Human agency beliefs affect interaction behaviours and task performance when learning with computerised partners

Citation for published version:
Crompton, CJ & MacPherson, SE 2019, 'Human agency beliefs affect interaction behaviours and task performance when learning with computerised partners', *Computers in Human Behavior*, vol. 101, pp. 60-67. https://doi.org/10.1016/j.chb.2019.07.006

Digital Object Identifier (DOI):
10.1016/j.chb.2019.07.006

Human agency beliefs affect interaction behaviours and task performance when learning with computerised partners

Catherine J. Crompton[1,2] & Sarah E. MacPherson[1,3]

[1]*Human Cognitive Neuroscience, Department of Psychology, University of Edinburgh, UK*

[2]*School of Informatics, University of Edinburgh, Edinburgh, UK*

[3]*Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, UK*

Correspondence to:
Dr Catherine J. Crompton
Centre for Clinical Brain Sciences
University of Edinburgh
Kennedy Tower, Royal Edinburgh Hospital
Morningside Terrace
Edinburgh
EH10 5HF
Tel: +44 (0)131 537 6505
E-mail: catherine.crompton@ed.ac.uk

Abstract

Computer-mediated systems can support aging in-place, although little is known about how older adults interact with these systems and how they learn from them. Using a Wizard-of-Oz paradigm, this study compared how older adults interacted and learned with a system that they believed was a human, and with a system they believed was a computer. While both systems were identical, the human system used natural speech and the computer system used synthetic speech. In a within-subjects design, twenty-four older adults aged 60-85 years completed a collaborative learning task with both the human and computer systems. The task involved negotiating and learning referential labels for abstract tangram shapes. A learning effect was observed in both conditions. However, participants took longer to complete the task when they believed they were interacting with a computer, were less accurate in their answers, changed their answers more, and recalled them with less detail after a delay, compared to when they believed they were interacting with a human. These findings suggest that beliefs about agency affect how efficiently and how accurately older adults learn with technology, which has implications for computer mediated support in aging.

Keywords: aging; social interaction; learning; memory; collaborative learning

Introduction

With a shift in the demographic structure of the population characterised by adults living longer, technology can play a key role in supporting aging in-place (Haub & Yanagishita, 2011; Wiles, Leibing, Guberman, Reeve, & Allen, 2012, Elers et al., 2018). Developing technologies to facilitate and support independent living can be both economically and practically beneficial, and increase quality of life for older adults (White et al., 1999). However, there are a number of reasons why older adults may struggle using certain technologies (Damant & Knapp, 2015). Firstly, older adults may be less familiar with computers; as computer technology has advanced exponentially over the last 30 years, older adults are less likely to have encountered or engaged with computers in the ubiquitous way that younger adults have (Fozard & Wahl, 2012, Franklin & Myneni, 2018). Secondly, older adults may be less interested in learning how to use new technologies, and consider computers irrelevant to their everyday lives (Gatto & Tak, 2008; Wagner, Hassanein, & Head, 2010). Thirdly, age-related cognitive decline (e.g., working memory, attention) in older adults may reduce their ease of technology use (Fletcher-Watson, Crompton, Hutchison, & Lu, 2016), and cause difficulty when navigating and engaging with a system (Czaja et al., 2006; Hawthorn, 2000; Zajicek, 2004). Finally, it is common for older adults to become anxious around computers, and computer anxiety is an important predictor of computer use (Czaja et al., 2006; Yoon, Jang & Xie, 2016, Vaportzis, Giatsi Clausen & Gow, 2017).

Yet, technology has become a part of everyday life, with more reliance on smartphones, tablets, and laptops, and the increasing reach of internet connectivity has meant that more services have become computerised (Selwyn, Nemorin, Bulfin, & Johnson, 2017). Advanced computer systems are now relatively inexpensive and therefore increasingly used by people, organisations, and corporations (Caruana, Spirou, & Brock, 2017). In turn, there is a risk that older adults experience digital exclusion, which could further isolate older adults over time (Choi & DiNitto, 2013; Morris, 2007). Higher levels of computer literacy and internet use in older adults are significantly predictors of psychological well-being, reduced loneliness, and higher life satisfaction (González, Ramírez & Viadel, 2015; Heo, Chun, Lee, Lee & Kim, 2015; Gardiner, Geldenhuys & Gott, 2018).

Voice-based and spoken-dialogue systems may minimise the difficulties older adults have with computer interaction (Wolters, Georgila, Moore, Logie, et al., 2009; Wolters, Georgila, Moore, & MacPherson, 2009). These systems attempt to produce a human-computer interaction that is similar to a human-human spoken dialogue (Wolters, Georgila, Moore, & MacPherson, 2009). These systems can offer efficient and natural human-computer interactions, easy access to applications such as email, calendars, and navigation systems (Demberg, Winterboer, & Moore, 2011) and may be particularly useful for situations where the user's eyes and hands are engaged in another task such as driving or using equipment (Demberg et al., 2011; Hieronymus & Dowding, 2004; Pon-Barry, Weng, & Varges, 2006). Furthermore, voice-based interaction systems are increasingly being applied as embodied computerised social partners with the goal of reducing social isolation (Bickmore, Caruso, Clough-Gorr, & Heeren, 2005; Leite, Martinho, Pereira, & Paiva, 2009; Vardoulakis, Ring, Barry, Sidner, & Bickmore, 2012), including systems that provide companionship, play games, promote exercise and wellbeing, and facilitate connections with friends and family (Vardoulakis et al., 2012). They may also be particularly beneficial for older adults, avoiding the use of graphical user interfaces, keyboards, mice, and touchpads, which can be susceptible to age-related fine motor problems (Findlater, Froehlich, Fattal, Wobbrock, & Dastyar, 2013; Smith, Sharit, & Czaja, 1999). Voice-based systems depend on a naturalistic, conversational interaction that older adults are used to, and thus may break down the barriers and anxieties that older adults have when using computers (Wolters, Georgila, Moore, & MacPherson, 2009).

Early work investigating individuals' responses to social interactions with computers explored computers as social actors (Nass, Steuer, & Tauber, 1994). These studies found that, while people do not believe computers are human, or even human-like, their interactions with computers are fundamentally social in nature. While humans are aware that computers do not have emotions, intentions, or motivations, they continued to apply the same social rules and expectations they have for human partners to computers (Reeves & Nass, 1998). Individuals often showed polite behaviours and social reciprocities, although this was not evident in their self-reports of how they would interact with computers (Nass & Moon, 2000). This suggests that some aspects of social behaviour are not dependent on mental state beliefs regarding the interactive partner (i.e., whether they are a human or a computer), as understanding that a computer does not have emotions does not preclude people interacting with computers in a polite, social way (Branigan, 2003; Branigan, Pickering, Pearson, & McLean, 2010).

More recent work, however, has shown that while computers have an ever-present role in society, we do not have the same social expectations for computers as for humans, and we do not treat them as equal social partners (Cross, Ramsey, Liepelt, Prinz, & Hamilton, 2016). Beliefs about an interlocutor's agency and perceived humanness can affect social perception and interaction (Cross et al., 2016; Gowen, Stanley, & Miall, 2008; Klapper, Ramsey, Wigboldus, & Cross, 2014; Stanley, Gowen, & Miall, 2007; Stenzel et al., 2012; Tsai & Brass, 2007). For example, Cross et al. (2016) showed participants two sets of identical movements and told participants one set was based on human motion capture, and the other was based on computer-generated movements. Participants rated the human-generated movements as significantly smoother and more pleasant to watch than the computer-generated ones. Neuroimaging findings mirror these results; when participants believed movements were created by a human, there was greater engagement in neural regions associated with theory of mind and person perception, specifically the right inferior occipital gyrus and fusiform gyrus (Cross et al., 2016). Similarly, Liepelt and Brass (2010) found that motor priming is stronger when participants believe movements are made by a human hand than a wooden hand. Therefore, the belief that participants are engaging either with human-generated or computer-generated stimuli affects perception at both the behavioral and neural level, and pre-conceived beliefs and prior knowledge about the humanness of a partner shapes important aspects of interaction.

Differences in how agency beliefs affect interaction on a psycholinguistic level have also been found. During human conversation, people tend to verbally align, converging on similar vocabulary and syntax (Pickering & Garrod, 2004). While alignment is a largely automatic process, it can be in part mediated by beliefs about the interaction partners' knowledge and understanding (Branigan et al., 2010; Pickering & Garrod, 2004). Linguistic alignment can also occur in human-computer interactions and can be more pronounced than in human-human interactions (Branigan et al., 2010), perhaps because participants believe that computer systems are less sophisticated and have restricted capabilities compared with human interaction partners (Pearson, Hu, Branigan, Pickering, & Nass, 2006). People strategically align with computers, rather than in the automatic way they do with human partners, to ensure the computer understands their meaning and the communication is successful (Branigan et al., 2010).

Recently, research has shown that a learning context emphasising social interaction and collaboration reduces age-related episodic memory changes. Older adults performed a learning task over multiple trials with similar speed and efficiency compared with younger adults, despite having poorer episodic memory than the younger adults when assessed individually (Derksen et al., 2015). People often include others in their goals and strategies for everyday problem solving (Berg, Johnson, Meegan, & Strough, 2003; Strough & Margrett, 2002), and collaborating with others on problem-solving tasks can enhance cognitive performance in older adults (Bassuk, Glass, & Berkman, 1999; Stine-Morrow & Parisi, 2008).

If we do not treat computers as 'human' social partners, even if they behave in a human-like way, how might our agency beliefs affect our collaborative learning with computers, and how the learned information is later remembered? This question is of importance as technologies become easier and less costly to produce and purchase, and they have become increasingly embedded in our lives and the lives of older adults. In particular, if older adults believe that the system they are using does not have agency, will they interact with and learn from the system less accurately? Technology developers attempting to create social interactions may need to consider the effect of agency beliefs on user interactions, and whether they affect how users achieve their goals and learn from a system (Caruana et al., 2017).

In the current study, we investigated whether beliefs about human agency have a direct impact on how older adults learn and recall information. We address the question of whether perceived agency of a learning partner affects how efficiently and accurately older adults complete learning tasks, and their delayed recall of this information. Participants interacted with a partner while performing a collaborative learning task (i.e., the Barrier Task; Derksen et al., 2015; Duff, Hengst, Tranel, & Cohen, 2008). In one condition, they believed that they were interacting with a human, and in the other condition, they believed they were interacting with a computer. In both learning conditions, participants were in fact interacting with computer systems with synthetic and natural speech used to emulate computer and human interlocutors respectively. The task involved identifying tangram cards based on referential labels (for example, "the one that looks like the man crouching", and "the one that looks like a spiky plant") and learning the card-label pairings over multiple trials in order to arrange the cards in a specific order. Unlike previous studies investigating the impact of human agency beliefs on performance, this study investigated both how participants interacted with their partner linguistically, and how well they learned and recalled the information presented in both conditions.

Methods

Participants. Participants were 24 right-handed independently living adults aged 60-85 years (Mean = 70.46 years, SD = 7.64 years, 18 female). Participants were recruited through the University of Edinburgh Psychology volunteer research database. All participants were native English speakers, and reported having no neurological or psychiatric conditions listed in the Wechsler Adult Intelligence Scale IV selection criterion (WAIS-IV; Wechsler, 2008). Participants had a mean of 15.21 years of full-time education (SD = 3.66, range = 10-22). Ethical approval was granted by the University of Edinburgh Philosophy, Psychology and Language Sciences Ethics Committee, and all participants provided written informed consent.

Materials and procedure. The Test of Premorbid Functioning (TOPF; Wechsler, 2009) was administered to provide a measure of full scale IQ, the Rey Osterreith Complex Figure Test (Meyers, 1995) was administered to assess visuospatial memory, and the Digit Span subtest from the WAIS-IV (Wechsler, 2008) was administered to assess working memory.

The Computerised Barrier Task. To compare how participants behaved and performed when they believed they were interacting with a human and a computer, a Wizard-of-Oz (WoZ) simulation was used (Green & Wei-Haas, 1985). In this paradigm, the participant believes that they are interacting with a computer system (and also in this case, a human) that automatically responds to their input, when they are in fact interacting with a system that is being manipulated or semi-manipulated by a human operator. A WoZ paradigm creates the illusion of an intelligent computer system, as a human operator intercepts the user's input, and manipulates appropriate responses in real-time, usually through the medium of keyboard shortcuts (Green & Wei-Haas, 1985). Using a WoZ set-up allowed the creation of two directly comparable human and computer conditions where every aspect of the interaction was identical and constant, with the only manipulations being participants' beliefs about their interaction partner, and whether information was presented using natural or synthetic speech.

The natural speech used in the human condition was provided by a female speaker with a Scottish accent. The synthetic speech used in the computer condition was provided by the Festival Speech Synthesis System (http://www.cstr.ed.ac.uk/projects/festival), and a female voice ("Nina") to match the gender of the natural speech voice. The command-line utility Normalize was used to equalise the sound levels across all sound files (http://normalize.nongnu.org).

The Computerised Barrier Task was based on the Barrier Task paradigm previously used in human interaction studies without the manipulation of agency (Derksen et al., 2015; Duff et al., 2008). Participants sat with a 6 x 2 grid in front of them on a table with 12 numbered spaces, and a set of 12 cards which featured a black tangram shape on a white background. The computer system described the card to be placed in each space, starting with the first tangram and then subsequent tangrams in ascending order. After arranging all 12 cards in order, participants were instructed to move the cards to the side of their boards, and the task was repeated with the computer system describing the same cards to be arranged in a different order. Each complete re-ordering of the 12 cards was defined as one trial, and participants completed nine trials with both the human and computer partners in a single session (total = 18 trials). Different sets of tangram cards were used with the human and computer partners and these tangram card sets were counterbalanced across participants to prevent stimuli effects. The two sets of tangrams were piloted in human pairs to ensure equal difficulty before commencing the WoZ study.

The phrases used to describe each of the tangram shapes were generated in an earlier human-human Barrier Task study (Crompton, 2017). The four most commonly used phrases to describe each card were accessible as descriptors, with the least common of the four presented initially. If participants could not recognise the card based on the description provided, they could ask for a different description, and the second-least commonly used phrase was then provided. This procedure continued until participants placed a card in the space or all four descriptions were presented. Participants were instructed not to miss out a card, and that they would not be able to go back and switch cards once they had been placed. In subsequent trials, participants were provided with the description that they finally matched the card with (regardless of whether it was correct or incorrect), and previously unsuccessful descriptions were not subsequently provided.

At the start of the session, participants were told they would be completing a card sorting task twice, once with a research assistant named Kirsty, and once with a voice-based computer system. Before completing the task with Kirsty, participants were told that she was in an adjacent lab and they would be communicating with one another using headphones and microphones. Before completing the task with the computer system, participants were told that they would be interacting with a voice-based computer system that would recognise what they said and reply verbally, and that they should communicate using the microphone and speakers. In reality, participants were interacting with the WoZ system run by the experimenter in an adjacent room in both conditions. Participants were told that they would not be timed and they should treat the task as a game.

The Computerised Barrier Task yields two dependent variables relating to interaction with the system; time taken to complete the task and number of interactive turns taken, aligning with previous Barrier Task research (Derksen et al., 2015). Time taken was defined as the amount of time in seconds it takes for a participant to complete a trial, and a turn is defined as the end of a participant's utterance, delineated by an utterance from the computer system, or the end of a computer utterance delineated by a human one, indicating the amount of back-and-forth interaction between interlocutors. In line with human Barrier Task research, in order to minimise inter-trial noise, the nine trials were collapsed into three trial bins, each representing three consecutive trials. Bin 1 included trials 1-3, bin 2 included trials 4-6, and bin 3 included trials 7-9 (Derksen et al., 2015; Crompton, 2017).

Accuracy was measured as the total number of cards in the correct grid location at the end of each trial and was averaged across each trial bin, with a higher score indicating greater accuracy. A fluctuation score was calculated to measure how often participants changed their mind across trials. The score was the sum of the absolute difference between each participant's score on each trial (i.e., trial 2 – trial 1 + trial 3 – trial 2 + … trial 9 – trial 8). The lower the fluctuation score, the greater the consistency in performance across trials.

Participants completed both the human and computer conditions of the Computerised Barrier Task (half performed the computer condition first and half performed the human condition first) before completing a short battery of neuropsychological assessments. After one hour, participants were presented with a surprise delayed recall task where they had to provide the description that they had matched with each of the tangram cards in the earlier task. Participants were presented with each tangram and asked to provide the description associated with it. Participants received one point for each referential label correctly provided. Finally, participants were debriefed about the true nature of the study and were asked whether they believed they were interacting with Kirsty, a research assistant in the next room.

Data analysis. All data were analysed using R (R Core Team, 2018). To investigate the effects of perceived agency on interaction and the effect of trial number on interaction, data were analysed using a linear mixed model approach using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). Condition (human versus computer, treated as a factor) and Trial Bin (1, 2 and 3, treated as an ordered factor) were included as fixed effects, and were entered as main effects and then as interactions during forward stepwise model comparisons,. All models included the random effects structure of (1+Condition|Participant), and models were standardised using the 'standardize' function in the arm package version (Gelman & Su, 2018). The threshold for statistical significance was $|t| < 2$. Effect sizes were calculated using Cliff's Delta, a non-parametric effect size measure which ranges between -1 and 1 (Macbeth, Razumiejczyk & Ledesma, 2011). For positive integers, $d < 0.33$ indicated small effect sizes, $d < 0.47$ indicated medium effect sizes and $d > 0.47$ indicated large effect sizes. For negative integers, $d > -0.33$ indicated small effect sizes, $d > -0.47$ indicated medium effect sizes and $d < -0.47$ indicated large effect sizes (Romano, Kromrey, Coraggio, & Skowronek, 2006).

Results

Table 1 shows participants' performance on the background neuropsychological tests, with participants performing within the expected range for cognitively healthy individuals in this age-range.

------ Insert Table 1 around here ------

Computerised Barrier Task with perceived human and computer partners. Figures 1a and 1b illustrate the time taken and number of turns taken to complete the Computerised Barrier Task in the human and computer conditions. Both the analysis of time taken and turns revealed significant main effects of condition and trial bin, and time taken revealed a significant condition * trial bin interaction (see Table 2a and 2b). In terms of time taken, participants became significantly quicker at completing the trials between bins 1 and 2, but not between bins 2 and 3 (see Table 2a). The interaction between the early trials and condition suggests that while initial interactions with the computer partner are quicker, participants show a greater overall decrease in the time taken to complete trials in the human partner condition. In terms of turns taken, participants used significantly fewer turns between trial bins 1-2, although this pattern did not continue between trial bins 2-3 as the number of turns taken did not significantly decrease (see Table 2b). Participants took significantly fewer turns in the computer condition compared with the human condition; and there was no interaction between condition and trial bin, indicating that the difference in number of turns taken in the human and computer conditions remained relatively stable over the trials.

------ Insert Figures 1a and 1b and Tables 2a and 2b around here ------

Post-hoc contrasts were used to compare the number of turns taken in the human and computer conditions in the final trials to investigate whether participants interacted similarly in both conditions. Participants used a similar number of turns in both conditions ($V = 15$, $p = 0.72$), but were significantly slower in the computer condition compared to the human condition, $t(1, 23) = 5.45$, $p < 0.0001$.

Performance accuracy, fluctuation, and delayed recall performance. Figure 2 demonstrates the mean number of cards correctly placed by trial bin and condition. Wilcoxon signed-ranks tests with False Discovery Rate (FDR) corrections (Benjamini & Hochberg, 1995) showed that across all trial bins, accuracy was significantly higher in the human condition (bin 1: $V = 169.5$, $p < 0.01$, $d = 0.36$, bin 2: $V = 159.5$, $p < 0.05$, $d = 0.45$ and bin 3: $V = 112$, $p < 0.05$, $d = 0.45$). The Cliff's Delta effect size was calculated as 0.25, equating to a small effect.

------ Insert Figure 2 around here -----

Figure 3 demonstrates the total fluctuation scores for trial accuracy between conditions. Participants showed a significantly higher rate of fluctuation in the computer condition ($Z = 26.5$, $p < 0.01$, $d = 0.78$), suggesting that participants changed their answers significantly more when the computer was perceived to be providing the descriptions, or that participants had not learned the descriptions as accurately, leading to an increase in between-trial score fluctuations. The Cliff's Delta effect size was calculated as -0.45 indicating a medium effect.

----- Insert Figure 3 around here ------

Figure 4 illustrates delayed recall accuracy across the two conditions. Wilcoxon signed-ranks ($V = 17$, $p < 0.05$, $d = 0.55$) revealed that, one hour later, participants recalled significantly more tangram descriptions in the human condition compared to the computer condition. The Cliffs Delta effect size was calculated as 0.41 indicating a medium effect size.

----- Insert Figure 4 around here ------

Agency beliefs of the interlocutor. Deception was successful for all participants. After completion of the task, when told that they had been interacting with a human using a computer system in both conditions, all participants stated that they were convinced they were interacting with another human and a computer system through the audio-system. Participants were then remunerated for taking part.

Discussion

The current study investigated the effect that human agency beliefs have on interaction behaviours and learning and memory performance during a collaborative learning task. Participants believed that they were interacting with a human partner in one condition and a computer partner in another condition and their interactive behaviours, performance, and later recall were assessed. The verbal content of the two conditions was identical; the only differences were the manipulation of participant belief, and that the human condition used natural speech and the computer condition featured synthetic speech. Results indicated that beliefs about the human agency of a learning partner affect the learners' interaction behaviour, learning, and the accuracy with which they recall the descriptions after one hour.

While participants were initially quicker to complete the trials when they believed they were interacting with a computer, the final trials in the human condition were significantly faster. Participants took longer to complete the task with a computer in the middle and final trials, but this was not reflected by taking more interactive turns; participants took fewer turns than with the human partner. This suggests that participants were taking longer to make their card selections but were not seeking additional interaction from the computer partner to assist with their choices. An alternative interpretation might be that participants were less interested in interacting with a computer; they did not consider it a social activity and simply slowed down. Previous work comparing responses to instructions given by a human or a computer found participants' responses were significantly shorter and less variable in their utterances when interacting with computers (Siegert, Bock, Wendemuth, Vlasenko, & Ohnemus, 2015; Tenbrink, Ross, Thomas, & Dethlefs, 2010).

Participants matched the tangram card to the referential label more accurately when they believed the description was created and provided by a human partner. They also provided more consistent answers over the nine trials. When participants believed they were learning with a computer, they recalled significantly less detail after a delay compared with when they believed they were interacting with a human interlocutor. These findings align with recent behavioural and neuroimaging studies exploring effects of perceived agency on human and computer interactions (Caruana et al., 2017; Cross et al., 2016). Caruana et al. (2017) found that belief that the interaction was with a computer affected eye contact and response times. Cross et al. (2016) found that informing participants that movements were created by computers resulted in those movements being rated as less smooth than human-created counterparts. Wykowska et al. (2014) argue that when participants believe they are interacting with a human, they view their behaviours are created by an intelligent mind with intention; however, when participants interact with computers, they assume a "design stance", interpreting the behaviours of the system as being created by an engineer and change how they communicate with the system. It is clear that beliefs about agency affect how we respond to, interact with, and learn with systems.

Additionally, older adults may find computer-based contexts artificial and unfamiliar, leading them to perform more poorly on computerised versions of tasks compared to non-computerised versions (Kosowicz & MacPherson, 2017). Perceiving that a task is computer-based may result in additional difficulty compared to when older adults believe they are interacting with a human partner. Therefore, although many designers are attempting to make systems that older adults consider more user-friendly for social and practical applications, if older adults still perceive them as being ‘a computer’, they will interact with them differently, and may not learn as accurately with them. It is important to note that the opposite pattern has been found when people are interacting with virtual and human therapists; they believe the virtual therapists to be less judgemental (e.g., Rizzo et al., 2005). However, these results have largely been found in young and middle-aged adults, and may not reflect older adults’ experiences. Furthermore, VR therapy applications involve participants disclosing personal information which may have different implications for system development compared to learning systems.

The current study has some limitations. Firstly, the experiment used a simulated WoZ system rather than a real automated dialogue system. A fully automated dialogue system may be poorer at comprehending the participant compared with a human “wizard”, leading to more system errors, which may affect the speed and accuracy of learning (Fraser & Gilbert, 1991; Knursen, Le Bigot & Ros, 2017). Secondly, although the study was designed to explore the effect of agency beliefs, in order to create believable human and computer partners, synthetic and natural speech were used. Synthetic speech may place heavier demands on the memory system, and this may be particularly problematic for older adults (Luce, Feustel, & Pisoni, 1983; Smither, 1993). In addition, delayed recall was only assessed after one hour, and was not followed up at later timepoints. There may be differences in the recall of this information after a longer delay, and future research should focus on the influence of human and computer learning partners on longer-term recall. It is also important to note that the social manner of the computer may play a role in how people interact and learn from them, and increased computer sociability may create a more human-like alliance with an agent (Vardoulakis et al., 2012). Additionally, the human interlocutor in our study had a friendly, approachable tone. Future studies may focus on whether the approachability and friendliness of perceived human interlocutors has a role in how well older adults interact and learn. Furthermore, while our study was sufficiently powered to reliably detect medium-sized effects, our sample size prevented the reliable estimation of small effects and the inclusion of solely older, mainly female, participants recruited via convenience sampling may limit the generalizability of our results to males and other ages as well as older participants more widely. Despite these limitations, our results indicate that beliefs about agency play an important role in human-computer dialogue and highlight the need for future research in this area.

Computer-mediated support is a rapidly expanding industry and may promise innovative applications for aging in-place for older adults. In this study, we directly compared whether beliefs about a learning partner's agency significantly affects the way in which older adults behave and learn, with results indicating that agency beliefs do have an effect. If older adults believe they are interacting with a computer, their learning outcomes are poorer. As beliefs about agency have an impact on how older adults interact with and learn from systems, researchers and software designers should take this into account when creating systems designed to interact with and assist older adults. Research should explore how the wider social context of technology use may affect users' interactions with systems to maximise usability and user outcomes.

References

Bassuk, S., Glass, T., & Berkman, L. (1999). Social disengagement and incident cognitive decline in community-dwelling elderly persons. *Annals of Internal Medicine, 131*(3), 165–173.

Bates, D., Maechler M., Bolker, B. & Walker, S.(2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software,* 67(1), 1-48.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological),* 57(1), 289-300.

Berg, C. A., Johnson, M. M. S., Meegan, S. P., & Strough, J. (2003). Collaborative problem-solving interactions in young and old married couples. *Discourse Processes, 35*(1), 33–58.

Bickmore, T. W., Caruso, L., Clough-Gorr, K., & Heeren, T. (2005). 'It's just like you talk to a friend': relational agents for older adults. *Interacting with Computers,* 17(6), 711–735.

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Nass, C. (2003, July). Syntactic alignment between computers and people: The role of belief about mental states. In *Proceedings of the Twenty-fifth Annual Conference of the Cognitive Science Society* (pp. 186-191).

Branigan, H. P., Pickering, M. J., Pearson, J., & McLean, J. F. (2010). Linguistic alignment between people and computers. *Journal of Pragmatics*, 42(9), 2355–2368.

Caruana, N., Spirou, D., & Brock, J. (2017). Human agency beliefs influence behaviour during virtual social interactions. *PeerJ,* 5, e3819.

Choi, N. G., & DiNitto, D. M. (2013). The digital divide among low-income homebound older adults: Internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. *Journal of Medical Internet Research,* 15(5).

Cross, E. S., Ramsey, R., Liepelt, R., Prinz, W., & Hamilton, A. F. de C. (2016). The shaping of social perception by stimulus and knowledge cues to human animacy. *Philosophical Transactions of the Royal Society – B.* 371(1686), 20150075

Crompton, C.J. (2017). *Investigating human-human and human-computer collaborative learning and memory in healthy ageing: the role of collaborator identity and social cognition* (Unpublished doctoral dissertation). University of Edinburgh, UK.

Czaja, S. J., Charness, N., Fisk, A. D., Hertzog, C., Nair, S. N., Rogers, W. A., & Sharit, J. (2006). Factors Predicting the Use of Technology: Findings From the Center for Research and Education on Aging and Technology Enhancement (CREATE). *Psychology and Aging,* 21(2), 333–352.

Damant, J., & Knapp, M. (2015). What are the likely changes in society and technology which will impact upon the ability of older adults to maintain social (extra-familial) networks of support now, in 2025 and in 2040? *Future of ageing: evidence review,* Government Office for Science, UK, London, UK

Demberg, V., Winterboer, A., & Moore, J. D. (2011). A strategy for information presentation in spoken dialog systems. *Computational Linguistics,* 37(3), 489–539.

Derksen, B. J., Duff, M. C., Weldon, K., Zhang, J., Zamba, K. D., Tranel, D., & Denburg, N. L. (2015). Older adults catch up to younger adults on a learning and memory task that involves collaborative social interaction. *Memory*, *23*(4), 612-624.

Duff, M. C., Hengst, J. A., Tranel, D., & Cohen, N. J. (2008). Collaborative discourse facilitates efficient communication and new learning in amnesia. *Brain and Language,* 106(1), 41–54.

Elers, P., Hunter, I., Whiddett, D., Lockhart, C., Guesgen, H., & Singh, A. (2018). User requirements for technology to assist aging in place: qualitative study of older people and their informal support networks. *JMIR mHealth and uHealth*, *6*(6), e10741.

Findlater, L., Froehlich, J. E., Fattal, K., Wobbrock, J. O., & Dastyar, T. (2013). Age-related Differences in Performance with Touchscreens Compared to Traditional Mouse Input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 343–346). New York, NY, USA: ACM.

Fletcher-Watson, B., Crompton, C. J., Hutchison, M., & Lu, H. (2016). Strategies for enhancing success in digital tablet use by older adults: A pilot study. *Gerontechnology,* 15(3), 162–170.

Fozard, J. L., & Wahl, H.-W. (2012). Age and cohort effects in gerontechnology: A reconsideration. *Gerontechnology. 11*(1), 10-21.

Franklin, A., & Myneni, S. (2018). Engagement and Design Barriers of mHealth Applications for Older Adults. In *Proceedings of the Technology, Mind, and Society* (p. 9). ACM.

Fraser, N. M., & Gilbert, G. N. (1991). Simulating speech systems. *Computer Speech and Language*, *5*(1), 81-99.

Gardiner, C., Geldenhuys, G., & Gott, M. (2018). Interventions to reduce social isolation and loneliness among older people: an integrative review. *Health & Social Care in the Community*, *26*(2), 147-157.

Gatto, S. L., & Tak, S. H. (2008). Computer, Internet, and e-mail use among older adults: Benefits and barriers. *Educational Gerontology,* 34(9), 800–811.

Gelman, A & Su, Y (2018). arm: Data AnalysisUsing Regression and Multilevel/Hierarchical Models. R package version 1.10-1 https://CRAN.R-project.org/package=arm

González, A., Ramírez, M. P., & Viadel, V. (2015). ICT learning by older adults and their attitudes toward computer use. *Current Gerontology and Geriatrics Research*.

Gowen, E., Stanley, J., & Miall, R. C. (2008). Movement interference in autism-spectrum disorder. *Neuropsychologia,* 46(4), 1060–1068.

Green, P., & Wei-Haas, L. (1985, October). The rapid development of user interfaces: Experience with the Wizard of Oz method. In *Proceedings of the Human Factors Society Annual Meeting* (Vol. 29, No. 5, pp. 470-474). Sage CA: Los Angeles, CA: SAGE Publications.

Haub, C., & Yanagishita, M. (2011). World Population Data Sheet. Washington DC: Popultaion Reference Bureau.

Hawthorn, D. (2000). Possible implications of aging for interface designers. *Interacting with Computers*, 12(5), 507–528.

Heo, J., Chun, S., Lee, S., Lee, K. H., & Kim, J. (2015). Internet use and well-being in older adults. *Cyberpsychology, Behavior, and Social Networking*, *18*(5), 268-272.

Hieronymus, J., & Dowding, J. (2004). Clarissa spoken dialogue system for procedure reading and navigation.

Klapper, A., Ramsey, R., Wigboldus, D., & Cross, E. S. (2014). The Control of Automatic Imitation Based on Bottom–Up and Top–Down Cues to Animacy: Insights from Brain and Behavior. *Journal of Cognitive Neuroscience*, 26(11), 2503–2513.

Knutsen, D., Le Bigot, L., & Ros, C. (2017). Explicit feedback from users attenuates memory biases in human-system dialogue. *International Journal of Human-Computer Studies*, *97*, 77-87.

Kosowicz, M., & MacPherson, S. E. (2017). Improving multitasking assessment in healthy older adults using a prop-based version of the breakfast task. *Applied Neuropsychology: Adult*, *24*(3), 252-263.

Leite, I., Martinho, C., Pereira, A., & Paiva, A. (2009). As Time goes by: Long-term evaluation of social presence in robotic companions. *In RO-MAN 2009 - The 18th IEEE*

International Symposium on Robot and Human Interactive Communication (pp. 669–674)

Liepelt, R., & Brass, M. (2010). Top-down modulation of motor priming by belief about animacy. *Experimental Psychology,* 57(3), 221–227.

Luce, P., Feustel, T., & Pisoni, D. (1983). Capacity Demands in Short-Term Memory for Synthetic and Natural Speech. Human Factors: *The Journal of the Human Factors and Ergonomics Society*, 25(1), 17–32.

Macbeth, G., Razumiejczyk, E. & Ledesma, R. D. (2011). Cliff's Delta Calculator: A non-parametric effect size program for two groups of observations. *Universitas Psychologica, 10(2),* 545-555.

Meyers, J. E., &. Meyers, K. R. (1995). Rey Complex Figure Test and recognition trial professional manual. *Psychological Assessment Resources.*

Morris, A. (2007). E-literacy and the grey digital divide: a review with recommendations. *Journal of Information Literacy*, 1(3), 13–28.

Nass, C. I., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Computer-Human Interaction (CHI) Conference: Celebrating Interdependence* 1994, p72–78.

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues,* 56(1), 81–103.

Pearson, J., Hu, J., Branigan, H. P., Pickering, M. J., & Nass, C. I. (2006). Adaptive language behavior in HCI. *In Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06* (pp. 1177–1177).

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences,* 27(2), 169–190.

Pon-Barry, H., Weng, F., & Varges, S. (2006). Evaluation of content presentation strategies for an in-car spoken dialogue system. In *Ninth International Conference on Spoken Language Processing.* 1930-1933.

R Core Team (2018). R: A language and environment forstatistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. New York: Cambridge University Press.

Rizzo, A., Pair, J., McNerney, P. J., Eastlund, E., Manson, B., Gratch, J., Hill, R., & Swartout, B. (2005). Development of a VR therapy application for Iraq war military personnel with PTSD. *Studies in Health Technology and Informatics*, *111*, 407-413.

Romano, J., Kromrey, J. D., Coraggio, J., & Skowronek, J. (2006). Appropriate statistics for ordinal level data: Should we really be using t-test and cohen's d for evaluating group differences on the NSSE and other surveys? Annual meeting of the Florida Association of Institutional Research.

Selwyn, N., Nemorin, S., Bulfin, S., & Johnson, N. F. (2017). Left to their own devices: the everyday realities of one-to-one classrooms. *Oxford Review of Education,* 43(3), 289–310.

Siegert, I., Bock, R., Wendemuth, A., Vlasenko, B., & Ohnemus, K. (2015). Overlapping speech, utterance duration and affective content in HHI and HCI - An comparison. *6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom*) (pp. 83–88).

Smith, M. W., Sharit, J., & Czaja, S. J. (1999). Aging, Motor Control, and the Performance of Computer Mouse Tasks. *Human Factors,* 41(3), 389–396.

Smither, J. A. A. (1993). Short term memory demands in processing synthetic speech by old and young adults. *Behaviour & Information Technology*, *12*(6), 330-335.

Stanley, J., Gowen, E., & Miall, R. C. (2007). Effects of agency on movement interference during observation of a moving dot stimulus. *Journal of Experimental Psychology.* Human Perception and Performance, 33(4), 915–926.

Stenzel, A., Chinellato, E., Bou, M. A. T., Del Pobil, Á. P., Lappe, M., & Liepelt, R. (2012). When Humanoid Robots Become Human-Like Interaction Partners: Corepresentation of Robotic Actions. *Journal of Experimental Psychology: Human Perception and Performance,* 38(5), 1073–1077.

Stine-Morrow, E., & Parisi, J. (2008). The effects of an engaged lifestyle on cognitive vitality: a field experiment. *Psychology and Aging, 23*(4), 778-786.

Strough, J., & Margrett, J. (2002). Overview of the special section on collaborative cognition in later adulthood. *International Journal of Behavioural Development, 26* (1), 2-5.

Tenbrink, T., Ross, R. J., Thomas, K. E., Dethlefs, N., & Andonova, E. (2010). Route instructions in map-based human–human and human–computer dialogue: A comparative analysis. *Journal of Visual Languages & Computing*, *21*(5), 292-309.

Tsai, C. C., & Brass, M. (2007). Does the human motor system simulate Pinocchio's actions? Coacting with a human hand versus a wooden hand in a dyadic interaction. *Psychological Science,* 18(12), 1058–1062.

Vardoulakis, L., Ring, L., Barry, B., Sidner, C., & Bickmore, T. (2012). Designing relational agents as long term social companions for older adults. In *Intelligent virtual agents* (pp. 289–302). Springer.

Vaportzis, E., Giatsi Clausen, M., & Gow, A. J. (2017). Older adults' perceptions of technology and barriers to interacting with tablet computers: a focus group study. *Frontiers in Psychology*, *8*, 1687.

Wagner, N., Hassanein, K., & Head, M. (2010). Computer use by older adults: A multi-disciplinary review. *Computers in Human Behavior,* 26(5), 870–882.

Wechsler, D. (2008). Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). *NCS Pearson.*

Wechsler, D. (2009). Test of Premorbid Functioning. San Antonio, TX: *The Psychological Corporation.*

White, H., McConnell, E., Clipp, E., Bynum, L., Teague, C., Navas, L. & Halbrecht, H. (1999). Surfing the net in later life: A review of the literature and pilot study of computer use and quality of life. *Journal of Applied Gerontology*, 18(3), 358–378.

Wiles, J. L., Leibing, A., Guberman, N., Reeve, J., & Allen, R. E. S. (2012). The meaning of "aging in place" to older people. *The Gerontologist,* 52(3), 357–366.

Wolters, M., Georgila, K., Moore, J. D., Logie, R. H., MacPherson, S. E., & Watson, M. (2009). Reducing working memory load in spoken dialogue systems. *Interacting with Computers,* 21(4), 276–287.

Wolters, M., Georgila, K., Moore, J. D., & MacPherson, S. E. (2009). Being old doesn't mean acting old: How older users interact with spoken dialog systems. *ACM Transactions on Accessible Computing (TACCESS)*, *2*(1), 2.

Wykowska, A., Wiese, E., Prosser, A., & Müller, H. J. (2014). Beliefs about the minds of others influence how we process sensory information. *PLoS One*, *9*(4), e94339.

Yoon, H., Jang, Y., & Xie, B. (2016). Computer use and computer anxiety in older Korean Americans. *Journal of Applied Gerontology*, *35*(9), 1000-1010.

Zajicek, M. (2004). Successful and available: interface design exemplars for older users. *Interacting with computers*, *16*(3), 411-430.

Table 1: Mean and standard deviations for participants' neuropsychological test performance.

	Mean	SD
Test of Premorbid Functioning IQ (Max = 125)	111.50	6.32
Rey Osterrieth Complex Figure – Immediate recall (Max = 36)	16.62	7.22
Rey Osterrieth Complex Figure – Delayed Recall (Max = 36)	15.71	7.67
Digit Span Forwards Score (Max = 16)	10.75	2.69
Digit Span Backwards Score (Max = 16)	9.49	2.78
Digit Span Sequence Score (Max = 16)	8.17	2.33

Table 2a: Time to complete: beta, standard errors and t-values for fixed effects, and variance and residual for random effects, model fit by REML.

		B	**S.E.**	**T**
Fixed effects				
Intercept		158.64	4.88	32.49
Condition (computer)		-1594	6.54	-2.44*
Bin 1-2		-85.48	4.51	-12.96**
Bin 2-3		-8.05	4.51	-1.79
Interactions				
Condition (computer) * Bin 1-2		32.22	9.02	3.57**
Condition (computer) * Bin 2-3		2.49	9.03	0.28
Random effects			Variance	Standard Deviation
Participant	Intercept		328.03	18.11
	Condition (Computer)		50.27	7.09
Residual			48.48	22.10

* t > 1.96; ** t > 2.58

Table 2b: Number of turns taken: beta, standard errors and t-values for fixed effects, and variance and residual for random effects, model fit by REML.

		B	S.E.	T
Fixed effects				
Intercept		30.74	0.47	65.01
Condition (computer)		-2.03	0.67	-3.04**
Bin 1-2		-4.46	0.46	-9.65**
Bin 2-3		-0.62	0.45	-1.35
Interactions				
Condition (computer) * Bin 1-2		1.75	0.92	1.89
Condition (computer) * Bin 2-3		0.19	0.92	0.21
Random effects			Variance	Standard Deviation
Participant	Intercept		2.80	1.67
	Condition (Computer)		0.43	0.65
Residual			5.13	2.26

* t > 1.96; ** t > 2.58

Figure Captions

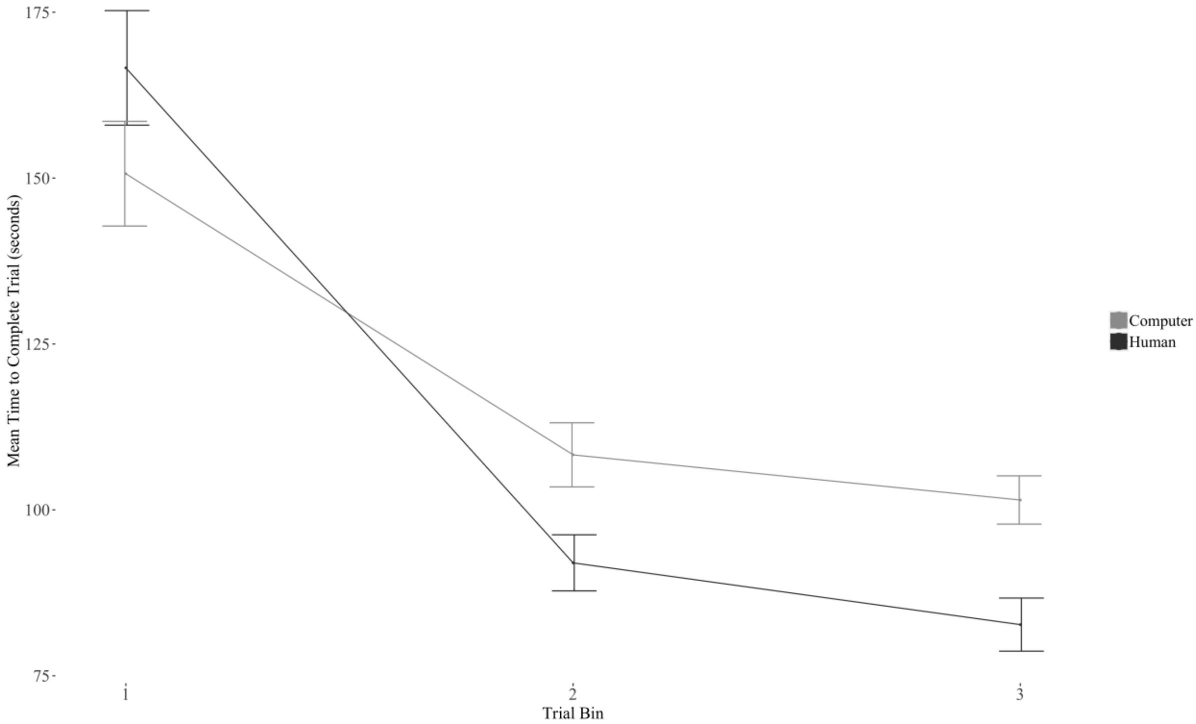
Figure 1a: Mean and standard error of the mean for time to complete the task by trial bin and condition.

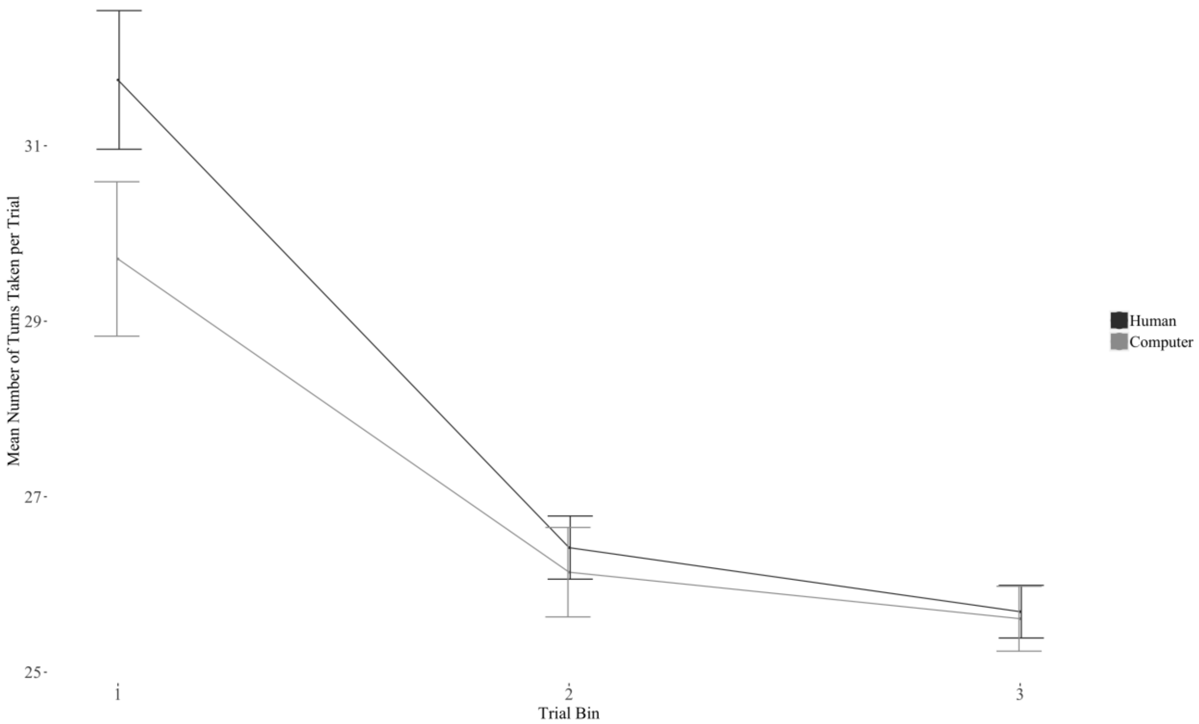
Figure 1b: Mean and standard error of the mean for number of turns taken to complete the task by trial bin and condition.

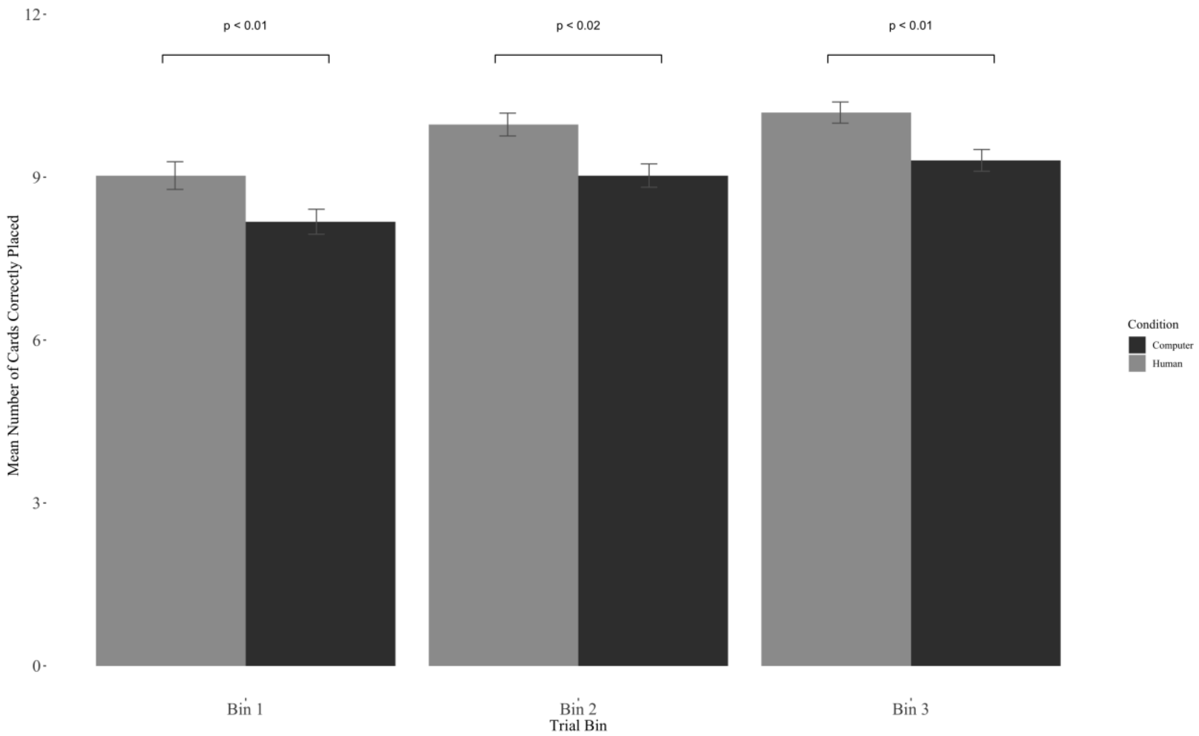
Figure 2: Means and standard errors of the mean for number of cards correctly placed by trial bin and condition, with FDR-corrected p-values.

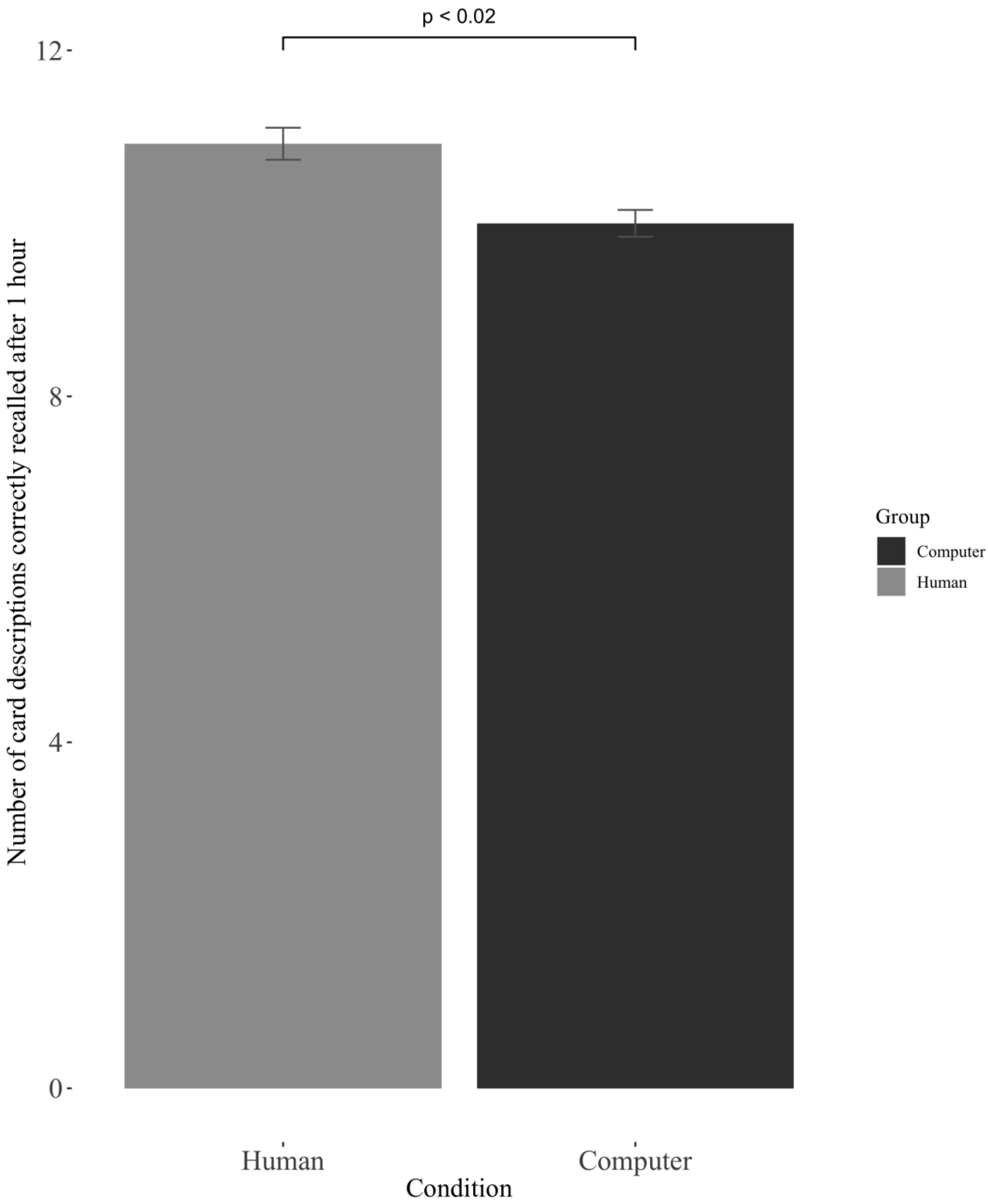
Figure 3: Means and standard errors of the mean for fluctuation scores in the human and computer conditions.

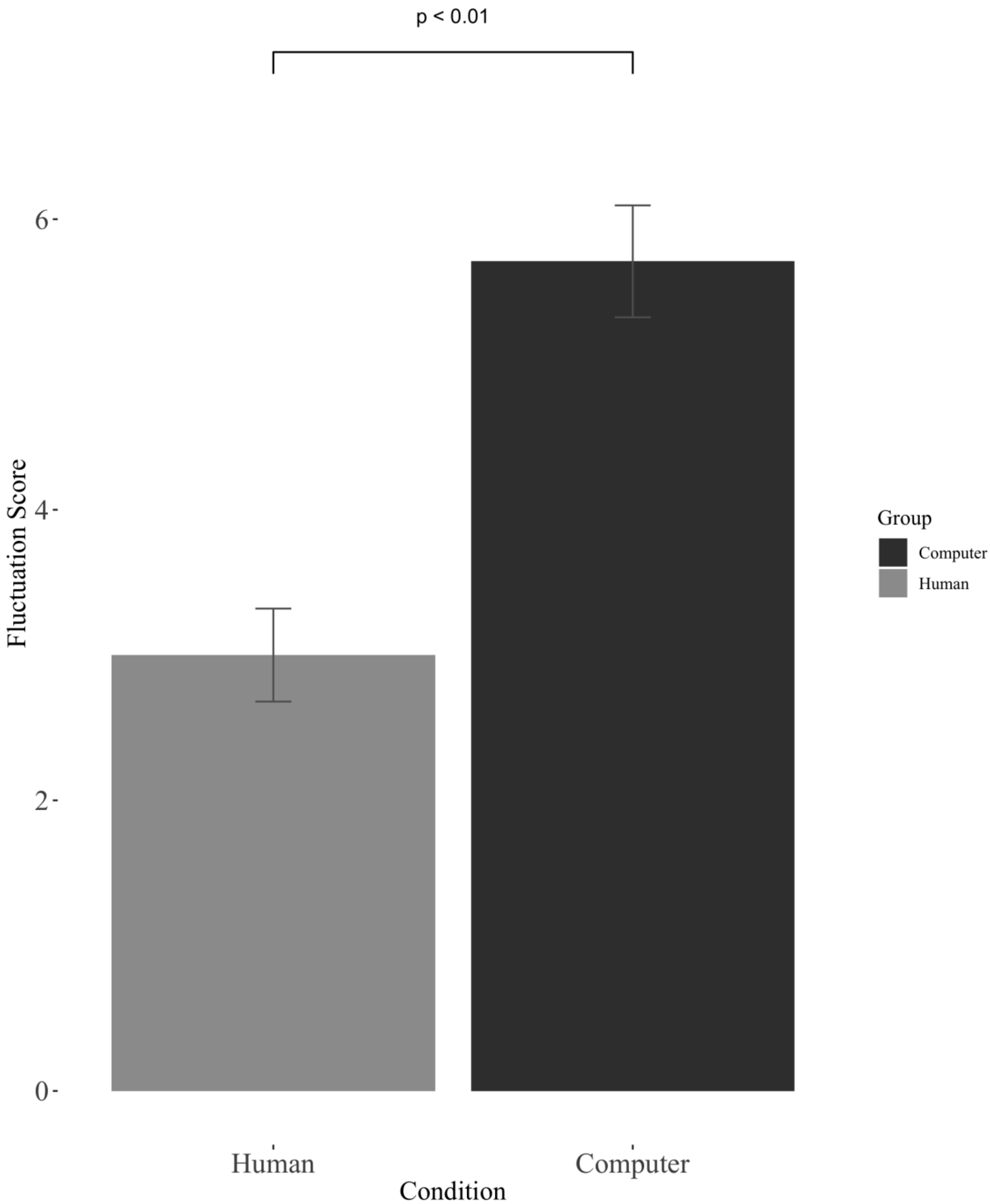
Figure 4: Mean and standard errors of the mean for delayed recall accuracy in the human and computer conditions.


175
150
125
100
75
Mean Time to Complete Trial (seconds)
1
2
3
Trial Bin
Computer
Human


31
29
27
25
Mean Number of Turns Taken per Trial
1
2
3
Trial Bin
Human
Computer


p < 0.01
p < 0.02
p < 0.01
12
9
6
3
0
Mean Number of Cards Correctly Placed
Bin 1
Bin 2
Bin 3
Trial Bin
Condition
Computer
Human


p < 0.02
12
8
4
0
Number of card descriptions correctly recalled after 1 hour
Human
Computer
Condition
Group
Computer
Human


p < 0.01
6
4
2
0
Fluctuation Score
Human
Computer
Condition
Group
Computer
Human