Correlation Between Gaze and Hovers During Decision-Making Interaction

Pierre Weill-Tessier
School of Computing and Communications,
Lancaster University, UK
p.weill-tessier@lancaster.ac.uk

Hans Gellersen
School of Computing and Communications,
Lancaster University, UK
h.gellersen@lancaster.ac.uk

ABSTRACT

Taps only consist of a small part of the manual input when interacting with touch-enabled surfaces. Indeed, how the hand behaves in the hovering space is informative of what the user intends to do. In this article, we present a data collection related to hand and eye motion. We tailored a kiosk-like system to record participants' gaze and hand movements. We specifically designed a memory game to detect the decision-making process users may face. Our data collection comprises of 177 trials from 71 participants. Based on a hand movement classification, we extracted 16588 hovers. We study the gaze behaviour during hovers, and we found out that the distance between gaze and hand depends on the target's location on the screen. We also showed how indecision can be deducted from this distance.

CCS CONCEPTS

• **Human-centered computing** → *HCI theory, concepts and models*;

KEYWORDS

Hand-Eye Coordination, Touch Devices, Gaze behaviour

ACM Reference Format:
Pierre Weill-Tessier and Hans Gellersen. 2018. Correlation Between Gaze and Hovers During Decision-Making Interaction. In *ETRA '18: 2018 Symposium on Eye Tracking Research and Applications, June 14–17, 2018, Warsaw, Poland.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3204493.3204567

1 INTRODUCTION

Exploration of touch and gaze correlation shows that gaze precedes touch [Weill-Tessier et al. 2016]. However, to our knowledge, other parts of the hand movement involved in the tapping process are not yet studied. We focus on the gaze behaviour during the preparation of the taps: the hand's stationary position (*hover*). Both gaze and hand accompany the human cognitive process, in memory retrieval in particular [Johansson and Johansson 2014; Tempel and Frings 2016]. So understanding how gaze and hand behave before a tap can provide the machine indications on the user's cognitive process and anticipate the adequate following steps.

ETRA '18, June 14–17, 2018, Warsaw, Poland

ACM ISBN 978-1-4503-5706-7/18/06...$15.00
https://doi.org/10.1145/3204493.3204567

We treat hovers as an indicator of selection consideration or choice. We are interested to know where, during hover, the hand is located according to the location on the display the user is gazing at. We also investigate how this relationship between gaze and hand can inform on the indecision the user experiences while selecting targets. In this article, we consider "indecision" as the cognitive state of not being able to make a clear choice, without serious impact to the user or her actions (whereas "indecisiveness", described in [Patalano et al. 2010], indicates a state where the user experiences "decision delay, worry and regret").

In this article, we present a collection of gaze and hand positions while playing "Memory game". The choice of this game, as explained later, has been particularly made to study how gaze behaves while users hover at the tablet. Besides it offers a way to generate and focus on the cognitive process of the users. We then describe the spatial relationship between gaze and hand during hovers for the different parts of the screen, and explain how this relationship changes when the user faces indecision.

2 RELATED WORK

2.1 Hesitation detection based on hand gesture

Observations associated with Fitts' law related studies informed that during continuous target selection, the hand realises "dwell times" between two consecutive movements [Fitts and Radford 1966]. Meyer et al. [1990] highlighted the role of hesitation in hand dwell time. Interpreting human hesitation has been studied in Human-Robot interaction. Moon et al. [2011] investigated hesitation characteristics in a conflict targeting inter-human collaborative activity, to later implement this behaviour to a robot. They modelised one type of hesitation (retract) based on acceleration to evaluate non verbal communication with robots. Their study, however, does not focus on more than one target, and they showed that their model could not work with the "pause" type of hesitation.

Nevertheless, their work is also used in HCI, as explained by Vodlan et al. [2015], who made a clear explanation on how Social Signals Processing can be used for intelligent HCI (HCI^2), and in particular how gestures can indicate human hesitation to a machine. They classified {hesitation | no hesitation} based on user observation, and proposed a logistic regression model relying on the most significant observed features [Vodlan and Košir 2015].

Time indicators between stimulation and response can show hesitation and therefore be used by machines too [Mu et al. 2010; Vodlan et al. 2015]. Our work adds-up with this research trend by proposing a method to evaluate hesitation based on different input channels (gaze and touch).

2.2 Decision-making study activities

The decision-making process during computer target selection involves both the human system (cognition), and the context of the stimuli on the machine [Zachary and Ryder 1997]. So decision-making activities are found in several fields, depending on which part of the process the research in focusing on. Psychology and medicine research focus on the human system, such as the architecture of a decision-making process [Calderon et al. 2015] or the impact of ADHD on decision making [Coghill et al. 2014]. Eye tracking is also used to assess the level of indecisiveness of individuals in [Lufimpu-Luviya et al. 2013], where the choice of different alternatives in a given context is presented to the participants.

Centred on computing, decision-making studies either evaluate ways to assess or avoid indecisiveness, or explain what triggers it. For instance, detecting frustration is presented in [Alabdulkarim 2014] using sensors to detect typical hand features. Gonzalez [1996] investigated the role of animation in user interfaces in decision-making. The impact of the stimuli and potential distractors on the hand movement are analysed in [Chapman et al. 2010; Song and Nakayama 2009]. Our work spans through both worlds as we describe the gaze and hand behaviour that characterises decision-making.

Figure 1: Data collection apparatus in a public space.

3 DATA COLLECTION

We designed a system to collect and analyse data, to understand the gaze and hand correlation during the hovering part of the target selection process on a touch device. We paid attention to propose an application that would require the participants to make decisions.

3.1 Content

Our data collection includes: 1) the eyeballs position and gaze samples provided by the eye tracker, 2) the hands position provided by Leap Motion, 3) the tap samples from our tracking application based on the Microsoft Raw Input API and 4) the game event information logs (i.e. when a tile has been flipped, when a pair has been matched).

3.2 Context

For the context of our data collection, we implemented a "memory game": 12 shuffled pairs of pictures, shown face down, the player has to match by flipping them at touch. This choice was driven by our interest in understanding the users' decision-making process, while maintaining a joyful and motivating user experience. This game meets these two criteria, and solely relies on memory. Besides, it consists of a very simple interface. Having the same interface across participants, as well as limiting the scope of actions (tap to flip a tile, match a pair) help with framing a clear reference for further data exploration. The tiles (304×304 pixels) were arranged in 6 columns by 4 rows. When a pair was found, it did not flip back and remained in the game.

3.3 Apparatus

The game was played on a Microsoft Surface Pro 4 (screen dimensions 260.28×173.52 mm, 1824×1216 pixels). The eyeballs and gaze position was collected using Tobii EyeX sampling at 60 Hz. The hands position was collected using Leap Motion running at approximately 110 Hz. We designed a 3D-printed support to hold the tablet and both sensors in place. The support laid on a table (90 cm height) and has been conceived so that each sensor can track without interfering with each other (infra-red emissions) and so that their respective fields of view cover the targeted body parts during playing. We verified the data quality of the sensors working together in a pilot study including 6 participants. We asked them to perform only one game, using their dominant hand index. We later asserted the data coherence by observing a replay of the estimated gaze points and index positions in a representation of the tablet display. We designed a C# application to manage the sensors and retrieve their logs, as well as to launch the game. Each sensor's API provides timestamps that we synchronised with the system clock via the manager application. Figure 1 illustrates the apparatus deployed in a public space. Participants typically stood about 66 cm away from the tablet centre.

3.4 Protocol

Before playing the game, the participants filled a consent form, and we assessed their hand laterality and their dominant eye (triangle test[1]). The participants were then introduced to the game with a demonstration version (3×2 abstract figures in larger pictures).

A 5-point eye tracker calibration was performed before the data collection (accuracy of $0.73° \pm 1.9$). Three increasing difficulty levels[2] were played. The participants' hands movements were also video-recorded. To finish, a 5-point accuracy test was run ($0.79° \pm 4$).

3.5 Participants

In total, 117 participants played the game (49 female, age 26±8.6). Most of them were right-handed (103) and their right eye was dominant (83). After discarding the trials with poor data collection (either from Leap Motion or the eye tracker) we kept 177 trials across 71 participants.

4 DATA ANALYSIS

4.1 Eye movements classification

In a post hoc step, we extracted the fixations from the gaze data by running a dispersion algorithm. The temporal threshold used in the algorithm was 100 ms for all participants. However, the spatial threshold we set for the dispersion detection varied among them.

[1] http://www.allaboutvision.com/resources/dominant-eye-test.htm (April 2018)
[2] Level 1 showed pictures of various objects or landscape easily recognisable from each other. Level 2 only included pictures of trees in different landscapes. Level 3 only contained pictures of close-up sea surfaces.

Table 1: Classification of the hovers

Leading to tap {L}	Hovered tile {LT}	(6192, 37.33%)
	Outbound {LO}	(2871, 17.31%)
Not leading to tap {NL}	Hovered tile {NLT}	(4764, 28.72%)
	Outbound {NLO}	(2761, 16.64%)

We computed the equivalent length on the screen of 2° of visual angle, based on the average distance, collected during the game, between the tablet and the participant.

4.2 Hand movements classification

We focused on the stationary events of the hands (*hovers*) that reflect the potential choices the participants considered. Hovers were detected post hoc using a velocity-based algorithm on the hands motion data. Since our data set included the finger tips' position, we first needed to extract the pointing finger, to work with a single point. We did so by finding the closest tip to the tablet screen. In case Leap Motion tracked both hands, we selected the closest hand to the screen. Then, we performed the velocity-based algorithm over this filtered data set. We chose a temporal threshold of 100 ms and a velocity threshold of 80 mm/s based on the pause velocity reported by Vogel and Balakrishnan [2005]. Once the hovers were detected, we classified them as described in Table 1.

The classification was based on the temporal order between the hovers and the taps (**L**/**NL**) and the projection of the pointing finger onto the tablet display. This projection was computed as the intersection of the tablet plane with the line passing by the dominant eye and the hover position (finger tip).

5 RESULTS

Our dataset contains 16588 hover samples and 46812 fixation samples. We constructed our study based on the assumption that hovers happen before taps. They may not lead to the tap straight forward: for instance, if a user hesitates to tap, the hand hovers, then moves and probably hovers again before tapping. In the following, when mentioning anything related to hover position, we will assume, if not stated otherwise, that it means the position of the *hover's projection* as described in Section 4.2.

5.1 Relationship between gaze and hover

When hovering above the tablet, we intuitively did not expect the participants' finger to be aligned with the gaze to avoid occlusion. Therefore, we wanted to understand where participants kept their hand during the data collection. We studied the median average position of the hover relative to the gaze. We wanted to understand how these values varied with the screen's part that is being looked at. In a first step, we only considered the tile position, and only kept hovers during which gaze stayed on a same tile (71.6% of all hovers). Figure 2 illustrates the median distance value between gaze and hovers for each tile. It shows that the distance increased radially from the bottom centre of the screen from 301±310 pixels to 757±605 pixels (top left) and 532±509 pixels (top right). We explain this radial distribution from the participants' tendency to keep their hand at a minimal distance from their position, certainly to prevent arm fatigue. It also indicates that the participants used a "manual

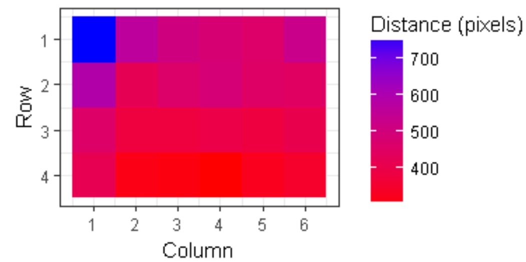


Figure 2: Relative median position between gaze and hover per tile, showing a radial distribution.

mapping" of the screen that was smaller than the actual projection of the screen at the hover depth level, and better aligned at the bottom centre of the screen.

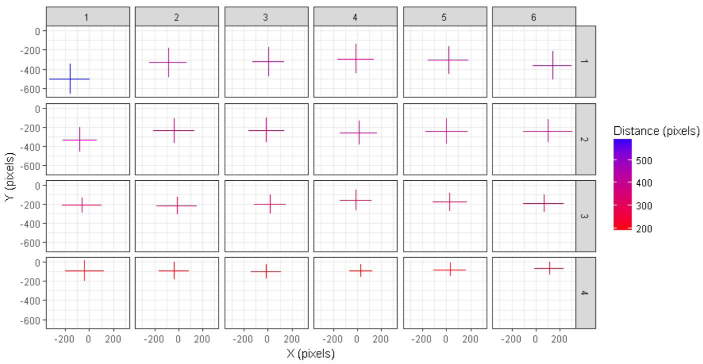
In a second step, we only focused on hovers inside the volume above the tablet's screen (**LT+NLT**, expecting them to be closer to gaze). The aforementioned distance radial distribution over the screen was observed for those hovers. We noticed that the median difference between gaze and hand positions, on the horizontal axis, increased at the edges and shifted approximately at the middle of the screen. On the vertical axis, this difference increased when the participants were looking towards the top border of the tablet. However, even if the hand was systematically below the gaze position, in the case of **LT+NLT** hovers (Figure 3a), the difference was more important towards the top corners of the screen. We interpret this as a tendency for the participants to favour horizontal hand movements over vertical hand movements when the hand was at a resting hovering position. The vertical position and distance between gaze and **LT+NLT** or **LO+NLO** hovers were significantly different for each tile of the screen (Wilcoxon rank-sum test[3], $p < 0.01$ for every tile).

We did not find a systematic pattern nor a significant difference between gaze and hovers depending if they are **L** or **NL** hovers.

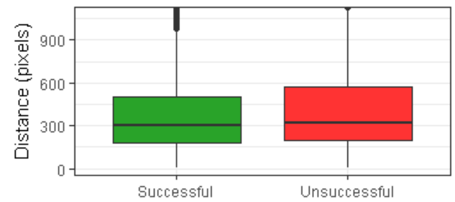
5.2 Indecision and gaze/hover relationship

We wanted to understand if gaze during hover presented characteristics that reveal how participants were confident about their choices. We evaluated indecision via the coarse approximation of pair matching failure on seen elements. We only focused on **L** hovers (because it indicated the participants were planning to tap), for tiles that had been seen before (to discard the exploratory phase of the game, when participants randomly flipped tiles to start the game) and that was the second element of the pair matching (to characterise the taps as "successful" or "unsuccessful"). We expected that, when participants were facing indecision, the duration or the number of fixations during hover was particular because they were reflecting their memory recall [Micic et al. 2010]. However, we did not observe a difference in the average number and duration of the fixations during hover leading to successful or unsuccessful taps (resp. 1.42±0.64 fixations for 252±182 ms and 1.47±0.86 fixations for 235±164 ms). Nevertheless, in the spatial domain, we observed a significant difference in the vertical position (Wilcoxon rank-sum

[3]We did not assume normality of the data, therefore we used the non-parametric Wilcoxon rank-sum test to compare groups.



(a)



(b)

Figure 3: (a) Relative median position between gaze and hover per tile for hovers inside the screen volume. (b) Median distance between gaze and hover positions, shorter for hovers leading to successful than for unsuccessful taps.

test, W = 878860, p-value < 0.05) and distance (Figure 3b, Wilcoxon rank-sum test, W = 986870, p-value < 0.05) between the gaze and hover depending on the tap success or failure. The hand's distance and vertical position were closer to the gaze point (ΔY = -265±365 pixels, distance = 383±367 pixels) for hovers that led to successful taps compared to unsuccessful taps (ΔY = -289±374 pixels, distance = 412±376 pixels).

6 DISCUSSION

Our work contributes to the understanding of the gaze/hand correlation in the context of touch devices. We retrieved hand hovers, to supplement existing work solely focused on taps [Weill-Tessier et al. 2016]. We found that the relationship between gaze and hand during the hovering stage of target selection is closely dependant on the target's location, and that users keep their hand closer to them in the vertical dimension while they preferably move in the horizontal dimension. More insights regarding the gaze/hand relationship during tapping should be obtain by exploring the role different screen sizes, orientations and target dimensions play in the visuomotor mechanism.

Integrating intelligence in machines to understand human cognitive clues is a challenge [Fischer 2001]. We aimed at finding how indecision can be inferred from the gaze and hand correlation. Approximating the decision-making cognitive states {decisive/indecisive} by the success of the tile pair matching on seen tiles, we found that contrary to our expectations, the number and duration of fixations during hover cannot reveal indecision. However, we noticed that during hover, the hand is closer to the point of gaze when the user is decisive, and that the vertical component of this distance brings this closeness. Surely, better indicators for indecision can be used to get a more accurate estimation of the users' state of mind. Nevertheless our approach enables a first coarse estimation that may serve as a basis for future intelligent systems.

As the tiles were shown facing down when they were not flipped or paired, we can assume that the tiles did not intrinsically play a role in the gaze movement: players did not search for a concrete picture to flip when they browsed the screen. Instead we can consider the gaze movements were directly related to the mental map the players were involved with [Allen 1997; Isola et al. 2011]. However, the role of the revealed paired tiles may be interesting to query, since they became spatial cues for the players to retrieve the tiles that have not yet been matched.

For our data analysis, we did not take into account personal differences despite being already acknowledged in gaze/hand correlation [Weill-Tessier et al. 2016]. Observing the participants playing, we saw that some of them did not move the hand unless for tapping on the tile, whereas some others were more likely to browse the screen with their finger. Categorisation of the participants based on their manual and visual behaviour (personal differences for indecisive vs. decisive groups were found in [Lufimpu-Luviya et al. 2013; Patalano et al. 2010]) should be taken into account towards implementing intelligent systems.

7 CONCLUSION

We have conducted a data collection that encompassed gaze and hand motion data, on a touch tablet while playing a memory game. Our objective was to understand how the hand and the eyes correlate before the taps are performed, particularly during the hovers, when the hand is in a standby position. We observed that the distance between gaze and hand depends on where the user looks at on the tablet. This distance increases radially from the bottom centre of the screen, and the distance variation between gaze and hand is more important in the horizontal axis.

We also wanted to estimate how the correlation can inform about the participants' cognitive process. We compared the gaze/hand relationship for hovers leading to successful tile pair matching with hovers leading to unsuccessful tile pair matching to approximate the participants' indecision. We found that the number and the length of fixations do not depend on the indecision, and that the distance between the finger and the eyes is larger when a decision has been taken with uncertainty.

We endeavour to explore the correlation in a more detailed approach by understanding how it differs on the personal level. We suspect the personal hand motion and/or gaze behaviour to have an impact on the correlation which can provide a finer detection of the different cognitive process stages.

REFERENCES

Arwa Alabdulkarim. 2014. Towards Hand-Gesture Frustration Detection in Interactive Systems. In *2014 3rd International Conference on User Science and Engineering (i-USEr)*. 153–157. https://doi.org/10.1109/IUSER.2014.7002694

Robert B. Allen. 1997. Mental Models and User Models. In *Handbook of Human-Computer Interaction*. Elsevier, 49–63. https://doi.org/10.1016/B978-044481862-1.50069-8

Cristian B Calderon, Tom Verguts, and Wim Gevers. 2015. Losing the Boundary: Cognition Biases Action Well After Action Selection. *Journal of Experimental Psychology-General* 144, 4 (2015), 737–743. http://dx.doi.org/10.1037/xge0000087

Craig S. Chapman, Jason P. Gallivan, Daniel K. Wood, Jennifer L. Milne, Jody C. Culham, and Melvyn A. Goodale. 2010. Reaching for the Unknown: Multiple Target Encoding and Real-Time Decision-Making in a Rapid Reach Task. *Cognition* 116, 2 (aug 2010), 168–176. https://doi.org/10.1016/j.cognition.2010.04.008

David R. Coghill, Sarah Seth, and Keith Matthews. 2014. A Comprehensive Assessment of Memory, Delay Aversion, Timing, Inhibition, Decision Making and Variability in Attention Deficit Hyperactivity Disorder: Advancing Beyond the Three-Pathway Models. *Psychological Medicine* 44, 9 (jul 2014), 1989–2001. https://doi.org/10.1017/S0033291713002547

Gerhard Fischer. 2001. User Modeling in Human–Computer Interaction. *User Modeling and User-Adapted Interaction* 11, 1 (01 Mar 2001), 65–86. https://doi.org/10.1023/A:1011145532042

Paul M. Fitts and Barbara K. Radford. 1966. Information Capacity of Discrete Motor Responses under Different Cognitive Sets. (1966), 475–482 pages. https://doi.org/10.1037/h0022970

Cleotilde Gonzalez. 1996. Does Animation in User Interfaces Improve Decision Making?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '96)*. ACM, New York, NY, USA, 27–34. https://doi.org/10.1145/238386.238396

Philip Isola, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. 2011. What Makes an Image Memorable?. In *CVPR 2011*. 145–152. https://doi.org/10.1109/CVPR.2011.5995721

Roger Johansson and Mikael Johansson. 2014. Look Here, Eye Movements Play a Functional Role in Memory Retrieval. *Psychological Science* 25, 1 (2014), 236–242. https://doi.org/10.1177/0956797613498260 arXiv:https://doi.org/10.1177/0956797613498260 PMID: 24166856.

Yannick Lufimpu-Luviya, Djamel Merad, Sebastien Paris, Véronique Drai-Zerbib, Thierry Baccino, and Bernard Fertil. 2013. A Regression-based Method for the Prediction of the Indecisiveness Degree Through Eye Movement Patterns. In *Proceedings of the 2013 Conference on Eye Tracking South Africa (ETSA '13)*. ACM, New York, NY, USA, 32–38. https://doi.org/10.1145/2509315.2509320

David E. Meyer, J. E. Keith-Smith, Sylvan Kornblum, Richard A. Abrams, and Charles E. Wright. 1990. Speed-Accuracy Tradeoffs in Aimed Movements: Toward a Theory of Rapid Voluntary Action. In *Attention and Performance 13: Motor Representation and Control*. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US, 173–226.

Dragana Micic, Howard Ehrlichman, and Rebecca Chen. 2010. Why Do We Move Our Eyes While Trying To Remember? The Relationship Between Non-Visual Gaze Patterns and Memory. *Brain and Cognition* 74, 3 (2010), 210 – 224. https://doi.org/10.1016/j.bandc.2010.07.014

AJung Moon, Chris A. C. Parker, Elizabeth A. Croft, and Hendrik F. M. Van der Loos. 2011. Did You See It Hesitate? - Empirically Grounded Design of Hesitation Trajectories for Collaborative Robots. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 1994–1999. https://doi.org/10.1109/IROS.2011.6094605

Xiangwei Mu, Yan Chen, Jian Yang, and Jingjing Jiang. 2010. An Improved Similarity Algorithm Based on Hesitation Degree for User-based Collaborative Filtering. In *Proceedings of the 5th International Conference on Advances in Computation and Intelligence (ISICA'10)*. Springer-Verlag, Berlin, Heidelberg, 261–271. http://dl.acm.org/citation.cfm?id=1926680.1926712

Andrea L. Patalano, Barbara J. Juhasz, and Joanna Dicke. 2010. The Relationship Between Indecisiveness and Eye Movement Patterns in a Decision Making Informational Search Task. *Journal of Behavioral Decision Making* 23, 4 (2010), 353–368. https://doi.org/10.1002/bdm.661

Joo-Hyun Song and Ken Nakayama. 2009. Hidden Cognitive States Revealed in Choice Reaching Tasks. *Trends in Cognitive Sciences* 13, 8 (2009), 360–366. https://doi.org/10.1016/j.tics.2009.04.009

Tobias Tempel and Christian Frings. 2016. How Motor Practice Shapes Memory: Retrieval But Not Extra Study Can Cause Forgetting. *Memory (Hove, England)* 24, 7 (aug 2016), 903–915. https://doi.org/10.1080/09658211.2015.1059858

Tomaž Vodlan and Andrej Košir. 2015. Methodology for Transformation of Behavioural Cues into Social Signals in Human-Computer Interaction. In *Handbook of Research on Interactive Information Quality in Expanding Social Network Communications*, Francisco V Cipolla-Ficarra (Ed.). IGI Global, 104–118. https://doi.org/10.4018/978-1-4666-7377-9.ch007

Tomaž Vodlan, Marko Tkalčič, and Andrej Košir. 2015. The Impact of Hesitation, a Social Signal, on a User's Quality of Experience in Multimedia Content Retrieval. *Multimedia Tools Appl.* 74, 17 (sep 2015), 6871–6896. https://doi.org/10.1007/s11042-014-1933-2

Daniel Vogel and Ravin Balakrishnan. 2005. Distant Freehand Pointing and Clicking on Very Large, High Resolution Displays. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology (UIST '05)*. ACM, New York, NY, USA, 33–42. https://doi.org/10.1145/1095034.1095041

Pierre Weill-Tessier, Jayson Turner, and Hans Gellersen. 2016. How Do You Look at What You Touch?: A Study of Touch Interaction and Gaze Correlation on Tablets. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications (ETRA '16)*. ACM, New York, NY, USA, 329–330. https://doi.org/10.1145/2857491.2888592

Wayne W. Zachary and Joan M. Ryder. 1997. Chapter 52 - Decision Support Systems: Integrating Decision Aiding And Decision Training. *Handbook of Human-Computer Interaction* (1997), 1235–1258. https://doi.org/10.1016/B978-044481862-1.50118-7