How Physical Interaction Helps Performance in a Scrabble-like Task

Morgan Fleming (mfleming2@ucmerced.edu) Paul P. Maglio (pmaglio@ucmerced.edu)
Cognitive and Information Sciences
University of California, Merced
5200 N Lake Rd
Merced, CA 95343

Abstract

An experiment tested the hypothesis that people sometimes take physical actions to help themselves solve problems. The task was to generate all possible words that could be formed from seven Scrabble letters. In one condition, participants could use their hands to manipulate the letters, and in another condition, they could not. Quantitative results show that more words were generated and lower frequency words were generated with physical manipulation than without. Qualitative results suggest that participants who could manipulate the letters tended to subdivide the task into smaller tasks (focusing on fewer letters at a time). Overall, our results can be explained in terms of an interactive search process in which external, physical activity effectively complements internal, cognitive activity, providing a reliable way to simplify search, explore the space of letter combinations, and identify potential words.

Keywords: Interactive skill, Scrabble, word games

Introduction

People actively reorganize and interact with their environments for cognitive purposes (Clark, 1997; Hutchins, 1995a, 1995b; Kirsh & Maglio, 1994; Kirsh, 1995b). Pilots use the cockpit to help calculate speeds and execute timely actions (Hutchins, 1995b). People reorient Tetris pieces rather than mentally rotating them (Kirsh & Maglio, 1994; Maglio & Kirsh, 1996). People group things physically while counting and doing arithmetic (Carlson, Avraamides, Cary, & Strasberg, 2007; Kirsh, 1995a). These are just some of many cases in which interaction with the external environment can amplify or simplify cognitive processing. But exactly how and under what conditions does physically arranging and interacting with the external environment help?

Here we examine performance and behavior in an word generation task like the one solved by Scrabble[1] players. In the board game Scrabble, players form words by arranging tiles with letters printed on them. When trying to come up with words in this game, people can either mentally rearrange the letters or physically rearrange the letters. Based on the idea that people routinely set up their environments to make their cognitive jobs easier, it is reasonable to suppose that it is easier to form words by physically moving the letters than by simply imagining their rearrangement (Clark & Chalmers, 1998; Kirsh, 1995b). In fact, prior research has shown that in specific situations, people are more effective at finding words in a set of Scrabble letters when they can move the letters than when they cannot (Maglio, Matlock, Raphaely, Chernicky, & Kirsh, 1999; Vallée-Tourangeau & Wrightman, 2010).

Moving letters around in the environment may support the cognitive process of forming words in a number of ways, for instance, by grouping letters to alleviate the need to rely on memory of recently created words, or by testing out letter combinations in the environment to "see" potential words rather than having to imagine them. Some studies suggest that physical interaction with the task environment can provide effective scaffolding for learning in a spatial task (Smith, 2001); in this case, participants who had trained to do a visual-spatial task by physically interacting with the environment, performed better on a non-interactive version of the same task after training than participants who did not interact or who only observed interactions during training. The general hypothesis is that interaction in this Scrabble task provides participants the opportunity to physically manage the problem of finding words, possibly simplifying their internal processes by relying on external action. If this hypothesis is true, then the general question is how exactly does taking physical action support the task of finding words? A more specific question is are there any lasting effects of physical interaction on agent-internal procceses in the Scrabble task?

Performance does not improve with physical interaction in all cases (Maglio et al., 1999; Vallée-Tourangeau & Wrightman, 2010), and people do not always choose to interact (Destefano, Lindstedt, & Gray, 2011). There may be tradeoffs in time and processing costs, and assessing the tradeoffs in real-time may be difficult to adequately compensate for. After all, humans are not the best at deciding when information seeking actions are worth the tradeoff (Fu & Gray, 2006). The Scrabble task in particular has not always shown benefits of physical interaction with the task environment (Fusaroli et al., 2014). There are many possible confounds, including the frequencies or complexity of words that can be found in a set of letters and individual experience with similar tasks or with word games generally. Controlling effectively for word frequency will be critical, as low frequency words are harder to find than high frequency words in anagram-solving tasks (Mayzner & Tresselt, 1958). Prior work on the Scrabble task in particular found performance benefits of interaction only for letter sets in which there were relatively low-frequency words, suggesting interaction with the letters helps mainly when the task is relatively difficult (Maglio et al., 1999).

In what follows, we describe our experiment in detail. Our results are both quantitative and qualitative. In the end, we discuss implications for a general theory of interactivity.

[1] Scrabble is a registered trademark of Hasbro Inc.

Method

The goal of our experiment is to examine performance and interactive strategies in a word formation task using Scrabble letters. Based on prior work, we expected that people would generate more words from a set of Scrabble letters when allowed to manipulate the letters than when not allowed to manipulate the letters (Maglio et al., 1999; Vallée-Tourangeau & Wrightman, 2010). We also expected those who manipulated letters would perform better over time, that is, they would get better with practice faster. Moreover, we expected those who manipulated the letters during practice would perform better on a post-test, even when they could no longer manipulate the letters.

A total of 18 undergraduates from the University of California, Merced, participated in the experiment for course credit. Each participant completed a pre-test, two blocks of five experimental trials each, and a post-test. Participants were split into two conditions (between-subjects design): *Non-interactive*, in which they could not move the letters, and *Interactive*, in which they could move the letters. The pre-test and post-test were *Non-interactive* trials for all participants. All participants received the same letter sets in the same order (see Table 1).[2]

To create reasonably matched stimuli, that is, letter sets with similar word productivity, we searched systematically for letter sets that could generate approximately the same numbers of words with approximately the same word frequencies. Specifically, after generating sets of seven letters, we examined all the words that could be made from two or more of the letters in each set and assessed the frequencies of each word (see Table 1), relying on the corpora available from the English Lexicon Project for word lists and frequency data (Balota et al., 2007).

The experiment was run using custom software on an iPad. The letters for each trial were displayed as Scrabble letters arranged horizontally in a row on the iPad screen. Each trial lasted 3 minutes. Participants were told to produce as many words as they could in that time, speaking and spelling each word they found. The iPad recorded the verbal responses. For *Interactive* trials, participants could drag the letters to arbitrary locations on the screen using their fingers. Participants in the *Interactive* condition were encouraged to move the letters on the screen. The iPad recorded all interactions with a resolution of about 10 milliseconds. For *Non-interactive* trials, the letters remained static. See Figure 1.

Results

We first completed a statistical analysis of the experimental data to see whether in fact physical interaction has an effect on performance, and then we looked more closely at individual behavior to learn exactly when and how individuals relied on physical interaction.

Table 1: Stimuli were created by finding letter sets with matching word productivity and word frequencies. Using data available from the English Lexicon Project, we calculated productivity (number of words found in the corpus) and frequencies (*log* HAL frequency) for each set (Balota et al., 2007).

Block	Letters	Avg Freq	Tot Wds
Pre-test	N M G E R D A	9.40	80
Block 1	E S W T L B A	8.12	115
	L R S E H C A	8.8	94
	T E M S H C A	8.93	113
	O S E L R C A	8.60	102
	T I C S E D A	8.80	96
Block 2	N T K S E D A	8.90	85
	K I L S E D A	8.51	96
	L F T E S D A	8.26	107
	E I M T S D A	8.60	123
	O H S R C E A	8.58	109
Post-test	L I E F S D A	8.7	100

Quantitative Analyses

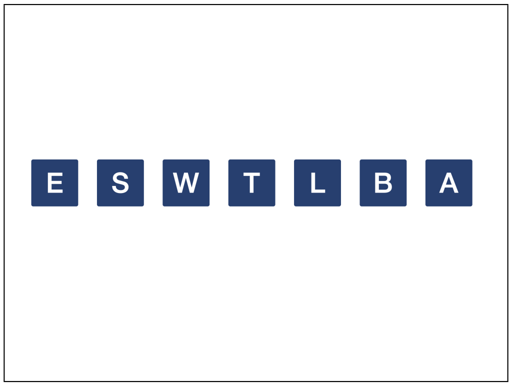
We first transcribed the recordings to find the words generated by each participant. There was an average of 12.2 words generated on the pre-test, with a standard deviation of 5.3 words. One participant who generated 31 words on the pre-test in the *Non-interactive* condition was removed from further analysis because the score was more than three standard deviations from the mean, leaving 8 participants in the *Non-interactive* condition and 9 in the *Interactive* condition. The remaining pre-test scores ranged from 6 to 14.

Because of wide variability in the number of words generated between participants, we divided the number of words produced on each trial and on the post-test by the number of words produced on the pre-test, resulting in a *normalized word-generation score* for each trial, with scores higher than 1 representing improvement over the pre-test. We first examined word generation by comparing performance on the post-test across conditions, and found no difference (*Interactive* $M = 1.2$, $SD = 0.35$, *Non-interactive* $M = 1.0$, $SD = 0.33$), $t(15) = 1.24$, $p = .23$.

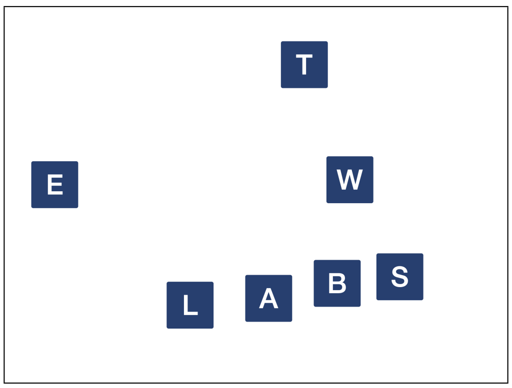
We next examined word generation by comparing performance in *Interactive* and *Non-interactive* conditions across *Block 1* and *Block 2*, see Figure 2. Participants in the *Interactive* condition generated more words overall ($M = 1.2$, $SD = 0.23$) than those in the *Non-interactive* condition ($M = 1.0$, $SD = 0.16$), $F(1,15) = 5.1$, $p = 0.040$.

We then examined the frequency of words generated (using word-frequency data as described previously) by comparing performance in *Interactive* and *Non-interactive* con-

[2] Note, of course, that sampling from our student population may limit generalizability of our results, as our participants may not reflect the diversity of the population, even though University of California, Merced, students are the most ethnically, racially, and economically diverse college students in the University of California system (see http://ipa.ucmerced.edu/student.htm).



(a) *Non-interactive* trial



(b) *Interactive* trial

Figure 1: The image in (a) shows a *Non-interactive* trial, with letters arranged statically in a row. The image in (b) shows an *Interactive* trial in which letters have been dragged around the screen arbitrarily.

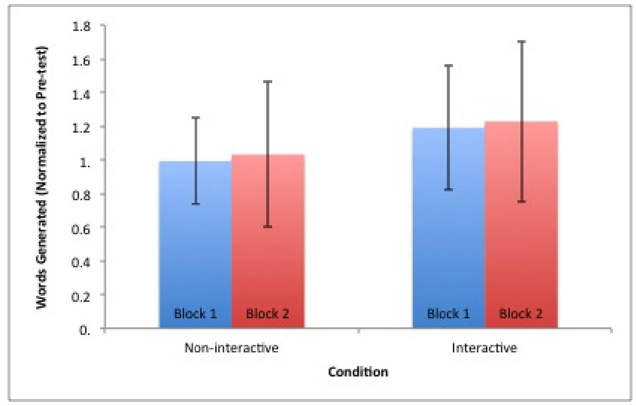


Figure 2: More words were generated overall in the *Interactive* condition than in the *Non-interactive* condition.

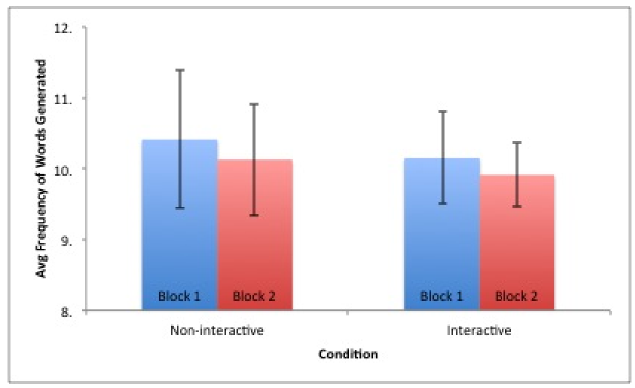


Figure 3: Lower frequency words were generated overall in the *Interactive* condition than in the *Non-interactive* condition.

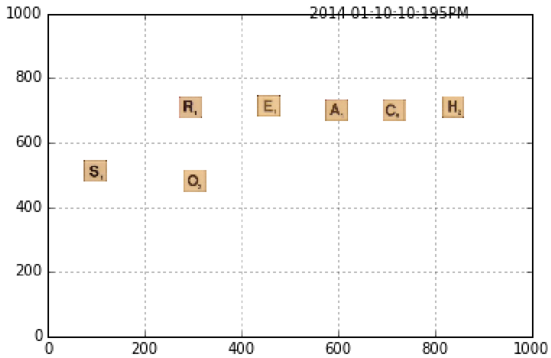
ditions across *Block 1* and *Block 2*, see Figure 3. Participants in the *Interactive* condition generated on average lower frequency words ($M = 10.1$, $SD = 0.37$) than those in the *Non-interactive* condition ($M = 10.3$, $SD = 0.28$), $F(1,15) = 14.5$, $p = 0.002$. Participants also generated on average lower frequncy words in *Block 2* ($M = 10.1$, $SD = 0.28$) than in *Block 1* ($M = 10.3$, $SD = 0.37$), $F(1,15) = 7.8$, $p < 0.013$.

We also examined the frequency of words generated on the pre-test versus those generated on the post-test by condition and found that participants generated on average lower frequency words on the post-test ($M = 9.8$, $SD = 0.57$) than on the pre-test ($M = 10.6$, $SD = 0.80$), $F(1,15) = 16.1$, $p = 0.001$.

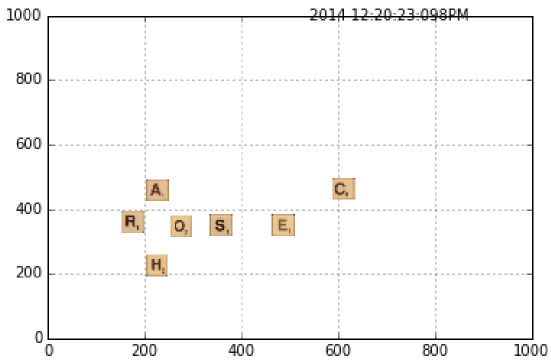
Descriptive Analyses

To examine what participants were doing in the *Interactive* condition, we used the movement and letter position data collected by the iPad to reconstruct and playback participant behavior aligned with the audio recordings of the words generated. Thus, we were able to examine in detail the behavior of participants. This sort of informal observation is a well-known and appropriate method for guiding theory-building (Simon, 1989). Specifically, we report here on behavior of participants during the last trial of *Block 2* in the *Interactive* condition. We sampled the last trial here only to ensure participants were fully familiar with the task and test environment. Because of errors in the data recording, we could not reconstruct playback for one participant, leaving a total of 8. In all, we analyzed 24 minutes of activity, during which participants found 94 words.

Overall Observations. We observed all participants using their hands to move the letters around, and all participants constructed at least some of the words they found interactively on the screen. For the most part, participants constructed words on the screen horizontally, with letters read left-to-right (91.58% of all the words found). On occasion, they also constructed words vertically, with letters read bottom-to-top (2.11% of all the words constructed on screen). When they did arrange the letters to form words on the screen, they constructed the words in their entirety, occasionally moving all extraneous letters out of the way (49.42% of all the words constructed on screen). Often, they moved

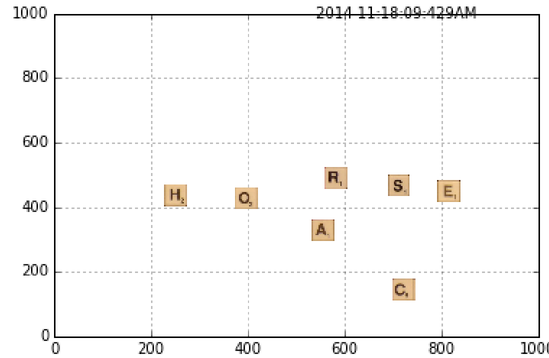


(a) Multi-level group

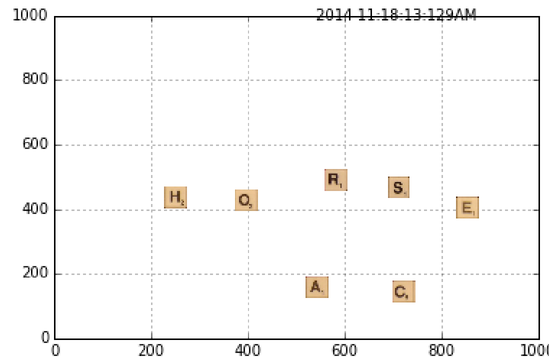


(b) Pooled group

Figure 4: The image in (a) shows an example of multi-level grouping in which two groups of letters are arranged horizontally, and the image in (b) shows an example of a pooling in which the word in the center, "rose", is surrounded by other letters.



(a) Before clean up



(b) After clean up

Figure 5: The image in (a) shows a multi-level group before letters are moved away or cleaned up, and the image in (b) shows the resulting configuration after clean up.

the letters into position in the words formed in order, meaning they would position the first letter, then the second, and so on (32.18% of all the words constructed on screen). In all cases, the first letter moved was part of the first word generated. Although there was plenty of movement overall, we also observed long periods during some trials in which there were no words generated and no action taken at all.

Interaction Details. We observed participants routinely grouping letters when constructing words on screen (about 35% of the time). We identified two consistent patterns of grouping behavior: *multi-level grouping* and *pooling* (see Figure 4). Multi-level grouping refers to a series of actions that result in two different groups of letters on the screen, each arrayed horizontally along a different level of the y-axis. Pooling refers to a series of actions that result in a single group of letters arrayed horizontally with additional letters scattered around them. On average, each participant spent 22% of the time using the multi-level grouping strategy, and 15% of the time using the pooling strategy.

We observed multi-level grouping and pooling differ mainly in the ways in which the letters are used *after* the structures are set up. Multi-level groups are characterized by multiple distinct letter sets in which the letters from each set are not used together to generate words. Pooled groups, by contrast, organize the screen around a focal word, with an inner and an outer structure in which the inner structure is used as the area to construct a word, and the periphery provides additional letters that are sometimes combined with letters from the focal word to generate a new word.

In addition to these two consistent methods of grouping letters to form words, we observed participants often *cleaning up* their groups on their way to constructing new groups. These cleaning actions do not result in new words arranged on the screen, but seem to be taken to segregate letters under consideration for word generation from letters that are not. In the case of a pooled group, clean-up often involves moving letters away from the word at the center. In the case of a multi-level group, participants often segregate the letters used to construct a word on one half of the screen, and then move all the remaining letters to the other half (see Figure 5).

The two different grouping strategies have different dynamics. Multi-level groups tended to be stable over time, with an *on-stage* area where words are actively constructed and an *off-stage* area where they are not. For the most part, the position of the first letter moved established the on-stage area, with additional letters either moved to the same horizontal line (on-stage) or moved to a different line to establish the off-stage area. Once the stage is set, letters are then moved from off-stage to on-stage and on-stage to off-stage to form new words on the screen.

In contrast to the relatively stable structure of multi-level groupings, the pooling strategy creates more transient structures. For example, we observed participants with multi-level groups quickly merge their on-stage and off-stage areas around a focal word to create pooled groups, creating words horizontally between the existing areas, and then just as quickly, move the letters back to their multi-level groups. In fact, we observed many instances in which participants moved letters around not to form words or parts of words visually, but to maintain their groupings, particularly multi-level groupings.

Overall, we observed consistent behaviors across all participants in the *Interactive* condition. All used their hands to move the letters on the screen to form words, mainly horizontally and left-to-right. They used strategies consistently to segregate sets of letters (multi-level groups) and to merge sets of letters (pooled groups), and even took actions that did not lead to new words but served only to maintain or establish groupings.

Discussion

In summary, we found (a) an effect of *Interactive* trials on productivity in the number of words produced during practice (normalized by pre-test score), (b) no effect of condition on the number of words produced on the post-test relative to the pre-test, (c) an effect of condition on the frequency of words produced (with *Interactive* trials resulting in less frequent words), and (d) an effect of block, with less frequent words produced in *Block 2* than in *Block 1*. Thus, participants generated more words when they could interact with the letters and they also generated lower frequency words when they could interact with the letters. However, there was no transfer from the *Interactive* trials during training to the *Non-interactive* post-test.

By observing in detail the behavior of participants in *Interactive* trials, we identified a number of consistent strategies that participants naturally developed and that simplify the problem of searching for possible words. By grouping letters in the environment, participants can effectively focus attention on subsets of letters. In a set of 7 letters, there are 13,692 different possible combinations of 2-to-7 letters. In our stimuli, each set of letters could be arranged to generate only 85-135 words. Moving letters into groups of 4 decreases the theoretical search space from 13,692 combinations to 84 combinations. Both the pooling and multi-level strategies segregate subsets of letters from the rest, creating visual displays with effectively fewer letters and potentially constraining the search for letter combinations.

It seems clear to us that the consistent interactive strategies we observed across all participants facilitate performance mainly by helping to constrain the sets of letters under consideration at any one time. Note, however, that establishing subsets of letters on the screen does not change *what* information is being displayed, only *how* it is being displayed. The letters themselves do not change as a consequence of a participant's actions, only their relative position changes. In a very basic sense, then, interaction cannot change the abstract task, it can change only how elements of the task are displayed. In the end, we believe moving the letters around serves an attentional function for participants, enabling them to focus on a subset of letters by organizing letters into groups in the visual environment.

Obviously, participants in both the *Non-interactive* and *Interactive* conditions are solving the same abstract problem: Finding words by rearranging letters in a given set. In the non-interactive case, participants could rely only on internal processing, presumably some kind of search through the space of letter and sound combinations (Maglio et al., 1999). In the interactive case, participants can interleave external actions with internal processing, presumably changing the sort of search they need to perform. By interleaving internal and external actions, participants may sometimes simplify the search by restricting the set of letters. Though it is possible to restrict the letters under consideration in the non-interactive case internally, our results suggest that it is in fact easier to do this in the interactive case externally.

Summary and Future Work

Participants who could interact with letters in this Scrabble-like task found more words and found lower frequency words than those who could not interact. These differences may have resulted from the interactive group's ability to simplify the problem of word-finding by spatially grouping different subsets of letters. In a process of searching through the space of letter combinations, restricting the set of letters under consideration at any given time seems to facilitate search.

Yet the question of what role *interaction* with the task environment is playing is not fully answered. It is possible that the organization of items on the screen and not the act of organizing the items on the screen may have the biggest impact on performance. This could be tested by comparing the performance on an interactive version of the task with performance on a non-interactive version in which letters are arranged on the screen in a multi-level group or similar fashion.

Our result on whether training in the *Interactive* condition facilitates future performance in non-interactive settings is inconclusive. Whereas we did not find transfer from interactive training to non-interactive test, it is possible we did not provide enough training or that the test stimuli were not properly controlled. For instance, here, we controlled for difficulty by ensuring that each letter set was highly productive and contained words of about the same frequency and length, but manipulating frequency may shed light on differences on what is being learned in interactive or non-interactive training.

Acknowledgments

We thank Lamar Williams for help constructing the iPad software, Cesar Gamez for help running subjects, Carolyn Jennings and Bodo Winter for helpful discussions of the study and results, and Teenie Matlock for use of her lab, helpful discussions and feedback, and comments on an early draft of this paper.

References

Balota, D., Yap, M., Hutchison, K., Cortese, M., Kessler, B., Loftis, B., ... Treiman, R. (2007). The english lexicon project. *Behavior Research Methods*, *39*(3), 445-459.

Carlson, R. A., Avraamides, M. N., Cary, M., & Strasberg, S. (2007). What do the hands externalize in simple arithmetic?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 747–756.

Clark, A. (1997). *Being there: Putting body, brain, and world together again*. Cambridge, MA: MIT Press.

Clark, A., & Chalmers, D. J. (1998). The extended mind. *Analysis*, *58*, 7-19.

Destefano, M., Lindstedt, J. K., & Gray, W. D. (2011). Use of complementary actions decreases with expertise. In *Proceedings of the 33rd annual conference of the cognitive science society.* Austin, TX: Cognitive Science Society.

Fu, W.-T., & Gray, W. D. (2006). Suboptimal tradeoffs in information seeking. *Cognitive Psychology*, *52*(3), 195 - 242.

Fusaroli, R., Ostergaard, S., Raczaszek-Leonardi, J., Bjørndahl, J. S., Stjernfelt, F., & Tylen, K. (2014). Doing words together: assessing joint problem solving in a scrabble task. In *Proceedings of the thirty-seventh annual conference of the cognitive science society.* Austin, TX: Cognitive Science Society.

Hutchins, E. (1995a). *Cognition in the wild*. Cambridge, MA: MIT Press.

Hutchins, E. (1995b). How a cockpit remembers its speed. *Cognitive Science*, *19*.

Kirsh, D. (1995a). Complementary strategies: Why we use our hands when we think. In *Proceedings of seventeenth annual conference of the cognitive science society* (pp. 212–217). Hillsdale, NJ: Lawrence Erlbaum.

Kirsh, D. (1995b). The intelligent use of space. *Artificial Intelligence*, *73*, 31–68.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, *18*, 513-549.

Maglio, P. P., & Kirsh, D. (1996). Epistemic action increases with skill. In *Proceedings of the eighteenth annual conference of the cognitive science society.* Mahwah, NJ: Lawrence Erlbaum.

Maglio, P. P., Matlock, T., Raphaely, D., Chernicky, B., & Kirsh, D. (1999). Interactive skill in scrabble. In *Proceedings of twenty-first annual conference of the cognitive science society* (p. 326-330). Mahwah, NJ: Lawrence Erlbaum.

Mayzner, M. S., & Tresselt, M. E. (1958). Anagram solution times: A function of letter order and word frequency. *Journal of Experimental Psychology*, *56*, 376-379.

Simon, H. A. (1989). The scientist as problem solver. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of herbert a. simon* (p. 376-398). Hilsdale, NJ: Lawrence Erlbaum.

Smith, G. (2001). Interaction evokes reflection: Learning efficiency in spatial visualization. In C. Montgomerie & J. Viteli (Eds.), *Proceedings of world conference on educational multimedia, hypermedia and telecommunications 2001* (pp. 1755–1760). Norfolk, VA: Association for the Advancement of Computing in Education (AACE).

Vallée-Tourangeau, F., & Wrightman, M. (2010). Interactive skills and individual differences in a word production task. *AI & Society*, *25*(4), 433-439.