SPECIFICITY OF TRANSFER EFFECTS FROM VIDEO GAME TRAINING

SPECIFICITY OF TRANSFER EFFECTS FROM VIDEO GAME TRAINING

By JULIANA LOUREIRO-KENT, B.SC.

A Dissertation submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2015) Hamilton, Ontario (Psychology, Neuroscience and Behaviour)

TITLE: Specificity of Transfer Effects from Video Game Training AUTHOR: Juliana Loureiro-Kent, B.Sc. (McMaster University) SUPERVISOR: Dr. S. Watter NUMBER OF PAGES:

Lay Abstract

Playing video games may improve people's cognitive skills, but the current research in this field is mixed and prone to confounds. This dissertation developed and tested a better methodology to study how training on video games can lead to the transfer of learned skills from video games to other tasks and situations. We collaborated with Canadian video game company Telos International to modify and develop the commercial games used in this research. Over a series of studies, many participants all played the same commercial video game for several weeks. Within the game, the degree of difficulty of certain elements was manipulated, while the rest of the game was consistent for all players. By using the game as its own control, we were able to eliminate a variety of confounds, and make the transfer effects we found much more attributable to the particular well-controlled feature variations between training groups.

Abstract

Although the transfer of cognitive performance from video game training is a real possibility, the current literature on the topic is mixed, and prone to a range of confounds. This dissertation developed and tested an improved method for studying transfer of skills from video game training that uses the same game as its own active control. An industrial research collaboration with Canadian video game company Telos International provided a number of commercial games that were able to be modified towards these research goals. In Chapter 2, an initial proof-of-concept study assessed specific near-transfer effects from training on "Membrain", a 3D spatial memory game, versus training with Sudoku, a traditional number puzzle game. In Chapters 3 and 4, participants played the same commercial video game over several weeks, where the relative proportions of particular easier or harder game elements were manipulated to give different groups of participants more or less experience with particular kinds of game features. In Chapter 3, the "Paint the West" game (a speeded shooting gallery-style game) showed a range of specific transfer effects from increased distractor stimulus similarity, number and crowding within the game, to both improved performance on resisting the influence of nearby distractor stimuli in cognitive tasks, and also to changes in speed/accuracy criterion performance in other speeded tasks. In contrast, in Chapter 4, the "Orphlings" game (a problem-solving spatial puzzle game) showed no convincing transfer effects with a range of working memory and spatial tasks. We suggest that these methods allow for a better estimate of the true effect size of game-specific training improvements, and that the transfer of training effects observed in this research is more directly and reliably attributable to the particular well-controlled feature variations between training groups within the same game context.

Acknowledgements

I would like to thank Dr. Humphreys and Dr. Watter for all of their support, I could not have done this without their help. Thanks to Dr. Milliken for his help editing this manuscript and thanks to Dr. Karle for his collaboration in developing the experiments and the games used for all the experiments. Thanks to Telos Inc. for providing us with funding and developing the games used in our experiments. I also wish to thank Esther Manoian for her invaluable help running the experiments. And a final thanks to my family and friends for supporting me emotionally through this whole process.

Table of Contents

List of Tables and Figures

Declaration of Academic Achievement

I declare that the thesis entitled "Specificity of Transfer Effects from Video Game Training" submitted by me for the degree of Doctor of Philosophy is the record of work carried out by me between June 2011 and May 2015 under the guidance of Dr. Scott Watter. Data was collected by Esther Manoian and myself, and analysed by Dr. Watter and me. The games used in this experiment were developed by Telos Inc, with the assistance of Dr. Watter and Dr. Karle. Dr. Humphreys, Dr. Milliken and Dr. Watter provided invaluable help editing the manuscript and all materials obtained from other sources have been properly acknowledged in this manuscript.

Chapter 1 - Introduction

The idea that the adult brain is plastic and changeable is not a new one. This idea has been pursued in many guises, from demonstrations of training resulting in new learning in a wide variety of species, including humans, from general ideas about learning to more specific ideas about reorganization or tuning of specific behavioural processes. Other approaches have taken a more directly neurological approach, showing that that training could cause a significant reorganization of brain function. For example, Elbert (1995) showed that musicians with more than 11 years of experience playing a string instrument have a larger cortical representation for the fingers of their left hand than that of a control group of non-musicians. A different study showed that therapy following a stroke can result in cortical reorganization of the motor cortex (Liepert, Bauder, Wolfgang, Miltner, Taub & Weiller, 2000).

While it is unarguable that brains, even adult brains are plastic, critical questions remain as to what kinds of training can result in precisely what kinds of changes. One type of training that has received increasing amounts of attention as a potential way to manipulate neuroplasticity has been video game playing, which has become an increasingly ubiquitous pasttime over recent decades. Studies comparing expert video game players (EVGP) to novice video game players (NVGP) have shown that EVGPs have better hand eye coordination (Griffith, Voloschin, Gibb and Bailey, 1983), are better at mental rotation tasks (Sims & Mayer, 2002), and have faster reaction times than NVGPs (Castel, Pratt and Drummond, 2005).

We had two goals in mind when we first developed the experiments discussed in this dissertation. First, we wanted to question whether video games actually improve skills, or to what extent are the findings in previous studies a result of self-selection. To answer this question we used a training methodology in which we trained non-expert video game players on video games, and tested their skills on a variety of cognitive tasks before and after the training to assess the effects of the games. It is clear that practice on a particular video game makes one better at that game, but to what extent does the training on a game transfer to other types of task? Can transfer of skills occur with any task, or are there skills that are resistant to change? The second and possibly more important goal of this dissertation was to develop and demonstrate a better testing methodology for video game training that is relatively free from the confounds that plague the current video game training literature. Many of the differential improvements seen in previous studies could be easily attributed to confounds such as difference in arousal levels between groups, or expectations of training. We developed a methodology that uses the same game for all participants in the experiment, but with varying degrees of difficulty in specific training-relevant features of the game, to better control for these confounds. Training effects from this kind of design can be more directly interpreted as a result of different degrees of training with specific within-game experiences, and should give a more realistic estimate of effect sizes due to game-specific training effects.

Video Game Expertise

Griffith and his colleagues (1983) set out to research the potential benefits of video games and to establish a relationship between video game use and hand-eye coordination. These authors argued that not only do video games require people to have better hand-eye coordination, but that playing video games may also improve hand-eye coordination. To prove their hypothesis, Griffith et al. tested subjects on a rotary pursuit task. In this task participants used their hands to track a light moving at various rotations per minute (rpm) in a rotary pursuit unit. EVGPs were able to better track the light stimulus at faster rpms than NVGPs. The authors also found a favorable correlation between the number of months a participant had been playing video games and the overall performance on the task.

Green and Bavelier have been leading researchers in highlighting the differences between EVGPs and NVGPs. In one of their earliest studies on the topic of video games, they investigated the differences between the two groups' selective attention capabilities and the effect that video games had in altering the visual system of an individual (Green & Bavelier, 2003). To test these hypotheses the authors used the flanker task, requiring participants to respond to whether a stimulus presented in a circle of six rings was a diamond or a square shape, while ignoring a distractor presented outside the area of the six rings. Green and Bavelier added one to five extra distractor shapes in the outer circle. The logic of the paradigm is that as the task becomes more difficult, subjects are left with less attentional resources to be spared and therefore do not process the distractor item. When novice players reached a level where all of their attentional resources were depleted (i.e. distracting stimuli were no longer interfering with the processing of the target stimulus), expert players continued to be affected by the irrelevant distractors. The authors concluded that EVGPs had more attentional resources to spare than NVGPs.

Green and Bavelier (2003) also found that EVGPs can process a rapid stream of visual information more efficiently than NVGPs using an Attentional Blink task. In the Attentional Blink task, two target stimuli are presented within a rapid stream of stimuli. Following successful identification of the first target, participants often fail to detect or identify the second target stimulus if it is presented shortly after the target letter – this task 2 decrement is known as the Attentional Blink effect. The results showed that EVGPs have a relatively smaller attentional blink effect than NVGPs. This result suggests that VGPs take less time to recover from processing the first target stimulus, and shows that EVGPs are able to process information more quickly than NVGPs.

EVGPs have better visuospatial attentional capabilities than NVGPs. Green and Bavelier (2003) used a Useful Field of View Paradigm, a task that has been used to assess the driving fitness of the elderly, and modified it to show how attention is distributed across the visual field (Ball, Beard, Roenker, Miller and Griggs, 1988). Participants had to identify the location on the screen of briefly presented triangle shaped targets that could appear at any one of three target eccentricities (10, 20, or 30 degrees from the centre of the screen). First person shooter games typically require players to focus on targets within the 10 degree area, while the 20 degree range can be considered a boundary area of "training" within these games. Most of the stimuli in a first person shooter game will appear at or within the 10 degree area, but stimuli will still appear regularly out to a 20 degree area, requiring participants to devote some of their attention to the boundary of training. The 30 degree target area falls outside of the normal area of focus for video game players, and stimuli requiring immediate attention will rarely, if ever, appear in this area in a game. This 30 degree area was used to measure whether video game play can create changes outside of the area of focus in video games. Targets could be presented with square shaped distractors or without distractors. The distractor present trials are supposed to reflect the same conditions as a normal visual search. Expert players were faster and more accurate in identifying the location of the target at all eccentricities used in the experiment. This suggests that video games alter spatial attention throughout all areas of the visual field and not just the areas involved in the typical playing visual field of a video game.

Green and Bavelier (2006b) did a follow-up of their study and included conditions where participants had an extra, secondary task to complete. In one condition, participants were required to determine whether a triangle or a diamond was presented in the centre of the screen. In a second condition, this extra central task was removed. This was done to determine whether the enhanced peripheral localization that they found in their earlier study came at the cost of central localization. As with the first study, expert players continued to be faster and more accurate at localizing the target in all three eccentricities of the visual field. Expert players were also faster at localizing the targets when the task conditions included the secondary task. Finally, expert players were faster and more accurate at determining the identity of the stimulus on the secondary task compared to novice players. This demonstrates that EVGPs have enhanced visuospatial attentional capabilities at peripheral areas of vision that do not come at the cost of the spatial attention resources devoted to central areas of visual processing, at least under these typical performance loads.

In an enumeration task, participants are required to rapidly and accurately determine the number of items that are briefly presented on the screen. Performance in this task is very fast and accurate until participants reach about four items on the screen, also known as the subitizing range. As the number rises past four items, participants tend to either count or estimate the number of items on the screen. This results in longer reaction times and a decrease in accuracy. Subitizing is thought to be an automatic process that is limited to a small number of items, while counting seems to require a different process. Both serial attention and the capacity to count items using working memory are processes that are implicated in counting. Most importantly, the accuracy breakpoint for subitizing, where accuracy starts to drop, provides an estimate of the number of items that can be attended to simultaneously. This provides a good estimate of the capacity of the attentional system. Green and Bavelier (2006a) used a variation of this task in which participants were tested both on a narrow field of view (5 degrees square from the centre of the screen), and a wide field of view (20 degrees square from the centre of the screen). This was done to test whether playing video games improved peripheral vision disproportionately to central vision, or whether the benefits of video games were evenly distributed across the visual system. Green and Bavelier found that expert players had an accuracy breakpoint for subitizing that was two items larger than novice players. Expert players also continued to be more accurate than novice players as the number of items on the screen increased. Expert players were better than novice players at not underestimating the number of items presented. However, expert players became slower than novice players as the number of items increased. This result indicates that novice players suffer more from a speed/accuracy trade-off than expert players. This may be linked to novice players having a faster loss of information in working memory than expert players.

In the same study as described above, Green and Bavelier (2006a) used a Multiple Object Tracking (MOT) paradigm (Pylyshyn & Storm, 1988) to determine whether expert players performed better than novice players when keeping track of objects that were moving simultaneously. In a MOT task, subjects need to keep track of circles that were cued at the start of the task as they randomly move around a screen. This paradigm provides an index of the number of items that can be attended to simultaneously and allows for the measurement of whether EVGPs are on average better able to track items.

Expert video game players and novice players had similar accuracy when tracking one moving circle, but as the number of circles increased, EVGPs' accuracy improved compared to NVGPs. This gap in accuracy levels of both groups continued to increase as the number of circles that had to be tracked increased to five. At six items and higher, accuracy levels for NVGPs and EVGPS became similar again.

Green and Bavelier (2007) tested whether EVGPs have better spatial resolution of the visual field than NVGPs by measuring the difference in the size of the crowding region for both groups. Crowding refers to the diminished ability of the visual system to identify a target when it is surrounded by distracting objects as opposed to when the target is displayed alone. Participants in this experiment were asked to identify the orientation of a T-shape flanked by two T-shape distractors that were either congruent or incongruent with the target T-shape. The shapes were presented at one of three possible eccentricities (0, 10, and 25 degrees) of the normal game playing field of view. Across all eccentricities, EVGPs' crowding regions were smaller than those displayed by NVGPs. This suggests that expert players have better visual acuity thresholds and smaller regions of spatial interaction.

A meta-analytical review paper (Dye, Green, & Bavelier, 2009) found that EVGPs were faster than NVGPs across a variety of tasks, such as those used to measure Spatial Cuing, Inhibition of Return, Simon, Flanker, N-Back, and Attention Network effects. The increased speed displayed by expert players did not come at a cost of accuracy, or as a result of expert players being 'trigger happy'. Overall, video games may be involved in giving players many skills that are valuable in the real world. The results from the Useful Field of View and Attentional Blink studies, among others, suggest that EVGPs may be better at noticing a "needle in a haystack" than NVGPs. EVGPs may be better at avoiding collisions while driving by being better able to detect animals or children that may be running towards the car. Expert video game players may also make better athletes or better air traffic controllers because of their better visuospatial capacity than NVGPs.

However, there is a possibility that self-selection plays a role in the above tests and that a pre-existing ability drives people to become EVGPs. It is possible that people with better spatial attention skills, or better hand-eye coordination are drawn to video games and find the activity enjoyable, causing them to continue playing the games. The people whose spatial attention and hand-eye coordination skills are not quite as good may find video games to be too challenging and are not likely to enjoy the activity. Of course, this self-selection could interact with actual training effects as well – those with better pre-existing visuospatial abilities may self-select to play games more, and in turn receive further benefits of game training. An important question, then, is what the relative contribution of individual differences versus game training is, for all of these observed EVGP versus NVGP differences described above, and what evidence exists to distinguish the relative contribution of these factors.

Self-Selection and Training Studies

Boot et al. (2006) suggested that at least some of the effects shown by other studies on gamers' enhanced cognitive abilities may be due to some pre-existing differences between them and people who do not play video games, thus creating the question of 'do video games improve cognitive abilities?', or 'do people with particular cognitive abilities get drawn to video games?'. The pre-existing condition argument has been countered by studies on brain plasticity throughout development that would indicate that extensive practice of video games throughout childhood by a person would lead to larger areas of the brain being dedicated to cognitive functions than if that person started playing at a later age in life (Dye et al., 2009). And although the brain does experience a critical period in life in which plasticity occurs at an optimal level, it is not impossible for the brain to undergo some plasticity throughout adulthood (Bavelier, Levi, Li, Dan, Hensch, 2010). It is possible, and likely, that video game expertise effects are a result of both pre-existing abilities and practice, but training studies are needed to establish a causal link between video games and improved cognitive abilities.

Green and Bavelier (2003) carried out a training experiment to test whether video game expertise effects could be induced in NVGPs. The training experiment consisted of a pre-test, training and post-test regimen that included Attentional Blink as one of the tasks to be tested. They trained a group of NVGPs on the game Medal of Honor (action video game group) for one hour a day over a period of ten days. Another group of NVGPs were trained on the game Tetris (control group) for the same period of time. Medal of Honor was the chosen game because it is similar to those often played by EVGPs. Green and Bavelier argued that this game required visual-manual effort and coordination from the players, and required that players must also switch their attention around numerous objects across the visual field. Tetris was chosen due to its similar requirements for visual-manual control, but a lack of demand on the attentional system, requiring participants to only look at one object at a time. Even with only a short duration of training on action video games, the action video game group was able to improve their performance on the Attentional Blink task, showing a reduction in the amount of time it took them to recover from the attentional blink. In contrast, the control group showed a similar level of performance relative to their pre-test. In the same experiment as above, Green and Bavelier (2003) tested participants on the Useful Field of View paradigm. Participants who were trained on Medal of Honor improved their accuracy by 20% in all eccentricities post-test, showing significant improvement in their performance.

Green and Bavelier performed a follow up study (2006b) to determine whether the enhancement of peripheral vision observed in their earlier study came at the cost of performance in central vision. In the follow up study, the action video game group showed significant improvement in their performance at the post-training phase. The follow-up experiment also included conditions with and without a secondary task. This secondary task required participants to determine whether a triangle or a diamond was presented in the centre of the screen. Expert players were faster at localizing the targets when the task conditions included the secondary task than novice players. In the training part of the experiment, NVGPs trained on an action video game were again faster and more accurate at localizing the target in all three eccentricities of the visual compared to the performance of the NVGPs that played Tetris. Although EVGPs were shown to be faster at determining the identity of the stimulus in the secondary task, both groups of NVGPs in the training study failed to improve their performance on the secondary task. This suggests that some of the benefits of video game playing may require a longer period of training.

Participants who played Medal of Honor (action game) compared to those who played Tetris showed an improvement in the enumeration task (Green and Bavelier 2003, 2006a). Following training, the action video game group was able to rapidly and accurately enumerate an average of 1.7 more items than in their pre-test performance. This shift in the accuracy breakpoint for subitizing was achieved with only 10 hours of training. Action video game participants also showed an improvement in performance past the subtizing range, whereas participants from the control group showed no such improvements.

Green and Bavelier's training study (2007) on crowding tested whether training novices on action games would result in better spatial resolution of the visual field than training novices on Tetris. The results showed that crowding regions were smaller across all eccentricities for players in the action video game group compared to those displayed by the control group. This suggests that video games can be used to improve visual acuity thresholds and alter the regions of spatial interaction.

Training and Transfer of Skills

Despite these many positive training results, training participants in video game play is not the equivalent of a magic pill that will give the user superior abilities in all areas of mental processing. It is also not likely that all cognitive and perceptual processes will undergo the same level of improvement. Different aspects of various video games will make different demands on different cognitive processes, so it stands to reason that each unique video game genre, such as first person shooter games or racing games, may differentially affect transfer of training to different cognitive abilities.

Nelson and Strachan (2009) explored how different genres of video games may affect different aspects of executive processing. They also investigated how gameplay may affect the strategies a player uses (i.e., in terms of speed and accuracy). The results show that for spatial localization tasks players who trained on the video game Unreal Tournament (action game) used a speed based strategy on the cognitive tasks, resulting in faster response times. In contrast, players who trained on the game Portal (puzzle game) used an accuracy based strategy on the cognitive tasks.

Many of the studies mentioned above have used a range of commercially available video games to train participants. All of those games possess multiple characteristics that may affect cognitive skills differently, thus stopping us from knowing specifically which characteristics of the video game caused improvements in a specific cognitive ability, such as spatial attention or executive control.

Before considering the experimental question of how specifically or generally training obtained by playing video games might transfer to other tasks, it is appropriate to discuss what is currently understood about training and transfer of cognitive skills more generally, and the types of transfer that can occur. Near transfer refers to the transfer of performance between tasks that are very similar. Far transfer refers to learning in a task and later applying that learning to a task that greatly differs across various dimensions (e.g. including location, time and modality).

Barnett and Ceci's (2002) paper on "Taxonomy for Far Transfer" provides a set of rules and nine dimensions that help distinguish between the possible types of transfer. To date no training programs or experiments focused on video game research have produced effects would be classified as far transfer of skills based on the dimensions described by Barnett and Ceci. The authors indicate that it is unlikely that any form of training would fall in the far spectrum of the transfer scale. When it comes to video game training, the only far context transfer that occurs is in the functional dimension when participants go from playing video games to being tested on a battery of cognitive tasks. The knowledge domain would not be applicable to these kinds of studies, as we are looking into cognitive processes that change with video game practice and not knowledge that can be gained from playing video games. The other context factors, such as the temporal, physical, social context and the modality dimensions, all fall in the near transfer spectrum of Barnett and Ceci's scale.

In general, we might expect training on video games to show transfer to tasks that share similar surface or structural features – so-called near transfer. Far transfer, on the other hand, where transfer occurs in the absence of these similarities, is rarely (if ever) observed. Far transfer would suggest effects mediated by training changes to fundamental processes common to a wide array of tasks. We note that there is currently a parallel focus in training and transfer in the working memory / fluid intelligence literature, where near transfer is easily observed, but far transfer is much less convincing with careful methodological controls (e.g., Harrison, Shipstead, Hicks, Hambrick, Redick & Engle, 2013).

The Present Experiments

The first study in this thesis (Chapter 2) presents an initial smaller-scale project, designed as an initial proof-of-concept study, to test whether relatively brief training on a non-action video game could induce game-specific training and transfer effects. We compared the results of training individuals on “Membrain”, a 3D spatial memory game, versus training on the traditional number puzzle game Sudoku. Both the Sudoku game and the “Membrain” memory game were played on the same iPod touch hand-held devices, to equate possible transfer of skills related to manual response mapping and general visual-manual coordination when interacting with these devices. As an initial approach, this study investigated specific near transfer that in principle might occur from playing Membrain. Considering the requirements and cognitive demands of this game, we investigated changes from pre- to post-test mental rotation and working memory (n-back) performance, as well as a more general test of long-term associative memory.

The second and third studies (Chapters 3 and 4) in this thesis developed and applied a more carefully controlled training-and-transfer methodology, where all participants across a set of different game training conditions played the same commercial video game. Within this same game, particular elements of gameplay were adjusted to give higher versus lower proportions of easier versus harder kinds of target game stimuli, distractor information, working memory loads, and other factors. Importantly, these differences occurred within a consistent and universally challenging game experience for all participants. Within this design, the question of training and transfer relates to differences in the proportion of harder versus easier elements experienced throughout game training – for example, does experiencing a higher proportion of visually similar versus dissimilar distractor stimuli during gameplay lead to better filtering of distractor information in laboratory tasks?

This relatively straightforward design allows a much more controlled and selective approach to testing for training and transfer effects. In general, training effects will have smaller magnitudes, because all participants are playing the same game – for example, if distractor similarity within the game gives some kind of training to participants, the transfer effects we see as training group differences measure the relative difference between two different amounts of this same kind of training. While these measured transfer effects may be relatively small, those that we do observe can be more directly linked to particular process-specific factors within the game. Two such studies are presented here, one focused on a speeded target response game ("Paint the West" in Chapter 3), and the other using a spatial puzzle game ("Orphlings" in Chapter 4). A larger number of pre- and post-test tasks were used in these studies to try and observe relevant transfer of training effects.

References

Ball, K. K., Beard, B. L., Roenker, D. L., Miller, R. L., & Griggs, D. S. (1988). Age and visual search: Expanding the useful field of view. *Josa a, 5*(12), 2210-2219.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological bulletin, 128*(4), 612-637.

Bavelier, D., Levi, D.M., Li, R.W., Dan, Y. & Hensch, T.K. (2010). Removing brakes on adult brain plasticity: from molecular to behavioral interventions. *The Journal of Neuroscience,* ***30*****(45)**, 14964–14971.

Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Psychologica, 129*(3), 387-398.

Dye, M. W., Green, C. S., & Bavelier, D. (2009). Increasing speed of processing with action video games. *Current Directions in Psychological Science, 18*(6), 321-326.

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. *Science*, *270*(5234), 305-307.

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*(6939), 534-537.

Green, C. S., & Bavelier, D. (2006). Enumeration versus multiple object tracking: The case of action video game players. *Cognition, 101*(1), 217-245.

Green, C. S., & Bavelier, D. (2006). Effect of action video games on the spatial distribution of visuospatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 32*(6), 1465-1478.

Green, C. S., & Bavelier, D. (2007). Action-video-game experience alters the spatial resolution of vision. *Psychological Science, 18*(1), 88-94.

Griffith, J. L., Voloschin, P., Gibb, G. D., & Bailey, J. R. (1983). Differences in eye-hand motor coordination of video-game users and non-users. *Perceptual and Motor Skills, 57*(1), 155-158.

Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychological Science,* 2409-2419.

Liepert, J., Bauder, H., Miltner, W. H., Taub, E., & Weiller, C. (2000). Treatment-induced cortical reorganization after stroke in humans. *Stroke, 31*(6), 1210-1216.

Nelson, R. A., & Strachan, I. (2009). Action and puzzle video games prime different speed/accuracy tradeoffs. *Perception, 38*(11), 1678-1687.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism*. *Spatial Vision, 3*(3), 179-197.

Sims, V. K., & Mayer, R. E. (2002). Domain specificity of spatial expertise: The case of video game players. *Applied Cognitive Psychology, 16*(1), 97-115.

Chapter 2 – "Membrain"

As an initial project in this series of video game training and transfer studies, we sought to test a commercial video game that might offer a suitable training experience that could be detected with typical cognitive tasks. We established a research collaboration with Canadian video game study Telos International, supported by a Natural Sciences and Engineering Research Council (NSERC) Engage industrial research grant, and considered which of their existing games might be suitable. We selected "Membrain," a 3D spatial pair-matching memory game, shown in Figure 1. Membrain is a classic pairs memory game, where a large set of items are shown with identities obscured. Participants select two items to reveal, searching for matching pairs. If the pairs match, they are removed from the game; if they don't match, the two items' identities are obscured again, and participants select another two items to reveal. Participants need to remember where previous specific items were in order to match newly revealed items to previously revealed items, to match all the pairs in the fewest number of moves. Membrain implements this memory game with items arranged in a 3D spatial array, which participants must rotate around in order to access and see all potential items – this requires participants to represent and encode item-location information relative to a particular location on a moving 3D object, rather than in a static location in 2D space.

Membrain seemed to emphasize a number of cognitive requirements in its gameplay which are generally seen to show improvement with practice in related laboratory tasks. The need to represent and keep track of different parts of a larger 3D shape while that shape is being rotated in 3D space aligns with the canonical performance strategy of the Shepard and Metzler (1971) (paired match/non-match decision) mental rotation task. The need to maintain a good representation of recently observed items in the face of interference from other items may align with running working memory tasks, for example the n-back task. Both mental rotation and n-back performance show considerable practice effects within the laboratory, and we might predict that if Membrain requires similar kinds of task requirements or abilities, these might be trained through playing Membrain, and transfer to standard mental rotation and n-back task performance in the laboratory, outside of the gameplay context.

As the first study in this larger series of training and transfer projects, Membrain offers an opportunity for a proof of concept study, where we should have fairly direct predictions about how transfer might occur from gameplay to laboratory task performance. As a contrasting control task, we chose a computerized implementation of the classic Sudoku number puzzle game as an active training control condition. In contrast to Membrain, we would expect Sudoku to have relatively little 3D spatial representational/manipulation demand, though there is a considerable 2D demand to search 2D spatial arrays for numbers. Similarly for n-back working memory-like demands, Sudoku requires an algorithmic (or sometimes trial-and-error) search through possible number solutions for the puzzle, but the numbers presented on screen essentially inform the participant what remaining digits must be used; we suggest Membrain may push participants to explicitly and actively use working memory in service of the game. As a complement and contrast, we also included a long term memory task to the set of laboratory tasks. In general terms (in this study and those in subsequent chapters), we ask what characteristics of a video game influence specific cognitive abilities, and whether differential changes in post-training performance might be specifically related to video game training.

Methods

Participants

Twenty-five participants (fourteen females), ages ranging from 18 to 30 years, (mean of 21.5 years) took part in the experiment and were paid $100 for their participation. No restrictions were placed regarding participants' first language or regarding their handedness (5 participants were left handed). All participants reported normal or corrected to normal vision. One participant did not finish the experiment (female, right handed) and her data were excluded from the experiment, leaving a total of 24 participants in the study. Participants were recruited through the Experimetrix online recruitment system used by McMaster University. Our recruitment notice requested that participants have little to no experience playing video games. Participants with considerable video game experience were not entered into the experiment based on their answers to a pre-experiment screening questionnaire. The cut off for video game experience was 4 or more hours per week playing video games over the last 6 months preceding the experiment. Participants were randomly assigned to the Membrain group or the Sudoku group (twelve participants per training group).

Apparatus

Participants were tested in a computer laboratory room at McMaster University in groups of up to 5 people, while the experimenter supervised from a chair just outside of the room in the open doorway. Presentation software (www.neurobs.com) was used to create and run a set of experimental tasks, which were presented on 22 inch LCD monitors using typical PC computers running Windows 7. Participants completed the same set of cognitive experimental tasks (mental rotation, n-back, and picture memory) on three separate days, with a number of gameplay training sessions in between (see below for study procedure).

Video Game Training

The game training portion for all participants was conducted on iPod Touch devices. Sudoku is a traditional number puzzle game, where the numbers 1 to 9 need to be arranged in a series of grids so that each digit only appears once on each vertical and horizontal line. We used a commercial iPod Touch version of Sudoku from the Apple App Store that provided a standard version of the Sudoku puzzle game, with colourful and engaging colours and graphical design.

"Membrain" is a 3D spatial symbol-matching memory game developed by video game company Telos International Inc. The work in this dissertation (here and in other chapters) was supported by a Natural Sciences and Engineering Research Council (NSERC) Engage industrial research grant, to independently evaluate the capacity of a number of Telos games to produce training and transfer effects in various cognitive and perceptual domains. Figure 1 shows an example of the Membrain game, and how it is played. Participants have to discover matching pairs of symbols, arranged (and hidden) on a set of spatially arranged cubes in 3D space. Participants can move the display viewpoint around to inspect the cubes from all angles. Participants select two cubes to reveal their hidden symbols – if they do not match, the symbols are hidden again after a brief delay (approximately 1 second after the second symbol is revealed); if they do match, the participant is rewarded with points in the game, those two cubes disappear, and the game continues. The goal is to remember where various symbols have been temporarily discovered, and match pairs of symbols with the minimum number of turns in the game. The 3D nature of the game requires participants to move around the stimuli in 3D space, keeping track of where previously revealed symbols were placed, including positions occluded as the display is moved around. The game becomes progressively harder over time, with larger display sizes, and a transition to the game rotating the display rather than this being under the participant's control.

Experimental Tasks

Participants completed the same set of experimental laboratory tasks on three separate occasions, in the same order. In order, participants completed: 1) the study portion of the picture memory task; 2) mental rotation task; 3) n-back task; 4) test portion of the picture memory task; and 5) a mental rotation performance questionnaire.

Picture Memory task: Participants were shown four sets of 10 pairs of objects (one set per page, on paper), and given one and a half minutes to try to memorize each set. An example picture set is shown in Figure 2. Picture sets at study consisted of 10 pairs of interacting foreground and background objects. After completion of the mental rotation and n-back tasks, participants were presented with a modified version of the picture sets they had originally seen, showing only the background objects without the foreground objects, and were asked to recall the missing paired items. Twelve different 10-item sets of picture pairs were used, with counterbalanced presentation of 4 unique sets for each participant for each of the three experimental sessions.

Mental Rotation Task: Our mental rotation task was akin to the Shepard and Metzler (1971) paradigm, where two pictures of similar block figures were presented simultaneously at varying degrees of rotation, and participants responded whether the two stimuli represented the same object or not. We used a set of high-resolution line drawing stimuli developed for this kind of mental rotation paradigm by Peters and Battista (2008).

The Peters and Battista (2008) stimuli represent 16 shapes and their mirror images, with stimulus pictures rotated at increments of 5 degrees along X, Y or Z axes. In the present study, we used only stimuli rotated on the Z axis (rotated around a vertical axis). For each trial, a stimulus pair was determined as follows: first, we randomly selected one of the basic stimulus shapes from the set, and randomly selected a starting orientation for that stimulus between 0-360 degrees in increments of 5 degrees, but excluding those within the range of 15 degrees from cardinal directions (0, 90, 180 and 270), so as to avoid visual ambiguity and occlusion of stimulus elements. To create Match trials, stimulus 2 was the same stimulus type with a rotation of 20, 60, 100 or 140 degrees higher than that of stimulus 1. To create Non-Match trials, we first generated a pair of Match trial stimuli, and then substituted the complementary opposite-handed version of one of these two stimuli from the Peters and Battista (2008) stimulus set. Within both match and non-match pairs, stimuli 1 and 2 in each pair were presented in randomized left versus right positions on screen for the matching task.

Participants were required to press the left mouse button if they believed that the stimuli on the screen were a match, and to press the right mouse button if they believed that the stimuli on the screen were not a match. Participants were instructed to mentally rotate the pictures in their minds to compare them, to determine whether they were the same object or not. Participants were instructed to be as quick as possible selecting their response and to maintain high accuracy. The two stimuli were presented side by side on the computer screen, and remained onscreen until the participant made a response. The subsequent trial started after a 1000ms blank screen. Participants were instantly notified if they had made an incorrect response for each trial with the word “wrong” in red, capital letters appearing on the screen for 1000ms followed by a 1000ms break before the appearance of the next stimuli. Participants were presented with a total of 120 pairs of stimuli divided into three blocks of 40 pairs each, with a participant controlled break period in between each block. At the end of each block, participants were told their accuracy scores and their average reaction time. At the end of each testing day (following the picture memory test phase), participants completed a short questionnaire to determine what strategies were used during the performance of the mental rotation task (see Appendix A).

N-Back task: The particular n-back task used for this experiment was a verbal working memory task in which participants were required to monitor a continuous sequence of digits randomly presented one at a time at a central location on the computer screen, and asked to respond on whether the identity of the current digit matched the identity of the digit presented n items previously. Participants were instructed to press the left mouse button if they believed the current number matched the number presented n items before, or to press the right mouse button if they believed the digits were not a match.

A total of six n-back blocks were completed, alternating between 1-back blocks (where the stimulus on the screen was compared to the stimulus that preceded the current trial) and 2-back blocks (where the current stimulus was compared to the stimulus presented 2 trials before the current trial on screen). Each block consisted of 30 trials, using digits that ranged from 1 to 9. Digits were presented serially and at the centre of the screen, in black, Arial font, size 36, on a gray background. Each stimulus remained on the screen for 1500ms with an interval of 1000ms given between each stimulus. Participants' response to each stimulus had to occur in the 2500ms allotted to each trial or it would count as a missed response. No feedback was given as to whether participants' responses were correct or not at the end of each trial, but at the end of each block participants were shown their accuracy score and average reaction time.

Procedure

Participants were tested in groups of up to five people. Experimental task sessions were performed on the first, second and eighth days of the experiment. On the first day of the experiment, participants completed a questionnaire on their video gaming habits, and only participants meeting inclusion criteria were entered into the experiment. Participants were excluded from completing the rest of the experiment if their answers showed four or more hours of video game playing time per week over the six months preceding the experiment. Included participants then completed an experimental tasks session on this first day.

On the second day of the experiment, participants played either the Membrain or Sudoku game for 50 minutes, after which they completed a second set of experimental tasks, minus the final mental rotation questionnaire. After the second day, participants completed a further five non-consecutive days of game playing, each day consisting of 50 minutes of play, spread over a nine day period, for a total of 300 minutes of video game playing time. The final day (day 8) of the experiment consisted of a third experimental tasks session, including the final mental rotation questionnaire again.

Results

Mental Rotation

Two participants from each group were excluded from analyses due to accuracy below 50% on one or more conditions in the experiment.

Mean reaction time data for mental rotation Match trials are shown in Figure 3, separated by stimulus angle and group. Data were analyzed via repeated measures ANOVA, with within-subjects factors of Day (1, 2 and 8), and Angle (20, 60, 100, 140 degrees), and a between-subjects factor of gameplay group (Membrain, Sudoku).

Results typical of mental rotation tasks were observed. A strong main effect of angle was observed, $F(3, 54) = 97.59$, $p < 0.001$, $\eta_p^2 = 0.84$, reflecting progressively longer RTs with increased difference in stimulus angle. A strong main effect of day was also observed, $F(2, 36) =$

41.45, $p < 0.001$, $\eta_p^2 = 0.70$, reflecting progressively faster RTs over days of performance. The angle by day interaction was also significant, $F(6, 108) = 7.81$, $p < 0.001$, $\eta_p^2 = 0.30$, reflecting a relatively smaller difference in RT across angles as days progressed.

While all participants improved their mental rotation performance over time, those in the Membrain group appeared to show a relatively greater improvement over days compared to Sudoku participants. This observation was supported by a significant interaction of participant group with angle and day, $F(6, 108) = 2.25$, $p = 0.044$, $\eta_p^2 = 0.11$, that modified the marginal interaction of group and day, $F(2, 36) = 2.90$, $p = 0.068$, $\eta_p^2 = 0.14$. The main effect of participant group was not significant, $F(1, 18) = 0.045$, $p = 0.834$, nor was the interaction of group and angle, $F(3, 54) = 1.58$, $p = 0.204$. To more directly assess differences between game groups, we calculated the minimum significant difference between means ($p < 0.05$), based on the error term for the significant 3-way interaction of day, angle and training group. Comparing pairs of means for Membrain versus Sudoku performance across days and stimulus angles (see Figure 3), mental rotation performance was significantly faster for Membrain participants in 60 and 100 degree stimulus conditions, on both Day 2 and Day 8; in comparison, on Day 1, performance for Membrain participants was significantly slower in 60 and 140 degree conditions, all 1-tailed.

Considering the very long reaction times involved with mental rotation, and potential concerns about extreme RT values contaminating results, we re-analyzed our data limiting very long RTs to a maximum of 8000 ms (trials with correct responses and reaction times over 8000 ms had their reaction times replaced with an 8000 ms value in the raw data set). This analysis approach may reduce effects of angle by compressing RT distributions more for larger-angle trials, but should provide a stricter test of differences between groups by limiting the influence of very long outlier trials.

Very similar effects were observed, with strong main effects of angle, $F(3, 54) = 155.21$, $p < 0.001$, $\eta_p^2 = 0.90$, and day, $F(2, 36) = 75.62$, $p < 0.001$, $\eta_p^2 = 0.81$, and a significant angle by day interaction, $F(6, 108) = 8.75$, $p < 0.001$, $\eta_p^2 = 0.33$. Membrain participants again showed more improvement over days versus Sudoku participants, supported by a significant interaction of gaming group and day, $F(2, 36) = 3.32$, $p = 0.048$, $\eta_p^2 = 0.16$, with a marginal interaction of gaming group, day and angle, $F(6, 108) = 1.99$, $p = 0.073$, $\eta_p^2 = 0.10$.

To assess accuracy in mental rotation performance, we calculated d-prime measures based on Match and NonMatch trial performance, and analyzed them as per RT data above. These d-prime data are shown in the bottom half of Figure 3, and corresponding mean accuracy data for both match and non-match trials is shown in Table 1. Participants showed overall good accuracy and sensitivity as reflected by d-prime scores. Participants were more accurate at smaller stimulus angles, with a strong main effect of angle, $F(3, 54) = 38.28$, $p < 0.001$, $\eta_p^2 = 0.68$. Participants also improved their sensitivity over days, with a main effect of day, $F(2, 36) = 3.31$, $p = 0.048$, $\eta_p^2 = 0.16$. There was no interaction of angle by day, $F < 1.4$. D-prime scores were numerically larger for the Membrain group across most conditions, but this effect did not reach significance, $F(1, 18) = 2.79$, $p = 0.112$, $\eta_p^2 = 0.13$, with interactions of group with other factors also not significant, Fs < 1.5.

Questionnaire data asking about mental rotation performance and strategy use was collected at the end of experimental sessions on day 1 and day 8. The questionnaire is included in Appendix 1. Three of the total five questions asked about methods used to aid performance of the task (visualization, verbalization, external aids), while the other questions focused on consistency of methods used and confidence in answers. Participants from both Membrain and Sudoku groups had extremely consistent and similar answers. Very few participants (0/12 Membrain, 1/12 Sudoku) reported using verbalization, very few reported using any external aids (2/12 Membrain, 1/12 Sudoku), and a similar number per group reported trying varying strategies (4/12 Membrain, 3/12 Sudoku) while almost all others used a consistent strategy (7/12 Membrain, 9/12 Sudoku). Most participants reported feeling vaguely confident of their answers before moving on to the next problem (10/12 Membrain, 10/12 Sudoku). Only answers to question 1 (visualization) initially seemed to differ between groups. All twelve participants in the Sudoku group picked the statement "I rotated the whole figure in my mind while making the comparison" while only seven participants in the Membrain group picked that statement. One Membrain group participant picked the statement "I rotated a section of the figure in my mind while making the comparison, and the four remaining participants picked "other, describe". These participants' descriptions all reported using a combination of rotating the figure as a whole and rotating sections of the figure in their mind, suggesting a very similar approach to those in the Sudoku group. All of these responses were extremely consistent between initial and final testing sessions, with no indication of changes in strategy or approach to the task over practice or game experience.

N-Back

Data for one participant in each group were excluded based on poor task performance (less than 50% accuracy on one or more days). Data from one additional participant in the Sudoku group were lost due to computer malfunction.

Data analysis focused on RT performance on Target and Filler (non-Target) trials, and accuracy performance as calculated by d-prime measures from Target and Filler trials. RT data analyses for correct trials excluded a small number of trials with RTs faster than 300 ms or slower than 1500 ms. Mean RT data are shown in Figure 4. RT data were analyzed via repeated measures ANOVA, with within-subjects factors of memory load (1-back, 2-back), trial type (Target, Filler) and day (1, 2 and 8), and a between-subjects factor of game training group (Membrain, Sudoku).

Data typical of the n-back paradigm were observed. All participants showed faster RTs for 1-back versus 2-back trials, $F(1, 19) = 38.81$, $p < 0.001$, $\eta_p^2 = 0.67$, and correct responses were made more quickly over days, $F(2, 38) = 17.99$, $p < 0.001$, $\eta_p^2 = 0.49$. RTs for target trials were faster than filler trials in some conditions, with greater differences in 1-back versus 2-back trials, $F(1, 19) = 5.17$,$p = 0.035$, $\eta_p^2 = 0.21$, and on Day 1 versus later days, $F(2, 38) = 5.09$, $p = 0.011$, $\eta_p^2 = 0.21$. The main effect of target versus filler trial type was not significant, $F(1, 19) = 2.34$, $p = 0.143$.

Game training group differences were observed within this pattern of n-back task effects. In the 1-back task, both gaming groups appeared to improve over days in a similar fashion. In the 2-back task, the Membrain group appeared to improve progressively over days, while the Sudoku group showed minimal improvement in 2-back RTs. This observed pattern of data was supported by a significant interaction of game group, day and memory load, $F(2, 38) = 5.91$, $p = 0.006$, $\eta_p^2 = 0.24$. Day by game group and memory load by game group interactions were not significant, Fs < 1.4, strengthening the interpretation of development of this game group difference in 2-back performance over time. To further test this apparent selective training effect in the 2-back task, we reanalyzed these RT data separately for 1-back and 2-back tasks. In the 1-back data, there was no effect of gaming group by day, $F(2, 38) = 0.36$, $p = 0.702$. In the 2-back data, we observed a significant day by gaming group interaction, $F(2, 38) = 3.42$, $p = 0.043$, $\eta_p^2 = 0.15$.

Accuracy data for target and filler trials are shown in Table 2, separated by day, memory load and gaming group, along with calculated d-prime scores for each condition. D-prime scores were analyzed as per RT data above, minus the trial type (target, filler) variable used in calculating d-prime scores, and are also shown in the bottom half of Figure 4. There was a general improvement in accuracy over days, $F(2, 38) = 13.17$, $p < 0.001$, $\eta_p^2 = 0.41$, and participants were more accurate overall in 1-back compared to 2-back tasks, $F(1, 19) = 28.97$, $p < 0.001$, $\eta_p^2 = 0.60$. From d-prime data in Figure 4, participants in the Sudoku group were less accurate on day 1 compared to Membrain participants, with both groups achieving extremely similar performance accuracies across both n-back tasks on day 2 and day 8. This observation was supported by an interaction of day and gaming group, $F(2, 38) = 4.26$, $p = 0.021$, $\eta_p^2 = 0.18$, modifying a non-significant interaction of gaming group, $F(1, 19) = 2.92$, $p = 0.104$, $\eta_p^2 = 0.13$. Within this pattern of data, the numerically larger difference between Membrain and Sudoku d-prime performance on day 1 showed only a non-significant interaction between memory load, day, and gaming group, $F(2, 38) = 2.19$, $p = 0.126$, $\eta_p^2 = 0.10$. Interactions between memory load and group, and between memory load and day, were not significant, Fs < 0.6.

The observed poor n-back accuracy in the Sudoku group on day 1 is primarily being driven by two participants who apparently took longer than other participants to understand the task – limiting day 1 data to the second half of performance greatly reduces this day 1 d-prime difference between groups. This suggests that the day 1 accuracy differences are not representative of the whole Sudoku vs Membrain group, and likely are not the cause of the RT differences seen by day 8 for the 2-back task.

Picture Memory Task

Data for the picture memory task are shown in Figure 5. Memory performance was best on day 8 for all participants, with variable performance across groups on days 1 and 2, reflected by a main effect of day, $F(2, 44) = 6.45$, $p = 0.003$, $\eta_p^2 = 0.23$. Differences between gaming groups over days were not reliable, with no significant interaction of day and gaming group, $F(2, 44) = 1.38$, $p = 0.26$, and no main effect of gaming group, $F < 1$.

Discussion

This initial study examined whether training on the Membrain video game would show selective transfer (improved performance versus active control training on a Sudoku game) to a related set of cognitive laboratory tasks. Consideration of the nature of the Membrain game lead to the selection of mental rotation and digit n-back tasks, which might share similar cognitive requirements but have different apparent surface features and contexts. Sudoku, in contrast, did not in theory involve the same degree of these cognitive demands. We expected all participants to improve in all laboratory tasks over testing sessions, through increasing experience and practice with those laboratory tasks. A greater degree of improvement for the Membrain training group in specific areas of performance would suggest some degree of transfer of training from Membrain to our laboratory tasks.

A number of game-specific training effects were observed. The mental rotation results show that while both training groups improved their RTs over days, participants in the Membrain group had a greater improvement in performance over days, with relatively larger effects at greater stimulus angles. Limiting RT data to a maximum of 8000ms removed the interaction of gaming group, day and angle, instead showing a significant gaming group difference by day.

These differences in performance were not accompanied by a drop in accuracy for the task, and both groups had similar overall accuracy. We did not see any differences between groups on the mental rotation questionnaire, suggesting that both groups used similar processing strategies to complete the mental rotation task.

These findings may reflect transfer of learning from game performance to specific requirements of mental rotation, selectively for the Membrain game participants. Participants playing Membrain need to represent and keep track of different parts of a larger 3D shape while that shape is being rotated in 3D space, and to do so for novel feature-location-shape bindings, which we suggest is not something that people have much practice with on a daily basis. This manipulation of 3D objects in space and item-location binding might help with general 3D spatial manipulation practice which may be the basis for the greater reaction time improvement seen for the Membrain group. Previous studies have implied that RT improvement on spatial tasks such as mental rotation depend on the spatial abilities required to play the game (Ogakaki & Frensch, 1994), and we believe our data supports these implications.

These findings (and others below) suggesting transfer could represent a number of different mechanisms of improvement, from participants simply learning more efficient or consistent ways to represent spatial information, or the dropping out of less-useful approaches, to increasing the basic capacity of 3D visuospatial representation. A consistent message that will be argued within the present work is that the former is far more likely. This topic will be revisited as more data are considered; an in-depth discussion of these issues is reserved for the General Discussion.

The n-back results showed improved RT and accuracy for both Membrain and Sudoku groups on the 1-back task, but the Membrain group showed a substantial and selectively larger improvement in RT performance on the 2-back task. While the Sudoku group showed a greater accuracy improvement on the 2-back task, both groups had similar accuracy scores by the end of the experiment; d-prime analyses showed initial poorer performance by the Sudoku group that became equivalent to the performance from the Membrain group by day 2, and was comparable across 1-back and 2-back tasks. Additional analyses showing initial poor accuracy in the Sudoku group was largely due to early trials of just two participants, who subsequently achieved comparable performance to other participants by day 2, suggest that the RT differences seen over training for Membrain can be reasonably interpreted against comparable accuracy between the two groups.

The selective benefit on 2-back task RT for both target and filler trials for Membrain participants is particularly interesting, given the consistent and comparable improvement for both groups in the 1-back task. The 2-back task requires participants to keep track of items in sequence, update and reorganize several items in working memory, in addition to the monitoring and response decision demands for the immediate stimulus (Kirchner, 1958; Stigler, Lee, & Stevenson, 1986; Watter, Geffen & Geffen, 2001). The 1-back task embodies the same monitoring and response decision requirements, with a simpler requirement to remember only the most recent item, allowing participants to rely on simple perceptual memory without any manipulation or "working" aspect of working memory (Watter, Geffen & Geffen, 2001).

Membrain requires participants to keep several recent items in working memory and continuously update information about items and their location. We suggest the selective benefit of Membrain versus Sudoku training on 2-back performance may reflect selective transfer of training of this working memory requirement. We do note that the kind of working memory training offered by Membrain involves remembering where different visual pictures are located around a larger 3D shape, and that we measure a transfer effect onto a 2-back digit order identity task presented centrally (no spatial information to distinguish stimuli or order). This training and transfer effect may represent some degree of cross-modality or general/amodal transfer of working memory performance. Alternatively, the nameability of picture symbols used in the Membrain game may allow participants to use a primarily verbal working memory approach to representing items. Verbal strategies would not completely solve this problem though – positions in the larger 3D arrays would be less easy to describe with uniquely identifying verbal descriptions (e.g., "sun symbol, lower left corner of the pyramid") as the 3D arrays have symmetrical arrangements with confusable descriptions, and rotate in space. Selective transfer of Membrain training to 2-back digit working memory performance may represent some degree of near to moderate cross-modal transfer, perhaps depending on how participants approached picture representation in the Membrain game. In general, we suggest this transfer likely reflects improved approaches to task elements, improved chunking, or similar task practice effects, and is not likely due to improvements in fundamental working memory representational capacity.

In contrast to mental rotation and n-back tasks, no differences were observed between the training groups for the picture memory task. In the present study this is a somewhat limited but necessary manipulation check – it is important that transfer is not present for some tasks, but is for others, if one wants to be able to draw conclusions about the specificity and potential representation or mechanism underlying transfer and training effects. In subsequent chapters, approach to theory and methodology is expanded, to ask whether one can carefully isolate the direct influences of video game play on post-test cognitive task performance.

This initial study was designed to measure the extent to which video game training can transfer to certain cognitive skills. The literature on learning and transfer suggests that far transfer of skills occurs rarely, if at all (Barnett & Ceci, 2002). For a far transfer of skills to occur, training would need to improve a very fundamental cognitive process that is a common underlying element of many tasks. In those cases, we would expect increased efficiency and capacity of the fundamental processes that would translate to improvement in a great variety of tasks.

Rather than search for generalized far transfer, the focus of subsequent chapters here is to carefully search for mechanism-specific near transfer in particular cognitive domains, where we can exert careful methodological control to be sure that transfer effects we might observe are the result of our deliberate training. We discuss these studies in subsequent chapters. We also note that training and transfer effects do not need to be overall facilitation effects on speed or accuracy of performance. It is also possible that video game training could be altering decision or performance criteria within a system, which could result in different degrees of speed/accuracy tradeoff with different kinds of game training. These and other issues relating to training and transfer effects are discussed throughout subsequent chapters.

Appendix A – Mental Rotation Questionnaire

1. a. I rotated the whole figure in my mind when making the comparison
   b. I rotated a section of the figure in my mind when making the comparison
   c. I am not sure how I did it
   d. other (explain) ____

2. a. I thought through the steps verbally in my mind (i.e. "two cubes up and three down")
   b. I relied mainly on visualizing the figures and did not talk myself through the steps
   c. I am not sure

3. a. I used movements of my finger, hand, and/or pencil to help me with the task
   b. I did not use movements of my finger, hand, and/or pencil to help me with the task

4. a. I developed a specific approach to make my decision.
   b. I tried various approaches to make my decision.
   c. I had no specific approach

5. a. I double checked my answers before moving on to the next problem
   b. I was vaguely confident of my answers before I moved to the next one
   c. I was unsure of my answers before moving on to the next one
   d. I guessed most of the time

References

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612-637.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*(4), 352-358.

Ogakaki, L., & Frensch, P. A. (1994). Effects of video game playing on measures of spatial performance: gender effects in late adolescence. *Journal of Applied Developmental Psychology*, *15,* 33–58.

Peters, M., & Battista, C. (2008). Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. *Brain and Cognition*, *66*(3), 260-264.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects.

Stigler, James W., Shin-Ying Lee, and Harold W. Stevenson. "Digit memory in Chinese and English: Evidence for a temporally limited store." *Cognition 23.1* (1986): 1-20.

Watter, S., Geffen, G. M., & Geffen, L. B. (2001). The n-back as a dual-task: P300 morphology under divided attention. *Psychophysiology, 38*(06), 998-1003.

Chapter 3 – "Paint the West"

A great deal of research has been done on the possible benefits of playing video games over the last few years. Studies have been conducted to compare performance of expert video game players (VGPs) versus non video game players (nVGPs) on a range of cognitive abilities. Authors have also examined the effects of video game training on attention, visuo-spatial abilities, memory, and many other kinds of performance (Feng, Spence & Pratt, 2007; Li, Polat, Makous, & Bavelier, 2009). Research has not, however, focused as much on finding the specific aspects of video games that may be responsible for the observed improvements in mental abilities that video games may confer. This more precise approach to investigating some of the underlying mechanisms involved in training and transfer from video game play is the goal of the present study.

Expert VGPs have shown to have superior skills when compared to nVGPs in many tasks. It has been claimed that expert VGPs possess more attentional resources than nVGPs (Green & Bavelier, 2003), that they can process a rapid stream of visual information more efficiently than nVGPs (Green & Bavelier, 2003), and that expert VGPs also have better visual acuity thresholds and smaller regions of spatial interaction (Green & Bavelier, 2007). Expert VGPs also show faster reaction times across a variety of tasks, such as those used to measure n-back, Simon and Inhibition of Return effects. Most importantly, the faster reaction times across those tasks did not come at the cost of accuracy, suggesting that players are not "trigger happy", but do indeed possess better skills at those tasks than nVGPs (Dye, Green, & Bavelier, 2009). It is important to point out that while there were differences between expert VGPs and nVGPs, these experiments were unable to tell whether the differences were due to pre-existing abilities (which may predict whether someone chooses to play video games) or whether they were gained through years of video game playing. Training studies are needed in order to better determine whether video games can cause an increase in attentional and spatial abilities.

Training studies have shown that people can improve certain skills with the aid of video games. Feng, Spence and Pratt (2007) were able to reduce gender differences in spatial attention and mental rotation tasks with only 10 hours of video game training. A different study was able to improve contrast sensitivity function by having participants play over 50 hours of action video games (Li, Polat, Makous, & Bavelier, 2009). Green and Bavelier (2003, 2006a, 2007) also reported that training in action video games can improve performance in tasks that measure the attentional blink, useful field of view, enumeration, multiple object tracking, and crowding. However, not all studies support these findings. Boot and his colleagues were not able to replicate most of the results found by Green and Bavelier in a training study (Boot, Kramer, Simons, Fabiani & Gratton, 2008). They were also unable to replicate all the results when comparing expert VGPs to nVGPs.

Most of the studies discussed above used complex, commercially available video games that are obviously representative of the real world of video game playing, but do not allow for experimental control over elements within the gameplay, thus making it impossible to attribute any specific attribute of the game to a particular improvement in a specific mental ability. Previous studies compared groups that were trained on different video games, such as comparing training on puzzle games to training on first-person shooter games. Some studies even compared people that received no training on video games to those trained on first person shooter games. This type of methodology creates many confounds that could cause differences between experimental groups to be caused by differences in arousal levels that each game may create, or decreases in reaction times that could be caused by being trained on a similar machine to the one used during training (Boots, Simons, Stothart, & Stutts, 2013). The present study was done, in collaboration with Telos International Inc, in an attempt to pinpoint some of those specific attributes, by systematically varying the visual attention and cognitive control demands of a game that was designed to be both experimentally controlled and commercially viable. The experiment was designed in order to study the actual video game related improvements in cognitive ability while minimizing the number of possible confounds. The experiment used the game "Paint the West", a western style paintball shooting gallery game designed to test whether we can improve visual attention and cognitive control in participants. The design of Paint the West allowed us to create two different levels of visual attention demands (easy versus hard) and cognitive control demands (easy versus hard) for the players to be trained on, creating four different experimental groups.

Methods

Participants

Sixty participants (nine males), ages ranging from 19 to 27 years, (mean of 20.8 years) took part in the experiment, and were paid $120 for their participation. No restrictions were placed regarding participants' first language (17 non-native English speakers, though all fluent in English) or regarding their handedness (10 left handed participants). All participants reported normal or corrected to normal vision. Participants were recruited through McMaster University's online Experimetrix participant recruitment and scheduling system. Our recruitment notice requested that participants have little to no experience playing video games. Participants with considerable video game experience were not entered into the experiment based on their answers to a pre-experiment screening questionnaire. The cut off for video game experience was four or more hours per week playing video games over the last six months preceding the experiment. Participants were randomly allocated to one of four between-subjects training groups, with 15 participants per group. No difference in the average age of each group was observed (19.8, 19.1, 20.1, and 19.7 years), $F(3, 56) = 0.53$, p = n.s.

Apparatus

Presentation software (v.13, Neurobehavioral Systems, www.neurobs.com) was used to run the cognitive tasks part of the experiment and standard PC computers were used to run the laboratory tasks and the game, using 22 inch LCD monitors. Participants sat at a viewing distance of approximately 60 centimetres away from the computer screen for the laboratory tasks, and headphones were used by each participant to avoid noise interference from other participants' game and cognitive tasks.

Video Game Training – "Paint the West"

"Paint the West" is a first person view, western style shooting gallery game in which participants simulate shooting targets with paintball guns. The primary objective of the game is to shoot the Bandits (targets) with coloured paintballs while avoiding mistakenly shooting the Sheriffs and Townspeople (distractors), and dodging paintballs fired back at participants by the Bandits. Figure 1 shows an example gameplay screen from Paint the West. Participants were also given opportunities to shoot bottles during speed rounds, where no distractors were used and points were given for hitting as many bottles as possible during the round – for our purposes this was essentially a filler task, and was equivalent across all training groups.

The game required the use of a 3-button mouse, with each button firing a different colour paint from the paintball gun (red-left, yellow-middle, and blue-right). All game characters wore generally dark shaded clothing, with highlight colours (on belts, scarves, and hat bands) of one or two of the three possible paintball colours. Participants had to shoot the Bandits using a paint colour that was not seen on the Bandits' clothing, while avoiding hitting Sheriffs or Townspeople. For all participants, the game's overall speed difficulty (rate and duration of targets) increased over time, and was titrated according to how well a player performed at hitting targets, not hitting non-target characters, and avoiding the paintballs fired back at them. This was done in order keep the task at an optimal and engaging level of difficulty for all participants; this type of titration system is commonly used in many commercial games.

Through an industrial research collaboration with video game company Telos International (and with NSERC Engage program funding support), we collaborated to design Paint the West to emphasize game features that would be demanding of visual selective attention and cognitive control. We established four different training conditions as a 2 x 2 between-groups implementation of Easy versus Hard Visual Attention demands, and Easy versus Hard Cognitive Control demands. All participants played the same game, and all participants experienced both easy and hard exemplars of stimuli, distractors, and response demands, all described in detail below. Our experimental training manipulation was implemented as the *proportion* of simpler versus more demanding stimulus, distractor, cuing, response selection, second task monitoring, and other factors that participants experienced over the course of training, within Easy versus Hard Visual Attention and Cognitive Control training groups. For example, all participants saw displays that contained Bandits (targets), Sheriffs (high-similarity distractors), and Townspeople (low-similarity distractors); participants in the Hard Visual Attention groups were shown a higher proportion of Sheriff distractors versus Townspeople distractors than the participants in the Easy Visual Attention groups.

This design allows the use of the same video game as its own active control. If participants in specific Hard training groups show a greater amount of improvement on laboratory transfer tasks than participants in corresponding Easy training groups, we can much more directly attribute specific transfer effects to differences in the proportions of higher-demand versions of specific game characteristics.

Visual Attention parameters that were manipulated for level of difficulty consisted of Distractor Similarity, Stimulus Crowding and Stimulus Number. Figure 2 shows examples of game characters and demonstrates the Distractor Similarity manipulation. Bandits (targets) shared many more similar visual features with Sheriffs (high-similarity distractors; similar clothes, hats, coats, etc.), compared to Townspeople (low-similarity distractors). While all distractor types were seen by all training groups, the Hard Visual condition used a greater proportion of high-similarity vs. low-similarity distractors throughout the game compared to the Easy Visual condition. Stimulus crowding manipulated the distance between targets and distractors on the screen. While all groups saw a mixture of inter-stimulus distances, a greater proportion of close stimulus-distractor spacing was used in the Hard vs. Easy Visual condition. Stimulus number manipulated the average number of stimuli on screen at a time (stimuli would appear and then disappear after a short delay, prompting a speeded response from participants if they were to hit targets and earn points). In the Easy visual condition, one or two stimuli plus one or two distractors were present at one time, versus three or four stimuli and distractors for the Hard Visual condition.

Cognitive Control parameters consisted of Response Selection difficulty, Paintball Capacity (secondary monitoring task demand) and Cue Trial Informativeness. Response selection difficulty is illustrated in Figure 3. Participants could shoot Red, Yellow or Blue paintballs (via left, middle and right mouse buttons), and had to tag Bandits (targets) with a colour they were not wearing. Game characters all had costumes in primarily dark colours, with highlight colours (belts, hat bands, scarves, etc) in one or two of the three possible paintball colours. Some targets had only one highlight colour (e.g. blue, as in Figure 3, Panel A), and could be responded to (tagged with a paintball) of either red or yellow – this reflected a low response selection demand, where any colour other than the single target colour would work as a response. In a high response demand situation, targets wore two different highlight colours (e.g., blue and red; Figure 3, Panel B), where only a yellow paintball response would successfully tag the target. While all game groups experienced both kinds of stimuli, the Hard Cognitive Control condition presented a larger proportion of high vs. low response selection demand stimuli throughout the game.

The game required participants to monitor the number of shots used and to click in the lower right corner of the screen to reload their paintball gun regularly, imposing a concurrent monitoring task on participants within the continuous dynamic nature of the shooting gallery style game. Paintball Capacity required participants to reload every 12 or 6 shots in the Easy vs Hard Cognitive Control condition, respectively, imposing a differential concurrent monitoring and secondary task cost.

In 10% of all trials in the game, whether in the Easy or Hard Cognitive Control conditions, participants received a cue to the identity of an upcoming stimulus in the form of a hat emerging from behind a box before the rest of the target or distractor figure. We manipulated the proportion of targets, high- and low-similarity distractors as the parameter of Cue Trial Informativeness. Townspeople had visually distinctive hats from Bandits and Sheriffs, whose hats were extremely similar. For the Easy Cognitive Control condition, participants were shown a mix of 80% Townspeople, 10% Sheriffs and 10% Bandits in these cued trials, while the Hard Cognitive Control groups were shown a mix of 20% Townspeople, 40% Sheriffs and 40% Bandits. Participants in the Easy condition could easily reject a target 80% of the time from this information, while those in the hard condition had to continue monitoring and preparing for an impending decision with cue uncertainty in 80% of cases.

Experimental Tasks

Visual Search task: In the Visual Search task participants had to indicate whether the letter "b" was present among a variety of distractor letters. Stimuli were presented at random locations in an invisible 10 x 10 position grid measuring 28° vertically and 34° horizontally, with a central red fixation cross marking the centre of the display. Stimulus sets were composed of 4, 10, 18 or 26 items per trial, randomly determined. Distractor letters included "g", "h", "j", "l", "p" and "y". All letters were white, lowercase in the *Courier New* font, size 30 and were presented on a black background. Participants indicated that the letter 'b' was present by pressing the left mouse button and indicated that the letter 'b' was absent by pressing the right mouse button.

Participants completed one block of 180 trials for this task. At the start of each trial the central red cross appeared in the middle of the screen for 500 ms and was then joined by the visual search stimuli. Each visual search trial remained on the screen for 6000 ms or until a response was selected, whichever came first. If the participant did not make a response within 6000 ms, the stimuli disappeared from the screen and participants had to look at a blank, black screen until they selected a response, as the trial did not finish until a response was selected. Participants were instructed to look at the red cross for the initial 500 ms but were allowed to move their eyes once the search stimuli appeared on the screen. The subsequent trial started after a 1000 ms delay. Participants were instructed to respond as quickly as possible and to be accurate, but were not given feedback as to the correctness of their response during the trials or at the end of the Visual Search task.

Flanker task: Participants had to indicate the direction (left or right) of a central arrow surrounded by two flanking arrows on each side (e.g., > > < > >). The central target arrow in each trial was either congruent or incongruent with the flankers. Stimuli consisted of five black “<” and/or “>” characters, font size 48, shown on a gray background screen. Participants pressed the left versus right mouse buttons to respond to left and right targets, respectively.

Participants completed four blocks of 45 trials of the Flanker task. Stimuli were randomly generated with 50/50 probability of a left vs right target, and 50/50 probability of congruent versus incongruent flankers, with no other constraints. Stimuli were presented centrally on the screen for 1000ms each with a 500ms interval given between each trial. Response to the trial was expected during the 1500ms total duration of the trial otherwise the trial would constitute a miss. Participants were instructed to be as fast as possible but not to sacrifice accuracy in the process. At the end of each block, participants were given their average reaction time and accuracy scores.

Go/No-Go task: The Go/No-Go task required participants to respond by pressing the left mouse button when the target letter “M” appeared on the screen (80% probability), and withhold a response when the distractor stimulus, the letter “W” appeared on the screen (20% probability). The target and distractor stimuli were presented in black Arial font, size 36 on a gray background. Stimuli were shown at the centre of the screen for 500ms each, with a 500ms blank interval between trials. Response to each go trial was expected in the 1000ms total duration of the trial, or the trial counted as a missed response trial. Participants were instructed to be as fast as possible but not to sacrifice accuracy in the process. At the end of each block, participants were given their average reaction time and accuracy scores. Participants completed a total of 4 blocks, each containing 60 trials, for a total of 240 trials for the Go/No-Go task.

Mental Rotation task: The mental rotation task was the same version of the Shepard and Metzler (1971) paradigm as was used in the Membrain study described in Chapter 2 (see above), using stimuli from the Peters and Battista (2008) stimulus set.

N-Back task: The n-back task used here was the same 1-back and 2-back digit n-back task as was used in the Membrain study, described fully in Chapter 2, above.

Procedure

Participants were tested in groups of up to four people in a single laboratory room, seated at separate computer workstations at separate desks around the perimeter of the room. An experimenter was present and visible seated near the doorway of the room at all times, for both experimental laboratory tasks and game training. On the first day of the experiment, participants completed a questionnaire about their video game habits followed by a battery of five different cognitive tasks (visual search, flanker, go/no-go, mental rotation, and digit n-back). The tasks were completed sequentially and short breaks of up to one minute were given between each task so participants could hear the instructions to each upcoming task from the research assistant.

On the second day of the experiment, participants played the video game Paint the West for 50 minutes, and continued to play the game for a further 6 days spread over a 9 day period, for a total of 7 days (350 minutes) of video game playing time. The final day (day 9) of the experiment consisted of re-testing the participants on the battery of laboratory tasks.

Results

Visual Search

Data for three participants were excluded from the analyses due to extended poor performance. The criteria for exclusion consisted of accuracy below 50% for two or more of the conditions analyzed. Figure 4 shows visual search data separated by display size and group, for mean correct RT on target-present trials, corresponding Accuracy data for target-present trials, and d-prime data calculated from target-present versus target-absent performance. Correct RT data were analyzed via repeated measures ANOVA, with within-subjects factors of Day (1 and 9), and Display Size (4, 10, 18 and 26 items), and between-subjects factors of Visual Attention (Easy vs Hard) and Cognitive Control (Easy versus Hard) training. We limited the influence of extremely slow, but still correct trials by replacing RTs over 5000 ms with a 5000 ms value. RTs faster than 300 ms were excluded from analysis.

Results typical of visual search tasks were observed. A strong main effect of day was observed, $F(1, 53) = 53.01$, $p < 0.001$, $\eta_p^2 = 0.50$, reflecting faster RT performance over time. A strong main effect of set size was also observed, $F(3, 159) = 539.58$, $p < 0.001$, $\eta_p^2 = 0.91$, reflecting slower RTs with larger display sizes. An interaction between day and set size $F(3, 159) = 4.21$, $p = 0.007$, $\eta_p^2 = 0.07$ reflected a greater improvement in RT over days with larger display sizes. While all groups improved their visual search RT performance over time, an interaction between visual training, cognitive control training and day was observed $F(1, 53) = 4.78$, $p = 0.033$, $\eta_p^2 = 0.08$, with no 2-way interactions of day with either visual training or cognitive control training, *F*s < 1.6. From the top panel of Figure 4, the Visual-Hard+Cognitive-Hard training group showed the largest improvement in RT across display sizes compared to other groups. To better visualize this effect, Figure 5 shows RT data for both correct target-present and correct target-absent trials, collapsed over display size. While the largest improvement is observed in the Visual-Hard+Cognitive-Hard group, the second largest improvement is in the Visual-Easy+Cognitive-Easy group, which does not present a clean pattern of training-related improvement.

RT analyses were also performed for target absent trials. Reaction times in target absent trials are typically slower than those in target present trials in visual search, and our data reflected this as well (see Figure 5). A strong main effect of day was observed, $F(1, 53) = 55.38$, $p < 0.001$, $\eta_p^2 = 0.51$, showing general improvement in performance over time, and a strong main effect of display size was also observed, $F(3, 159) = 743.668$, $p < 0.001$, $\eta_p^2 = 0.93$, with slower RTs with larger display sizes (not shown). An interaction between day and set size $F(3, 159) = 29.28$, $p < 0.001$, $\eta_p^2 = 0.37$ reflected a greater improvement in RTs with larger display sizes. In contrast to target-present trials, no training differences were observed between gaming groups, Fs <1.5.

To assess accuracy in visual search task performance, we calculated d-prime measures based on target-present and target-absent trial performance, and analyzed them as per RT data above. These data are shown in Figure 4. Participants showed overall good accuracy as reflected by d-prime scores. No main effects were observed for days, display sizes, visual training groups or cognitive control training groups, Fs < 1.5. A between-groups interaction between visual training and cognitive control training was observed, $F(1, 53) = 7.86$, $p = 0.007$, $\eta_p^2 = 0.13$, with Visual-Easy+Cognitive-Easy and Visual-Hard+Cognitive-Hard groups having generally lower d-prime scores overall compared to the other two groups. There was no evidence that these between-groups differences changed differentially over training, Fs < 0.5.

Flanker

Figure 6 shows mean Flanker task data for correct RT trials and Accuracy data, separated by flanker congruency, and d-prime data calculated from congruent versus incongruent trial performance, all separated by day and training group. Correct RT data were analyzed via repeated measures ANOVA, with within-subjects factors of day (1 and 9), flanker congruency (congruent, incongruent), and between-subjects factors of Visual Attention (Easy vs. Hard) and Cognitive Control (Easy vs. Hard) training.

Results typical of the flanker task were observed. A main effect of day was observed, $F(1, 56) = 49.95$, $p < 0.001$, $\eta_p^2 = 0.47$, reflecting faster RT performance over time. A main effect of congruency was also observed, $F(1, 56) = 334.92$, $p < 0.001$, $\eta_p^2 = 0.47$, reflecting faster RTs on congruent trials than on incongruent trials. An interaction of day by congruency was observed, $F(1, 56) = 8.67$, $p = 0.005$, $\eta_p^2 = 0.13$, reflecting a general reduction in the flanker congruency effect (reduced difference between congruent and incongruent trials) over days. There were no overall effects of training group, *F*s < 0.5.

A significant interaction of day by congruency by visual group was observed, $F(1, 56) = 5.07$, $p = 0.028$, $\eta_p^2 = 0.08$, with participants in the hard visual training groups showing greater reduction of the flanker congruency effect (greater improvements in reaction times on incongruent trials than participants in the easy visual training groups). Cognitive control training did not show changes over day, *F*s < 1.4.

Considering the influence that sequential trial effects can have on flanker compatibility performance, we performed an additional analysis of these RT data with the additional factor of previous trial congruency, to better assess improvement on congruent versus incongruent trials relative to visual attention game training groups. For this analysis, we excluded all trials that followed an error trial, and a subset of trials that represent exact stimulus repetitions – this equates (removes) stimulus identity repetition from unduly influencing more abstract demands of congruency repetition. (A congruent trial following another congruent trial could occur by repeating the same stimulus, e.g. all left arrows, then all left arrows, or could occur as two different congruent stimuli, e.g. all left arrows followed by all right arrows. The same arrangement may occur for incongruent-incongruent sequences. In contrast, alternations of congruency can never have stimulus identity repeats. Removing identical stimulus repetitions allows a fairer comparison of attentional control across congruency sequences, independent of additional facilitation of exact stimulus repetitions.)

Sequential Flanker data for correct RT trials as described above are shown in Figure 7. The top panel shows data as in Figure 6, now conditionalized on prior trial congruency. The lower panels show data collapsed over visual training demands, to visualize the influence of visual training condition on flanker congruency over time. This analysis approach highlighted the influence of hard versus easy visual training conditions. Within this larger ANOVA, we still observe the interaction of visual training group with flanker congruency and day, $F(1, 56) = 4.05$, $p = 0.049$, $\eta_p^2 = 0.07$, with relatively greater improvement on incongruent trials in the hard visual training condition; improvement on incongruent trials in the easy visual training condition was comparable to the equivalent improvement on congruent trials across all conditions. A number of sequential effects on flanker performance were observed, but did not interact with game training, Fs < 0.5.

We analyzed d-prime data to assess accuracy performance over training; data are shown in Figure 6. Accuracy was generally quite high across the experiment. Participants showed no substantial changes in performance over day, and no training differences between groups were found, Fs < 1.5.

Go/No-Go

Data for one participant were excluded based on extended poor performance on all days of testing. Figure 8 shows mean RT, Accuracy and d-prime data for the go/no-go task, separated by day and training group. Correct RT data for go trials (Hits) were analyzed via repeated measures ANOVA, with within-subjects factors of Day (1 and 9), and between-subjects factors of Visual Attention (Easy vs. Hard) and Cognitive Control (Easy vs. Hard) training conditions. RT data for correct trials were filtered to include only those trials with RTs between 150ms to 1500ms, excluding trials unlikely to reflect genuine trial performance.

A main effect of day was observed, $F(1, 55) = 4.66$, $p = 0.035$, $\eta_p^2 = 0.78$, reflecting faster RTs for go trials over time. An interaction of day by visual training condition, $F(1, 55) = 8.10$, $p = 0.006$, $\eta_p^2 = 0.13$, and an interaction of day by cognitive control training condition, $F(1, 55) = 4.23$, $p = 0.04$, $\eta_p^2 = 0.07$, were both observed, with no three-way interaction, $F(1, 55) = 1.14$, $p = 0.291$. These results reflect an additive pattern of training-related speeding of RT for go trials, for both hard visual and hard cognitive control training conditions.

Assessing d-prime measures for go/no-go performance allows a more complete assessment of task behaviour in addition to go trial RTs. D-prime data were analyzed as per RT data above. Participants showed a general decrease in d-prime over days, $F(1, 55) = 10.74$, $p = 0.002$, $\eta_p^2 = 0.16$. A marginal interaction of day by visual training group was observed, $F(1, 55) = 2.89$, $p = 0.095$, $\eta_p^2 = 0.05$, suggesting a larger decrement in d-prime over days in the hard visual training groups. While this training effect is marginal, together with observed RT training effects in the visual training condition, these data are suggestive of a criterion shift or speed-accuracy tradeoff effect induced via game training.

Mental Rotation

Figure 9 shows mean correct RT and Accuracy data for match trials, and d-prime data calculated from match and non-match trial performance for the mental rotation task, separated by stimulus angle, day and training group. Correct mean RT data were analyzed via repeated measures ANOVA, with within-subjects factors of Day (1 and 9), and Stimulus Angle (20, 60, 100, 140 degrees), and between-subjects factors of Visual Attention (Easy vs Hard) and Cognitive Control (Easy versus Hard) training conditions. We focused our RT analyses on match trial data. To limit the excessive influence of very slow outliers, but retain relative variability between conditions, correct trials with RTs longer than 8000 ms were replaced with an 8000 ms value. RTs faster than 500 ms were excluded from analysis.

Results typical of mental rotation tasks were observed. A strong main effect of stimulus angle was observed, $F(3, 168) = 50.37$, $p < 0.001$, $\eta_p^2 = 0.47$, reflecting longer RTs with larger stimulus angles. A strong main effect of day was also observed, $F(1, 56) = 46.88$, $p < 0.001$, $\eta_p^2 = 0.46$, reflecting an overall improvement in RT performance over time. While all groups improved their mental rotation performance, no influence of game training condition was observed on RT, *F*s < 1.1.

To assess accuracy, we calculated d-prime measures based on match and non-match trial performance, and analyzed them as per RT data above. Overall, participants showed better performance at smaller stimulus angles, with a strong main effect of angle, $F(3, 168) = 101.89$, $p < 0.001$, $\eta_p^2 = 0.65$. Participants also improved their performance over time, with a main effect of day, $F(1, 56) = 11.88$, $p = 0.001$, $\eta_p^2 = 0.18$. There was no interaction of angle by day, $F < 0.6$. Overall differences in d-prime between groups were observed, with generally smaller d-prime scores in Visual-Easy+Cognitive-Easy and Visual-Hard+Cognitive-Hard groups, compared to the other two training groups, supported by the between-subjects interaction of visual and cognitive control training conditions, $F(1, 56) = 7.61$, $p = 0.008$, $\eta_p^2 = 0.12$. (This overall between-groups pattern of d-prime data mirrors overall d-prime performance in the visual search task.)

With respect to training effects, the pattern of d-prime changes across training groups in Figure 9 suggests greater numerical d-prime improvement for Easy Visual training conditions (left two groups) versus d-prime change in the Hard Visual training conditions (right two groups). However, this interaction of day and visual training condition did not reach significance, $F(1, 56) = 2.60$, $p = 0.113$. The same day by visual training group interaction in the Accuracy data for match trials (visualized immediately below the corresponding d-prime data in Figure 9) was marginally significant, $F(1, 56) = 3.26$, $p = 0.077$, $\eta_p^2 = 0.06$. No other training-relevant effects were observed, *F*s < 1.

N-Back

Data for two participants in each group (8 total) were excluded based on sub-50% accuracy in one or more conditions on at least one day of testing. Data analysis was done separately for 1-back and 2-back performance. Figure 10 shows mean correct RT and Accuracy data for Target (match) and Filler (non-match) trials, and d-prime data calculated from Target versus Filler trial performance for 1-back and 2-back tasks, separated by day and training group.

RT data for correct trials were filtered to include only those trials with RTs between 300ms to 2000ms for analysis. Correct RT data for 1-back and 2-back tasks were analyzed via separate repeated measures ANOVA, with within-subjects factors of trial type (Target vs. Filler), and day (1 and 9), and between-subjects factors of Visual Attention (Easy vs. Hard) and Cognitive Control (Easy vs. Hard) training conditions.

For the 1-back task, RTs for target trials were faster than for lure trials, $F(1, 48) = 55.44$, $p < 0.001$, $\eta_p^2 = 0.54$, overall performance improved over days, $F(1, 48) = 14.63$, $p < 0.001$, $\eta_p^2 = 0.23$, and target versus lure RT differences diminished over time, $F(1, 48) = 6.92$, $p = 0.011$, $\eta_p^2 = 0.13$. A marginal training effect of visual condition was observed, with larger RT improvement in Hard Visual versus Easy Visual groups, $F(1, 48) = 3.30$, $p < 0.076$, $\eta_p^2 = 0.06$. Considering only Target trial data, this visual condition training effect was considerably stronger, $F(1, 48) = 6.54$, $p = 0.017$, $\eta_p^2 = 0.11$. No other influences of game training groups were observed, Fs < 0.8.

For the 2-back task, general patterns of RT performance mirrored 1-back performance. RTs for target trials were faster than for lure trials, $F(1, 48) = 14.34$, $p < 0.001$, $\eta_p^2 = 0.23$, overall performance improved over days, $F(1, 48) = 104.06$, $p < 0.001$, $\eta_p^2 = 0.68$, and target versus lure RT differences diminished over time, $F(1, 48) = 8.76$, $p = 0.005$, $\eta_p^2 = 0.15$. In contrast to 1-back performance, 2-back data showed no evidence of any game training group effects, Fs < 0.4.

D-prime scores were analyzed as per RT data above. For the 1-back task, d-prime appeared to be consistent across days for hard visual training conditions, but to decline over days for easy visual conditions This interaction between day and visual training condition was marginal, $F(1, 48) = 3.12$, $p = 0.084$, $\eta_p^2 = 0.02$, with no other effects observed, Fs < 1.1. For the 2-back task, all groups improved over time, $F(1, 48) = 40.85$, $p < 0.001$, $\eta_p^2 = 0.46$, but with no suggestion of any game training group effects, Fs < 1.5.

Discussion

A number of results from this study suggest that training on the Paint the West game had positive transfer effects onto a number of standard cognitive laboratory tasks. For both the n-back task and the flanker task, performance on aspects of these tasks improved significantly more for participants who played Paint the West in the Hard Visual conditions (increased proportions of close vs. distant flanker spacing, increased proportions of high- vs. low- feature similarity flanker/distractor stimuli, and increased proportions of high vs. low number of stimuli on screen). Figure 11 highlights these key training results from our n-back and flanker tasks. In addition, for the go/no-go and mental rotation tasks, we observed changes for these same Hard Visual condition participants that suggested training-related effects on performance criterion (changes in speed-accuracy tradeoff over training) rather than an outright effect of improved performance on these tasks. These two general sets of results are discussed below, and considered more broadly in the General Discussion in Chapter 5.

For the n-back task, all training groups showed a significant and comparable amount of improvement on the 2-back task, in both RT and d-prime performance. Against this background, for the 1-back task, participants in the Hard Visual conditions showed improved RT and constant d-prime over training, compared to smaller or negligible improvement in RT and worsened d-prime over training for the Easy Visual groups. As discussed in Chapter 2 previously, successful 2-back task performance requires effortful manipulation of transiently represented stimulus information, including keeping track of items in sequence, and updating and reorganizing items in working memory (Stigler, Lee, & Stevenson, 1986; Watter, Geffen & Geffen, 2001). In contrast, the 1-back task primarily embodies less-demanding monitoring and response decision requirements, with a simpler requirement to remember only the most recent item, allowing participants to rely on simple perceptual memory without any manipulation or “working” aspect of working memory (Watter, Geffen & Geffen, 2001).

The training effects seen here, limited to 1-back and not 2-back performance, are particularly convincing because of their specificity. We suggest that Hard Visual training is improving the ability to focus on and speed response decision processes, so long as the target category or response/task set is well-established and does not require substantial decision or other demanding cognitive work. These effects are in line with other research that suggests similar kinds of video game training increase speed of visual processing (Li, Polat, Makous, & Bavelier, 2009). These benefits seem to be eliminated or obscured when a moderate working memory manipulation demand is introduced.

This pattern of improvement in n-back performance from game training is different from the improvement seen for participants in the Membrain study presented in Chapter 2. Participants playing Membrain (versus Sudoku) showed selectively greater improvement of performance on the 2-back task, with equivalent improvement across training groups for 1-back; here participants in the Hard Visual conditions of Paint the West showed selectively greater improvement on the 1-back task, with equivalent improvement across groups for 2-back performance. The demonstration that post-training performance can be selectively improved for either 1-back or 2-back performance, relative to gameplay involving recent memory for pairs of items (Membrain, 2-back benefit) versus speeded choice RT training with minimal working memory requirements (Hard Visual conditions in Paint the West, 1-back benefit), speaks to the selective and specific nature of training and transfer in general, and increases our confidence that observed training effects are genuine cases of transfer from game experiences.

The results of the Flanker task also support a theory of specificity of transfer effects from training with video games and a benefit in visual processing. While all groups improved their RT performance over training while maintaining consistent d-primes, participants in the Hard

Visual conditions showed greater reaction time improvements for incongruent trials compared to participants in the Easy Visual conditions. These selectively greater incongruent RT improvements did not come at the cost of accuracy.

Paint the West required participants to identify and selectively respond to (shoot/click on) Bandits (targets), while selectively not shooting Sheriffs (high-similarity distractors) and Townspeople (low-similarity distractors). Participants in Hard Visual conditions played with higher proportions of high-similarity distractors, more often spaced close together, and more often with greater numbers of targets and distractors on screen. We suggest that this experience gave participants more practice in specific attentional demands of selecting targets in the presence of featurally interfering distractors, and that this greater training in filtering out non-target distractors is observed as transfer to incongruent RT performance in the laboratory flanker task. We again focus on the specificity of this transfer effect relative to specific demands of manipulated game requirements. Transfer was seen only to incongruent trials in our Hard Visual conditions, we did not observed a generalized benefit (congruent trial performance increased comparably across all groups), and this incongruent training effect did not interact with sequential congruency effects.

Although participants in the Hard Visual groups appeared to show training and transfer from Paint the West to these tasks, we are very cautious about the claim (and do not really believe) that playing this video game increased the fundamental amount of attentional resources or working memory capacity available to these participants. There are many other mechanisms by which these training effects might occur, which we suggest are more likely. At the same time, we are not discounting that participants truly did improve their demonstrated performance in these tasks as a direct result of game training, just that these effects are much more likely to reflect fairly specific and limited cases of near transfer. These issues are discussed more fully in the General Discussion chapter.

In addition to these performance improvements in n-back and flanker tasks, other effects of training were observed that were less directly improvements in performance. In the go/no-go task, participants in the Hard Visual condition showed larger improvements in RT with training versus Easy Visual groups, but also showed a corresponding relative decrement in d-prime. While the measured d-prime effect was marginal, taken together the entire set of go/no-go data (faster RT, lower d-primes with inflated False Alarm rates) are highly suggestive of a training-induced criterion shift for Hard Visual participants. Similar effects may be present in the mental rotation data. All groups improved their RT performance similarly over training, but the Easy Visual groups showed marginally improved d-prime scores in contrast to Hard Visual groups.

Training in the Hard Visual condition may produce a greater degree of skilled automaticity in rapid target identification and choice response selection, and that over training participants in this training situation may come to rely on this automaticity within the context of the dynamic Paint the West game. We speculate that when these participants are asked to perform similar kinds of speeded choice response laboratory tasks, they may rely too much on this automaticity, or otherwise choose an inappropriate response criterion or speed-accuracy tradeoff setting. This may represent some degree of dissociation between participants' subjective sense of fluency or appropriateness of unmonitored behaviour, and the actual prior expertise or automaticity they actually have for that situation. This may have been encouraged or enhanced by having participants complete the laboratory tasks on the same computers in the same lab setting as all of their game training experience. On the other hand, this criterion shift might be less sensitive to specific context – perhaps gaining automaticity and some expertise with this kind of speeded game training leads participants to approach a wide array of only vaguely similar situations with similarly relaxed decision criteria. In this sense, what constitutes a matching “game” context might be very broad – in our case, other speeded computer tasks – but if broader could lead to unwanted real-world effects.

In contrast to these effects from the Visual training manipulations, the set of Cognitive Control manipulations did not show any substantial transfer effects. Some evidence of a possible go/nogo RT benefit was observed for Hard vs. Easy Cognitive Control conditions, but otherwise these manipulations had little influence on our laboratory tasks. This relative lack of transfer effects in comparison to effects from Visual training conditions may speak to the sensitivity or appropriateness of laboratory tasks to detect potential game training effects – we may have simply failed to use appropriate tasks to measure effects of training. On the other hand, this difference may illustrate the idea that some kinds of cognitive or perceptual abilities are more trainable than others, for various reasons. Some abilities may be relatively unpracticed and have room for easy improvement, or may show relative performance improvement when a new approach or strategy is taught or discovered; other abilities may be relatively insensitive to training effects, whether because they are already at close to ceiling performance, or because practice in those tasks or abilities is not well achieved with game play. These issues are considered more in the General Discussion.

We did observe some overall group differences (not training related) in d-prime for mental rotation and visual search experiments (overall lower d-primes for the Visual-Easy+Cognitive-Easy and Visual-Hard+Cognitive-Hard groups) – these differences may reflect generally poorer overall visuospatial ability for these groups. Importantly, these overall group differences were observed for an orthogonal set of groups than where we see our training effects in d-prime (both Visual-Hard groups versus both Visual-Easy groups); as such, these overall group differences are unlikely to reflect an underlying cause of our game training effects.

As we expected, training participants in video game play is not the equivalent of a magic pill and it does not improve all cognitive and perceptual processes. Experience playing Paint the West seems to speed up the selection of a response when the task is simple (1-back improvement), and to improve the ability to filter out conflicting non-target information (incongruent flanker improvement), demonstrating some very specific transfer. When the tasks become more difficult, or more dissimilar from the training provided by Paint the West, participants seem to undergo a change in criterion where speed is emphasized over accuracy. Importantly, these training and transfer effects are produced via a method of varying the proportions of particular hard or easy game elements (e.g. distractor similarity) for different training conditions, within a common overall game experience. While the size and extent of these observed transfer effects may seem small, they can be much more directly linked to differences in the amount of very specific experiences within the specific game.

In the final data chapter, we take the same training and transfer approach as in Paint the West, considering another game with different gameplay requirements, and ask whether differences in spatial attention and spatial memory requirements in gameplay might transfer to related laboratory tasks.

References

Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Psychologica, 129*(3), 387-398.

Boot, Walter R., et al. "The pervasive problem with placebos in psychology why active control groups are not sufficient to rule out placebo effects." *Perspectives on Psychological Science 8.4* (2013): 445-454.

Dye, M. W., Green, C. S., & Bavelier, D. (2009). Increasing speed of processing with action video games. *Current Directions in Psychological Science, 18*(6), 321-326.

Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science, 18*(10), 850-855.

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423*(6939), 534-537.

Green, C. S., & Bavelier, D. (2006a). Enumeration versus multiple object tracking: The case of action video game players. *Cognition, 101*(1), 217-245.

Green, C. S., & Bavelier, D. (2007). Action-video-game experience alters the spatial resolution of vision. *Psychological Science, 18*(1), 88-94.

Li, R., Polat, U., Makous, W., & Bavelier, D. (2009). Enhancing the contrast sensitivity function through action video game training. *Nature Neuroscience, 12*(5), 549-551.

Peters, M., & Battista, C. (2008). Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. *Brain and Cognition*, *66*(3), 260-264.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 701–703

Stigler, J. W., Lee, S. Y., & Stevenson, H. W. (1986). Digit memory in Chinese and English: Evidence for a temporally limited store. *Cognition*, *23*(1), 1-20.

Watter, S., Geffen, G. M., & Geffen, L. B. (2001). The n-back as a dual-task: P300 morphology under divided attention. *Psychophysiology, 38*(06), 998-1003.

Chapter 4 – "Orphlings"

The third study in this dissertation sought to extend the careful training and transfer methods used in the previous chapter with Paint the West, to a different kind of game performance that was less involved with speeded choice responding. Given general findings in the research literature that expert gamers show many speed-related improvements, this chapter sought to use a game that was still visuospatially dynamic and challenging, but where speeded choice response was not involved. We used the dynamic spatial puzzle game "Orphlings" (described at length below), with an overall methodological design akin to the study with the "Paint the West" game in Chapter 3. As a companion study to Paint the West, this study helps to ask how different kinds of games might show different kinds or degrees of transfer effects.

This study was again conducted in collaboration with Telos International Inc, in an attempt to pinpoint specific cognitive attributes that can be improved through video game play, by systematically varying the cognitively-relevant components of games that were designed to be both experimentally controlled and commercially viable. All participants played the same game, but were assigned to different difficulty levels of specific features within the game. As such, the game acts as its own control, to allow a more selective test for game training-related improvements. This methodology should lead to results that are much less biased by factors such as test-retest effects, arousal level effects and other undesirable confounds.

Methods

Participants

Sixty-eight participants (thirteen males) with ages ranging from 18 to 23 years (mean of 19.1 years), completed the experiment, and were paid $80 for their participation. (An additional four participants were originally recruited but dropped out of the study part way through, leaving four between-subjects groups of 16, 18, 17 and 17 participants.) No restrictions were placed regarding participants' first language (22 non-native English speakers, though all were fluent in English) or regarding their handedness (6 left handed participants). All participants reported normal or corrected to normal vision. Participants were recruited through McMaster University's online Experimetrix participant recruitment and scheduling system. Our recruitment notice requested that participants have little to no experience playing video games. Participants with considerable video game experience were not entered into the experiment based on their answers to a pre-experiment screening questionnaire. Just as in the "Paint the West" study, the cut off for video game experience was four or more hours per week playing video games over the six months that preceded the experiment.

Apparatus

The same laboratory room and computer workstation setup was used as for the Paint the West study. Presentation software (v.13, Neurobehavioral Systems, www.neurobs.com) was used to run the cognitive tasks part of the experiment and standard PC computers were used to run the laboratory tasks and the game, using 22 inch LCD monitors. Participants sat at a viewing distance of approximately 60 centimetres from the computer screen for the cognitive tasks, and headphones were used by each participant to avoid noise interference from other participants' game and cognitive tasks. Up to four participants were tested at a time, and an experimenter was present for all gameplay and laboratory task sessions.

Video Game Training – "Orphlings"

"Orphlings" is a third person view, three dimensional dynamic puzzle game in which participants need to figure out ways in which to get all of the orphling creatures to the end of a level (with various obstacles in their path) by using a set of tools that change the behaviour or properties of the creatures to let them avoid or overcome these obstacles. An example game level of Orphlings is shown in Figure 1 (top panel) illustrating the path that orphling creatures take through this level, with the correct tools in place to let them proceed past various obstacles; two close-up sections (lower panels) show details of the larger example. Orphling creatures wander through the maze on their own, with new creatures appearing at the entrance to the level at regular intervals. The game levels have interactivity in the depth dimension in addition to the overall 2D horizontal and vertical layout. In many sections of game levels, there were two levels of depth through which creatures could travel, and where players could choose to place the various in-game tools.

The primary objective of the game is to ensure that as many creatures as possible reach the end of the level, in as little time as possible, and with the least number of creatures encountering hazards (e.g. falling off ledges, etc) and coming to harm. Players must use a combination of orphling creature physical changes (tools that change the properties of the creatures – e.g., making them sticky so they can climb a wall, making them bouncy so they can survive a long fall, etc) and a series of in-game machines (tools that directly move the creatures - e.g., pistons that push creatures, “gravity wells” that attract falling creatures, etc) in order to accomplish this goal. The game offers a substantial but engaging problem-solving challenge to arrange a sequence of creature-guiding tools in real-time throughout a game level, while the creatures are travelling through and potentially encountering dangerous obstacles, before the player can figure out how to keep the creatures safe and progressing through the level. In addition to this primary gameplay, a spatial memory component was required at varying intervals. At pseudo-random intervals throughout the game, participants were warned of an impending blackout and given a very short time interval to memorize the location of the orphling creatures on the screen. The entire computer screen would then turn black and participants needed to click on the locations where the orphling creatures were last spotted to end the blackout without losing any creatures.

The game required the use of a standard 2-button mouse with a scrolling wheel. The left mouse button was used to select and place the various in-game tools, and the right mouse button was used to select the depth at which the tool would be placed. The game also required some use of the keyboard for various game commands. The overall difficulty of the game increased as participants progressed through each level, thus increasing the level of challenge for the participants in order to maintain engagement.

For our training study, we manipulated the relative difficulty of two elements – Depth interactivity, and Spatial Memory demand – within the game to create a 2 x 2 between-groups training design. This allows the overall game to be used as its own control, with the relative difficulty or demand of particular game features the only systematic difference manipulated between training groups. Participants were randomly allocated to groups of equal size (initially 18 participants per group, but final data only for 16, 18, 17 and 17 participants due to participant drop out). No difference in the average age of each group was observed (19.4, 18.6, 19.2, and 19.2 years, $F(3, 65) = 1.19$ p = n.s.).

To create Easy and Hard Depth conditions, we manipulated whether participants had to consider tool placement in different depth planes to solve game levels. In all training conditions, the orphling creatures could navigate between two different depth levels on the screen. Participants in the Easy Depth groups had no relevant need to control this feature, as the orphling creatures would always travel to the same consistent depth level (the default, perceptually nearest to the participant) in order to be affected by in-game tools, by the game making minor adjustments to critical decision points (necessary tool placement locations) to force this consistent outcome in the game. Participants in the Hard Depth groups needed to consider the depth location of the obstacles and the depth location of the creatures and make the necessary adjustments using the tools given to them to successfully navigate through the game.

To create Easy and Hard Spatial Memory conditions, we manipulated the number of orphling creatures on the screen when the occasional "blackout" and click-to-rescue spatial memory events occurred. Participants in the Easy Memory condition always had these events occur with two, three or four creatures on the screen, while participants in the Hard Memory condition always had these events occur with five, six or seven creatures on the screen. The game display returned after the player had clicked once per creature in the previous display. Any creatures whose location were not correctly identified disappeared and were sent back to the start of the level, slowing a player's progress.

Procedure

The general procedure and testing/training schedule was the same as used in the Paint the West study, though with a different video game and several different laboratory tasks. On the first day of the experiment the participants completed a questionnaire about their video game habits followed by a battery of five different cognitive tasks, consisting of a mental rotation task, visual search task, spatial working memory task, a digit n-back task, and a spatial n-back task. The tasks were completed sequentially and short breaks of up to one minute were given between each task so participants could hear the instructions to each upcoming task from the research assistant.

On the second day of the experiment, participants played the video game "Orphlings" for 50 minutes, and continued to play the game for a further 6 days spread over a 9 day period, for a total of 7 days (350 minutes) of video game playing time. The final day of the experiment consisted of re-testing the participants on the battery of laboratory tasks.

Experimental Tasks

Mental Rotation, Visual Search, and Digit N-Back tasks: These three tasks were the same as were used in the Paint the West study, described in Chapter 3. In this study, we identify the n-back task used previously (1-back and 2-back tasks for visually presented digits) as the "digit n-back" task, to distinguish it from a separate visuospatial position n-back task, described below.

Spatial N-Back task: The spatial n-back task used for our experiment used the same design and procedure as the digit 1-back and 2-back tasks used throughout the series of studies in this thesis, but used spatial location stimuli in place of centrally presented digits. Participants were required to monitor the locations of squares randomly presented one at a time on the computer screen and asked to respond whether the location of the current square matched the location of the square presented n trials previously. Participants were instructed to press the left mouse button if they believed the current location matched the location presented n trials before, or to press the right mouse button if they believed the locations were not a match.

A total of six spatial n-back blocks were completed, alternating between 1-back where the location of the stimulus on the screen was compared to the location of the stimulus on the screen that preceded the current trial; and 2-back blocks where the location of the current stimulus was compared to the location of the stimulus presented 2 trials before the current trial on screen. Each block contained 29 trials. Stimuli consisted of black squares shown one at a time on a gray background, each measuring approximately 1.2° by 1.2° of visual angle. Stimuli were presented serially at one of 36 locations within an invisible six by six grid centred on the screen. The invisible grid measured 24° by 24°, with the outer edges of the grid being presented at diagonal visual angles of 17° from the centre of the screen. Each stimulus remained on the screen for 1000ms with a blank interval of 1500 ms between each stimulus. Participants' response to each stimulus had to occur in the 2500 ms allotted to each trial or it would count as a missed response. No feedback was given as to whether participants' responses were correct or not at the end of each trial, but at the end of each block participants were shown their accuracy score and average reaction time.

Spatial Working Memory task: The spatial working memory task used for our experiment required participants to monitor the spatial location in which a sequence of squares was presented across a grid on the computer screen and to respond by replicating that spatial pattern. Stimuli consisted of 3.1° x 3.1° white squares being shown at random but non-repeating locations in a visible six by six white grid on a gray background, with one square presented at a time. The grid was centred on the middle of the screen and measured 18.6° by 18.6°. The outer edges of the grid were located approximately 13.5° diagonally from the centre of the screen.

The task consisted of four blocks, with each block containing 15 trials. The number of squares presented at each trial varied between either four or six items, with blocked presentation. (To generate stimulus locations, in the six by six grid, the six x-axis locations and six y-axis locations were randomly paired on each trial, assuring a reasonable distribution of locations across the screen, with no immediately horizontally or vertically adjacent items.) Each square stimulus remained on the screen for 1000ms with an interval of 1500ms given between each square. Once the 4- or 6-item stimulus sequence presentation ended, a red pointer for the mouse appeared in the centre of the screen, and participants used the mouse to click on the remembered stimulus locations. The empty grid display remained onscreen while participants made these location responses. The trial ended after participants clicked the mouse either four or six times, according to the number of squares presented on that trial. Participants were not constrained to respond to locations in presentation order, but were instructed to try to remember where all the items were as best as they could, and click within those grid positions to respond.

Results

Mental Rotation

Figure 2 shows mean correct RT and Accuracy data for match trials, and d-prime data calculated from match and non-match trial performance for the mental rotation task, separated by stimulus angle, day and training group. Correct mean RT data were analyzed via repeated measures ANOVA, with within-subjects factors of Day (1 and 9), and Stimulus Angle (20, 60, 100, 140 degrees), and between-subjects training factors of Depth (Easy vs. Hard) and Spatial Memory (Easy vs. Hard). RT analyses focused on match trial data. As in the Paint the West study, to limit the excessive influence of very slow outliers, but retain relative variability between conditions, correct trials with RTs longer than 8000 ms were replaced with an 8000 ms value. RTs faster than 500 ms were excluded from analysis.

Results typical of mental rotation tasks were observed. A strong main effect of stimulus angle was observed, $F(3, 192) = 83.04$, $p < 0.001$, $\eta_p^2 = 0.57$, with progressively longer RTs at larger stimulus angles. A strong main effect of day was also observed, $F(1, 64) = 149.98$, $p < 0.001$, $\eta_p^2 = 0.70$, reflecting general improvement in RTs over time. An interaction of day and angle was also observed, $F(3, 192) = 3.17$, $p = 0.026$, $\eta_p^2 = 0.05$, reflecting a greater improvement in RTs over time for larger stimulus angle trials. While all training groups improved their mental rotation performance over time, no effects were observed between different gaming groups, all Fs < 1.3.

To assess accuracy in mental rotation performance, we calculated d-prime measures based on match and non-match trial performance, and analyzed them as per RT data above. Participants were more accurate at smaller stimulus angles, $F(3, 168) = 99.06$, $p < 0.001$, $\eta_p^2 = 0.61$, and participants also improved their sensitivity over days, $F(1, 64) = 12.02$, $p = 0.001$, $\eta_p^2 = 0.16$. Improvement over time did not interact with stimulus angle, $F < 1.2$. A marginal interaction was observed between the Depth and Memory group variables, $F(1, 64) = 3.30$, $p = 0.074$, $\eta_p^2 = 0.05$, with overall lower d-primes in the Depth-Hard+Memory-Hard and Depth-Easy+Memory-Easy groups versus the other two groups.

This between groups interaction was modified by a marginal three-way interaction of these two training group variables and day, $F(1, 64) = 2.87$, $p = 0.095$, $\eta_p^2 = 0.04$, suggesting a potential training effect of relatively larger improvement in d-prime scores in these Hard-Hard and Easy-Easy groups versus the other two groups. However, the fact that these larger training improvements are selectively observed in conditions with overall lower scores may reduce our confidence in attributing them solely to training and transfer effects. Assessing this same effect in Accuracy data for Match trials shows a slightly different pattern, with a significant interaction of day by Depth and Memory conditions, $F(1, 64) = 6.29$, $p = 0.015$, $\eta_p^2 = 0.09$, driven mainly by selective improvement in accuracy in the Depth-Hard+Memory-Hard group.

Visual Search

Data for five participants were excluded from analysis due to accuracy below 50% for two or more task conditions. Figure 3 shows visual search data separated by display size and group, for mean correct RT and Accuracy data for target-present trials, and d-prime data calculated from target-present versus target-absent performance. Correct RT data were analyzed via repeated measures ANOVA, with within-subjects factors of Day (1 and 9), and Display Size

(4, 10, 18 and 26 items), and between-subjects factors of Depth (Easy vs. Hard) and Spatial Memory (Easy vs. Hard) training. We limited the influence of extremely slow, but still correct trials by replacing RTs over 5000 ms with a 5000 ms value. RTs faster than 300 ms were excluded from analysis.

Results typical of visual search tasks were observed. A strong main effect of day was observed, $F(1, 59) = 21.73$, $p < 0.001$, $\eta_p^2 = 0.27$, reflecting improvement in RT performance over time, along with a strong main effect of display size, $F(3, 177) = 662.97$, $p < 0.001$, $\eta_p^2 = 0.92$, with slower RTs as display size increased. An interaction between day and display size reflected greater improvement in RTs over time for larger display size trials, $F(3, 177) = 3.40$, $p = 0.019$, $\eta_p^2 = 0.05$.

An interaction between Memory condition, display size and day was observed $F(3, 177) = 2.69$, $p = 0.048$, $\eta_p^2 = 0.04$, suggesting a potential training effect, but no systematic difference in improvement over display sizes was observed (for example, while larger improvements were seen for the 26-item displays in Hard Memory conditions, the reverse pattern was seen for 18-item displays, with intermediate effects at smaller display sizes). The absence of any interaction of day with either of the training group variables, *F*s < 0.6, further suggests that this interaction likely does not reflect a meaningful visual search improvement due to game training. Analysis of target absent trials (not shown) revealed similar strong effects of day and set size, but showed no evidence of training differences between gaming groups.

To assess accuracy in visual search task performance, we calculated d-prime measures based on target present and target absent data, and analyzed them as per RT data above. Participants showed overall good accuracy performance, with better d-prime performance with smaller display sizes, $F(3, 177) = 14.20$, $p < 0.001$, $\eta_p^2 = 0.19$. While a marginal interaction between Depth and Memory training conditions suggested a potential overall performance difference between groups, $F(1, 59) = 3.56$, $p = 0.064$, $\eta_p^2 = 0.06$, d-prime did not change over day, with no interaction of day with game training conditions, Fs < 1.2.

Digit N-Back

Data for four participants were excluded based on below 50% accuracy performance on one or more days. Data analysis was performed separately for 1-back and 2-back tasks. Figure 4 shows mean correct RT and Accuracy data for Target (match) and Filler (non-match) trials, and d-prime data calculated from Target versus Filler trial performance for 1-back and 2-back tasks, separated by day and training group.

RT data for correct trials were filtered to include only those trials with RTs between 300 ms to 2000 ms for analysis. RT data for Target trials were analyzed via repeated measures ANOVA, with within-subjects factors of day (1 and 9), and trial type (Target vs. Filler), and between-subjects factors of Depth (Easy vs. Hard) and Spatial Memory (Easy vs. Hard).

For the 1-back task, participants generally improved their RT performance over time, $F(1, 60) = 28.91$, $p < 0.001$, $\eta_p^2 = 0.32$, and target trial RTs were generally faster than for filler trials, $F(1, 60) = 27.76$, $p < 0.001$, $\eta_p^2 = 0.32$. The RT difference between target and filler trials became less pronounced over time, supported by the interaction of day and trial type, $F(1, 60) = 12.03$, $p = 0.001$, $\eta_p^2 = 0.17$. No game training group differences or interactions with day were observed, Fs < 1. Only a non-significant interaction between day, trial type and Depth training condition was observed, $F(1, 60) = 2.63$, $p = 0.110$, if anything suggesting possible differences between target and lure performance over time.

For the 2-back task, participants generally improved RT performance over time, $F(1, 60) = 47.35$, $p < 0.001$, $\eta_p^2 = 0.44$, and showed faster RTs for target trials than filler trials, $F(1,60) =$

24.76, $p < 0.001$, $\eta_p^2 = 0.29$, with a marginal interaction of task type by day, $F(1,60) = 3.32$, $p = 0.073$, $\eta_p^2 = 0.05$, suggesting decreased target vs. filler differences over time. A significant effect of both Depth and Memory training groups, day and trial type was observed, $F(1, 60) = 4.73$, $p = 0.034$, $\eta_p^2 = 0.07$, showing a differential pattern of target vs. filler trial improvement over time across groups. However, the absence of any interaction between group training variables with day suggested no evidence of a general group training effect, Fs < 1. To more directly test this, we assessed 2-back target and filler RT data separately. For target trial data, no interactions between training group variables and day were observed, Fs < 1. For filler trial data, no interactions were seen for comparisons of Depth training condition by day, or Memory training condition by day, Fs < 0.6. The interaction of Depth and Memory conditions with day was not significant, $F(1, 60) = 1.95$, $p = 0.168$.

D-prime scores were analyzed as per RT data above. For the 1-back task, d-prime appeared to show some degree of greater improvement for Hard Memory training conditions over time, but this effect was not significant, $F(1, 60) = 2.34$, $p = 0.131$, with no main effect of day, $F(1, 60) = 1.89$, $p = 0.174$, and no other interactions, Fs < 0.2. Overall group differences were observed independent of training effects, with relatively lower d-prime performance in the Easy Depth training conditions, $F(1, 60) = 7.41$, $p = 0.008$, $\eta_p^2 = 0.11$, and a marginal effect of lower overall d-primes in the Hard Memory training conditions, $F(1, 60) = 3.18$, $p = 0.080$, $\eta_p^2 = 0.05$. These pre-existing differences may reflect a better opportunity to observe larger changes with practice, such as the marginal effect for Hard Memory conditions described above, independent of game training itself. For the 2-back task, d-prime performance did generally improve over time for all groups, $F(1, 60) = 10.25$, $p = 0.002$, $\eta_p^2 = 0.15$, with no interactions with training groups, and no overall differences between groups, Fs < 1.

Spatial N-Back

Data for four participants were excluded based on sub-50% accuracy performance on one or more days. Data analysis was performed separately for 1-back and 2-back tasks. Figure 5 shows mean correct RT and Accuracy data for Target (match) and Filler (non-match) trials, and d-prime data calculated from Target versus Filler trial performance for 1-back and 2-back tasks, separated by day and training group. Analyses were performed as for digit n-back, described above.

For the spatial 1-back task, participants showed improved RTs from day 1 to day 9, $F(1, 60) = 25.32$, $p < 0.001$, $\eta_p^2 = 0.30$, and RTs for target trials were faster than for filler trials, $F(1, 60) = 48.47$, $p < 0.001$, $\eta_p^2 = 0.45$, with no significant reduction of target vs. lure differences over day, $F(1, 60) = 2.69$, $p = 0.107$. Participants in the Hard Memory conditions showed marginally slower RTs overall, as compared to participants in the Easy Memory conditions $F(1, 60) = 2.84$, $p = 0.097$, $\eta_p^2 = 0.05$, but no training effects over day were observed, Fs < 1.1.

For the spatial 2-back task, participants again improved RTs over days, $F(1, 60) = 39.98$, $p < 0.001$, $\eta_p^2 = 0.40$, and showed faster RTs for target vs. filler trials, $F(1,60) = 30.42$, $p < 0.001$, $\eta_p^2 = 0.34$. Participants in the Hard Memory condition showed slower RTs overall than in the Easy Memory condition, $F(1, 60) = 6.53$, $p = 0.013$, $\eta_p^2 = 0.10$, but no evidence of group training effects was observed, Fs < 1.

D-prime scores were analyzed as per RT data above. For the spatial 1-back task, there was a general improvement in performance over time, $F(1, 60) = 25.87$, $p < 0.001$, $\eta_p^2 = 0.30$. Apparent differences in improvement over time between training groups were not reliable, with no interaction of either group training variable with day, Fs < 0.6, and no significant three-way interaction, $F(1, 60) = 2.50$, $p = 0.119$, $\eta_p^2 = 0.04$. D-prime scores appeared somewhat smaller overall for Hard than for Easy Memory conditions, but this effect was not significant, $F(1, 60) = 2.24$, $p = 0.140$, $\eta_p^2 = 0.04$, with no other between-subjects effects, Fs < 1. For the spatial 2-back task, there was strong general improvement over time, $F(1, 60) = 48.24$ $p < 0.001$, $\eta_p^2 = 0.45$, with no other significant effects observed, Fs < 1.6.

Spatial Working Memory

Accuracy data for the spatial working memory task are shown in Figure 6, separated by memory set size, day and training group. Accuracy data were analysed via Repeated Measures ANOVA, with within-subjects factors of day (1 and 9), and set size (4 vs. 6), and between-subjects factors of Depth (Easy vs. Hard) and Spatial Memory (Easy vs. Hard). Accuracy data here are reported and analyzed as the absolute number of items recalled per set rather than the proportion of items, as proportion will be necessarily smaller with increasing set size given limited working memory capacity.

Participants correctly recalled an overall larger number of items on 6-item vs. 4-item trials, $F(1, 64) = 281.06$, $p < 0.001$, $\eta_p^2 = 0.82$. The small overall numerical decrease in performance over time was not significant, $F(1, 64) = 2.22$, $p = 0.141$, $\eta_p^2 = 0.03$, and this change over time did not interact with Depth training condition, $F(1, 64) = 1.98$, $p = 0.165$, or Memory training condition, $F(1, 64) = 1.51$, $p = 0.223$, nor was there a 3-way interaction, $F < 0.6$. Overall performance appeared worse in the Hard memory training condition than in the Easy memory training condition, which produced a marginal main effect of Memory training group, $F(1, 64) = 3.83$, $p = 0.055$, $\eta_p^2 = 0.06$. This effect was most notable for 6-item trials, supported by a significant interaction of set size and Memory training condition, $F(1, 64) = 6.00$, $p = 0.017$, $\eta_p^2 = 0.09$. No other effects were observed, $Fs < 1.2$. Analysis of this dataset using proportion correct data showed equivalent effects as described above, though with no significant interaction of set size and Memory training group.

Discussion

While typical results were observed for a range of different laboratory tasks, little if any convincing evidence of transfer from game training manipulations to laboratory task performance was observed. What limited evidence we did observe (e.g., relative accuracy improvement for Depth-Hard+Memory-Hard group in mental rotation, and for Hard vs. Easy Memory groups in digit 1-back d-prime performance) occurred against a background of overall poorer performance for these same groups. In these situations, general practice might result in a relatively larger improvement (compared to a worse initial state), independent of selective effects from training in a particular game condition. Especially given the absence of any other convincing group-specific training effects, we interpret the current data very cautiously, and somewhat skeptically.

The results from "Orphlings" may seem a little disheartening at first, but the goal of this experiment was to develop a method for studying the effects of video games that is methodologically constrained and conservative. By creating such tight controls on methods we can study how much of the training on a video game transfers to other skills and we can be sure of the effect sizes and generalizability of our study.

Overall then, the results from the Orphlings study suggest that this particular game may not be well suited to be used as a cognitive skill training tool, at least for the kinds of tasks and related cognitive abilities we test here. One general possibility is that the game itself does not involve a sufficient degree of demand or challenge for those cognitive abilities compared to everyday experiences of participants, and so in practice the different training groups did not end up with very different amounts of practice on manipulated training conditions (i.e., general life experiences over this time were more than enough to effectively equate groups for any game-related differences).

In a similar sense, a related possibility is that the kinds of cognitive abilities targeted by training manipulations in the game may be relatively insensitive to practice – even if the game offered a substantial degree of demand or challenge, certain cognitive abilities may be much harder to improve than others. Taking this idea further, we might consider what kinds of training and transfer effects seem plausible with this or other kinds of training. For example, improving performance via refining a strategic approach to a task seems much more likely than improving the fundamental capacity of selective attention – while both could produce improvements in working memory performance (for example), the former seems more likely than the latter.

These and other issues of training and transfer from video game experience, plus comparisons between our different training experiments, are considered more fully in the General Discussion chapter (Chapter 5), in the following pages.

Chapter 5 - General Discussion

This dissertation presents three separate training studies, investigating transfer of particular cognitive skills from training on commercial video games. The first study, presented in Chapter 2, was an initial proof-of-concept study using the commercial spatial/memory puzzle game “Membrain.” This study tested for particular performance improvements (mental rotation and n-back working memory) that might be selectively caused by playing Membrain, compared to playing the traditional number puzzle Sudoku. Selective improvement on some tasks with equivalent improvement on others suggested transfer to post-training tasks from Membrain gameplay.

Chapters 3 and 4 presented two larger-scale studies, where we developed a more carefully controlled training-and-transfer methodology. In these studies, all participants played the same commercial video game, and we manipulated the relative proportions of particular easier or harder game elements to give different groups of participants more or less experience with particular kinds of game features. This design uses the game as its own control, with any potential transfer effects much more attributable to the particular well-controlled feature variations between training groups within the same game context. In Chapter 3, “Paint the West” showed a range of specific transfer effects from increased distractor stimulus similarity, number and crowding within the game, to both improved performance on resisting the influence of nearby distractor stimuli in cognitive tasks, and also to changes in speed/accuracy criterion performance in other speeded tasks. In Chapter 4, with the same kind of conceptual training and transfer design using a different kind of game, training with “Orphlings” showed minimal convincing transfer to a range of working memory and spatial tasks.

These studies were conducted with a number of goals in mind. The first thing we wanted to explore with these experiments is whether training on a video game could improve training-specific cognitive skills, and whether these specific transfer effects could be isolated well enough from other more general effects of task practice. Isolating these specific training and transfer effects allows a better estimate of the true effect size of game-specific training improvements on task performance, which we suspect is over-estimated in most video game research.

Riesenhuber (2004) questioned whether hours of playing “Where’s Waldo?” would result in striped sweaters jumping out to a person’s attention in crowded stores, or whether it would give baggage screeners the ability to detect dangerous items in a suitcase more efficiently. Psychologists have been asking for many years whether transfer of skills from one domain to another actually occurs and to what extent it may occur. Now we need to ask the question of whether skills practiced while playing video games transfer to other areas, and how general or specific these training and transfer effects are.

By following Barnett and Ceci’s (2002) set of rules for transfer, it becomes evident that many of the improvements seen in our experiments are specific to the attributes and characteristics of the games played, and that transfer between the games and tasks occurred mostly when the tasks had similar attributes to the game. It also appears that transfer may occur more readily for cognitive abilities or skills that are not performed at near optimal levels, or skills that do not receive regular practice.

The “Membrain” experiment findings suggest that some of the training received through the game may have improved performance on selective cognitive tasks by means of near transfer. Training with “Membrain” improved participants’ speed of mental rotation, without an accompanying drop in accuracy. The attention to particular locations and features of 3D objects in space and the demand to represent and remember them in a dynamic moving 3D situation, may have improved participants' abilities to represent and manipulate 3D mental rotation objects. It seems unlikely that this game practice altered a fundamental feature of spatial attention. Instead, we suggest here that "Membrain" offered participants a chance to practice a skill that is not used on a daily basis, thus improving performance for a cognitive system that is not functioning at already high levels of performance. The fact that this training-specific improved performance on mental rotation was not seen with the other two studies increases our confidence that this was a training-specific effect, as neither "Paint the West", nor "Orphlings" involved the kind of 3D spatial processing demands required by "Membrain".

Data from digit n-back tasks across several of these studies show interesting and convincing evidence for the specificity of transfer effects. In the "Paint the West" experiment, participants from the Hard Visual Attention groups showed greater improvement in reaction time performance for the 1-back task than participants from the Easy Visual Attention groups. These reaction time improvements did not come at the cost of accuracy. In contrast, all groups improved equally for 2-back performance. This particular pattern of data is convincing in that we can observe a game-related training effect on one condition of a task (1-back performance), where a set of within-task comparison conditions show equal improvement for all groups (2-back task). Consideration of the different task demands of 1-back and 2-back tasks, and the kinds of game training differences between Easy and Hard Visual Attention groups in the "Paint the West" game are suggestive of specificity in training and transfer from gameplay. This improvement could reflect an increase in the speed of visual processing (Li, Polat, Makous & Bavelier, 2009), or an increase in the speed of response selection within a relatively low-conflict or low-preparation task situation (1-back is a relatively simple repetition detection task). The equivalent 2-back improvement suggests this was not a training effect involving working memory ability.

Training with “Membrain” on the other hand, resulted in selective performance improvement on the 2-back task (faster reaction times, with no drop in accuracy) versus Sudoku training, but with equivalent improvement across both game groups for 1-back performance. “Membrain” required participants to remember identifying stimulus-specific information (here, both object and location information) for a continuous series of recently observed stimuli, in order to match them with newly presented items. These game requirements seem to be a close match to maintenance and updating of working memory requirements central to successful 2-back performance. In contrast, neither Membrain nor Sudoku required or induced a speeded response demand, and so we might not expect any kind of selective training for this element of task performance.

Given the specificity we observe for training experiences on particular task elements, and the subsequent selectivity of transfer to relevant subsets of laboratory task performance depending on the particulars of game training, the n-back results we discuss here would appear to represent a convincing case of training and transfer from video games. Improvements in n-back tasks are often (but perhaps wrongly) expressed as “improvements in working memory capacity” – improved n-back performance could certainly result from improved working memory capacity, but many other changes in elements of the working memory task, changes in task strategy, or changes in related cognitive skills (e.g., getting better at chunking sequences) could all improve n-back performance without any change to the fundamental capacity of a participant’s working memory system.

Of course, this brings up the issue of what is implied by the term "capacity" here – capacity of some more fundamental memory or attentional system, or the observed performance independent of underlying mechanism, or something in between? Here, this is discussed to highlight the fact that performance differences through training need not reflect improvement in fundamental cognitive mechanisms. Indeed, such improvement might be expected to show far transfer to many disparate tasks that shared that underlying mechanism; in contrast, observing only near transfer in our data suggests that training effects must be mediated by training of more limited and task-specific strategies and skills.

I do not wish to suggest here that either "Membrain" or "Paint the West" improved participants' overall working memory capacity. As discussed above, "Paint the West" results suggest an improvement of visual attention or response decision processes under low-conflict task demands, independent of any working memory differences. "Membrain" results could suggest an improvement in any one of the component elements of working memory task performance – simple primary memory, access to information in secondary memory, or the ability to control and allocate selective attention (Shipstead, Lindse, Marshall, & Engle, 2013). An improvement in any of those components could result in overall decreases in reaction times for tasks that require working memory, without actually improving working memory capacity. Improvements could also be the result of improved strategy (not necessarily explicit) or refinement of cognitive skills related to n-back performance, such as improved scheduling of attending to each stimulus element, or better practice at chunking of stimulus groups. In general, the point is that all of these potential mechanisms of improvement speak to near-transfer-like effects, and not improvement of more fundamental abilities which would otherwise show widespread improvement on many other kinds of performance.

The improvement seen in the flanker task of the “Paint the West” experiment is also indicative of task specific transfer effects. The distractor (flanker) effect in the flanker task is thought to be a measure of attentional capacity, and measures the influence or cost of a to-be-ignored distractor on reaction time. Participants in the Hard Visual Attention groups showed selectively greater reaction time improvements for incongruent trial performance than participants from the Easy Visual Attention group, with equivalent improvement for congruent trials across all gaming groups. This selectively greater improvement for incongruent flanker trials may be the most direct example of near transfer in the present studies – game practice with a higher proportion of distractors that were more featurally similar to targets, were presented spatially closer, and presented in greater numbers, may have given participants a great deal more practice filtering or resisting the influence of irrelevant distractor stimuli within a speeded choice performance task. Selective improvement on incongruent flanker trials for participants with greater degrees of this kind of game training demand would appear to be an excellent demonstration of near transfer.

In experiments comparing NVGPs versus expert VGPs, Green and Bavelier (2003) assessed flanker performance over a range of overall task difficulties. In easier trials, when sufficient attentional resources were available to both groups of players, both NVGPs and expert VGPs were unable to ignore distractors, causing a sizeable flanker congruency effect. As the task became harder, the attentional resources left to process the intruding distractors diminished at a faster rate for NVGPs, and their flanker congruency effect became smaller, while expert VGPs did not show a reduction in the size of this distractor effect, suggesting that they may have more attentional resources available to them than NVGPs. Although these general findings (expert VGPs show larger flanker effects) seem on the surface at odds with our current data (game training with more challenging distractors show reduced flanker interference costs), they represent two very different training situations. Expert VGPs may well have greater attentional resources to deploy (whether because of game training or other individual differences), and flanker interference is one way to assess these differences; in our present data, specific training on dealing with conflicting distractor information in our game transfers to very different materials in a laboratory flanker task, suggesting transfer of attentional skills or ability.

Several of the improvements seen in the cognitive tasks in these studies seem to have come as a result of a change in criterion as a result of training, and not as a change in ability or capacity of performance for a particular task. The reaction time improvements seen for “Paint the West” participants in the Hard Visual Attention group on the Go/No-go task came at a cost to their accuracy scores. This so-called “trigger happy” shift in performance criterion was similarly suggested for the same participants by marginal shifts in the speed/accuracy criterion for Mental Rotation and Visual Search results. For laboratory tasks that did not have close overlap with Paint the West gameplay, participants appeared to show a general tendency for speeded responding. This may reflect an example of inappropriate transfer – applying a previously learned and well-tuned response criterion or degree of control/automaticity (perhaps implicitly, or elicited from a general “speeded performance” task context) to a situation similar enough to prior experiences (another speeded response situation), but where participants had relatively little specific experience with the new task.

It is also interesting to note that the shift in criterion or change in speed/accuracy tradeoff with game training occurred only with the “Paint the West” game. Other studies have noted shifts in strategy where players select strategies to complete cognitive tasks that were similar to the attributes emphasized in the games they played. Nelson and Strachan (2009) noted that participants that were primed on puzzle games relied on accuracy based strategies to complete cognitive tasks, while participants primed on action games relied more on speed based strategies to complete cognitive tasks. We consider "Paint the West" to be a speed based shooter game, while both "Orphlings" and "Membrain" would be considered cognitive puzzle style games.

With this framing, it is also important to note that in Paint the West, we can distinguish data showing transfer of criterion or strategy from game training, and also for the same participants show selective improvement in absolute task performance for particular kinds of performance specifically supported by elements of game experience. Training or priming of speed versus accuracy based approaches cannot completely account for some of the near transfer effects we observe; that said, while they are reliable, the absolute size of these pure enhancements of performance from training are very small compared to typical effects of expert VGP versus NVGP effects often discussed in the literature.

Finally, we have not discussed the Orphlings study very much, as there were few (if any) convincing training effects observed. The few training differences were mainly observed in groups whose initial pre-test performance was lowest amongst groups, with final post-test performance equivalent to other groups. The absence of a comparable set of training effects in Orphlings as for the other studies is not necessarily a negative outcome. As discussed at the end of Chapter 4, certain games may not provide sufficient challenge or practice of cognitive elements that is sufficiently different or more intensive compared to everyday experiences, and as such no particular additional or special training is generated by the game. Alternatively, the laboratory tasks used to test for transfer may have been insensitive to any training effects bestowed by the game. It is also possible that the kinds of cognitive abilities targeted by this game may themselves not be very sensitive to this degree of additional practice. For example, for individuals who rarely engage in intensive visuospatial manipulation of 3D figures, practice on a game with these kinds of 3D spatial demands may show substantial transfer. In comparison, other cognitive abilities may already be exercised a great deal every day, and game training may offer relatively little additional influence – in a sense, some abilities may well operate at close to ceiling efficiency, and training here may have little opportunity to improve that efficiency.

Importantly, the general set of null effects for training and transfer in Orphlings sits against a robust and well-defined set of typical cognitive effects across the battery of cognitive tasks – these data show that all groups of participants were sensitive to manipulations of visual search size, verbal and spatial n-back memory load, and many other factors, and that all participants generally improved their task performance over days. Against that background, we do not observe convincing evidence that within-game training manipulations lead to differential transfer to our laboratory tasks. To some degree, this is a reasonable demonstration of another benefit of this methodological approach – this methodology is deliberately selective instead of sensitive (we will be more convinced by strong and direct transfer effects we detect through this method, but will perhaps not detect general effects of game training, as all participants play the game), and in the case of this study, we do not see convincing effects of these differential game elements. This relative lack of findings in Orphlings makes our systematic set of effects in Paint the West look even more appealing – these methods do make it harder to find any general training effects, but those that we do find are particularly meaningful.

The training effects found in these experiments are quite interesting, but the second, and possibly more important goal for these studies was the development of a methodology for studying video games that would reduce confounds, thus increasing our confidence in the results. An important recent paper on the effects of active controls in research (Boot, Simons, Stothart &

Stutts, 2013) brings these problems to the forefront. Although their study did not test improvement on games, it measured expectation of improvements. Participants who viewed a short clip on an action game were more likely to believe that video game training would improve performance on vision and attention tasks than participants who viewed clips of non-action games. These differential expectation effects could result in significantly skewed results in training studies, at a minimum inflating effect sizes, and at its worst, suggesting causal relationships where there are none.

Another problem with many methodologies used in video game studies comes from the differing arousal levels that different games can elicit. Arousal levels have long been known to influence learning, and action video games often elicit higher arousal levels from players than puzzle or strategy video games. This might lead to biased results when comparing improvements gained from training with action video games versus improvements gained from more sedate games.

Even if arousal effects are controlled for, priming of strategies from using different games as training and control can still be a problem, as discussed above in work by Nelson and Strachan (2009). The methodology used for “Paint the West” and “Orphlings” is relatively simple, but avoids or reduces many of these potential confounds. By using the same game for all participants in the study, we were able to create an active control group (or more correctly, a crossed set of more- versus less-trained groups with respect to certain variable game parameters), where the design is better protected from confounds created by differential expectations, differences in arousal levels or different strategy priming. This should increase the generalizability of our results, and furthermore, these methods can easily be adapted so that studies can be done with commercially available games. Although our results show only small levels of improvements gained from training on video games, we suggest that the observed results reflect a more accurate measure of the true influence of training and transfer effects coming from those particular differences in gameplay variations (e.g., distractor difficulty manipulations in Paint the West). Transfer effects we see are over and above any differences that playing the game in general may have – all participants played the game – this method allows a more stringent focus on training effects that stem from specific elements of gameplay.

Despite the relatively small size of these effects, it is not to say that video game play is an inadequate training tool, or that methods other than video game play would be a superior training tool. Video games can be used as a part of a more complex strategy that includes not only diverse styles of video games but also other tools that have been previously shown to aid in cognitive skills training, such as in the case of exercise (Cotman and Berchtold, 2002) and playing musical instruments (Wan and Schlaugh, 2010). Future research might focus on specific aspects of video games that target specific cognitive deficits, thus tailoring game play training to the areas in which a person requires cognitive remediation. The research done to date on video games is very encouraging, and does lead us to believe that video games can be used as a valuable tool in the improvement of cognitive skills, with potential implications for rehabilition purposes as well.

It is also important to note that the studies in this dissertation, and the large majority of other training studies in the literature, do not assess how durable these training effects may be. In order for training and transfer to be practically useful, beneficial transfer to real world tasks from gameplay would be a lot more useful if these gains were maintained over time, rather than being short-term effects that disappear without continued gameplay. The present work does not assess these issues, but we do note that this is an important practical question with respect to more general societal interests in cognitive training through video games.

The main message of this dissertation is that the real size of training and transfer effects from video games is probably small, at least for casual to moderate players, and much of the current literature likely over-estimates both the size and the extent of real transfer effects from playing video games. When we test carefully, we can observe convincing transfer effects in very specific domains from relatively small amounts of practice on commercial video games. Although we see reliable effects in some areas, the absolute size of these effects is small. When we look for absolute benefits in transfer from game training, we see only near transfer effects. Some other apparent transfer effects to more dissimilar tasks appear on the surface to be more generalized training effects, but on closer inspection these appear to be shifts in criterion or perhaps strategic or contextual effects on speed/accuracy performance. All of these results support well-established notions in cognitive psychology that near transfer is often possible with training, but that far transfer and/or real generalization from training is extremely rare, and certainly not observed in the current literature to date.

This dissertation stands as a demonstration of a high level of care to methodology, and that ultimately illustrates a distinction between real game-specific near transfer effects on the one hand, and criterion or strategic effects on the other hand. A major issue in the literature is whether training and transfer effects from video games are real. We suggest that this is not a very useful question – it should not surprise us that some degree of near transfer is possible with appropriately matched training and transfer tasks; better questions include how much of this transfer is really from elements of gameplay, and how should we usefully interpret these data for application in the real world. This dissertation suggests that we can measure these kinds of effects more directly and more accurately, and that doing so may show us that in many domains these effects are small. Importantly, this way of assessing transfer from game training allows us to have much higher confidence in the specific effects we do see, and allows a more useful and direct interpretation of observed training and transfer effects.

References

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612-637.

Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The pervasive problem with placebos in psychology why active control groups are not sufficient to rule out placebo effects. *Perspectives on Psychological Science, 8*(4), 445-454.

Cotman, C. W., & Berchtold, N. C. (2002). Exercise: a behavioral intervention to enhance brain health and plasticity. *Trends in Neurosciences*, *25*(6), 295-301.

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423*(6939), 534-537.

Li, R., Polat, U., Makous, W., & Bavelier, D. (2009). Enhancing the contrast sensitivity function through action video game training. *Nature Neuroscience*, *12*(5), 549-551.

Nelson, R. A., & Strachan, I. (2009). Action and puzzle video games prime different speed/accuracy tradeoffs. *Perception*, *38*(11), 1678-1687.

Riesenhuber, M. (2004). An action video game modifies visual processing. *Trends in Neurosciences, 27*(2), 72-74.

Shipstead, Z., Lindsey, D. R., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language*, *72*, 116-141.

Wan, C. Y., & Schlaug, G. (2010). Music making as a tool for promoting brain plasticity across the life span. *The Neuroscientist*, *16*(5), 566-577.

Appendix B – Tables and Figures

		Angle							
		20°		60°		100°		140°	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Match Accuracy									
Day 1	Membrain	0.99	0.02	0.87	0.16	0.90	0.10	0.88	0.12
	Sudoku	0.98	0.04	0.82	0.19	0.80	0.13	0.76	0.17
Day 2	Membrain	1.00	0.02	0.89	0.10	0.95	0.07	0.91	0.08
	Sudoku	0.98	0.04	0.87	0.09	0.88	0.12	0.93	0.06
Day 8	Membrain	0.99	0.03	0.92	0.10	0.95	0.06	0.93	0.07
	Sudoku	0.95	0.05	0.90	0.13	0.92	0.08	0.88	0.12
Non-Match Accuracy									
Day 1	Membrain	0.90	0.10	0.88	0.12	0.88	0.11	0.84	0.84
	Sudoku	0.80	0.13	0.76	0.17	0.88	0.12	0.74	0.16
Day 2	Membrain	0.95	0.07	0.91	0.08	0.92	0.09	0.91	0.09
	Sudoku	0.88	0.12	0.93	0.06	0.89	0.13	0.82	0.12
Day 8	Membrain	0.95	0.06	0.93	0.07	0.90	0.07	0.90	0.11
	Sudoku	0.92	0.08	0.88	0.12	0.91	0.10	0.86	0.12
D-Prime									
Day 1	Membrain	3.66	0.83	3.12	1.07	2.83	0.97	2.35	1.12
	Sudoku	3.13	0.75	2.05	0.83	2.84	1.21	1.79	1.14
Day 2	Membrain	4.23	0.58	3.18	1.10	3.18	0.99	2.66	0.71
	Sudoku	3.77	0.72	3.02	0.95	2.94	0.99	2.27	0.82
Day 8	Membrain	4.15	0.57	3.50	0.85	2.73	0.94	2.47	1.16
	Sudoku	3.40	0.80	3.14	0.93	2.72	1.07	2.27	0.88

Chapter 2, Table 1. Mean accuracy data for the mental rotation task.

		Memory Load			
		1-back		2-back	
		Mean	St. Dev	Mean	St. Dev.
Match Accuracy					
Day 1	Membrain	0.88	0.09	0.82	0.09
	Sudoku	0.80	0.17	0.70	0.22
Day 2	Membrain	0.90	0.05	0.86	0.05
	Sudoku	0.89	0.07	0.83	0.09
Day 8	Membrain	0.91	0.04	0.86	0.12
	Sudoku	0.92	0.07	0.87	0.05
Non-Match Accuracy					
Day 1	Membrain	0.93	0.05	0.90	0.06
	Sudoku	0.80	0.17	0.80	0.20
Day 2	Membrain	0.96	0.02	0.91	0.03
	Sudoku	0.95	0.03	0.91	0.03
Day 8	Membrain	0.96	0.02	0.92	0.04
	Sudoku	0.96	0.02	0.91	0.04
D-Prime					
Day 1	Membrain	2.92	0.81	2.26	0.36
	Sudoku	1.90	1.10	1.63	1.31
Day 2	Membrain	3.11	0.39	2.47	0.30
	Sudoku	2.98	0.43	2.41	0.51
Day 8	Membrain	3.17	0.30	2.71	0.71
	Sudoku	3.32	0.52	2.53	0.43

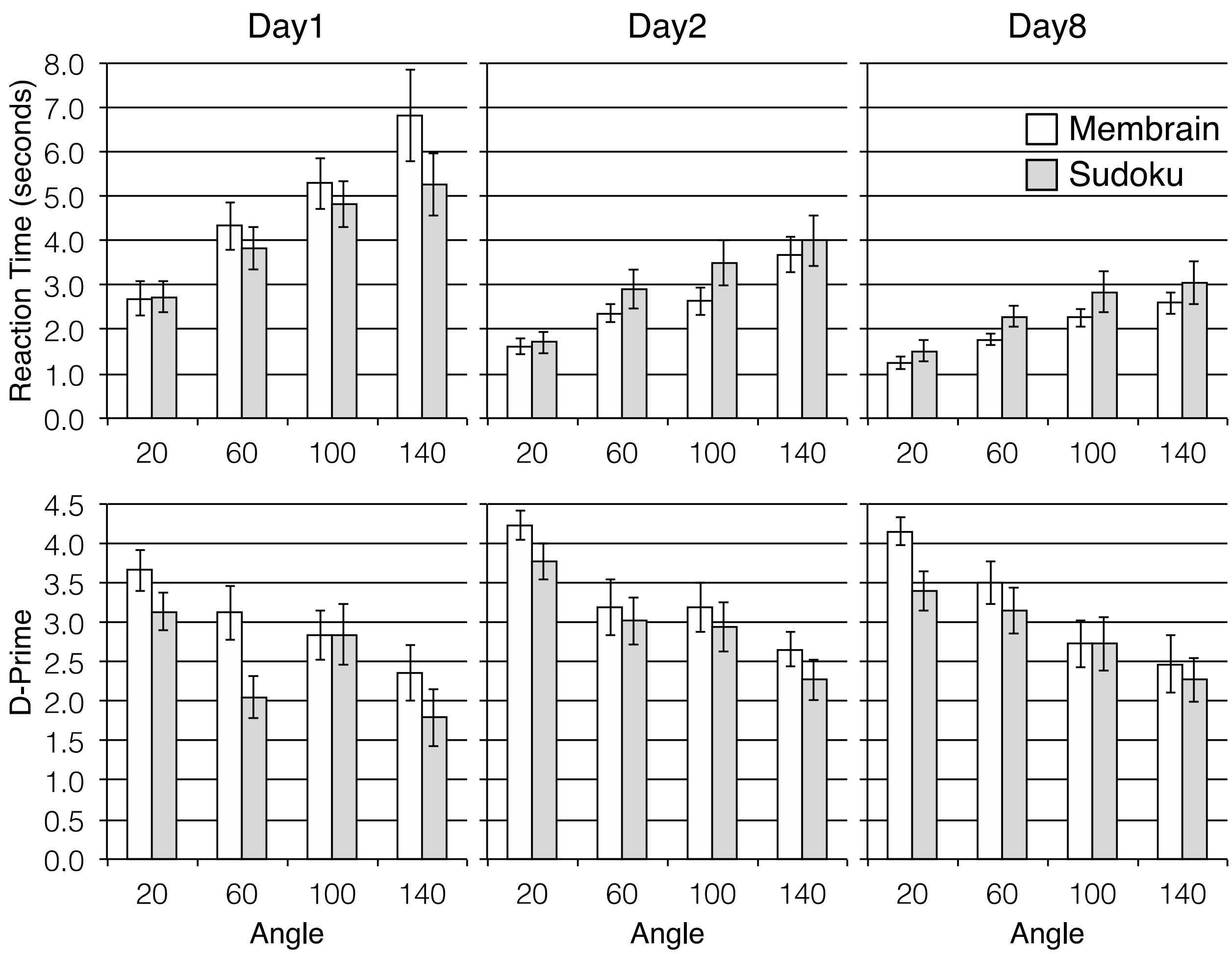
<u>Chapter 2, Table 2.</u> Mean accuracy data for the n-back task.

Chapter 2, Figure 1. “Membrain” – a 3D spatial symbol-matching memory game. Example screens from left to right show the initial state of an early level in the game (first panel), and a participant selecting two cubes to reveal their symbols, which do not match (second panel). Symbols disappear after a short delay, and the participant selects two items again, searching for a matching pair. In the third panel items match, the participant is awarded points, and those two matching cubes disappear (fourth panel). The participant continues to select pairs of cubes, looking for matching pairs. Participants can rotate the array of cubes to view from any angle. The goal is to reveal and match all hidden pairs in a minimal number of moves. Remembering which symbols are hidden on which cubes as the game progresses is the challenging memory aspect of this game.

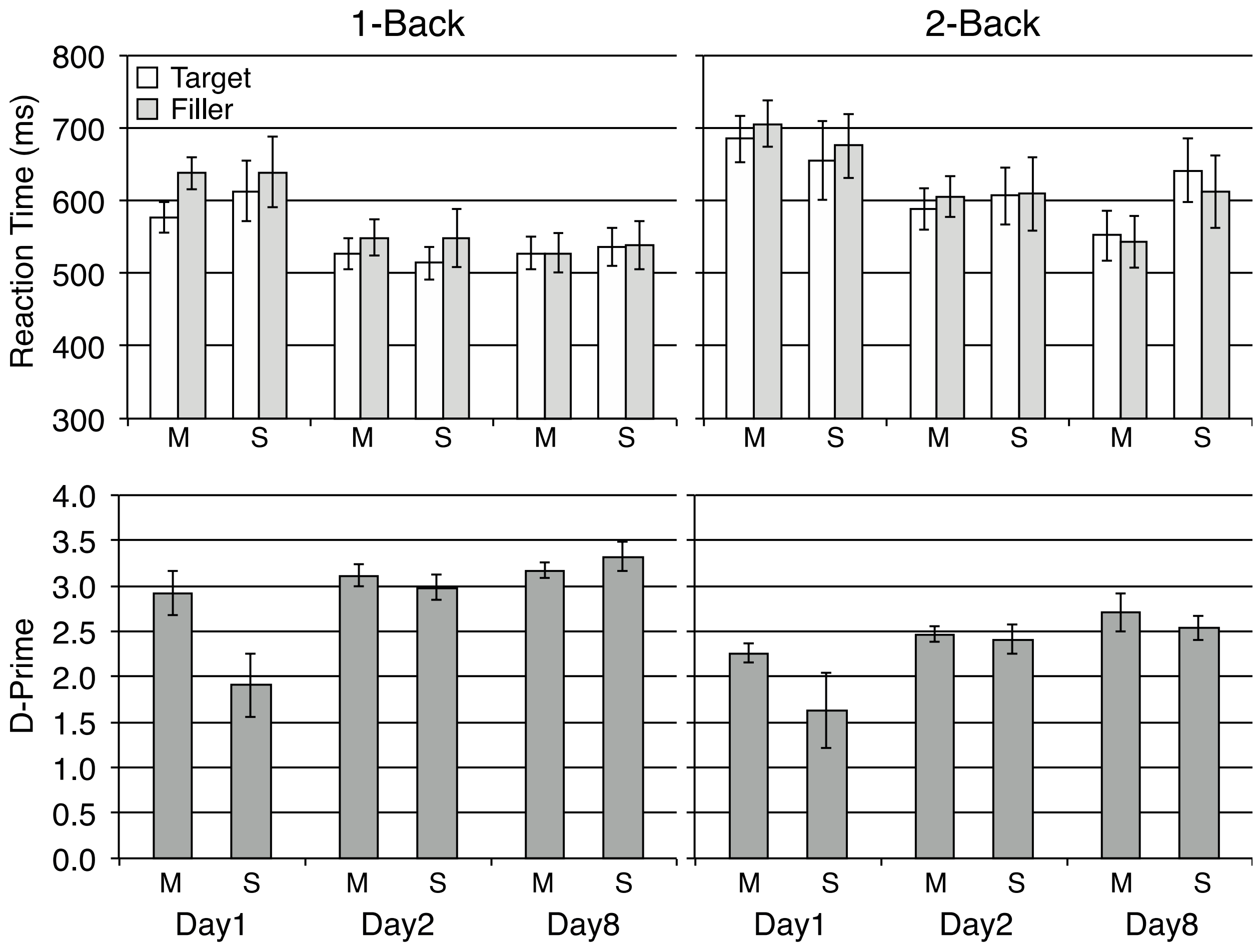
A.

B.

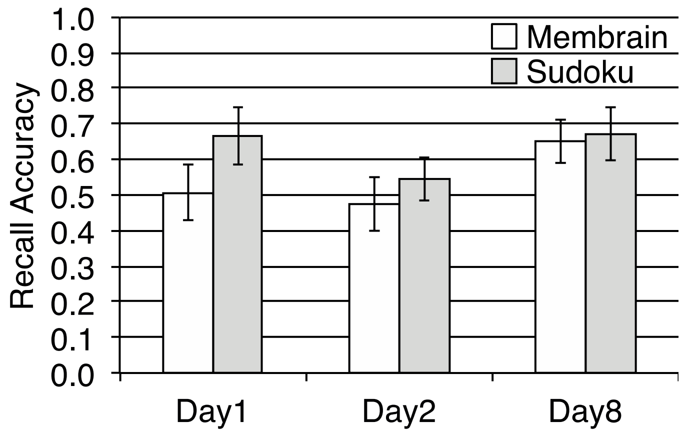
<u>Chapter 2, Figure 2.</u> Sample stimuli for the Picture Memory Task. Panel A (upper) show an initial study set of paired items. Panel B (lower) shows a test set of single background items, where participants had to recall the missing items originally associated with each test element.



Chapter 2, Figure 3. Mean reaction time (top panels) and d-prime data (bottom panels) for the Mental Rotation task, divided by day and stimulus angle (degrees). Membrain participants showed greater improvement in RT performance over days compared to Sudoku participants. Error bars represent standard errors.



Chapter 2, Figure 4. Mean reaction time (top panels) and d-prime data (bottom panels) for n-back tasks (1-back, left side panels; 2-back, right side panels), divided by day and trial type (target, filler). For 1-back task performance, participants in the Membrain (M) and Sudoku (S) training groups improved equally over days. Participants in the Membrain group appeared to show an advantage in 2-back performance (faster RT versus Sudoku participants, with equivalent d-prime sensitivity) as training progressed over days. Error bars represent standard errors.



Chapter 2, Figure 5. Mean accuracy data for the picture pairs memory task. Error bars represent standard errors.

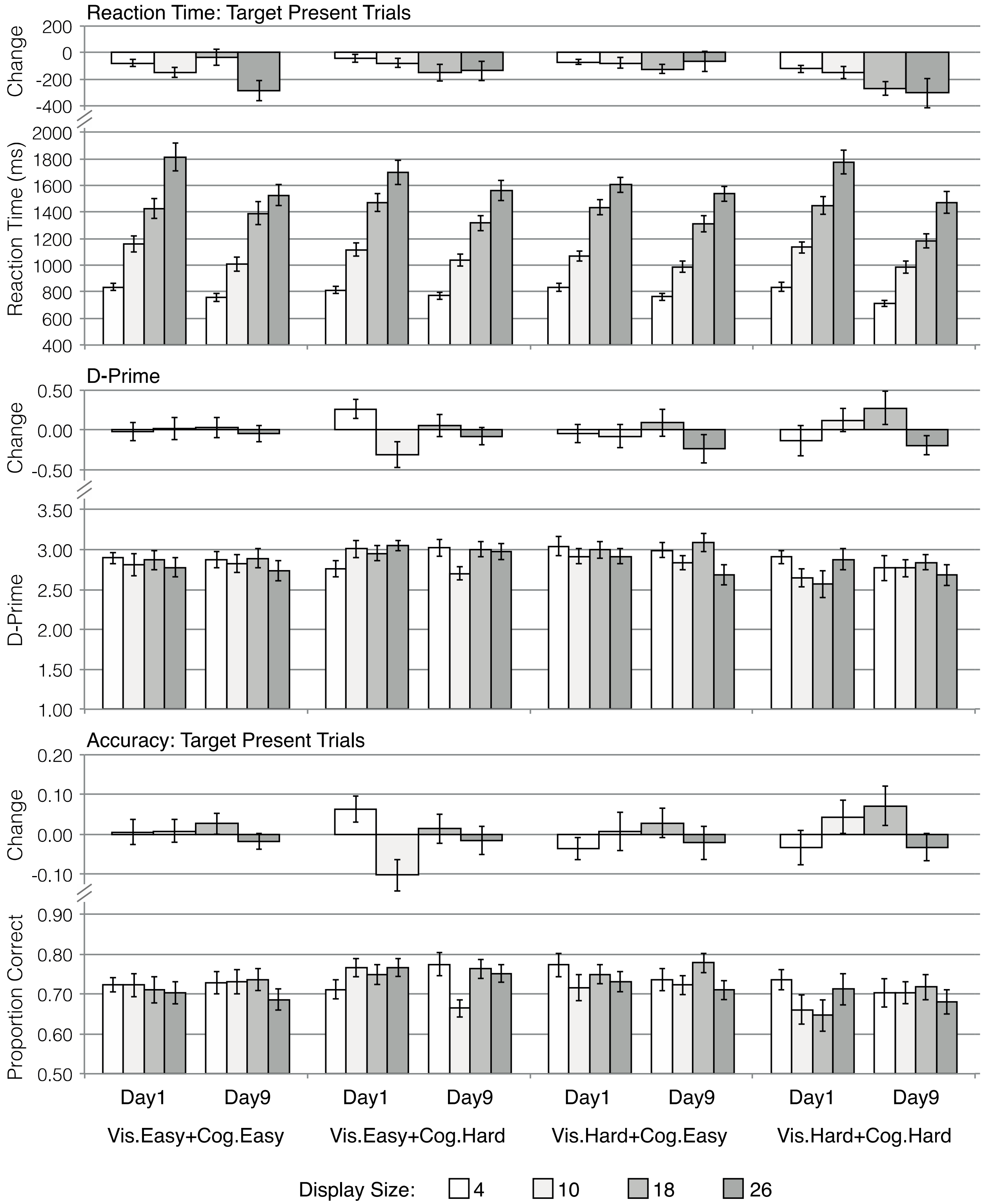
<u>Chapter 3, Figure 1.</u> Sample gameplay screen from the video game "Paint the West".

A.  B. 

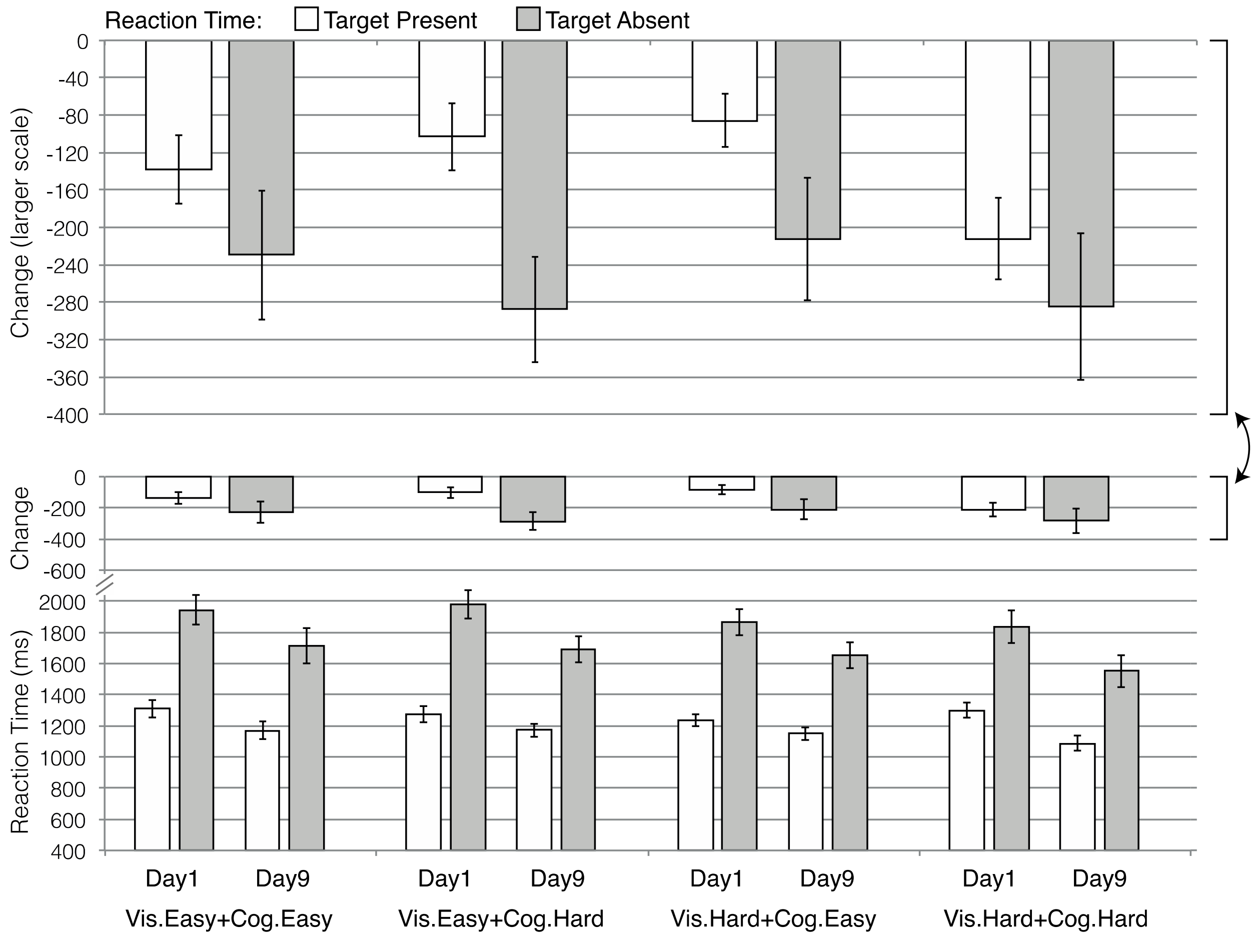
Chapter 3, Figure 2. Distractor similarity in Paint the West. Panel A shows a Townsperson (left) and a Bandit (right); Panel B shows a Sherriff (left) with a Bandit (right). Sherriffs (high-similarity distractors) have many more similar visual features to Bandits (targets) than do Townspeople (low-similarity distractors). Both Easy and Hard Visual training conditions used both kinds of distractor pairings; the Hard Visual training condition used a greater proportion of high-similarity pairings.

A.  B. 

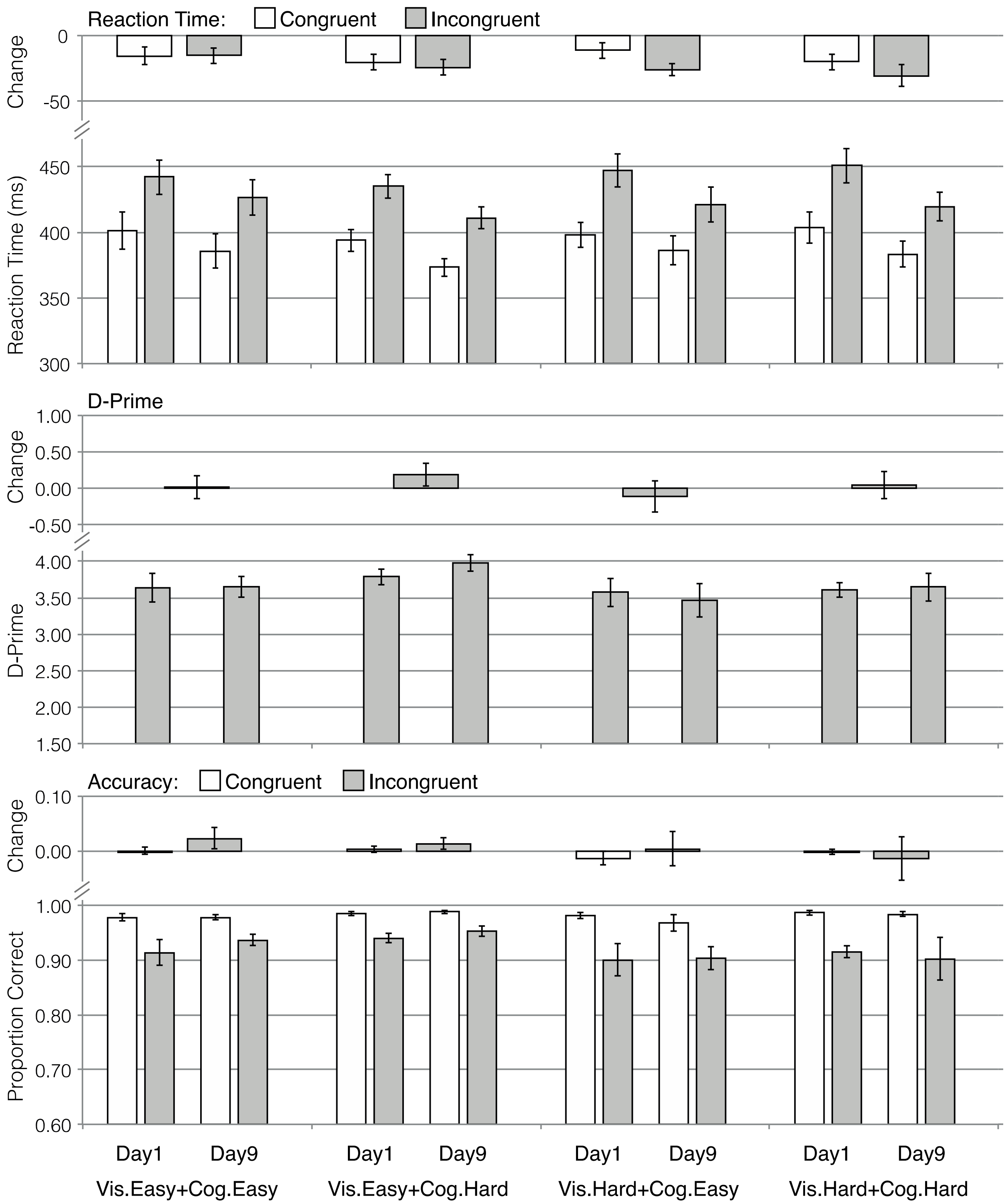
Chapter 3, Figure 3. Response selection difficulty in Paint the West. Participants could shoot Red, Yellow or Blue paintballs (via left, middle and right mouse buttons), and had to tag Bandits (targets) with a colour they were not wearing. In Panel A (low response selection demand), for a Bandit wearing only Blue, either Red or Yellow was a valid response; in Panel B (high response selection demand), only a Yellow paintball was an effective response to a Bandit wearing both Red and Blue. Both Easy and Hard Cognitive Control training conditions used both kinds of stimuli; the Hard Cognitive Control training condition used a greater proportion of high-demand stimuli.



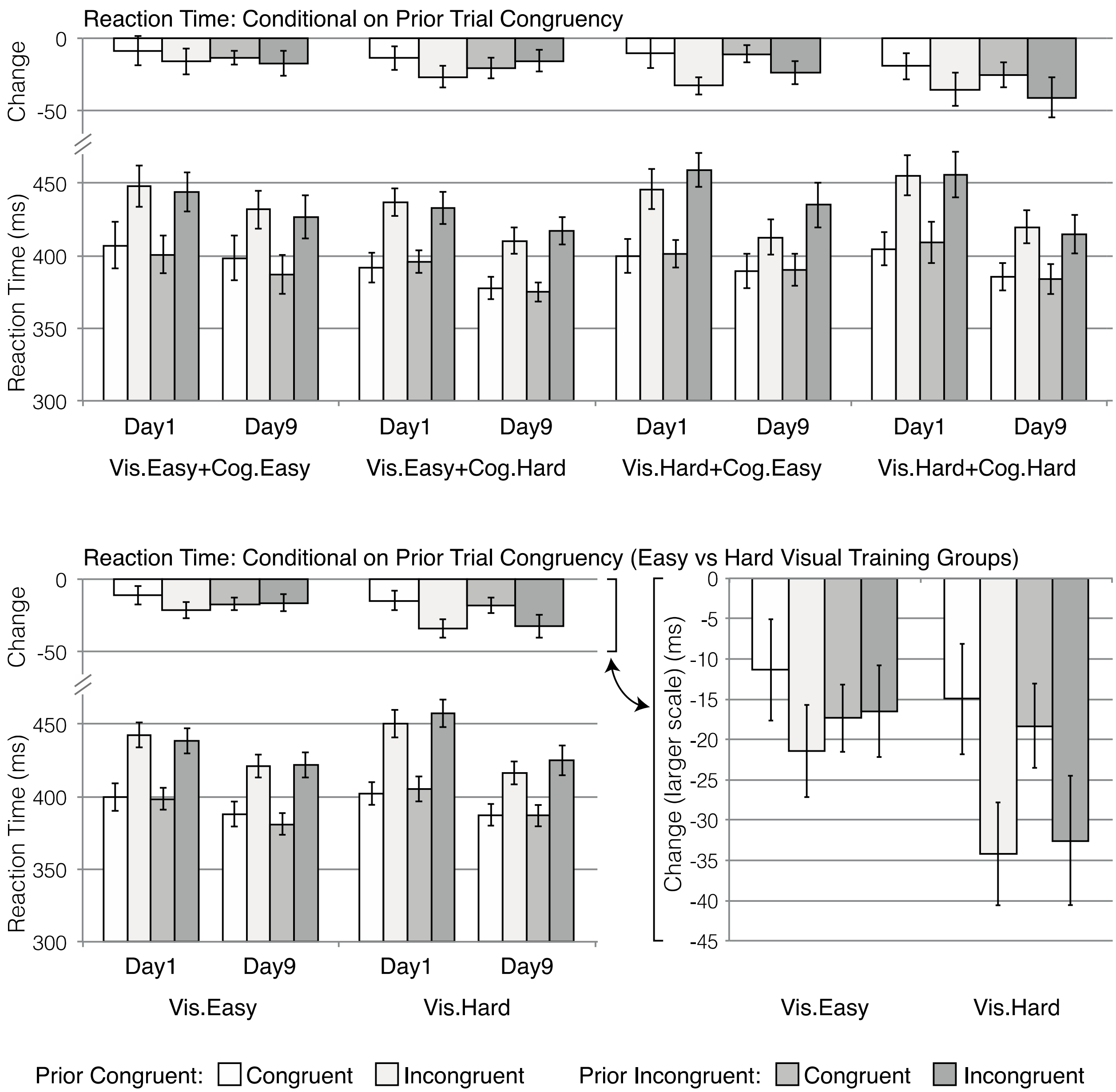
Chapter 3, Figure 4. Mean visual search data for Paint the West. Greater target RT improvement over training was seen in the Vis.Hard+Cog.Hard group. Vis.: Visual training type; Cog.: Cognitive Control training type. Error bars represent standard errors.



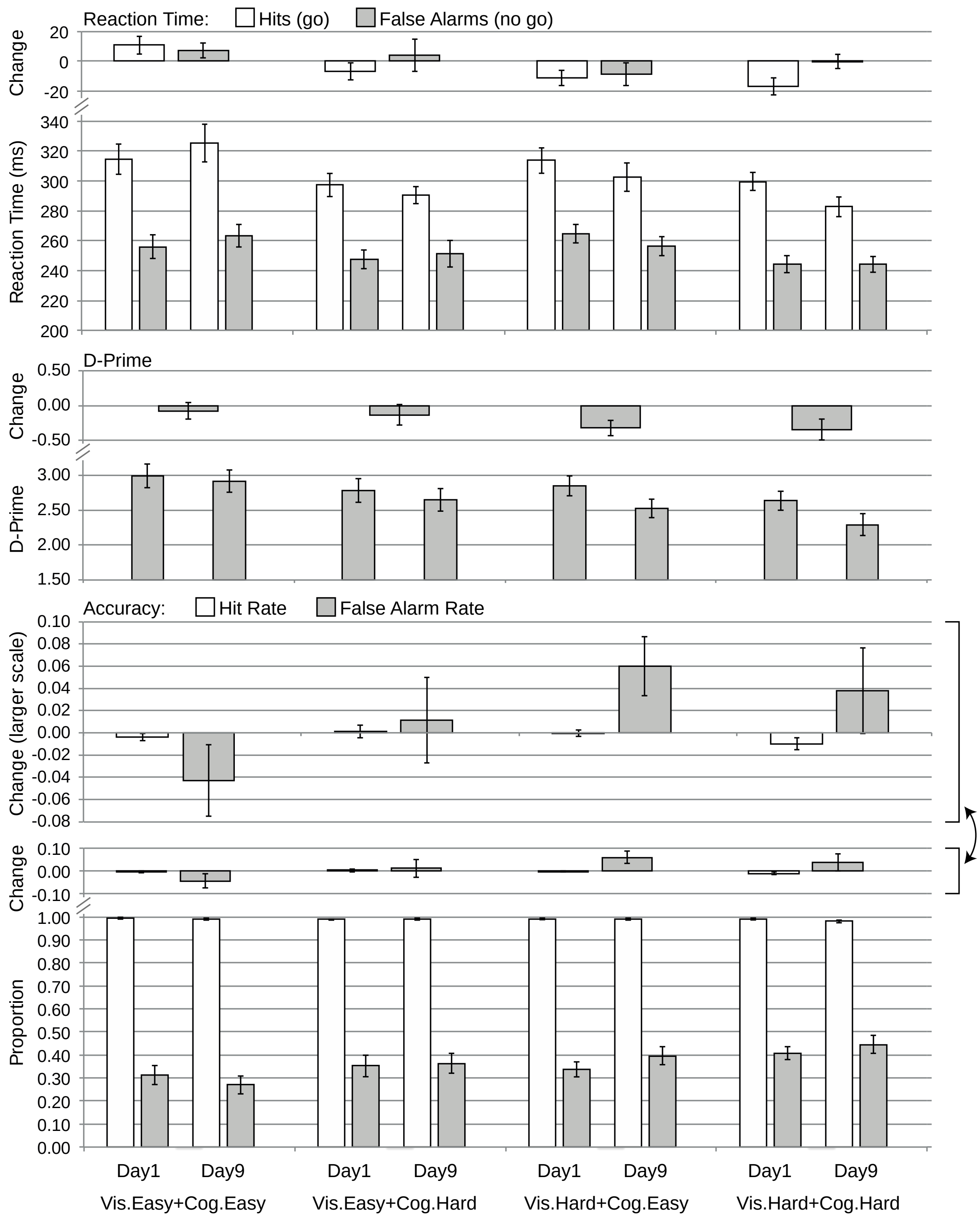
<u>Chapter 3, Figure 5.</u> Mean visual search RT data for Paint the West, for both target-present and target-absent trials, collapsed over display size. Greater target RT improvement over training was seen in the Vis.Hard+ Cog.Hard group; target-absent RT improvement is equivalent. Vis.: Visual training type; Cog.: Cognitive Control training type. Error bars represent standard errors.



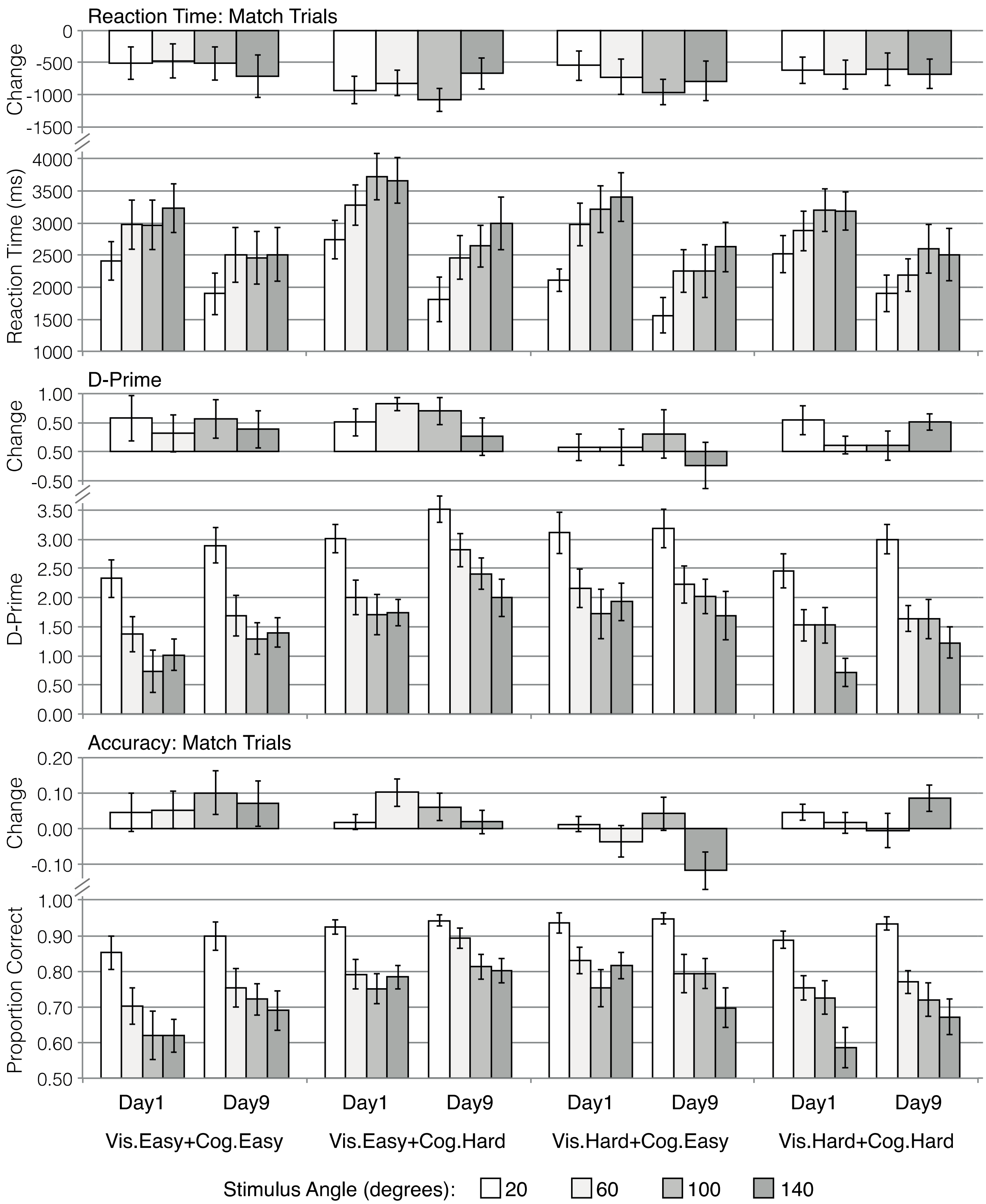
Chapter 3, Figure 6. Mean flanker task data for Paint the West. Participants in Vis.Hard groups showed larger reduction in flanker RT costs (congruent vs. incongruent change). Vis.: Visual training type; Cog.: Cognitive Control training type. Error bars represent standard errors.



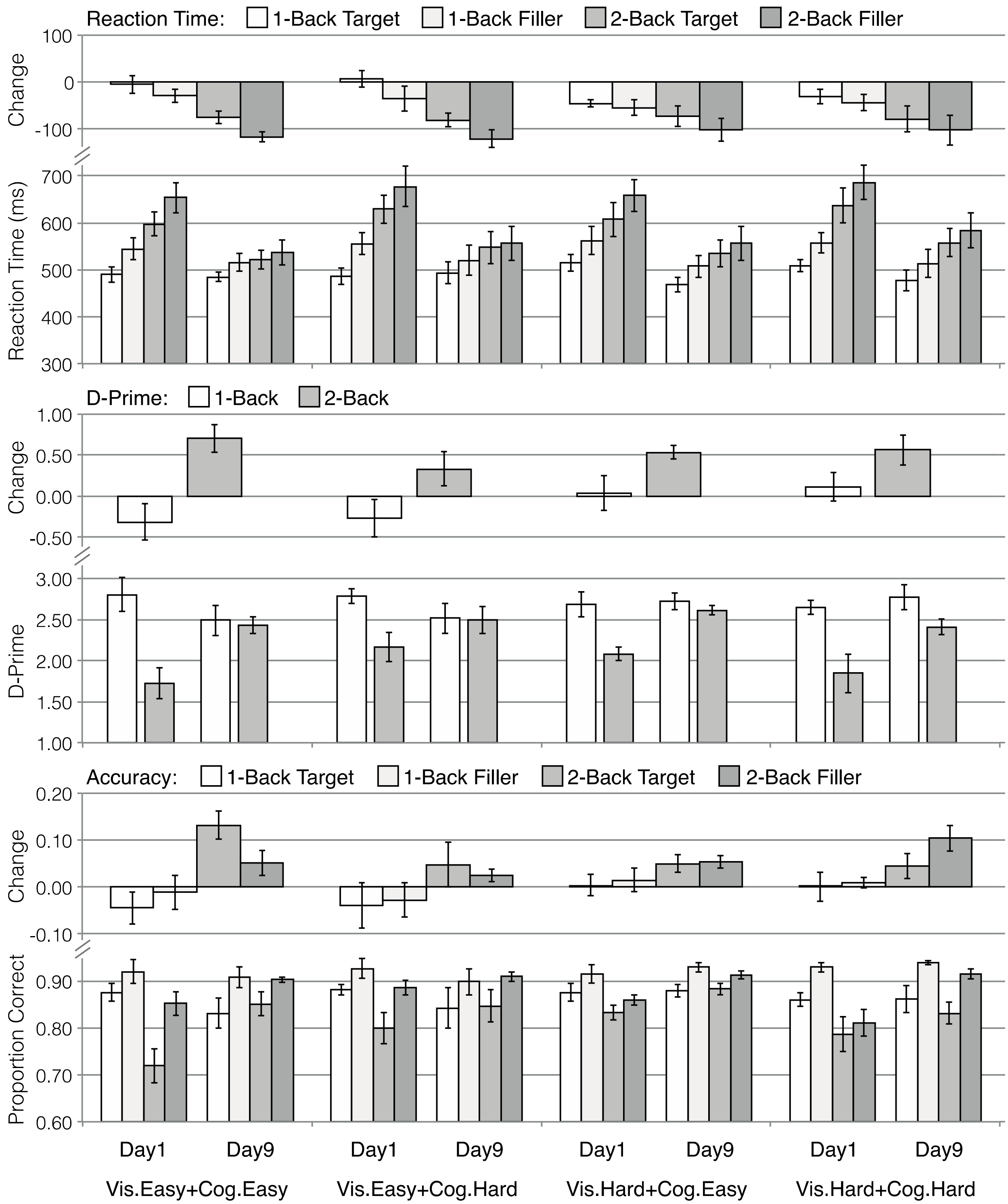
Chapter 3, Figure 7. Mean flanker task RT data for Paint the West, conditional on prior trial congruency. The upper panel shows all mean RT data; the lower two panels show these data collapsed over Cog. training groups, to visualize effects of Easy vs. Hard Visual training group performance. All groups reduced their flanker RT over time; participants in the Hard Visual training condition showed a selectively greater RT improvement over training for incongruent flanker trials. Vis.: Visual training type; Cog.: Cognitive Control training type. Error bars represent standard errors.



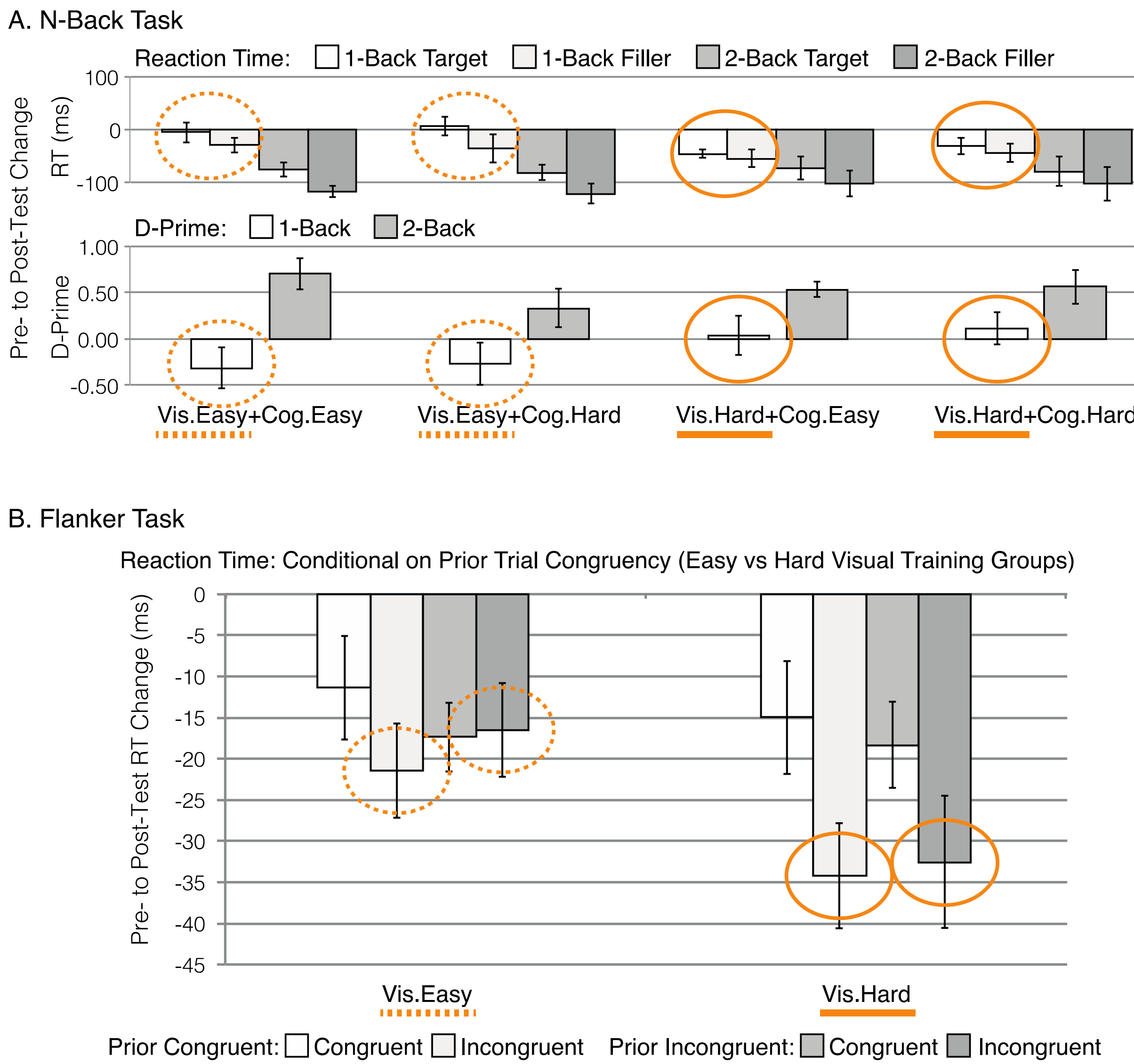
Chapter 3, Figure 8. Mean go/no-go data for Paint the West. Participants in the Hard Visual training condition showed greater improvement in hit RT, but a greater decline in d-prime over training. Vis.: Visual training type; Cog.: Cognitive Control training type. Error bars represent standard errors.



Chapter 3, Figure 9. Mean mental rotation task data for Paint the West. A marginal effect of improved accuracy and d-prime for Easy Visual training, but not for Hard Visual training is observed, with no RT training effects. Vis.: Visual training type; Cog.: Cognitive Control training type. Error bars represent standard errors.



Chapter 3, Figure 10. Mean n-back task data for Paint the West. In the 1-back task, Hard Visual training showed greater RT improvement with no decrease in d-prime, most notably for Target trials, compared to Easy Visual training. No training effects were seen in the 2-back task. Vis.: Visual training type; Cog.: Cognitive Control training type. Error bars represent standard errors.



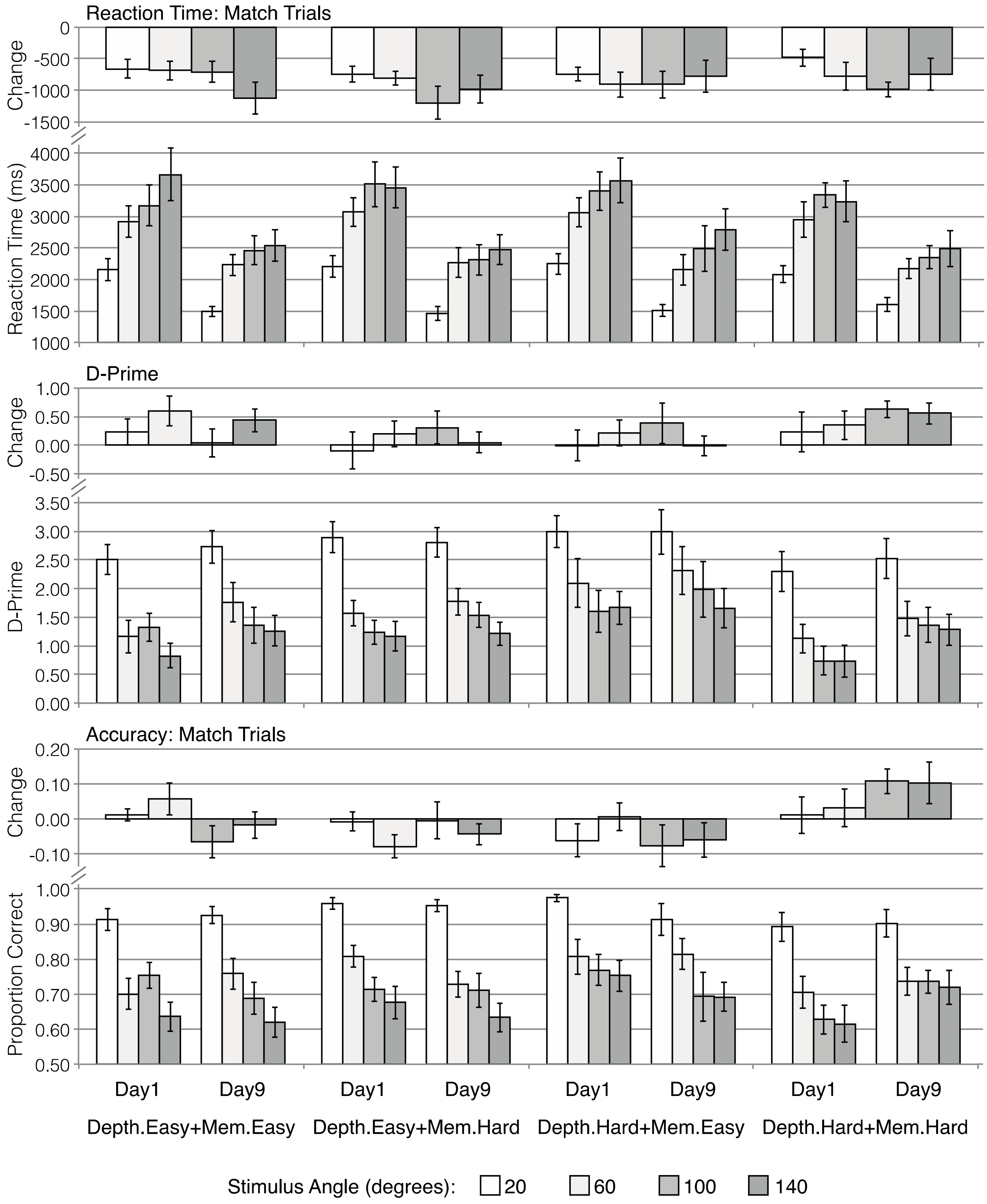
Chapter 3, Figure 11. Summary evidence of transfer effects from n-back (Panel A) and flanker task data (Panel B). Selectively better performance was seen for Hard Visual training groups (solid orange lines) compared to Easy Visual training groups (dotted orange lines). These training effects were seen as relatively faster RT and better d-prime in 1-back (contrasted with equivalent improvements in 2-back), and relatively faster incongruent flanker RT (contrasted with equivalent improvements in congruent flanker RT). Vis.: Visual training type; Cog.: Cognitive Control training type. Error bars represent standard errors.

A.

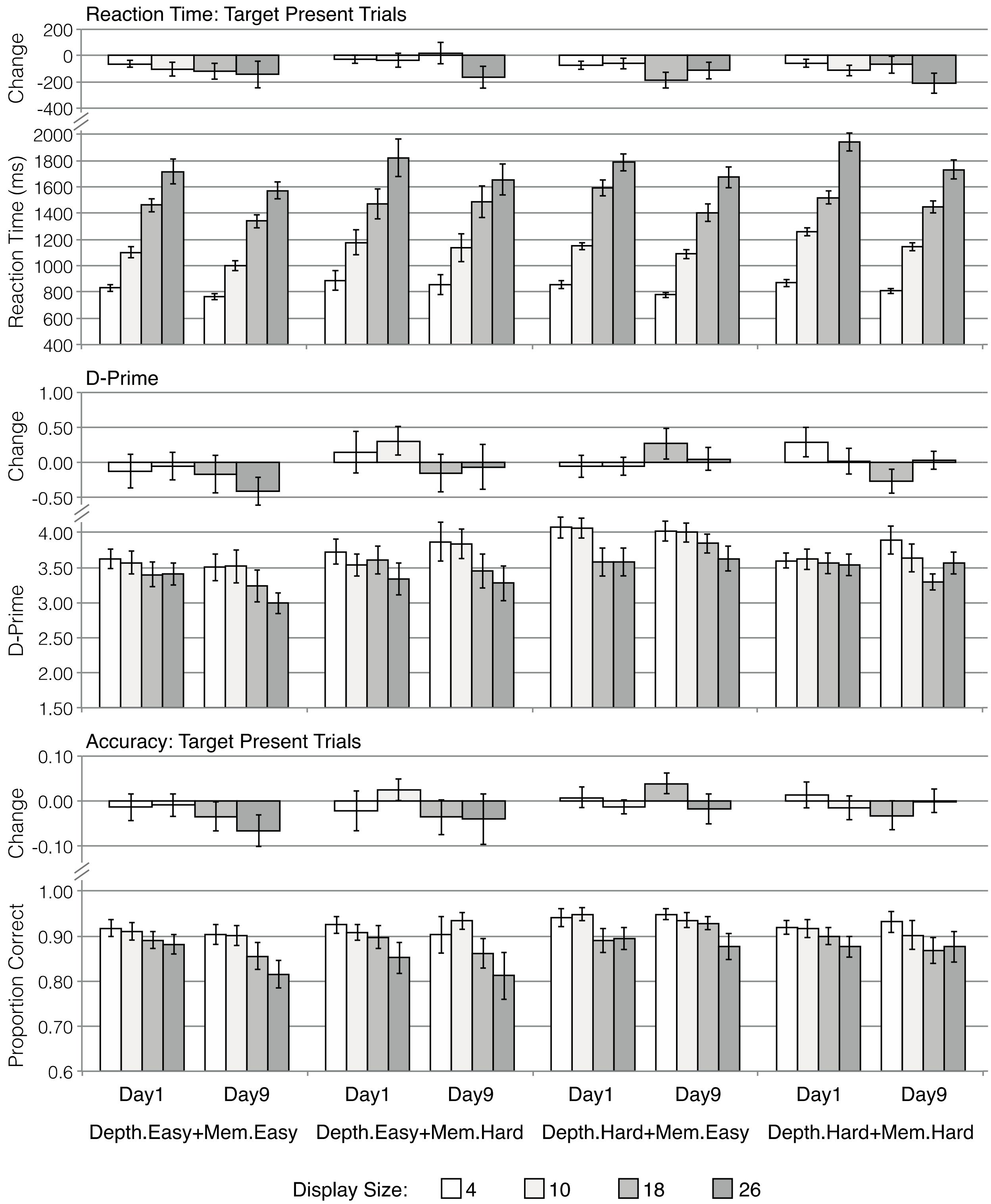
B.

C.

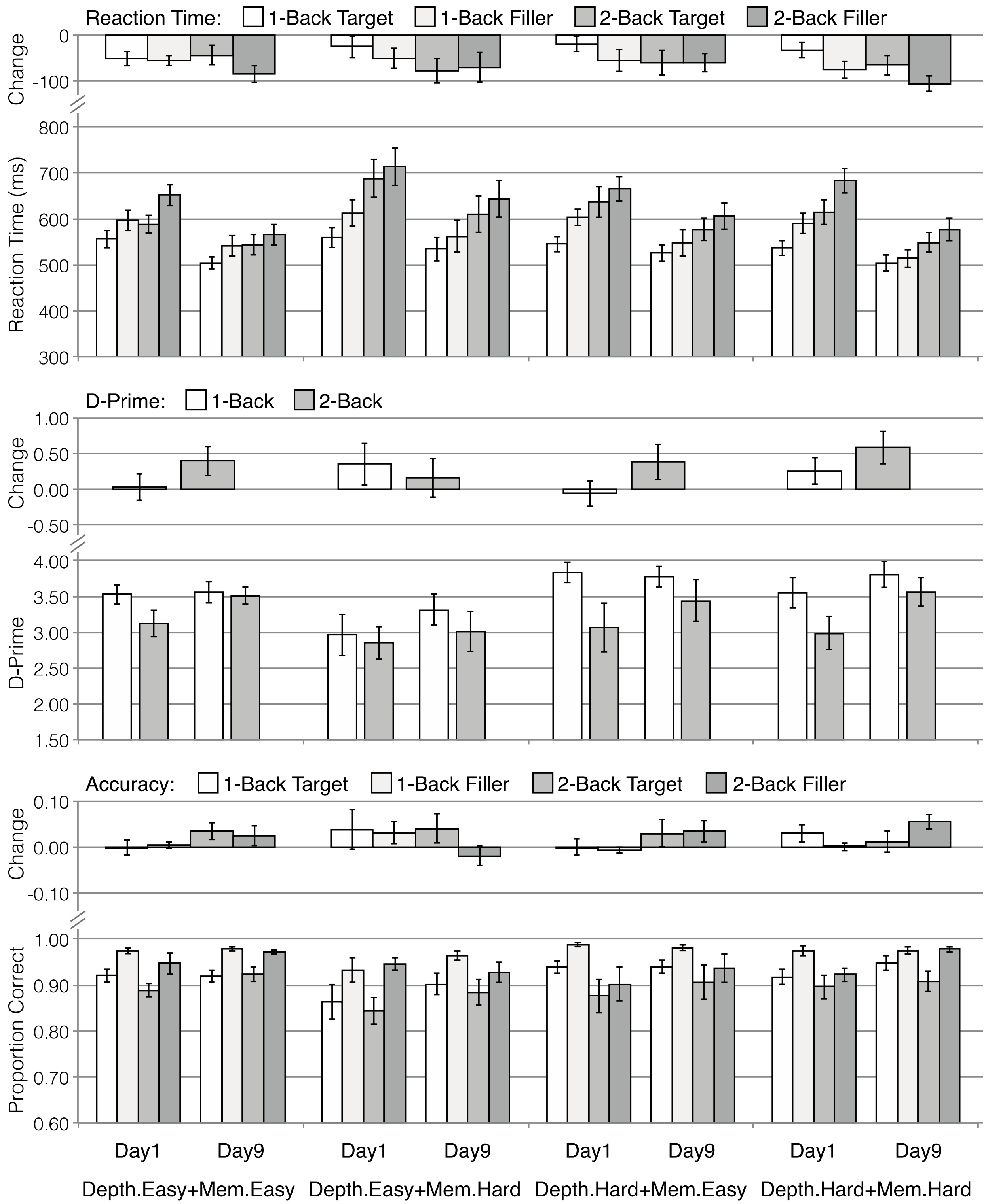
<u>Chapter 4, Figure 1.</u> Example gameplay for the "Orphlings" video game. Players have to guide the orphling creatures through a game level by placing tools that change the properties of creatures (e.g., making them sticky so they can climb a wall) or that directly move (pull, throw, etc.) creatures, while keeping creatures safe from hazards in the game. Panel A illustrates a level solution with tools added to the level; Panels B and C show close-up detail. Game training manipulations of depth and spatial memory are described in the text.



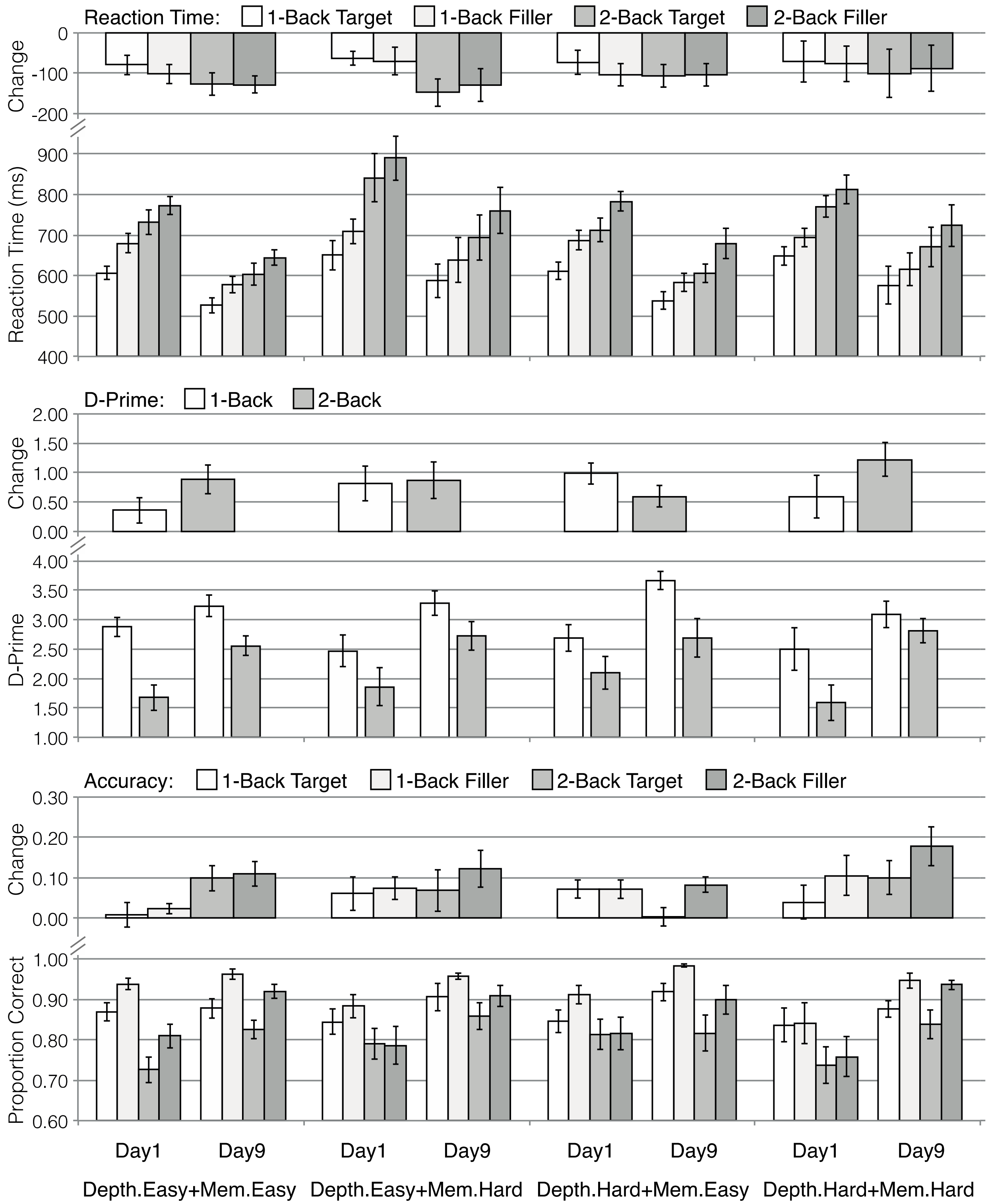
Chapter 4, Figure 2. Mean mental rotation data for Orphlings. Depth.: Depth training type; Mem.: Memory training type. Error bars represent standard errors.



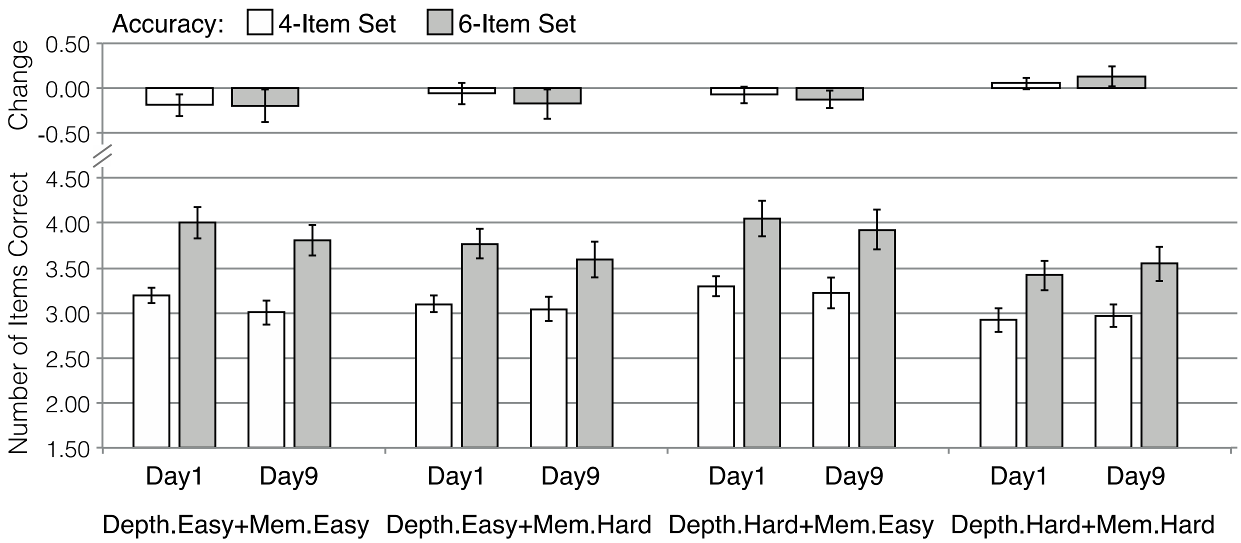
Chapter 4, Figure 3. Mean visual search data for Orphlings. Depth.: Depth training type; Mem.: Memory training type. Error bars represent standard errors.



Chapter 4, Figure 4. Mean digit n-back data for Orphlings. Depth.: Depth training type; Mem.: Memory training type. Error bars represent standard errors.



Chapter 4, Figure 5. Mean spatial n-back data for Orphlings. Depth.: Depth training type; Mem.: Memory training type. Error bars represent standard errors.



Chapter 4, Figure 6. Mean spatial working memory data for Orphlings. Depth.: Depth training type; Mem.: Memory training type. Error bars represent standard errors.