced-order MANOVA model was used that represented only the main and two-interaction effects, $\alpha = 0.05$. Whether the surface and risk budget constraints were respected was investigated using logistic regression models, detailed in a subsequent section.

For the MANOVA, all normality and homogeneity assumptions were met except for mean trust, where Levene's test for homogeneity of error variance was significant (*p* = .008). This required a square root transformation which stabilized the variance. For the between-subjects comparisons across the individual dependent variables, there were no significant main effects across any of the dependent variables for controllability or observability, but there were interesting main effects for maps, gender and gaming experience, as well as a significant interaction shown in Table 1.

The two operator factors from Figure 1 included in the model were gender and video game experience. As seen in Table 1, women in general were significantly less confident than the men (M_W = 3.66, SD_W = .08, M_M = 4.02, SD_M = .09). The video game covariate was a significant factor for self-confidence, with a Pearson correlation of ρ = .262, *p* = .002, indicating self-confidence moderately increased with video game experience. Not surprisingly, the overall mean perceived risk was lower for those with more gaming experience, ρ = -.230, *p* = .008, which aligns with the influence diagram in Figure 1, where situational trust, gaming experience and self-confidence are interrelated.

Table 1. MANOVA Results

	Map	Gender	Gaming Experience	Map * Observability
Path length	*F*(3, 114) = 5.94 *p* = .001	NS	NS	NS
Mean self-confidence	NS	*F*(1,114) = 6.79 *p* = .010	*F*(1,114) = 8.51 *p* = .004	*F*(1,114) = 3.38 *p* = .021
Mean trust	NS	NS	NS	*F*(1,114) = 4.61 *p* = .004
Mean risk perception	*F*(3,114) = 2.95 *p* = .036	NS	*F*(1,114) = 6.6 *p* = .011	NS

b. *The Map Effect*

Given the four different maps (Figure 2) that represent environmental variability, it was important to control for this in the model. Least significant difference post hoc comparisons indicated that the mean path length on Map B (*M* = 45.35, *SD* = 5.11) was significantly shorter from Map A (*M* = 48.49, *SD* = 3.56, *p* = .008), Map C (*M* = 50.24, *SD* = 3.06, *p* = .000), and Map D (*M* = 48.70, *SD* = 4.90, *p* = .009).

Map A was seen as the least risky, even though it had more obstacles than Map B, with overall mean ratings statistically lower than Maps C (*p* = .008) and D (*p* = .021). All other maps were not statistically different from one another at the factor levels. However, the significant interactions between the map and observability factor for both trust and self-confidence can be attributed to Map C, which had the highest number of obstacles. For this map, people *without* observability (M_{SC} = 4.08, SD_{SC} = 0.59; M_T = 4.07, SD_T = 0.54) had significantly higher confidence and trust ratings as compared to those who had the risk bar visualization (M_{SC} = 3.35, SD_{SC} = 0.67; M_T = 3.16, $SD_{T=}$0.85), *p* = .005 and .002 respectively. When examining just the collision trials, of which there were only 12 (8% of all trials), seven (58.3%) happened on Map C, four (33.3%) on Map D, one (8.3%) on Map A, and none on Map B.

c. Constraint Violations

To examine whether controllability and observability affected a participant's ability to stay within the risk and surfacing budgets, two logistic regression analyses were conducted, also including gender, map, and computer game experience as predictors. The logistic regression analysis for risk budget adherence demonstrated that the map variable was the only significant predictor (Wald $\chi^2(1) = 7.15$, *p* = .008). Of the four maps, Map D (Figure 3) caused more risk violations than the other three maps. People using Maps A-C only violated the risk budget 5-16% of the trials compared to 65% for Map D (Figure 3). Map D had the second highest number of obstacles and they were more evenly spaced than the other maps, which could account for the high number of risk violations. While this issue deserves more research, the more consistent spacing of the obstacles in Map D may have contributed to the acceptance of greater risk.

For those exceeding the surfacing budget, a slightly different picture emerged. The controllability factor was statistically significant (Wald $\chi^2(1) = 5.56$, *p* = .018). The percentage of trials that exceeded the surfacing budget was less with the Controllability interface (33%) compared to without it (58%) (Figure 4). The map variable was also significant (Wald $\chi^2(1) = 9.09$, *p* = .003), but the difficulty across maps varied (Figure 4). For Maps A and B, roughly 33% of participants across all trials did not stay under the surfacing budget, while for Maps C and D, 72% and 55% respectively did not stay within their surfacing budget.

d. Subjective Factors

Spearman rank-order correlations between mean risk perception, mean self-confidence, and mean trust were calculated (Table 2). These ratings were generated by questions after each leg was planned, and while most people made it from the start to goal in 6 or fewer legs, 47% of the trials went to 7 legs, with just 4 trials going to 8 legs. Thus, the means represent the average ratings for each person. In addition to the risk perception metric in Table 2 which shows how risky each leg appeared to be to every participant on average, the average risk each person chose to take per leg (i.e., leg risk) is also listed in Table 2. It should be noted that only 45% of participants had the ability to control the leg length (Figures 2c and 2d).

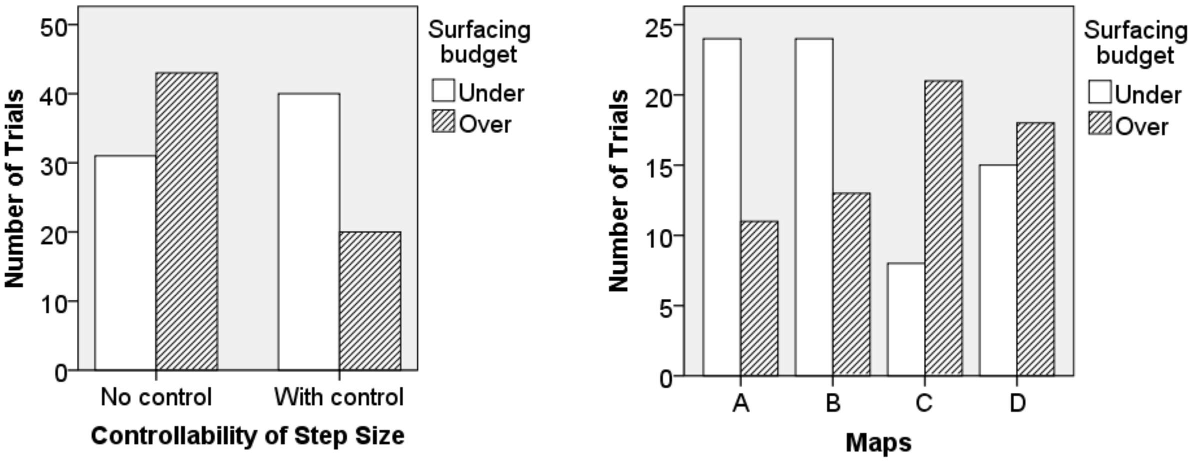


Figure 4: Impact of controllability (left) and different maps (right) on staying within the surfacing budget

Self-confidence and trust were significantly positively correlated (Table 2). As expected, risk evaluation negatively correlated with self-confidence and trust, with self-confidence as the stronger correlate. Overall as perceived risk in each scenario increased, trust and self-confidence declined. However, curiously the risk that each person chose to take for each leg in response to perceived risk, on average, did not correlate. However, there was a weak but significant correlation with the overall leg risk and trust, i.e., the higher a person's trust in the planner, the more risk taken per leg.

Table 2. Spearman rho correlation coefficients, means (M), and standard deviations (SD) for mean risk perception, leg risk, self-confidence, and trust ratings. Self-confidence and trust were measured on a 5-point scale, while risk evaluation and leg risk were measured on a 3-point scale.

	1	2	3	4	*M*	*SD*
1. Mean risk perception	1	.036 *p* = .682	-.426 *p* < .001	-.190 *p* = .028	1.65	.32
2. Mean leg risk	.036 *p* = .682	1	.103 *p* = .238	.230 *p* = .008	2.00	.33
3. Mean self-confidence	-.426 *p* < .001	.103 *p* = .238	1	.625 *p* < .001	3.85	.68
4. Mean trust	-.190 *p* = .028	.230 *p* = .008	.625 *p* < .001	1	3.72	.78

While Table 2 demonstrates the overall averages for trust, self-confidence, risk perception, and leg risk for each total path from the start to the goal, it does not indicate how these values changed over time. Figure 5 illustrates mode values for participant self-confidence, trust, risk perception and leg risk taken as a function of legs across the map. While overall trust and self-confidence remained relatively high for the majority of each mission, as participants came within 1-2 legs from the end, their self-confidence and trust rose to the highest levels. The exception is the four individuals who took 8 legs to cross. Their self-confidence and trust dropped in this last leg.

The plots of perceived risk versus leg risk taken over time in Figure 5 demonstrate how participants adapted their strategies as a function of where they were on the map. Generally participants both perceived and were willing to take moderate risk through the 4th leg. The majority of participants across both trials took six or more legs to get across each map (91%), and these participants typically took the most risk for the 6th leg (Figure 5). Once they made it past this point, most people dropped to the lowest risk level for path planning.

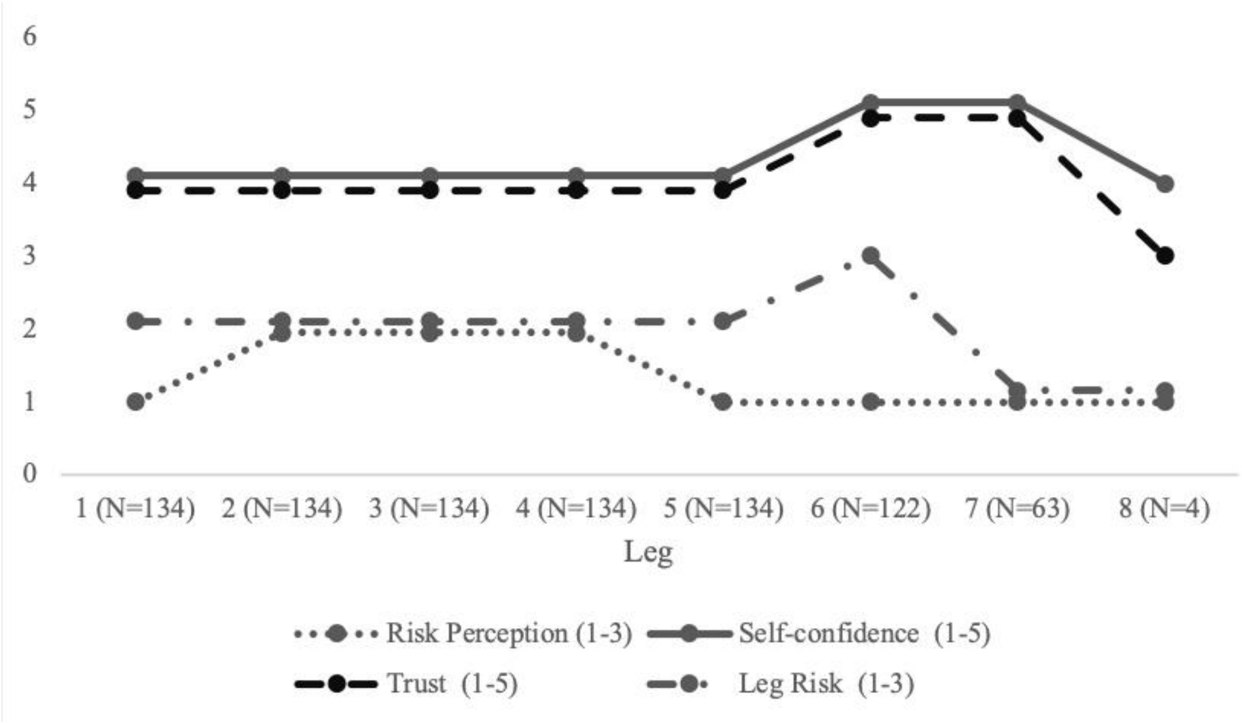


Figure 5. The mode trend of risk perception, self-confidence, trust, and leg risk. The Y axis represents question Likert ratings. Risk ratings ranged from 1 to 3, self-confidence and trust ranged from 1 to 5. The X axis represents legs of the UUV. It took 4 to 8 legs to reach the goal.

Curiously, even though participants were willing to take more risk at the point where surfacing budgets were of concern (recall the soft constraint was 6), in general at this same point, they reported their perceived risk at the lowest rating. Also, at this point self-confidence and trust ratings rose to their highest levels. Figure 5 demonstrates that at the 6th leg, participants perceived the situation risk to be low, and were willing to increase risk taking, with the highest self-confidence and the highest trust in the autonomous planner. It is also important to note that this rise happened one time step after people dropped their risk perception from moderate to low for the 5th leg.

Only 8% of the total number of trials resulted in collisions, and 66% of these trials (8 of 12) had observability and/or controllability elements. The per leg self-confidence, trust, risk perception, and risk execution mode ratings for these limited number of trials are plotted in Figure 6. With significantly fewer number of observations as compared to Figure 5, it is worth noting that the patterns between those who collided and those who succeeded were different. Most participants collided after the third leg, which generally would occur somewhere in the middle of the screen. These people tended to exhibit either high self-confidence or less trust in the autonomous planner than did those who succeeded. For example, the one individual who almost made it to the end before colliding with an obstacle took more risk earlier, which then led to increased trust for the autonomous planner for two more legs. But interestingly, this person’s trust dropped just prior to colliding with an obstacle.

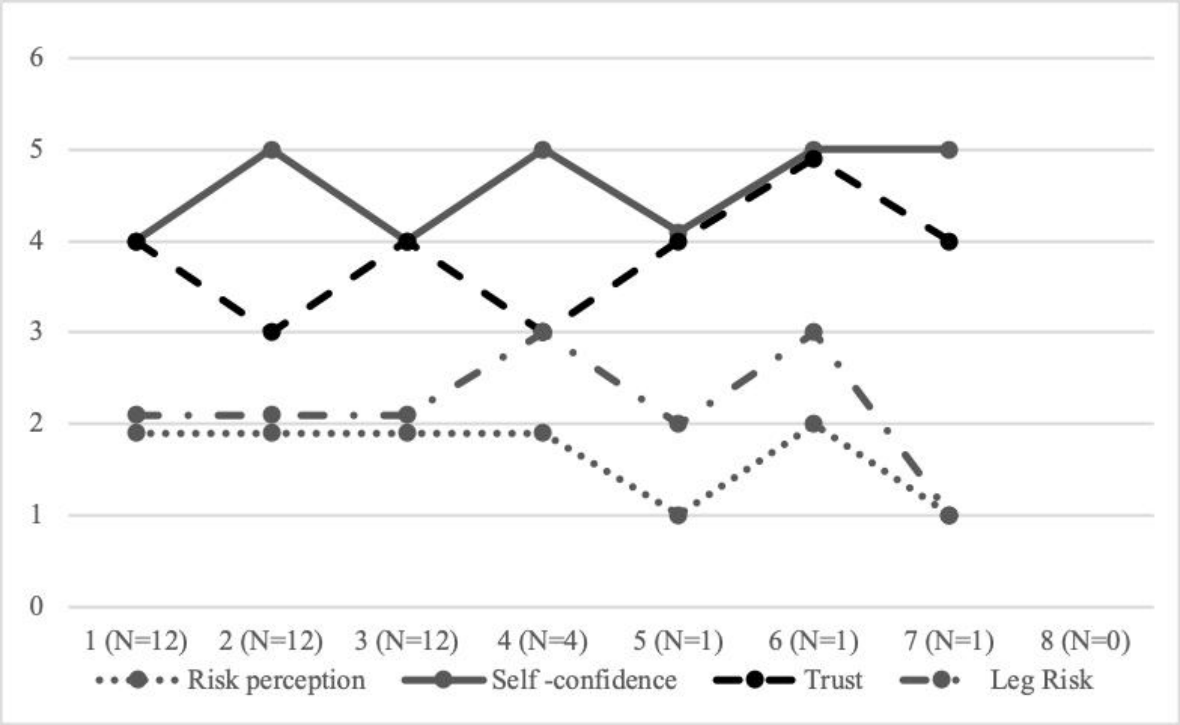


Figure 6. For the collision trials, the mode trend of risk perception, self-confidence, trust, and leg risk. The Y axis represents question Likert ratings. Risk ratings ranged from 1 to 3, self-confidence and trust ranged from 1 to 5. The X axis represents legs of the UUV. It took 4 to 8 legs to reach the goal.

4. DISCUSSION

The goals of this study were to determine whether providing explicit observability through a risk budget representation and providing the ability to control the path planning time horizon through setting the path leg length could significantly improve joint human-autonomous system performance as well as affect subjective ratings. More specifically, the study examined whether such aids could produce more efficient paths, as well as influence subjective perceptions of risk, trust, and self-confidence.

In terms of the two design features and their effect on performance, as well as the other influencing variables in Figure 1, the most influential variable was the environmental variable of mission complexity, represented by the map. It was the only variable that mattered for path planning length, as well as adhering to the risk budget. Controlling the time horizon did aid people in respecting their surfacing budget constraint so this design feature did help operators close the gulf of execution for this element of the mission. However, this tool did not measurably influence any of the subjective metrics.

As has been shown in other supervisory control studies (Cummings & Tsonis, 2006; Kirwan, Scaife, & Kennedy, 2001; Majumdar & Ochieng, 2002), the complexity of the environment can be the primary driver of operator performance, and in this study, produced the largest performance effects. Curiously, the map environment variable was not a significant predictor of trust or self-confidence, but it was for risk perception. This indicates that participants used map spatial elements to determine a vehicle's risk profile, but the map variability did not influence their sense of trust in the planner or their views of their own capabilities, which were relatively stable as seen in Figure 5. One issue to consider is that HAIER showed other concrete visual risk cues, such as red obstacles and yellow probability circles when generating paths on the map. So, it is possible that such concrete visual cues on a map are effective at communicating risk information, but trust and self-confidence are more nuanced and multi-dimensional, so are then harder to measure.

While the risk budget bar did not demonstrate any direct performance improvements, its observability influenced self-confidence and trust through *reduced* ratings when using more complex maps (i.e., more obstacle-rich). The reduced self-confidence and trust ratings suggest that participants with the risk budget bar were more cautious than those without this aid, indicating that the inclusion of such observability tools may be beneficial in cases where user overreliance is either likely or potentially dangerous.

While these results align with the influence diagram in Figure 1 that suggests that the risk budget bar (i.e., communication effectiveness) may influence dynamically learned trust, the diagram did not adequately represent any associations with self-confidence, which is related to situation trust. Moreover, it should be noted that the concept of a risk budget is relatively new with limited literature (Ono, 2012).

Further research is needed on whether people understood the risk budget bar as a probability of collision as opposed to other types of risks, and how it influences decision making under increased uncertainty.

One important corollary result in this study is that while previous work has demonstrated a positive correlation between trust in autonomy, self-confidence, and performance (Lee & Moray, 1994), trust and self-confidence were not significantly correlated with performance in this experiment, but were highly correlated with one another. While trust and self-confidence are theoretically different, this study suggests that such differences may be very difficult to disaggregate in practice.

Perhaps the most interesting result of this study was the increase in risk taking as both trust and self-confidence increased, and perception of environmental risk decreased, i.e., fewer obstacles. As seen in Figure 5, participants were willing to take more risk both after the planner appeared to be trustworthy, but also when they were close to reaching their goal. It is important to note that these events took place in the upper corners of the maps in Figure 3. In these quadrants, obstacle density was lower than in the middle of the maps (although not the lowest), and the uncertainty in whether participants would reach the goal was lower at this point. So while increasing risky behavior towards the end of the mission was generally an effective strategy, it could backfire as evidenced in Figure 6.

This propensity to increase one's acceptance of risk as the goal gets closer is a phenomenon known as the Perseveration Syndrome and is well known in aviation as get-home-itis, which can and has lead to fatal accidents (Dehais, Tessier, Christophe, & Reuzeau, 2010). Military and civilian aviation agencies mitigate this phenomenon through training and crew resource management techniques, but this is the first effort to document such an effect in unmanned vehicle remote operation. More research is needed to better identify the occurrence of such a bias and how to mitigate it in unmanned vehicle control.

One limitation of the study is that because there were no consistent correlations between any individual map and objective or subjective measures, it is likely that individual differences played a significant role in the perception of map difficulty. Follow-up research is needed with a larger group of subjects to elucidate the nature and impact of map complexity. Other important limitations of this effort include the small number of participants and a potential mismatch between display design and operator expections. The autonomous planner displayed quantitative risk usage, (e.g., 10%), whereas operators could only input their risk tolerance across three qualitative categories of low, medium and high, and it is not clear that there was clear alignment between these two pieces of information. Current work is underway to allow operators the ability to enter risk on an interval scale so that it is aligned with the risk-aware autonomy, but this mixed use of formats may have influenced results.

5. CONCLUSION

As autonomous planners proliferate in settings where humans supervise such systems, such as military, space, and air traffic control applications which can experience significant uncertainty, it is critical that designers better understand how and to what extent humans should collaborate in real-time with such planners. Risk is an inherent component of all autonomous systems, and so promoting human-autonomous system collaboration through risk-aware autonomy could aid in joint human-system performance.

A risk perception gap influence diagram was proposed that links autonomous planner characteristics with those in the environment as well as for an associated human machine interface and representative operator attributes. These four clusters of variables influence risk perception and facets of human trust, which can lead to a gap in the way people perceive autonomy and the calculations of risk generated by an autonomous planner. Such a risk perception gap could lead to inferior system performance.

To assess two different design features and their interaction with a path-planning algorithm typically used in autonomous systems on system performance and subjective operator attributes, an experiment was conducted in the software platform, *HAIER*. This study demonstrated that the concept of observability into the autonomy via a risk budget bar did not affect performance but did reduce people's trust and self-confidence under environmental complexity. Moreover, allowing people to control a path planning time horizon allowed people to respect soft constraints but did not significantly influence their subjective ratings of risk, self-confidence or trust.

Most strikingly, participants altered their perceptions of risk, self-confidence and trust as they reached the latter one third of the mission and were willing to take more risk as the goal was closer. The acceptance of increased risk as the goal gets closer is well established in manned aviation, but these results demonstrate that it also exists in unmanned settings as well. More work is needed to examine whether this adjustment of risk perception, which is a widening of the risk perception gap in Figure 1, may be more pronounced given the remote-control aspect of such operations and what design interventions could prevent this.

Further work is also needed to better understand people's interpretation of the risk budget bar, as well as whether they fully understood and intended to take more risk as the mission came close to the goal or whether they indeed were willing to take more risk to finish. Additional work is needed to determine if and how operators could benefit from having more precise control of risk budget consumption than in the current HAIER version. For example, instead of simply responding to the risk per leg that a planner generates for a proposed action, would it be more helpful for an operator to specify an exact level of risk per leg that the operator is willing to take? Lastly, this research demonstrated that the proximity and density of obstacles on a map may not be directly correlated with performance and so better understanding those variables that influence operator decisions for complex maps would be particularly beneficial.

REFERENCES

American Nuclear Society, & Institute of Electrical and Electronics Engineers. (1983). *PRA procedures guide: A guide to the performance of probabilistic risk assessments for nuclear power plants* (NUREG/CR-2300). Washington DC: US Nuclear Regulatory Commission

Clare, A., Cummings, M. L., & Repenning, N. (2015). Influencing Trust for Human-Automation Collaborative Scheduling of Multiple Unmanned Vehicles. *Human Factors, 57*(7), 1208-1218.

Cummings, M. L., Clare, A., & Hart, C. (2010). The Role of Human-Automation Consensus in Multiple Unmanned Vehicle Scheduling. *Human Factors, 52*(1), 17-27.

Cummings, M. L., & Tsonis, C. (2006). Partitioning Complexity in Air Traffic Management Tasks. *International Journal of Aviation Psychology, 16*(3), 277-296.

Dehais, F., Tessier, C., Christophe, L., & Reuzeau, F. (2010). The Perseveration Syndrome in the Pilot's Activity: Guidelines and Cognitive Countermeasures In P. Palanque, J. Vanderdonckt, & M. Winckler (Eds.), *Human Error, Safety and Systems Development. Lecture Notes in Computer Science* (Vol. 5962, pp. 68–80). Berlin: Springer-Verlag

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal Human-Computer Studies 58*, 697-718.

Erev, I., & Cohen, B. L. (1990). Verbal versus Numerical Probabilities: Efficiency, Biases, and the Preference Paradox *Organizational Behavior and Human Decision Processes, 45*, 1-18.

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423*, 534-537.

Hardin, R. (2006). *Trust*. Cambridge, UK: Polity Press.

Heitmeyer, C. L., & Leonard, E. I. (2015). *Obtaining trust in autonomous systems: tools for formal model synthesis and validation.* Paper presented at the Proceedings of the Third FME Workshop on Formal Methods in Software Engineering, Florence, Italy.

Hoff, K., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors, 57*(3), 407–434.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica, 47*(2), 263-292. Retrieved from http://links.jstor.org/sici?sici=0012-9682%28197903%3A2%3C263%3APTAAOD%3E2.0.CO%3B2-3

Kirwan, B., Scaife, R., & Kennedy, R. (2001). Investigating complexity factors in UK Air Traffic Managament. *Human Factors and Aerospace Safety, 1*(2), 125-144.

Lee, J., & Moray, N. (1994). Trust, self confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies, 40*, 153-184.

Lee, J. D., & See, K. A. (2004). Trust in technology: Designing for Appropriate Reliance. *Human Factors, 46*(1), 50-80.

Leotti, L. A., Iyengar, S. S., & Ochsner, K. N. (2010). Born to choose: The origins and value of the need for control. *Trends in Cognitive Sciences,, 14*(10), 457–463.

Lin, J., Wohleber, R., Matthews, G., & Funke, G. J. (2015). *Video Game Experience and Gender as Predictors of Performance and Stress During Supervisory Control of Multiple Unmanned Aerial Vehicles.* Paper presented at the 59th International Annual Meeting of the Human Factors and Ergonomics Society, Los Angeles, CA.

Majumdar, A., & Ochieng, W. Y. (2002). *The Factors Affecting Air Traffic Controller Workload: A Multivariate Analysis Based Upon Simulation Modeling of Controller Workload.* Paper presented at the 81st Annual Meeting of the Transportation Research Board, Washington, DC.

Mittu, R., Sofge, D., Wagner, A., & Lawless, W. F. (Eds.). (2016). *Robust Intelligence and Trust in Autonomous Systems*. New York: Springer.

Moray, N., Inagaki, T., & Itoh, M. (2000). Adaptive Automation, Trust, and Self-Confidence in Fault Management of Time-Critical Tasks. *Journal of Experimental Psychology: Applied, 6*(1), 44-58.

Murillo, M., Sánchez, G., Genzelis, L., & Giovanini, L. (2018). A Real-Time Path-Planning Algorithm based on Receding Horizon Techniques. *Journal of Intelligent and Robotic Systems, 91*(3-4), 445-457.

Nordgren, L. F., Van Der Pligt, J., & Van Harreveld, F. (2007). Unpacking perceived control in risk perception: The mediating role of anticipated regret. *Journal of Behavioral Decision Making, 20*(5), 533-544.

Norman, D. A. (1988). The Psychopathology of Everyday Things. In *The Design of Everyday Things* (pp. 257). New York: Doubleday.

Ono, M. (2012). *Robust, goal-directed plan execution with bounded risk.* (Doctorate). Massachusetts Institute of Technology, Cambridge, MA.

Ono, M., Williams, B., & Blackmore, L. (2013). Probabilistic Planning for Continuous Dynamic Systems under Bounded Risk. *Journal of Artificial Intelligence Research, 46*, 511-577.

Spiegelhalter, D. (2017). Risk and uncertainty communication. *Annual Review of Statistics and Its Application, 4*, 31-60.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124-1131.

Woods, D. D. (2016). The Risks of Autonomy: Doyle's Catch. *Journal of Cognitive Engineering and Decision Making, 10*(2), 131-133.

Authors' Bios:

Mary L. Cummings received her Ph.D. in Systems Engineering from the University of Virginia in 2004. She is currently a Professor at the Duke University Department of Duke Electrical and Computer Engineering Department. She is the director of the Duke Humans and Autonomy Laboratory.

Lixiao Huang received her Ph.D. in Human Factors and Applied Cognition from North Carolina State University Department of Psychology in 2016. She conducted this study during her position as a postdoctoral associate at the Duke University Humans and Autonomy Laboratory. She is currently an assistant research scientist at the Center for Human, Artificial Intelligence, and Robot Teaming at Arizona State University.

Masahiro Ono received his Ph.D. in Aeronautics and Astronautics from Massachusetts Institute of Technology in 2012. He is currently a research technologist in the Robotic Controls and Estimation Group at NASA Jet Propulsion Laboratory.