

How Shared Visual Attention Patterns of Pairs Unfold Over Time when Workload Changes

Shannon Patricia Devlin
University of Virginia
Charlottesville, Virginia, USA
spd7t@virginia.edu

Jake Ryan Flynn
Clemson University
Clemson, South Carolina, USA
jakef@g.clemson.edu

Sara Lu Riggs
University of Virginia
Charlottesville, Virginia, USA
sriggs@virginia.edu

ABSTRACT

Data rich environments rely on operator collaboration to manage workload changes. This work explores the relationship between operators' visual attention and collaborative performance during these workload changes. Percent gaze overlap and percent recurrence were calculated over time for best and worst performing pairs of participants who experienced low and high workload in an unmanned aerial vehicle command and control testbed. It was found that the best performing pairs had higher values for both metrics after workload changed. These results suggest successful collaborative performance is dependent on both continuous high levels of synchronized visual attention and coordinated sequences of visual attention. This work has the potential to inform the design of real-time technology.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; • **Applied computing** → *Psychology*.

KEYWORDS

eye tracking, collaboration, cross-recurrence quantitative analysis, workload transitions

ACM Reference Format:

Shannon Patricia Devlin, Jake Ryan Flynn, and Sara Lu Riggs. 2020. How Shared Visual Attention Patterns of Pairs Unfold Over Time when Workload Changes. In *Symposium on Eye Tracking Research and Applications (ETRA '20 Short Papers)*, June 2–5, 2020, Stuttgart, Germany. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3379156.3391339>

1 INTRODUCTION

Dynamic and data-rich domains rely on multiple operators to coordinate to complete multiple tasks. Operators may have difficulty working together to manage these tasks, especially when their demands, i.e. workload, change (e.g. unmanned aerial vehicle (UAV) command and control; [Williams 2006]). Of interest is identifying whether quantitative measures can be used to assess collaboration in real-time. Eye tracking technology is a promising means to do so given: (a) the majority of the information in these complex domains

include various visual displays and (b) visual attention has been informative on aspects of the operator, like cognitive load [Coral 2016] and task completion process (e.g., [Chierichetti et al. 2012]). However, more work is needed to understand whether eye tracking measures can capture aspects related to successful collaboration.

For this study, we assessed how well two collaborative visual attention measures—percent gaze overlap and percent recurrence—are indicators of performance. The novelty of this work is calculating these two measures to shed light on how visual attention changes over time. This study had 10 pairs of participants complete scenarios in a UAV command and control testbed at two different workload levels. We compared the performance and calculated visual attention metrics between the best and worst performing pairs. This work adds to the limited knowledgebase on examining eye tracking metrics over time with pairs. It has the potential to inform how eye tracking can be used in real-time to indicate when workload changes are affecting collaboration.

2 BACKGROUND

Studying the visual attention of multiple people in a shared environment has been completed in various contexts [Böckler et al. 2012; Brennan et al. 2008; D'Angelo and Begel 2017; Tien et al. 2012]. Research has shown that visual attention is influenced by the presence of another individual, but the impact depends on the context. For example, one commonly used metric to characterize pairs' visual attention is *percent gaze overlap*, which quantifies the amount of time multiple observers are concurrently viewing the same area [Pietinen et al. 2010]. Increases in percent gaze overlap have been shown with improved performance [Devlin et al. 2019; Gergle and Clark 2011; Hajari et al. 2016], but this is not always the case (e.g. joint visual search; [Brennan et al. 2008; Villamor and Rodrigo 2018]).

As a result, other measures may be more informative about collaboration success. For example, *cross-recurrence quantitative analysis* (CRQA) is growing in popularity in applied eye tracking research as it can detect temporal patterns. CRQA has been defined as, "the quantification of the patterns of co-visitation taking place" [Coco and Dale 2014]. CRQA analyzes two time series of separate entities in a dynamic system to quantify how the two entities visit states over time and the relation of those visits to one another [Marwan et al. 2007]. In the context of eye tracking, these states are areas of interest (AOIs). CRQA analysis includes several metrics, but the most common consists of calculating *percent recurrence*, i.e., how often two separate time series are visiting the same states in the same order over time.

For example, if two participants are collaborating, two time series in the system may consist of the scanpaths generated by each

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ETRA '20 Short Papers, June 2–5, 2020, Stuttgart, Germany

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-7134-6/20/06...\$15.00

<https://doi.org/10.1145/3379156.3391339>

Table 1: Hypothetical scanpath of two participants viewing a display with four AOIs

	5 s Time Window				
	1s	2s	3s	4s	5s
Participant 1	1	3	4	2	1
Participant 2	3	4	2	3	1
Participant 2' (1s lag)		3	4	2	3
Match between 1 & 2?	No	No	No	No	Yes
Match between 1&2'?	n/a	Yes	Yes	Yes	No

participant. Scanpaths are based on the order in which AOIs are visited and percent recurrence is calculated by taking the number of instances when the AOIs across the two time series match divided by the total number of time periods considered. Table 1 provides an example of two participants' theoretical scanpaths (i.e., time series) over a 5 s window, assuming a sampling rate of 1 s. The pair has 20% recurrence because there is one match out of five. Percent recurrence, however, is typically calculated using different lag times [Coco and Dale 2014]. Using the same example, if we shift participant 2's scanpath by 1 s, (i.e., one time period to the right), and compare this scanpath (called participant 2') to participant 1's scanpath. Here there is 75% recurrence (3 out of 4 matches). This example analysis shows two collaborators may have coordinated scan patterns differently than perfect synchronization, which percent gaze overlap would not capture. Previous research has shown that depending on the context, large [Dale et al. 2011; Hajari et al. 2016; Richardson and Dale 2005] and small [Coco et al. 2018] percent recurrence are associated with better performance outcomes. There has been limited work examining the relationship between pair's visual attention and performance when workload changes. The aforementioned metrics have the potential to capture how visual attention changes over time across pairs of collaborators. This research builds on previous work that examined whether shared visual attention affected the collaboration strategies of successful and unsuccessful pairs managing a workload change [Devlin et al. 2019]. The chosen application for this work is UAV command and control, as the Department of Defense looks to rely on quantitative measures to better understand the collaboration process [Resick et al. 2010] and to inform technology design to facilitate collaboration in real-time [Fiore and Wiltshire 2016; Fussell et al. 2000; Winnefeld and Kendall 2013]. Eye tracking has been specifically referred to potentially assist in this goal [Sibley et al. 2015].

3 METHOD

3.1 Participants

Ten pairs of undergraduate students (20 students total) from Clemson University were recruited for the study ($M = 21.3$ years, $SE = .24$ years). The experiment lasted 75-90 minutes and participants were compensated \$10/hour for their time.

3.2 Experimental Setup

The testbed was developed using Unity and based on the interfaces the U.S. Air Force uses to train for multi-UAV missions [Feitshans et al. 2008]; Figure 1). Pairs were collocated, but each participant

viewed separate monitors (28 inches, 2560×1440 screen resolution) and used separate mice to input responses. Although participants could view each other's screen, they likely did not as: (a) they were asked to focus on their own screen to ensure eye tracking data was properly collected and (b) the simulation was networked so the pair could see inputs from each other in real-time (e.g., when participant 1 rerouted a UAV, participant 2 could see his/her route choice in real-time). Pairs could communicate verbally, but this data is outside the scope of this study (see [Devlin and Riggs 2018] for more details). Two desktop mounted FOVIO eye trackers with a sampling rate of 60 Hz were used to collect point of gaze data, with one eye tracker placed below each monitor. Participants sat 66-71 cm from the monitor and the average degree of error for this eye tracker is 0.78° ($SD = 0.59^\circ$; [EyeTracking, Inc. 2011]).

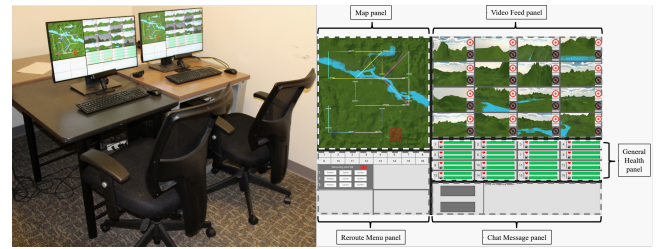


Figure 1: The experimental setup with the testbed shown on two networked computers. This is the same setup as [Devlin et al. 2019].

3.3 Tasks

Each pair was responsible for completing a primary task and three secondary tasks for up to 16 UAVs. Although all tasks were the pair's responsibility, only one participant from each pair had to complete each task. The primary task was the target detection task where pairs monitored UAV video feeds and indicated whether a target—i.e., a semi-transparent cube—was present. The secondary tasks included a rerouting task (avoiding the no-fly zone), fuel leak task (maintaining UAV health), and chat message task (responding to chat messages). These tasks and their structure emulate the multitasking, dynamic, and interdependent environment of UAV command and control [Feitshans et al. 2008; Sibley et al. 2015].

3.4 Workload

Workload was manipulated by varying the number of UAVs the pair needed to manage for the primary task. There were two workload conditions: low and high. For the low workload condition, the pair was responsible for 3-5 UAVs at all times and for the high workload condition they were responsible for 13-16 UAVs at all times. Pairs always completed the low workload condition before the high workload condition given the Department of Defense's aim to increase overall task load in command and control environments. Each condition was its own UAV testbed scenario.

3.5 Procedure

Participants of each pair read and signed the consent form and were then briefed about the study's goals and expectations. Participants

then independently completed a five-minute training session where they had to demonstrate they could achieve 70% accuracy for all tasks. The pairs were then informed on how the simulation was networked and were then provided three minutes to discuss anything they deemed necessary. Afterwards, pairs completed the low workload condition, were provided a short break, and then completed the high workload condition. Each scenario was 15 minutes in length. After, participants completed a debriefing questionnaire and were compensated for their time.

3.6 Experimental Design

The primary independent variables in this study were pair performance (best performing vs. worst performing). The best performing pairs were the top three highest scoring pairs and the worst performing pairs were the bottom three scoring pairs, where the score was the combined point total from the low and high workload scenario. Dependent variables included percent gaze overlap on the Video Feed panel (Figure 1) and maximum percent recurrence for each minute of both scenarios.

4 RESULTS

The gaze data were screened to meet data quality requirements as outlined in ISO/TS 15007-2:2014, which states that at most 15% data loss is acceptable for good quality data. Following this guideline, no participants were excluded from the study and the mean data loss was 9.23%. Previous work indicated best performing pairs had higher levels of percent gaze overlap on the area that was the source of the workload change [Devlin et al. 2019]. Therefore, only percent gaze overlap on the Video Feed panel was calculated for each minute in both the low and high workload scenario because this was the panel associated with the primary task and was the means in which workload was manipulated.

The best performing pairs' range of mean percent gaze overlap across both low (range = 17.4-33.0%) and high workload (range = 40.0-63.7%) was higher than the worst performing pairs' (low workload range = 1.0-7.3%, high workload range = 9.9-22.2%). Figure 2 shows the change in percent gaze overlap from low to high workload averaged for each minute of the scenario. The best performing pairs increased their percent gaze overlap from low to high workload throughout the scenario more than worst performing pairs, except for the last minute. Percent recurrence was also calculated to determine whether there were other shared visual attention patterns between the pair. Specifically, maximum percent recurrence was of interest as this would indicate when the pairs' visual attention sequences were most coordinated. Percent recurrence was then calculated for each lag, which was based on the eye tracker refresh rate, present in a ± 10 s window per recommendations by previous work [Dale et al. 2011]. Then the max percent recurrence value was found for each minute of the low and high workload scenario and then averaged for the best and worst performing pairs (Figure 3 and 4, respectively). Compared to worst performing pairs, the mean max percent recurrence was consistently higher for best performing pairs for the entirety of low and high workload (except for the last minute of low workload). Similar to percent gaze overlap, the best performing pairs max percent recurrence also increased from low (range = 22.2-37.4%) to high workload (range =

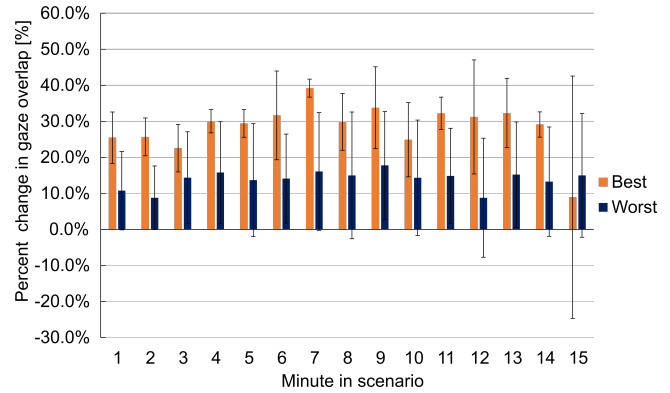


Figure 2: Mean change in percent gaze overlap over time between the low and high workload scenario for best and worst performing pairs. Error bars represent standard error of the mean

49.9-74.6%) whereas worst performing pairs did not increase at the same rate (low workload range = 3.9-36.8%; high workload range = 14.0-40.8%). This suggests the best performing pairs viewed AOIs in a similar pattern. Across all pairs, percent recurrence increased as the scenario progressed.

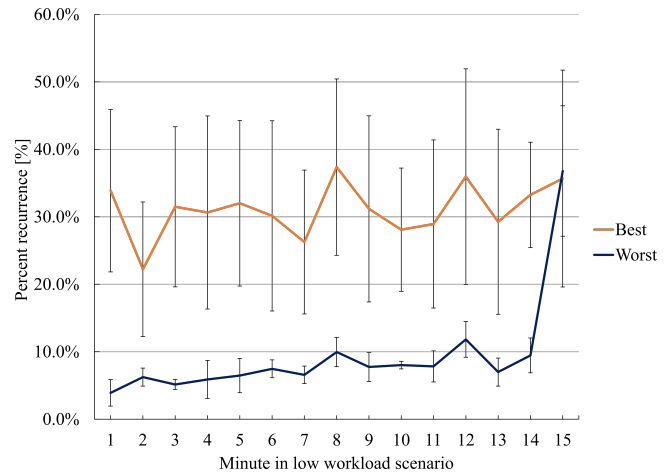


Figure 3: Mean max percent recurrence over time for best and worst performing pairs in low workload scenario. Error bars represent standard error of the mean

For every percent recurrence value, there was an associated lag time. In this work, the lag time indicates when visual attention patterns were most similar during the ± 10 s window: negative lag times are associated with leading behavior whereas positive lag times are associated with lagging behavior. For this study, the lag time variation can be captured by counting the amount of times the sign of the lag time changed (i.e., to and from a negative and positive value) over the 15 min scenarios. We found no trend in these lag times as their sign seemed to change frequently over the course of each scenario for both pair types. For reference, a lag

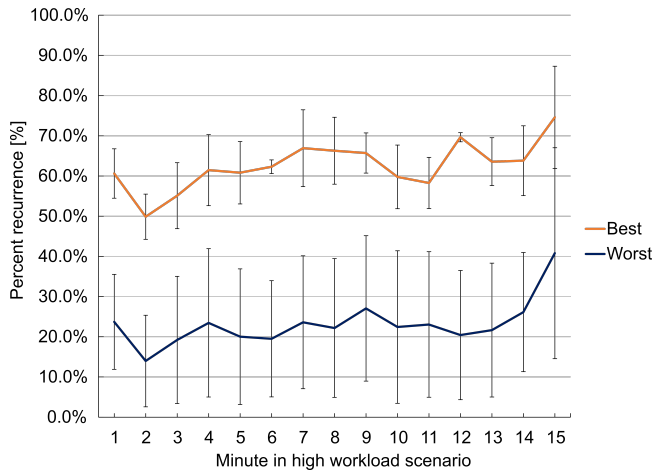


Figure 4: Mean max percent recurrence over time for best and worst performing pairs in high workload scenario. Error bars represent standard error of the mean

time could change sign a minimum of zero times (i.e., be the same sign the entire scenario) or a maximum of 14 times (i.e., switch every minute during the scenario). During low workload, the sign of the lag time changed, on average, 8.3 times for best performing pairs and 7.3 times for worst performing pairs. For high workload, the sign of the lag time changed, on average, 4.7 times for best performing pairs and 7 times for worst performing pairs. Although the number of switches is relatively large for both performance groups, the best performing pairs' average number of switches decreased by 44% when workload increased whereas the number of switches remained relatively the same for the worst performing pairs. This may suggest the best performing pairs began to establish shared visual attention allocation patterns, which in turn may have led to better performance.

5 DISCUSSION

This work aimed to understand how shared visual attention differed with performance when pairs experienced a workload change from low to high with a more granular and sensitive eye tracking analysis. The results support there is a positive relationship between gaze overlap and performance over time. This is expected from previous work that studies collaborations in a multitasking, dynamic, coordinated environment [Devlin et al. 2019; Gergle and Clark 2011; Hajari et al. 2016]. It further suggests better performance is associated with participants engaging in substantial increases in their shared visual attention earlier and sustaining those levels over time, even when they could divide tasks and responsibilities. This further underscores the importance of technology to include common information spaces for collaborators, regardless of task interdependence, so they can adapt to environmental changes quickly and effectively [Fiore and Wiltshire 2016]. This finding has the potential to directly inform technology design for environments experiencing varying workload levels, as it could rely on this metric to indicate when shared visual attention levels are ideal or need intervention.

Maximum percent recurrence was larger for best performing pairs, on average, compared to worst performing pairs, which is consistent with the percent gaze overlap results and is also similar to previous work [Coco et al. 2018; Hajari et al. 2016; Richardson and Dale 2005]. This suggests improved performance is not only dependent on “viewing the same location at the same time”, but also the sequence in which viewing is shared. These results further suggest poor performance is associated with a lack of coordinated visual attention as the worst performers consistently had much lower percent recurrence, even when various lag times were considered. However, this may not hold true for all cases, as [Coco et al. 2018] contends recurrence is dependent on environmental factors like task paradigm and the presence of feedback. Nevertheless, this work suggests percent recurrence of shared visual attention could also be used as a quantifiable indicator of successful collaborative performance in real-time as workload changes.

Our results also found the lag times of when percent recurrence was maximized were dynamic over time for both best and worst performing pairs. This may be attributed to the experimental setup as we did not define set roles for the pairs in the testbed, which is unlike previous work, but is expected to be the approach used for future missions [Sibley et al. 2015]. Incorporating lag times further explains how a pair is coordinating their visual attention and can potentially be used to inform how technology should intervene to increase coordination. We found here that a basic analysis of the lag time values associated with maximum percent recurrence may indicate when more established visual attention strategies were becoming adopted by pairs who had better performance during high workload. Although a more robust analysis is needed, this finding could be applied by designing technology that assists operators to engage in more effective collaboration when workload changes to avoid performance detriments. For example, maximum recurrence and its lag time could be continuously calculated so cues are presented to collaborators when needed and designed to assist them towards more efficient leading/lagging behavior.

6 CONCLUSION

Overall, this work suggests better performance during workload increases may be attributed to shared visual attention that is large in magnitude and consistent over time. The findings support the potential of technology to rely on these metrics to inform and improve collaboration. Future research needs to further explore how to effectively use this information and take into account other potential environmental features, e.g. the impact of seeing a partner's gaze in real-time [Schneider et al. 2018]. Nevertheless, this work shows the value of including quantitative and real-time measures in dynamic domains to better understand differences in collaboration success, which in turn can inform technology to effectively assist operators with changing workloads in real-time [Winnefeld and Kendall 2013].

ACKNOWLEDGMENTS

This study was supported in part by the National Science Foundation (NSF grant: #1750850, Program Manager: Dr. Andrew Kerne). The authors would like to thank Aakash Bhagat and Jawad Alami for developing and refining the testbed used in this study.

REFERENCES

- Anne Böckler, Günther Knoblich, and Natalie Sebanz. 2012. Effects of a coactor's focus of attention on task performance. *Journal of Experimental Psychology: Human Perception and Performance* 38, 6 (2012), 1404.
- Susan E. Brennan, Xin Chen, Christopher A. Dickinson, Mark B. Neider, and Gregory J. Zelinsky. 2008. Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition* 106, 3 (2008), 1465–1477.
- Flavio Chierichetti, Ravi Kumar, Prabhakar Raghavan, and Tamas Sarlos. 2012. Are web users really Markovian?. In *Proceedings of the 21st International Conference on World Wide Web*, 609–618.
- Moreno I Coco and Rick Dale. 2014. Cross-recurrence quantification analysis of categorical and continuous time series: An R package. *Frontiers in Psychology* 5 (2014), 510.
- Moreno I Coco, Rick Dale, and Frank Keller. 2018. Performance in a collaborative search task: The role of feedback and alignment. *Topics in Cognitive Science* 10, 1 (2018), 55–79.
- Melissa Patricia Coral. 2016. *Analyzing cognitive workload through eye-related measurements: A meta-analysis*. Master's thesis. Wright State University.
- Rick Dale, Anne S Warlaumont, and Daniel C Richardson. 2011. Nominal cross recurrence as a generalized lag sequential analysis for behavioral streams. *International Journal of Bifurcation and Chaos* 21, 04 (2011), 1153–1161.
- Sarah D'Angelo and Andrew Begel. 2017. Improving communication between pair programmers using shared gaze awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 6245–6290.
- Shannon Patricia Devlin, Jake Ryan Flynn, and Sara Lu Riggs. 2019. Examining the visual attention of pairs of operators during a low to high workload change. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 63. SAGE Publications, 201–205.
- Shannon Patricia Devlin and Sara Lu Riggs. 2018. The effect of video game experience and the ability to handle workload and workload transitions. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 62. SAGE Publications, 736–740.
- EyeTracking, Inc. 2011. *Hardware: Eye tracking systems*. EyeTracking, Inc. <http://www.eyetracking.com/Hardware/Eye-Tracker-List>
- Gregory Feitshans, Allen Rowe, Jason Davis, Michael Holland, and Lee Berger. 2008. Vigilant spirit control station (VSCS): The face of COUNTER. In *AIAA Guidance, Navigation and Control Conference and Exhibit*, 1–22.
- Stephen M. Fiore and Travis J. Wiltshire. 2016. Technology as teammate: Examining the role of external cognition in support of team cognitive processes. *Frontiers in Psychology* 7 (2016), 1531.
- Susan R Fussell, Robert E Kraut, and Jane Siegel. 2000. Coordination of communication: Effects of shared visual context on collaborative work. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, 21–30.
- Darren Gergle and Alan T Clark. 2011. See what I'm saying? Using dyadic mobile eye tracking to study collaborative reference. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, 435–444.
- Nasim Hajari, Irene Cheng, Bin Zheng, and Anup Basu. 2016. Determining team cognition from delay analysis using cross recurrence plot. In *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 3482–3485.
- Norbert Marwan, M Carmen Romano, Marco Thiel, and Jürgen Kurths. 2007. Recurrence plots for the analysis of complex systems. *Physics Reports* 438, 5-6 (2007), 237–329.
- Sami Pietinen, Roman Bednarik, and Markku Tukiainen. 2010. Shared visual attention in collaborative programming: A descriptive analysis. In *Proceedings of the 2010 ICSE Workshop on Cooperative and Human Aspects of Software Engineering*, 21–24.
- Christian J Resick, Toshio Murase, Wendy L Bedwell, Elizabeth Sanz, Miliani Jiménez, and Leslie A DeChurch. 2010. Mental model metrics and team adaptability: A multi-facet multi-method examination. *Group Dynamics: Theory, Research, and Practice* 14, 4 (2010), 332.
- Daniel C Richardson and Rick Dale. 2005. Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science* 29, 6 (2005), 1045–1060.
- Bertrand Schneider, Kshitij Sharma, Sebastien Cuendet, Guillaume Zufferey, Pierre Dillenbourg, and Roy Pea. 2018. Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups. *International Journal of Computer-Supported Collaborative Learning* 13, 3 (2018), 241–261.
- Ciara Sibley, Joseph Coyne, and Jeffrey Morrison. 2015. Research considerations for managing future unmanned systems. In *2015 AAAI Spring Symposium*, 454–459.
- Geoffrey Tien, M Stella Atkins, and Bin Zheng. 2012. Measuring gaze overlap on videos between multiple observers. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, 309–312.
- Maureen Villamor and Ma. Mercedes Rodrigo. 2018. Predicting successful collaboration in a pair programming eye tracking experiment. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*, 263–268.
- Kevin W Williams. 2006. *Human factors implications of unmanned aircraft accidents: Flight-control problems*. Technical Report. FAA Civil Aerospace Medical Institute.
- James A. Winnfeld and Frank Kendall. 2013. *Unmanned systems integrated roadmap FY 2013-2038*. Technical Report. Office of the Secretary of Defense.