11-4-2019

Changes in the Cognitive Dynamics of Problem-Solving

Shireena McGee
shireena.mcgee@uconn.edu

Changes in the Cognitive Dynamics of Problem-Solving

Shireena McGee

B.A., Rutgers University, New Brunswick, NJ, 2016

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

At the

University of Connecticut

2019

APPROVAL PAGE

Master of Science Thesis

Changes in The Cognitive Dynamics of Problem-Solving

Presented by

Shireena McGee, B.A.

Major Advisor____

Adam Sheya

Associate Advisor____

Sumarga Suanda

Associate Advisor____

James Dixon

University of Connecticut

2019

Abstract

How do humans develop the ability to approach and solve diverse, attentionally-demanding problems? Current work on problem-solving focuses on well-defined problems in which explicit task constraints directly inform the planning and execution of behaviors. However, in situations that might be more informative for naturalistic learning and problem-solving, what is necessary to solve problems when task constraints, and thus planning abilities, are not as easily accessible? Executive functions such as selective attention and working memory seem fundamental to these types of activities, yet there is not a clear account of how exactly the coordination of these cognitive processes contribute dynamically to this capability over the course of development. This study aimed to enhance the current characterization of problem-solving by further uncovering relationships among processes of executive functioning to better understand how this competence might emerge and self-organize throughout development. To first examine this phenomenon in adults, a sample of 86 people between the ages of 18-27 years participated in this phase of the study. The experiment consisted of a code-breaking problem-solving task and two working memory tasks; the Simon color sequence game, and the Dual N-Back. In the problem-solving task, participants attempted to solve a correct code using incremental feedback. Overall, results suggested that the ability to adequately coordinate information, and thus overall performance, decreases when working memory demands are imposed.

Changes in the Cognitive Dynamics of Problem-Solving

Problem-solving is ubiquitous in the everyday activities of a child. From learning how to tie shoes to learning math, problem-solving abilities underly the development of self-directed learning. At its core, problem-solving facilitates learning through the discovery of new information in our environment (see Newell & Simon, 1972). As adults, humans are quite efficient at solving a wide variety of problems that demand different amounts of exploratory behavior, or require different solution scopes; however, much of the developmental research on problem-solving has focused on a narrow range of problem types (see Keen, 2011 for review). Some researchers have emphasized the critical role social interactions play in the development of a more general problem-solving ability (Ellis & Siegler, 1994). Others have explored how strategy and executive function contribute to problem-solving (Siegler, 1999; Zelazo, Müller, Frye & Marcovitch, 2003). However, this research has not considered the potentially reciprocal relationships that exist between the development of executive function and aspects of cognition that underly the development of problem-solving. In order to understand how a general problem-solving ability develops in adults, research that focuses on how executive function processes are organized to a problem-solving task over development is needed.

Newell, Shaw and Simon (1958) proposed that theories of problem-solving should: predict performance in specified tasks, explain problem-solving via process and mechanism, and account for changes in both internal and external conditions that alter problem-solving activity. In order to predict performance and to propose processes and mechanisms that generate the performance, there needs to be a developmentally appropriate cognitive description of the solver. To date, much of the conceptualizations of problem-solving, even the developmental considerations (with the notable exception of the seminal works of Piaget, Vygotsky, and the dynamical systems approach of Thelen & Smith), have been formulated solely from the vantage of the adult system. This involves applying understandings of the adult cognitive system to children's behavior (see Luo & Baillargeon, 2005). This developmental work has generally disregarded developmental differences in sensitivities to the internal and external conditions that can alter the problems a child can solve. Oftentimes, the logic of these approaches assumes that using the same task at multiple points in development is sufficient for measuring differences in problem-solving. However, it is not clear, given the developmental differences in cognition that will emerge over these periods, that these types of approaches provide equivalent problems from both the child and adult perspective. Therefore, what is needed is a characterization of problem-solving that accounts for developmental differences in the structure of problems that can be solved. This consideration is related to the tension in the adult problem-solving literature between whether abilities are general or context/knowledge dependent (see Keating & Crane, 1990 for review). We propose the key process to measure for predicting and explaining performance in a problem-solving context is information gathering activity. This perspective suggests the activity of information gathering undergoes significant change over development, and is what underlies general problem-solving ability in both children and adults.

This research attempts to measure changes in cognition that underlie the development of problem-solving. In the following sections, perspectives on cognition and problem-solving are discussed first from an adult, and then subsequently from a developmental perspective. This is then followed by a discussion of the development of executive functions and its role in problem-solving. This discussion will make the case that in order to understand our general problem-solving ability, an understanding of the changes in the coordination of executive processes to problem-solving tasks across development is needed. As a first step in filling this gap in understanding, we consider a study where adults are given a challenging problem-solving task that requires participants to discover how to proceed to a solution.

Traditional Approaches to Problem-Solving

Traditional approaches to problem-solving often describe the phenomena in terms of a problem's initial state, solution or goal state, and the series of potential actions that are needed to reach that goal state (Newell & Simon, 1972). Sternberg (1986) added a few dimensions to this description, portraying problem-solving as a cycle in which there are seven stages: first recognizing or identifying the problem, then defining a mental representation of the problem, developing strategies, organizing information or knowledge of the problem, allocating mental and physical resources, monitoring progress, and lastly, evaluating error and accuracy towards the goal. According to Sternberg, the cycle is best navigated when one is able to flexibly, rather than sequentially, move between the known steps of the problem. Although this portrayal is a potential framework for characterizing how the same task might be a different problem at different points in development, it tends to be more useful for characterizing certain types of problems over others; namely, well-defined problems.

Much of the traditional work on problem-solving considers well-defined problems; problems in which there is a clearly defined representation of the problem, and consequently, the necessary strategies to solve the problem are available at the outset. Using only *well-defined* problems to attempt to understand *general* problem-solving abilities has led researchers to ignore critical information gathering aspects of problem-solving. III-defined problems are classified as those in which the problem space is insufficiently constrained by the available information. At the outset, the best strategy or step towards the goal state is not clear. The logic for understanding problem-solving through well-defined problems is similar to the logic of understanding development through problems that can be presented to multiple age groups; well-defined problems provide an opportunity to measure aspects of problem-solving that are believed to be independent of context or individual understandings. Although more difficult to study, ill-defined problems tend to be more representative of the problems faced in the everyday, such as discovering the social activities necessary to make new friends, learning to organize and manage schedules, or coming up with and executing a school/work project. A necessary first step towards understanding how ill-defined problems are solved is to understand what cognitive capacities underly the adult ability to gather information that can direct problem-solving.

Pretz, Naples and Sternberg (2003) suggest that in Sternberg's cycle of problem-solving, problem recognition, definition, and representation are essential processes in solving ill-defined problems. Within this framework, lower-level mechanisms of attention that direct visual gaze, as well as higher- level cognitive abilities involved in the encoding and retrieving of information from existing memory, aid in the processes of defining and representing a problem (Sternberg, 1986). Navigating through the potential solution paths to a problem depends on the information initially available, as well as the solver's ability to discover and encode new and relevant information. It is this process of information gathering (discovering, encoding and using) that is essential to the adult capacity to solve problems. Although this process has not been directly explored in the adult or in the developmental literature, insight into the development of this ability can be found in research on the development of Executive Function.

Development of Executive Function

Executive Function (EF) is a comprehensive term for the aspects of cognition involved in the coordination and organization of information in higher-order cognitive tasks that require effortful attention, such as problem-solving (Diamond & Lee, 2011; Lee, Ng & Ng, 2009; Zelazo, Carter, Reznick & Frye, 1997). Diamond (2013) describes the core components of executive function as selective attention, inhibitory control, working memory, and cognitive flexibility. Selective attention and inhibitory control involve the ability to direct focus and restrain behavior. Working memory (WM) is the ability to hold and process information in a mental workspace and cognitive flexibility is the ability to shift between concepts. From Diamond's (2013) perspective, reasoning and problem-solving emerge out of the fundamental functions of EF: selecting, inhibiting, and switching. Other conceptions of EF have focused more on the complexity of control required to complete tasks (Zelazo et al., 2003). From this perspective, reasoning and problem-solving emerge out of other fundamental functions of EF: representing, planning and evaluating. Under this framework, there are thought to be differences in the structure of rules that children can formulate and follow that are based on developmental changes in the abilities to represent, plan and evaluate.

Research on the development of EF has found substantial change over the preschool years (3-5 years of age) (Zelazo, Carter & Reznick, 1997). However, this development has been mostly measured by a single EF task (Dimensional Change Card Sort, or DCCS) and there is much debate as to what is driving success and failure in this task. For example, some suggest selective attention is driving this change (Brooks, Hanauer, Padowska & Rosman, 2003), whereas others have suggested rule formulating and following (Zelazo et al., 2003). It is generally acknowledged that EF cannot be described as a single function and that tasks like DCCS are just convenient measures that provide a crucial first step towards explanation (Zelazo et al., 2003). In order to use DCCS or any task as a route to understanding EF more generally, a general set of functions (represent, plan, evaluate vs. select, inhibit, switch) must be assumed; and which set is assumed determines how the measures are able to be interpreted. One point of agreement between different functional characterizations of EF is a conceptualization of its composition. It is therefore possible that an explanation of the development of EF might be found within the development of the components of EF.

Working memory, a key component of EF, undergoes significant development between the ages of 4 and 15 (Gathercole, Pickering, Ambridge & Wearing, 2004). However, there is much the same debate about the nature of WM development (and WM itself), as there is among those that have shaped the research on the development of EF. Although there is agreement about the general function of WM as a mental workspace, there are different ways of characterizing the components of WM. There are also different ways of characterizing the relationships that exist between processing, strategic, and attentional aspects in conjunction with storage capacity. Cowan (2010) has argued that there is a real storage capacity limit that determines how much can be held in the mental workspace; and this capacity limit is considered to be independent of any strategic or attentional process. Baddeley's expansive view of adult WM incorporates aspects of EF into WM, as opposed to viewing WM as a component of EF (Baddeley & Hitch, 1974). The fundamental premise here is that WM is limited by a tradeoff between resource limits in how much can be processed and how much can be stored. Increasing storage would, under this notion, place increased demands on information processing resources; and increasing processing resources would, in return, demand greater storage. In this way, WM is a mental workspace in which immediate information is processed with past knowledge under the guidance of a central executive to solve a problem.

Towse, Hitch, and Hutton (1998) argue that WM isn't structured by just this tradeoff between resources and storage capacity, but rather is shaped by retention. Different problems require holding different amounts of information in the mental workspace over time. Both the capacity to retain, as well as knowing what to forget, might be key determiners of WM performance. Chi (1978) investigated differences in the ability to remember chess positions between children who were experts and adult novices. The main finding was that the experts (children) were able to hold more positions in memory than the adult novices. What this indicates is that the developmental trajectory of WM (or EF more generally) might not be so direct and simple, but rather dependent on knowledge and familiarity with certain domains. Baurer (2014), Ornstein & Naus (1985) and others suggest that one's knowledge base, or a person's bank of experiences and learned information, is the most important organizer and predictor of working memory performance (DeMarie & Lopez (2014). From this perspective, the knowledge base coordinates the holding of information in WM and directs attention in a given context, in contrast to conceptions that focus on working memory's role in EF, which posit a general ability to selectively attend as what is most predictive of WM performance (Gazzaley & Nobre, 2012).

This conceptual tension might ultimately be reflecting a bi-directional influence over development between knowledge and Executive Functions such as selection or rule formulation. Indeed, increased sophistication in the ability to allocate attention is a precursor to development of strategies that are able to manipulate capacity (Cowan, 1988; Kane & Engle, 2003). What changes these limits are the ways in which we are able to organize and act on information in a given context. This is similar to Karmiloff-Smith's (1979) proposal that once a child successfully employs a representation of their knowledge in a cognitive task, this representation itself becomes a unit on which the cognitive system can act. This metaprocedural behavior doesn't require explicit awareness, but rather reflects a balance between information content and information processing (similar to the resource/storage tradeoff that motivated Baddeley's conceptualization of WM). This characterization of a general tendency in cognitive development towards the metaprocedural should thus be reflected in Baddeley's component models of WM as well.

Tasks that comprehensively assess Baddeley's components of working memory by individually testing each component via multiple sub-tests, such as the working memory version of the comprehensive assessment battery for children (Cabbage, Brinkley, Gray, Alt, Cowan, Green & Hogan, 2017) or the working memory test battery (Pickering & Gathercole, 2001), find that the components of working memory most predictive of performance in tasks like mathematical reasoning change over time. Meyer et al. (2010) found that the central executive and phonological loop were more predictive of 2nd grader's performance, whereas the visuospatial sketchpad was more predictive of 3rd grader's performance in mathematical reasoning tasks. This shift could reflect the metaprocedural developmental tendencies proposed by Karmiloff-Smith (1979). For instance, elementary school children are taught various visuospatial representations of mathematical problems, therefore this change in components might reflect the emergence of a visuospatial metaprocedure from, what is initially, a more linguistic representation. This is consistent with Gathercole, Alloway, Willis and Adams (2006) findings that show changes in WM components related to reading and mathematical comprehension during a similar time period. Further evidence for the developmental tendency towards the metaprocedural may be found in the development of what has been termed metacognition.

Metacognition refers to the ability to reflect, understand, and control learning (Schraw & Dennison, 1994). Karmiloff-Smith (1979) proposed that it is metaprocedural behavior that controls the balance between information content and information processing; in other words, the coordination of the known, not known, and needs to be known in a problem-space. Davidson & Sternberg (1998) and Gardner & Hatch (1989) suggested this ability involves forming a representation better suited to selecting plans and to identifying impediments to goals. This is also consistent with conceptualizations of metacognition from the perspective of EF (Fernandez-Duque, Baird & Posner, 2000; Roebers & Feurer, 2015; Bryce, Whitebread & Szucs, 2015). Early metacognitive abilities develop between 3 and 5 years of age (Kuhn, 2000). This is the same period in which significant advances in the acquisition of language (Dockrell & Campbell, 1986), in the representation of categorical knowledge (Sheya & Smith, 2006) and in rule formulating and following (Jacques, Zelazo, Kirkham, Semcesen, 1999) seem to emerge. It is possible that Executive Function, WM and metacognition might just be different ways of conceptualizing the process of controlling how information is represented, acquired and utilized in a problem context. What is needed is a systematic, direct study of this process. In order to directly characterize this process, a general problem-solving context in which performance does not appear dependent on specific or domain knowledge is needed.

The Current Study

In this experiment, participants were given one of three versions of a code-breaking problem-solving task in which the objective was to guess a four-item color code when given six color options, and incremental feedback after each attempt. Versions of this task have been used to look at problem-solving previously (e.g., Best, 2000; Laughlin & Barth, 1981), yet none have used eye-tracking methodologies, nor have used this task to characterize the cognitive processes associated with the task. The three task conditions differed in levels of imposed cognitive load by altering the number of previous attempts and their feedback that were visible, thus shifting what would feasibly need to be processed by the cognitive system's working memory in order to be successful in the task. The imposition of this cognitive load would theoretically constrain working memory's ability to both store these attempts as well as process ongoing strategy and incoming information simultaneously.

In the *standard condition* of this task, participants were able to see all available information, attempts, and feedback on those attempts, when working towards a solution. This condition provided information regarding the ways in which an adult would act on a problem and use information when working memory was presumably not additionally taxed. In the *two-trial* condition, participants were only able to see their previous two attempts starting on game three of five. In the *zero-trial* condition, participants were not able to view any past attempts when making a current attempt. These two conditions were meant to tax working memory to a lower and higher degree, respectively.

Participants then completed two separate working memory tasks. These tasks observed separate contexts of working memory processing. The Dual N-Back task was used to measure the phonological loop, visuospatial sketchpad, and the central executive simultaneously in order to view how all aspects of WM together contribute to problem-solving. Then, a color-sequence recall task (Simon game) was used to observe how short-term storage more specifically plays a role in problem-solving. It was hypothesized that information gathering strategies would differ as a result of changing cognitive demands on working memory, and that WM systems that engage both storage and processing would be more necessary when increasing the amount of to-be-remembered information. We also aimed through exploration to gain a better sense of the dynamic relationship between executive cognitive function and activity in problem-solving in order to get an overall sense of how these cognitive capacities coordinate in their activity.

Method

Participants

The sample consisted of 96 (60 female, M_{age} = 19.38, *SD* = 1.79, range = 18-27 years) participants recruited from the University of Connecticut's participant pool. A total of 10 participants were excluded due to language barrier in understanding directions, technical difficulties or experimenter error, leaving a total of 86 participants (54 female, M_{age} = 19.28, *SD* = 1.60 range = 18-27 years). Participants were randomly assigned to a problem-solving condition prior to their arrival (standard condition: *n* = 29; two-trial condition: *n* = 29; zero-trial condition: *n* = 28). Participants received in-course credit for participating.

Apparatus

Participants were dressed in head-mounted eye-tracking glasses upon arrival. The data were recorded via scene and pupil cameras; however the analysis for this will not be discussed in the current paper. All tasks were played on a wall-mounted television monitor (86 in. diagonally) where participants sat at a table that was placed directly in front of the screen (approximately 30in.). All tasks were run on an Alienware computer that was connected through HDMI to the television monitor. The code-breaking task and Simon game were both created and run through a Unity platform. The Dual N-Back was gathered from an open-source version created by BrainWorkshop through Jaeggi and Colleagues (2008). Participants responded to all tasks using a Logitech gaming mouse.

Procedure

Participants were first dressed in head mounted eye-trackers. They sat at a table in front of a large wall-mounted monitor and were given a wireless computer mouse for game-response. Participants then fixated their gaze on a handful of small icons on a screen to calibrate the eyetracker. Afterward, participants were provided directions for their first task, the code-breaking game. The Dual N-Back or Simon games were presented in counterbalanced order following the code-breaking task.

Code-Breaking Task. The code-breaking game was always presented first. All participants were read scripted directions, detailing all necessary information. In the two-trial condition, participants were told "*you might start to notice that some of your guesses will start to disappear*. *In this case, just do your best to guess the correct code*". In the zero-trial condition, participants were told "*you will only be able to view one guess at a time, so do your best to pay attention*". Next, a presentation of example feedback with symbols (instead of colors) was provided and participants were asked what certain types of feedback would tell them about their response (e.g., "*If you chose these four symbols and received four white feedback pins, what would that tell you about the symbols you chose?").* This was done to ensure participants comprehended the directions. If participants did not comprehend, experimenters repeated verbiage from the scripted directions to alleviate confusion. Explanation continued until participants correctly answered what each type of feedback in the given examples implied. Participants were told that experimenters were only interested in viewing how they played, not how well they played and were also told to do their best but not worry about performance. Participants were then told they could begin five games, where they were to attempt to solve five different codes.

Dual N-Back Task. In the Dual N-Back, participants were read scripted instructions and then presented with demonstrations of potential Dual N-Back scenarios. Experimenters asked participants which button they would press in both a 1- and 2-back condition, given three trials of stimuli (consisting of a symbol on a grid, accompanied with a stated letter underneath). After addressing questions from the interactive example, the experimenter told participants to *"do your* *best, but do not worry about performance"*. Participants then played five full rounds of the Dual-N Back.

Simon Game. In the Simon game, participants were read scripted directions but were not given any visual examples. They performed one game with experimenter feedback, if needed, to reduce any technical challenges regarding game accessibility. Participants then played five games of Simon, with each round ending due to an error in repeating the pattern.

After finishing all three tasks, participants removed their glasses and were asked to report their age, school major or focus, if they had ever played any of these games before, and/or if they have ever played any similar games.

Experimental Tasks

Code-Breaking Task. An adapted computerized version of a popular children's game was used to assess problem-solving capacities (shown in Figure 1). The objective of this task was to solve a four-item color code given six color options (red, orange, green, aqua, dark blue, black). Participants had 10 trials to solve a correct code and were given incremental feedback as to the accuracy of each attempt. After choosing four colors for a code (the repeat of colors was acceptable), they pressed a "check-move" button to receive feedback on that attempt. Yellow feedback pins appeared if a participant had placed any correct colored pegs in a correct position, and white feedback pins appeared if any colors selected were correct but were in the wrong position. For example, in Figure 1a the first attempt made was: red, green, orange, aqua. After clicking check move, three white feedback pins appeared which indicated three of the colors chosen were correct (1 incorrect), but none were in the correct location (lack of yellow pins). The location of given feedback pins was not intended to correspond to the location of the colors in the correct code; therefore, feedback referred to the attempt overall.

The participant continued to make attempts until either ten attempts were made, or the code was correctly solved. Each participant had five codes to solve, in which 10 trials were given to solve each code (50 trials possible in total). In the standard condition (same as Figure 1a), all five rounds were played as normal. In the two-trial condition (Figure 1b) in rounds three through five, participants were only able to view their two previous attempts when choosing colors for a new attempt. In other words, previous attempts and their feedback disappeared as one moved through the task, as to only leave two previous attempts available to a participant. In the zero-trial condition (Figure 1c) participant's attempts disappeared when a color was selected for a new attempt in all rounds. Due to this, participants could still view feedback for their current attempt as long as they desired; however, they were never able to view any attempts or feedback when actively making an attempt. Participants took 20 minutes on average to complete this task. Since this was the only self-paced task included in the protocol, if a participant reached 30 minutes in-task and were not beyond round 3, participants were reminded *"we are not concerned with performance, just how you play, so do your best, but don't worry about performance"*.

Figure 1a.

Figure 1b.

Figure 1c.

Figure 1. Above are three images from the code-breaking task. Participants must use incremental feedback to solve a four-item color code. Figure 1a shows the standard condition, where all attempts and feedback are visible throughout the task. Figure 1b shows the two-trial condition; on rounds 3 through 5, two attempts were visible at a time. Figure 1c shows the zero-trial condition after a participant had started a new attempt, where no attempts are visible. In all conditions, there was no time limit for viewing available feedback.

Dual-N-Back. The Dual N-Back (Figure 2) is a computerized working memory task that involves simultaneously seeing and hearing a sequence of visual and auditory stimuli while deciding whether the stimulus you are looking at or seeing is a match to one that you were presented with (n) times ago (range: 1-3). The methods and program used were developed by Jaeggi et al. (2008). Participants were given 20 stimulus presentations, with 3 seconds to view each stimulus before the next one was shown. Stimulus presentations consisted of a blue square flashing on a grid, and this was accompanied by a letter sound. There were nine possible positions for the blue square and eight possible letter sounds.

In the Dual 1-Back, participants had to decide if the current blue square was in the same position as it was 1 stimulus prior, while simultaneously deciding if the current letter sound was a match to the sound heard one stimulus ago. If what the participant was seeing was a position match, they would press the left button, and if what they were hearing was an auditory match, they would press the right button. After finishing a set of 20 presentations for a given N-Back (always beginning with 1-Back in game 1), a score of 68% correct responses or higher would move them onto the next numerical “N”-Back round (e.g., a 2-Back). A score of lower than 68% moved a participant to the next game (e.g., if a participant received a score of 50% on game 1, 1-Back, they would then move to the next game, starting again at 1-Back (game 2, 1-Back)). The task is played 5 full times, with each round consisting of a 1-Back, and then a potential 2- or 3-

Back depending on the individual participants' scores (3-Back being the highest N that was able to be reached). Participants played the Dual N-Back for 20 minutes on average.

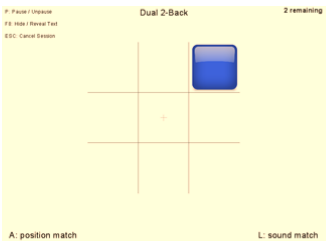

Figure 2a. Stimulus 1

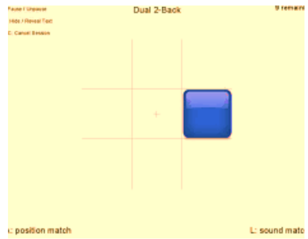

Figure 2b. Stimulus 2

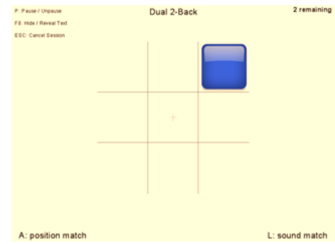

Figure 2c. Stimulus

Figure 2. In the Dual N-Back, participants are asked to recall a sequence of visual and auditory stimuli. Above is an example of three visual stimuli, presented one at a time, in the Dual N-Back (9 possible square positions). Not pictured are letter sounds (8 possible) that were presented simultaneously with the visual stimuli as they appeared. The objective of the Dual N-back was for participants to decide if the blue square position, letter sound, or both being presented was a match the one presented N stimuli prior (with N starting at one and increasing incrementally by one each round).

Simon Game. In this computerized version of a popular children's game (Figure 3), the objective is to recall a sequence of colors in the correct order. On the screen, participants saw a circular stimulus with four color sections (red, yellow, green, and blue). In this task, tone was eliminated to focus on just one aspect of storage (visuospatial) due to its necessity in the code-breaking game. In each round, participants saw an incrementally larger sequence of colors flash on the screen. The presentation started with just one-color flashing for 1 second in the first round. Once the color flashed on the screen, participants needed to repeat the color they saw by clicking on its position. If they correctly repeated that color, a "go" button appeared. To move on to the next step in the sequence, the go button was pressed, and the sequence began again, growing to two color items in the second round; the previous color would remain the same, and a new color would be added on to the end of the pattern. Colors could be repeated in the sequence up to two times in a row. The sequence kept growing until the participant either made a mistake in repeating the pattern or ran out of time in the task. The timer was set for 5 seconds per number of colors in the sequence and would begin counting down at the start of the sequence presentation. When participants made a mistake, a "game over" screen would appear. Participants also played this task five times.

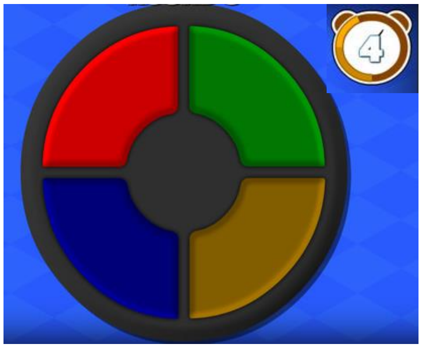

Figure 3. The Simon game

Figure 3. In the Simon game, the participants' objective was to repeat the pattern of colors presentenced to them on the screen. When the pattern was presented, buttons would look like they were being pressed down and darkened. In the first round, one color would be presented, then the participants' computer mouse would re-appear on the screen, indicating for them to click on the color they had seen. If repeated correctly, the following rounds would repeat the same color pattern, with a new color being added to the end of the sequence with each round. The game continued to increase in rounds until the participant made a mistake in repeating the pattern.

Analyses and Coding

Code-Breaking Task. To measure problem-solving success overall, an average number of wins (out of five games) was calculated for each participant. Then, the average number of attempts (or trials) needed to reach a solution was calculated for each participant across rounds, and then across condition.

Efficiency of Information Gathering. An efficiency of information gathering score was coded to determine the level to which an attempt made provided new or useful information for the solution. Operationally, this was determined by identifying whether the attempt provided redundant information (i.e., did the attempt provide information that had already been discovered?)

and/or if the response did not adequately consider previous information given (e.g., information given implies no pins were in the correct position, but positions from that attempt were subsequently repeated). Efficiency scores were coded for each attempt made (excluding the first attempt, as there was no previous information to consider) with each receiving a score of a 0, 1, or 2 for the overall response. A response was scored a 0 if all four selected items provided redundant information. A score of 1 described an attempt that either (a) included some but not all redundant information, or (b) ignored some but not all previous information. A score of 2 was assigned to attempts that included all new information and/or adequately considered previous information. An average efficiency score was calculated across the five games. In order to also get a better description of the breakdown of average activity within a game, a proportion of each score type (0, 1, 2) was also calculated. Two coders blind re-coded a random subset of ten videos (not formerly coded by themselves) in order to calculate reliability. A score of 93.76% overall match existed between coders prior to agreement being found.

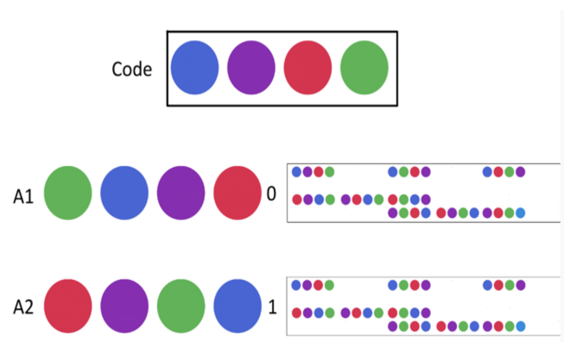

Figure 4. Above is an example visual of a simplified version of this task, where there are only four color options, and no possibility of repeating colors in the code. Pictured is a target correct code with two attempts, and a number next to each attempt indicating how many positions were correct. Directly next to the feedback is a visual representation of the possible solution space, given the feedback received. In the first attempt, we see no color positions are correct. In the second attempt, the participant does not repeat any color positions, gaining new information and further narrowing the scope of possible solutions. Attempt two would therefore receive a 2 for an efficiency of information gathering score; all previous information was considered, and new information was gained where appropriate.

Dual N-Back. In traditional working memory tasks, an average across all Dual N-Back trials (N-1, N-2, and N-3) is calculated as a description of performance. However, each N-Back trial has inherently different working memory demands; thus, it was essential to determine an average of each individual N-Back trial. Due to the fact that the code-breaking task was played first, initial analyses examined whether there was an interaction between code-breaking task condition and scores in each 1, 2, or 3-Back. Afterward, associations between working memory scores in the Dual N-Back and problem-solving performance, attempts, and efficiency were investigated.

Simon Game. This task collected a relative measure of short-term item-span capacity. The sum of the last sequence of items the participant was able to repeat was taken and averaged across five tasks for all participants. Initial analyses examined if an interaction between code-breaking task condition and scores in the Simon game existed. Afterward, associations between working capacity scores in the Simon game and problem-solving performance, attempts, and efficiency were investigated.

Results

Individual ANOVAs were performed to compare dependent measures of attempts made, number of wins, and efficiency scores, with problem-solving (PS) condition types as the between subjects factor. It was hypothesized that those in the highest demand condition (zero-trial) would have lower efficiency and number of wins, as well as a higher number of attempts as compared to those in the standard and two-trial conditions.

Next, an ANOVA investigating differences between scores in the Dual N-Back by PS group condition was performed on overall Dual N-Back scores, as well as N-1, N-2, and N-3 individually. No differences between groups on overall performance or between tasks in the Dual

N-Back were hypothesized. Additionally, correlations were performed to investigate the relationship between working memory in the Dual N-Back task and PS attempts, number of wins, and efficiency.

Next, an ANOVA comparing scores in the Simon game with group condition as the between-subjects factor was performed to investigate if any group differences that could be due to either individual differences or recent exposure to the taxing task existed. No group difference in this task was hypothesized. Then, correlations were performed to investigate the relationship between short-term storage in the Simon game and PS attempts, number of wins and efficiency. Lastly, a correlation was performed to see the extent to which Simon game and Dual N-Back tasks were associated with one another. With each analysis performed, Hedges g was used as a measure of effect size. This was chosen due to the fact that this is a novel task that does not have any standardized population means.

Code-Breaking Task

An ANOVA revealed there was a significant difference between problem-solving groups in number of wins (out of five) [$F(2,83) = 5.94$, $p = .004$]. Post-hoc analyses revealed the condition with the lowest imposed cognitive load demand (standard) condition won more games ($M = 3.21$, $SD = 1.52$) than the highest cognitive load demand (zero-trial) condition ($M = 1.89$, $SD = 1.10$) [$g = 1.31$, $p = .001$]. The intermediate cognitive load demand (two-trial) condition ($M = 2.62$, $SD = 1.64$) did not differ significantly from the standard [$g = .59$, $p = .34$] or zero-trial [$g = .73$, $p = .13$] condition.

An ANOVA regarding differences in the number of attempts made towards the solution between the three cognitive load conditions revealed there was a significant difference between group conditions [$F(2, 83) = 5.24$, $p = .007$]. Post-hoc analyses revealed that the standard condition

(*M* = 7.90, *SD* = 1.28) required significantly fewer attempts to achieve a correct solution than the zero-trial condition (*M* = 8.84, *SD* = .80) [*g* = .94, *p* = .005]. The standard condition did not differ from the two-trial condition (*M* = 8.20, *SD* = 1.20) [*g* = .31, *p* = .61] and the two-trial condition also did not differ significantly from the zero-trial condition [*g* = .63, *p* = .06].

*Efficiency of Information Gathering***.** An ANOVA regarding efficiency scores (range 0-2) also revealed a significant difference between group conditions [$F(2, 83) = 17.66$, $p < 0.001$]. The standard condition (*M* = 1.40, *SD* = .18) scored significantly higher than the two-trial condition (*M* = 1.29, *SD* = .16) [*g* = .11, *p* = .04] and the zero-trial condition (*M* = 1.16, *SD* = .11) [*g* = .24, $p < 0.001$]. The two-trial condition also scored significantly higher than the zero-trial condition [*g* = .13, *p* = .002].

An ANOVA was then run to investigate differences between the average proportion of times each specific efficiency score was observed between conditions. For efficiency scores of 0 (least efficient: no previous information used, no new information gained), there was a significant difference between group conditions in proportion of attempts that scored this score [$F(2, 83) = 7.39$, $p = .001$]. The standard condition (*M* = 1.43%, *SD* = 3.04%) made inefficient attempts (score 0) significantly less than the zero-trial condition (*M* = 5.06%, *SD* = 3.98%) [*g* = 0.036, *p* = .001]. The standard condition did not differ from the two-trial condition (*M* = 3.02% , *SD* = 3.64%) [*g* = .02, *p* = .18], and the two-trial condition also did not differ from the zero-trial condition [*g* = .02, *p* = .12]

For efficiency scores of 1 (intermediate efficiency, some but not all previous information used or new information gained) there was a significant difference between conditions [$F(2, 83) = 11.15$, $p < .001$]. The standard condition (*M* = 62.41%, *SD* = 16.81%) made attempts with intermediate efficiency less often than the zero-trial condition (*M* = 78.46%, *SD* = 8.27%) [*g* = .16, *p* < .001] and the two-trial condition (*M* = 69.79%, *SD* = 11.84%) used less intermediate efficiency attempts than the zero-trial condition [*g* = .09, *p* = .006]. The standard condition did not differ from the two-trial condition [*g* = .07, p = .140].

For efficiency scores of 2 (highest efficiency; attempt made use of all current information, gained all new information, or a combination of the two) there was a significant difference between all group conditions [*F*(2, 83) = 14.90, *p* < .001]. Those in the standard condition (*M* = 36.16%, *SD* = 16.36%) attempted efficiently marginally significantly more than those in the two-trial condition (*M* = 27.20%, *SD* = 12.09%) [*g* = .09 *p* = .055] and those in the standard condition attempted efficiently more often than those in the zero-trial condition (*M* = 17.54%, *SD* = 8.95%) [*g* = .19, *p* < .001]. Those in the two-trial condition also attempted efficiently more often than those in the zero-trial condition [*g* = .10, *p* = .003].

In sum, the data were in agreement with the main hypotheses. Those in the high cognitive load demand condition (zero-trial) were less successful and less efficient at finding a code, and also used more attempts to do so; those in the lowest demand condition (standard) solved the code more often, had fewer attempts and overall better efficiency, as predicted *(see Table 1).*

Dual N-Back task

An ANOVA revealed a significant difference between problem-solving task condition groups on overall Dual N-Back task score [*F*(2, 83) = 3.46, *p* = .04]. The standard condition (*M* = 57.99%, *SD* = 9.62%) scored significantly lower than the zero-trial condition (*M* = 64.31%, *SD* = 9.35%) [*g* = .06, *p* = .04]. The standard condition did not score differently than the two-trial condition (*M* = 60.54%, *SD* = 8.29%) [*g* = .03, *p* = .53], and the two-trial condition did not score differently than the zero-trial condition [*g* = .04, *p* = .25]

Although there were no significant group differences in the N-1 or N-3 portions of the task [N-1: $F(2, 83) = 2.43$, $p = .09$; N-3: $F(2, 83) = 1.11$, $p = .34$], there was a difference between groups on performance in the N-2 task [$F(2, 83) = 5.10$, $p = .008$]. A post-hoc test revealed that the standard condition (*M* = 38.04%, *SD* = 15.70%) scored significantly lower the zero-trial condition (*M* = 50.26%, *SD* = 12.74%) [g = .12, $p = .006$]. The standard condition did not differ from the two-trial [$g = .08$, $p = .15$] and the two-trial did not differ from the zero-trial [$g = .05$, $p = .42$]

Dual N-Back and Code-Breaking Task. There was no correlation between N-1 scores and attempts made in standard condition (r = -.10, $p = .62$), or the two-trial condition (r = .10, p = .61). There was, however, an association between Dual N-Back scores in the N-1 and attempts made in the zero-trial condition (r = -.47, $p = .01$), indicating that higher WM scores were associated with less PS attempts in this condition. In the N-2, scores and attempts made also had no correlation in the standard condition (r = -.18, $p = .36$), or the two-trial condition (r = -.04, $p = .85$). There was, however, a correlation in the zero-trial condition (r = -.49, $p = .008$). Scores in the N-3 and attempts made did not correlate in the standard condition (r = -.08, $p = .67$), the two-trial condition (r = -.17, $p = .39$) or in the zero-trial condition (r = -.22, $p = .27$).

There was no correlation between N-1, N-2 or N-3 scores and number of wins in the standard condition (N-1: r = .17, $p = .40$; N-2: r = .08, $p = .67$; N-3: r = .01, $p = .98$) or the two-trial, (N-1: r =.06 , $p = .74$; N-2: r = .14, $p = .48$; N-3: r = .23, $p = .23$). There was a correlation in the zero-trial in the N-1 and N-2 (N-1: r = .44, $p = .02$; N-2: r = .48, $p = .01$), but not the N-3 (r = .183, $p = .35$).

There was no correlation between N-1 scores and efficiency in standard condition (r = .12, $p = .54$) and the two-trial condition (r = -.09, $p = .64$). There was, however, a correlation in the zero-trial condition (r = .39, p = .04). In the N-2, scores and efficiency were not related in the standard condition (r = .16, p = .39), the two-trial condition (r = .19, p = .33) or in the zero-trial condition (r = .18, p = .36). In the N-3, scores and efficiency were also not related in the standard condition (r = .09, p = .62), the two-trial condition (r = .19, p = .32) or in the zero-trial condition (r = .01, p = .96).

In sum, those in the zero-trial condition scored higher in the Dual-Back as compared to those in the standard condition. Although this was not hypothesized, correlations found between Dual N-Back and problem-solving attempts as well as efficiency in specifically the zero-trial condition also support this association *(Table 2).*

Simon game

An ANOVA revealed there were no significant differences between problem-solving task condition groups on Simon game performance [$F(2, 83) = 1.19$, p = .31]. The standard condition (M = 7.29, SD = 1.54), two-trial condition (M = 8.28, SD = 2.79) and zero-trial condition (M = 7.84, SD = 2.82) all scored very similarly, remembering about 7-8 colors in a sequence on average *(see table 2).*

Simon Game and Code-Breaking Task. Correlations revealed that the association between score in the Simon game and average attempts made was marginally significant (r = -.209, p = .053). This implies that those who scored higher on the Simon game often made fewer attempts in the code-breaking game. There was, however, a slightly stronger association between number of wins and Simon scores (r = .234, p = .030). In terms of efficiency, there was not a significant correlation between the Simon game and average efficiency scores in the problem-solving task overall (r = .12, p = .27).

In sum, the Simon game was associated with overall problem-solving success but was only slightly correlated with average number of attempts, and not at all with efficiency.

Simon and Dual N-Back

There was a significant positive correlation between Simon performance and overall score in the Dual N-Back (r = .44, $p < .001$). This is also true when viewing scores in just the N-1 (r = .39, p <. .001) and N-2 (r = .46, $p < .001$), but not the N-3 (r = .165, $p = .35$).

Overall, this implies that performance on the two memory tasks are associated with each other. This might also suggest that the nature of the N-3 portion of this game might involve the largest amount of difference in cognitive processes necessary to execute the task.

PS Task Condition	Wins *M* (*SD*)	Attempts *M* (*SD*)	Efficiency score *M (SD)*	Proportion of 0 Efficiency Scores *M (SD)*	Proportion of 1 Efficiency Scores *M (SD)*	Proportion of 2 Efficiency Scores *M (SD)*
Standard (*n* = 29)	3.2 (1.52)	7.90 (1.28)	1.40 (.18)	1.43% (3.05%)	62.41% (16.81%)	36.16% (16.36%)
Two-trial: All Games (*n* = 29)	2.62 (1.63)	8.21 (1.20)	1.29/2 (.15)	3.02% (3.64%)	69.79% (11/84%)	27.20% (12.09%)
Zero-trial (*n* = 28)	1.89 (1.10)	8.84 (.80)	1.16/2 (.11)	5.06% (3.98%)	78.46% (8.27%)	17.54% (8.95%)

Table 1. Performance in task by condition in the code-breaking task

Two-trial: Games 3-5 (*n* = 29)	1.34 (1.01)	8.60 (1.30)	1.25 (.18)	3.54% (4.70%)	72.09% (13.13%)	24.37% (12.98%)

Table 1a. Two-trial condition scores in the code-breaking task on exclusively rounds three through five, when information started to disappear.

	Simon Score *M* (*SD*)	N-Back Score (overall) *M* (*SD*)	N-1 Score *M* (*SD*)	N-2 Score *M* (*SD*)	Percentage of Participants to play N-2 *M* (*SD*)	N-3 Score *M* (*SD*)	Percentage of Participants to play N-3 *M* (*SD*)
All conditions (*n* = 86)	7.84 (2.39)	60.91% (9.36%)	77.86% (14.98%)	44.49% (15.17%)	-	31.89% (12.58%)	-
Standard (*n* = 29)	7.29 (1.54)	57.99% (9.62%)	74.44% (15.20%)	38.04% (15.70%)	67.59% (21.34%)	29.17% (7.63%)	8.28% (5.77%)
Two-trial (*n* = 29)	8.28 (2.79)	60.54% (8.29%)	76.57% (13.98%)	45.58% (14.73%)	74.67% (21.68%)	28.65% (11.41%)	14.00% (7.60%)
Zero-trial (*n* = 28)	7.84 (2.82)	64.31% (9.35%)	82.74% (14.98%)	50.26% (12.74%)	80.71% (13.74%)	35.14% (14.38%)	18.52% (9.34%)

Table 2. Performance in memory tasks overall, as well as by condition

Discussion

Participants in the low memory demand condition (standard) performed best on the code-breaking task, finding more solutions overall, as well as finding them with the highest level of efficiency in their information-gathering activity. Working memory scores in the Dual N-Task were highest in the high demand condition (zero-trial), although this was not an expected result. Scores in the Simon game, a measure of short-term visuospatial sketchpad capacity, did not differ by condition, with all groups being able to remember roughly 7-8 items on average. Average wins, attempts, and efficiency of information gathering scores in the code-breaking task were significantly associated with working memory scores for participants in the high demand (zero-trial) but not significantly in the intermediate (two-trial) or low demand (standard) conditions. Across the whole sample scores in the Simon game were associated with overall success in the problem-solving task, but not strongly associated with efficiency or number of attempts.

These findings provide evidence for a link between information gathering in a problem-solving context and demands on working memory (whether these demands are storage, resource, or retention based). Efficiency in the code-breaking task requires participants to structure guesses so that they maximize information gain. The findings show that increasing the demand for what is held in the mental workspace decreases performance on the code-breaking task, but also seems to improve working memory performance. The implications of these findings for our understanding of how cognitive load (demands on mental workspace) shapes problem-solving performance are considered in the next section.

Problem-Solving Task Performance and Cognitive Load

As predicted, there were observed differences in efficiency of information gathering between groups, with those in the zero-trial condition having the lowest efficiency and those in the standard condition having the highest efficiency. This supports many of the findings found in the cognitive load literature suggesting that it is difficult to learn effectively (e.g., strategies for efficiency) when multiple systems are competing for processing (Ayres & Paas, 2012; Sweller, 1988). However, it might also be the case that the exact situation of cognitive demands in this context are just essentially not well experienced, therefore the knowledge base is not well organized to task demands. Due to the fact that the problem presented in this task required the coordination of multiple mental resources, as well as ongoing processing, it is possible that this demand constrained participants in the two or zero-trial condition in such a way that the resources available to them to execute the task were inefficiently organized to the problem.

When information in a problem-space is not immediately available, demands on processes such as working memory increase in order to store and coordinate relevant information that one might be searching for. At the same time, it seems as if the balance of cognitive resources by which these higher-order tasks are accomplished might necessarily shift to compensate for this change in demand, thus changing the resources by which one might execute the task. In the highest cognitive load demand condition (zero-trial) and intermediate demand (two-trial) condition, the ability to reach a solution, and the overall efficiency of reaching it, were significantly hampered when compared to the lowest demand (standard) condition. It is possible that this change in association of working memory and task performance between condition might reflect somewhat of a trade-off system for how certain cognitive resources are allocated given the context of the problem (Daneman & Carpenter, 1980).

Increased Cognitive Load, Better Working Memory

Scores for those in the highest demand condition (zero-trial) were significantly correlated with all measured aspects of problem-solving, although scores for those with no imposed demands, or an intermediate amount, were not significantly related (standard and two-trial conditions). A potential interpretation of these results might be that the task demands presented in the high demand condition likely solicited the use of working memory more robustly in order to be able to properly plan and execute steps in the given problem-space. Given the fact that the code-breaking task was always presented first, and that those in the high demand condition scored highest on the Dual N-Back task, there are several possible explanations. There is no direct evidence that high cognitive load tasks improve subsequent WM functioning generally speaking, however, the opposite has been found (Anderson, Reder & Lebiere, 1996). One possible explanation for this discrepant finding is that the nature of the zero-trial task activated

WM to an extent that participants in this condition were somewhat more primed for better performance in the Dual N-Back more specifically. In other words, it is possible that the nature of both the zero-trial PS task and the Dual N-Back task were such that the demand in the PS task was in a way priming the organization capacities necessary for success in the subsequent working memory task. Another possibility is the pre-existence of group differences underlying the difference in Dual N-Back scores following the problem-solving task. This can be addressed in follow-up studies by counter-balancing all three tasks and observing if this difference in groups persists. If the difference between conditions on Dual N-Back scores does not persist when the problem-solving task is played *after* the Dual N-Back, then there would be reason to support a potential activation hypothesis.

Code-Breaking Task and Working Memory

The Dual N-Back task was chosen due to its inclusion of all three of Baddeley's components of working memory (phonological loop, visuospatial sketchpad, central executive) and was therefore predicted to be associated with all aspects of PS, regardless of condition. This was hypothesized because WM is needed to coordinate sources of incoming information with ongoing strategy while also evaluating performance and planning future steps; something presumably needed to be successful in the code-breaking task. However, no such association was found. This lack of association in the standard and two-trial conditions implies that the nature of the problems presented in these contexts were variably demanding of WM to an extent that the standard and two-trial conditions didn't rely on WM much, if at all, to complete the task. Under this notion, one could consider that overall measures of working memory are more associated with problem-solving when there is no easily accessible or automated strategy for organizing information about attempts. This would support Meyer et al.'s (2010) developmental work that suggests the central executive's processing aspects of WM are more related to solving mathematical word problems when one is less experienced at organizing units of information, and less related when one does not need robust processing to execute a problem.

Simon scores were related to overall number of wins across all conditions. This is consistent with a general finding that short-term storage-based tasks are related to overall success in problem-solving (Bull & Johnston, 1997). Since no previous work has used this specific memory task in relation to problem-solving, the current findings enhance what is understood about the role of this ability in more specific measures of PS activity (attempts and efficiency of information gathering). The Simon game involved the storage of visuospatial information as well as mechanisms of rehearsal (a strategy well used and known to adults), both of which are necessary for success in varying conditions of the code-breaking game.

Developmental Implications: Connecting Micro to Macro

How a problem is specified, and what type of problem it might be, changes as we gain experience in the problem's domain. Many of the problems that children encounter in the natural world are likely ill-defined, given their limited knowledge. Therefore, how children gather information is crucial to how they solve problems. The problem presented in the code-breaking task was well-defined, albeit unfamiliar to adults, and thus, required information gathering in order to organize an effective strategy. Because this same problem can be used with kids, the information gathering process in children and adults can be directly compared. From Karmiloff-Smith's (1979) view, even though adults will have different metaprocedures available, the microlevel process of adjusting representation to the problem-solving context is separable from these metaprocedures. How visual attention is allocated during the task might provide some insight into how cognitive load is managed and changes as adults and children acquire code-breaking skills. Visual attention might be independent of higher-level strategies, such as rehearsal, which does not tend to be as familiar to young children as it is to adults. Visual attention might also reflect a kind of visuospatial rehearsal. There isn't a great deal known about microlevel aspects of problem-solving in children, therefore the adult data in the code-breaking task provides a suitable starting point for generating hypotheses about the very nature of this process.

Conclusion

The code-breaking task provides a direct measure of information gathering in a problem-solving context. This is important due to how little is understood about how adults generally solve problems. Due to the fact that most of the research to date has focused on well-defined problems that do not require information gathering, the current findings provide insight into how participants coordinate information in a mental workspace when less information is initially known. Performance in the code-breaking task demonstrated the relative influence cognitive load has in an information gathering context. Associations between standard working-memory tasks and performance in the code-breaking task indicated that cognitive load did indeed have an impact on information gathering; however, this cannot be simply interpreted as a storage limit. This code-breaking task also provided a platform to both characterize the general adult capacity to solve problems, as well as to consider developmental differences in the microlevel processes that organize information gathering to a problem.

References

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology*, 30(3), 221-256.

Ayres, P., & Paas, F. (2012). Cognitive load theory: New directions and challenges. *Applied Cognitive Psychology*, 26(6), 827-832.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47-89). New York, NY: Academic press.

Baurer, P.J. & Fivush, R. (2014). The development of memory: multiple levels and perspectives. In *The Wiley handbook on the development of children's memo*ry. (Vol 1, pp. 1-13). West Sussex, UK: John Wiley & Sons.

Best, J. B. (2000). The role of context on strategic actions in Mastermind. *The Journal of general psychology*, 127(2), 165-177.

Brooks, P. J., Hanauer, J. B., Padowska, B., & Rosman, H. (2003). The role of selective attention in preschoolers' rule use in a novel dimensional card sort. *Cognitive Development*, 18(2), 195-215.

Bryce, D., Whitebread, D., & Szűcs, D. (2015). The relationships among executive functions, metacognitive skills and educational achievement in 5 and 7 year-old children. *Metacognition and Learning*, 10(2), 181-198.

Bull, R., & Johnston, R. S. (1997). Children's arithmetical difficulties: Contributions from processing speed, item identification, and short-term memory. *Journal of Experimental Child Psychology*, 65(1), 1-24.

Cabbage, K., Brinkley, S., Gray, S., Alt, M., Cowan, N., Green, S., & Hogan, T. P. (2017). Assessing working memory in children: The Comprehensive Assessment Battery for Children–Working Memory (CABC-WM). *JoVE (Journal of Visualized Experiments)*, (124), e55121.

Chi, M. T. H. (1978). Knowledge structures and memory development. In R.S. Siegler (Ed.), *Children's thinking: What develops*? (pp. 73-96). Hillsdale NJ: Lawrence Erlbaum Associates, Inc.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why?. *Current Directions in Psychological Science*, 19(1), 51-57.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological bulletin*, 104(2), 163.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450-466.

DeMarie & Lopez (2014). Memory in schools. In *The Wiley handbook on the development of children's memo*ry. Vol 2. (pp. 836-864). West Sussex, UK: John Wiley & Sons.

Davidson, J. E., & Sternberg, R. J. (1998). Smart problem-solving: How metacognition helps. *Metacognition in Educational Theory and Practice*, 47-68.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135-168.

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, 333(6045), 959-964.

Dockrell, J., & Campbell, R. (1986). Lexical acquisition strategies in the preschool child. In *The development of word meaning* (pp. 121-154). Springer, New York, NY.

Ellis, S., & Siegler, R. S. (1994). Development of problem solving. In T*hinking and Problem Solving* (Vol 2, pp. 333-367). Cambridge, MA: Academic Press.

Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive attention and metacognitive regulation. *Consciousness and Cognition*, 9(2), 288-307.

Gardner, H., & Hatch, T. (1989). Educational implications of the theory of multiple intelligences. *Educational Researcher*, 18(8), 4-10

Gathercole, S. E., Alloway, T. P., Willis, C., & Adams, A. M. (2006). Working memory in children with reading disabilities. *Journal of Experimental Child Psychology*, 93(3), 265-281.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental psychology*, 40(2), 177.

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: bridging selective attention and working memory. *Trends in Cognitive Sciences*, 16(2), 129-135.

Towse, J. N., Hitch, G. J., & Hutton, U. (1998). A reevaluation of working memory capacity in children. *Journal of memory and language*, 39(2), 195-217.

Jacques, S., Zelazo, P. D., Kirkham, N. Z., & Semcesen, T. K. (1999). Rule selection versus rule execution in preschoolers: An error-detection approach. *Developmental psychology*, 35(3), 770.

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, 18(4), 394-412.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of experimental psychology*: General, 132(1), 47.

Karmiloff-Smith, A. (1979). Micro-and macrodevelopmental changes in language acquisition and other representational systems. *Cognitive Science*, 3(2), 91-117.

Keen, R. (2011). The development of problem solving in young children: A critical cognitive skill. *Annual review of psychology*, 62, 1-21.

Keating, D. P., & Crane, L. L. (1990). Domain-General and Domain-Specific Processes in Proportional Reasoning: A Commentary on the" Merrill-Palmer Quarterly" Special Issue on Cognitive Development. *Merrill-Palmer Quarterly*, 411-424.

Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Science*, 9(5), 178-181.

Laughlin, P. R., & Barth, J. M. (1981). Group-to-individual and individual-to-group problem-solving transfer. *Journal of Personality and Social Psychology*, 41(6), 1087-1093.

Lee, K., Ng, E. L., & Ng, S. F. (2009). The contributions of working memory and executive functioning to problem representation and solution generation in algebraic word problems. *Journal of Educational Psychology*, 101(2), 373-387.

Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science,* 16(8), 601-608.

Meyer, M. L., Salimpoor, V. N., Wu, S. S., Geary, D. C., & Menon, V. (2010). Differential contribution of specific working memory components to mathematics achievement in 2nd and 3rd graders. *Learning and Individual Differences*, 20(2), 101-109.

Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review,* 65(3), 151-166.

Newell, A., & Simon, H. A. (1972). Human problem solving (Vol. 104, No. 9). Englewood Cliffs, NJ: Prentice-Hall

Ornstein, P. A., & Naus, M. J. (1985). Effects of the knowledge base on children's memory strategies. In *Advances in child development and behavior* (Vol. 19), 113-148.

Pickering, S., & Gathercole, S. E. (2001). Working memory test battery for children (WMTB-C). Psychological Corporation.

Pretz, J. E., Naples, A. J., & Sternberg, R. J. (2003). Recognizing, defining, and representing problems. *The Psychology of Problem Solving*, 30(3-29).

Roebers, C. M., & Feurer, E. (2016). Linking executive functions and procedural metacognition. *Child Development Perspective*s, 10(1), 39-44.

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science*, 26(1-2), 113-125.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19(4), 460-475.

Sheya, A., & Smith, L. B. (2006). Perceptual features and the development of conceptual knowledge. *Journal of Cognition and Development*, 7(4), 455-476.

Siegler, R. S. (1999). Strategic development. *Trends in Cognitive Sciences,* 3(11), 430-435.

Sternberg, R. J. (1986). Critical thinking: Its nature, measurement, and improvement. In F. R. Link (Ed.), E*ssays on the intellect* (pp. 45-65). Alexandria, VA: Association for Supervision and Curriculum Development.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285.

Towse, J. N., Hitch, G. J., & Hutton, U. (1998). A reevaluation of working memory capacity in children. *Journal of Memory and Language*, 39(2), 195-217.

Zelazo, P. D., Carter, A., Reznick, J. S., & Frye, D. (1997). Early development of executive function: A problem-solving framework. *Review of General Psychology*, 1(2), 198-226.

Zelazo, P. D., Müller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood: VI. The development of executive function: Cognitive complexity and control--revised. *Monographs of the Society for Research in Child Development*. 68(3), 1-157.

Appendix: Supplementary Information

Figure S1: Bar graph demonstrating the average number of wins in each condition. Bars indicate a significant difference between the standard and zero-trial conditions.

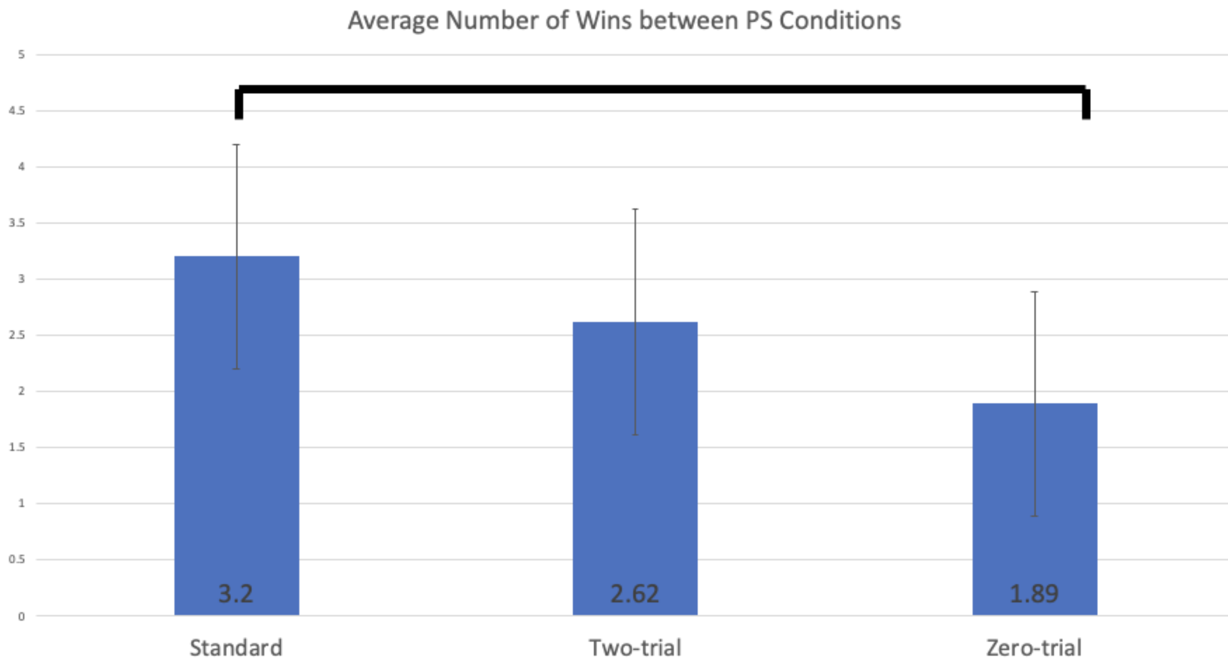

Figure S2: Bar graph demonstrating the average attempts in each condition. Bars indicate a significant difference between the standard and zero-trial conditions.

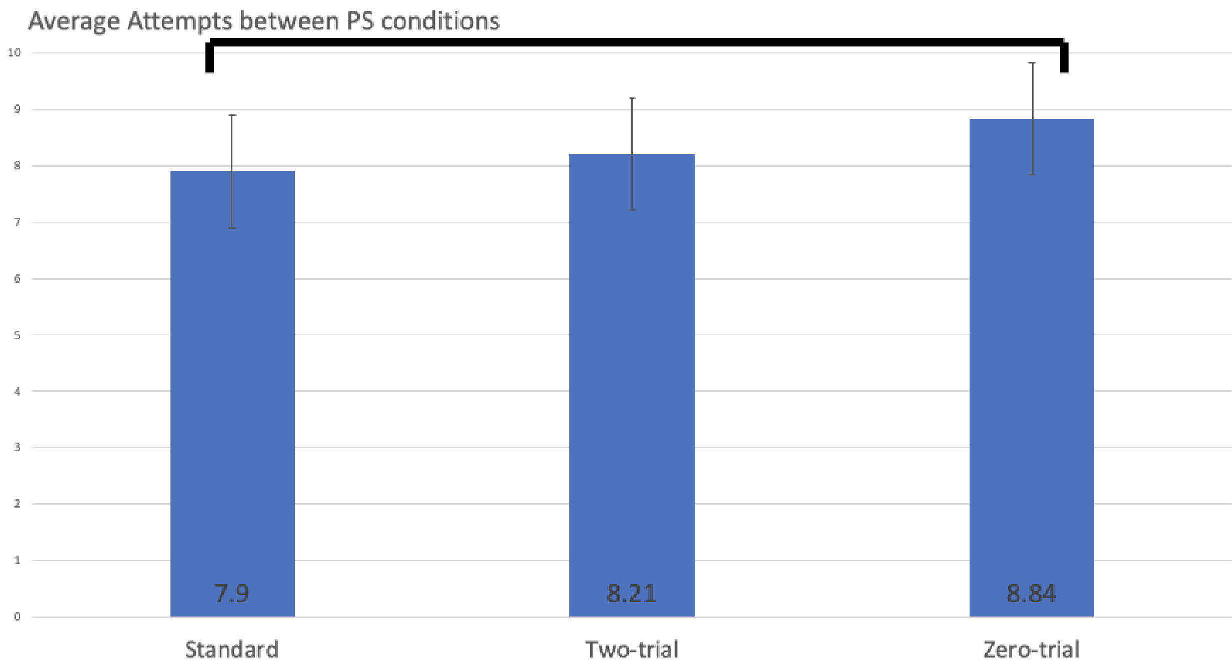

Figure S3: Bar graph of efficiency scores in each condition. Bars indicate a significant difference between all conditions.

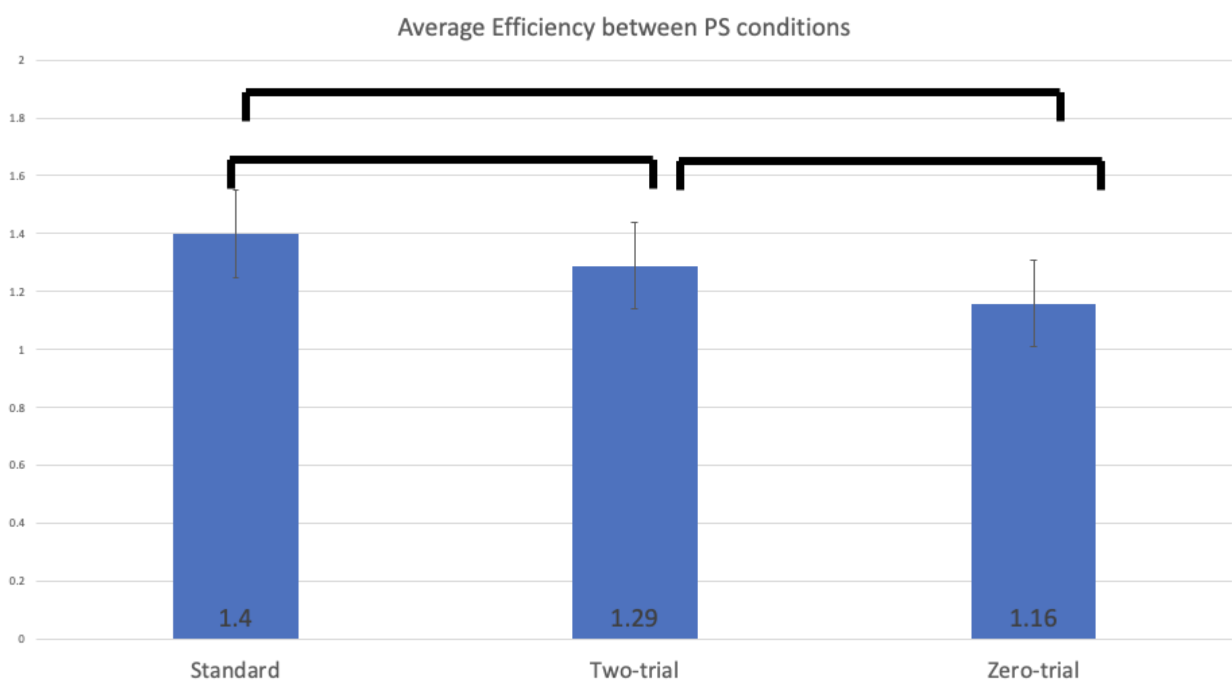

Figure S4: Proportion of each efficiency score used within each condition. The legend indicates which color each score is represented by. The corresponding p-values indicate if there was a significant difference between groups in the average proportion of times the attempts received each score (i.e., 0, 1 and 2).

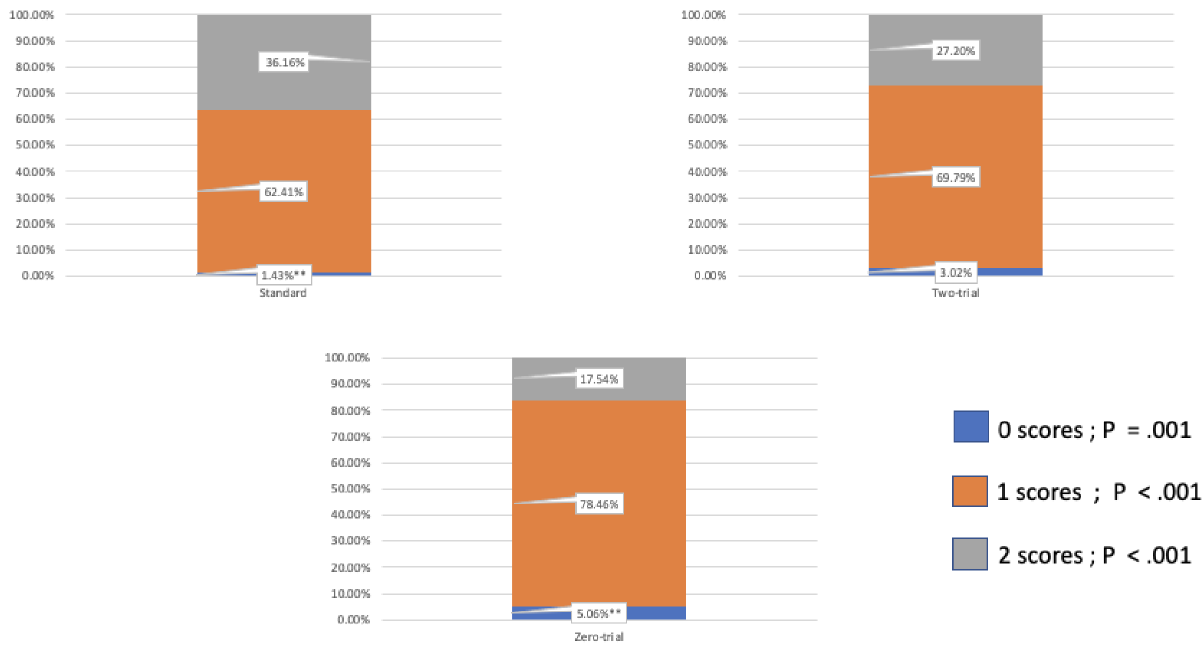

Figure S5: Bar graph of Simon game scores by condition.

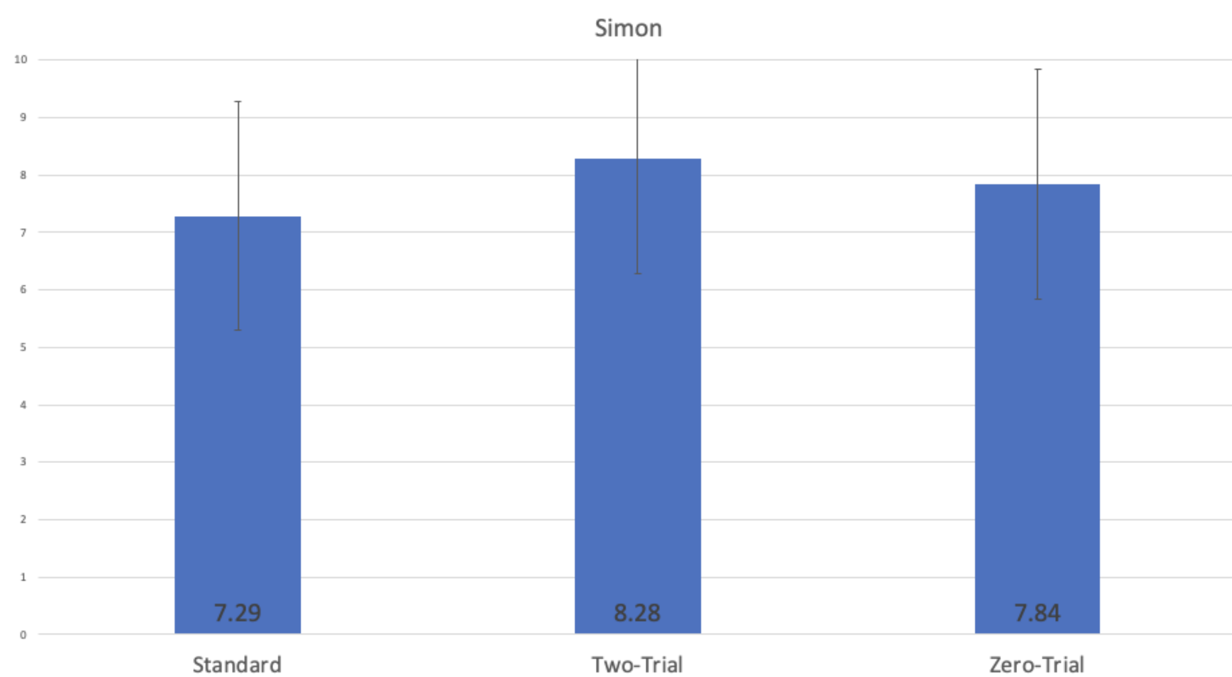

Figure S6: Scatter plot showing associations between Simon and overall wins.

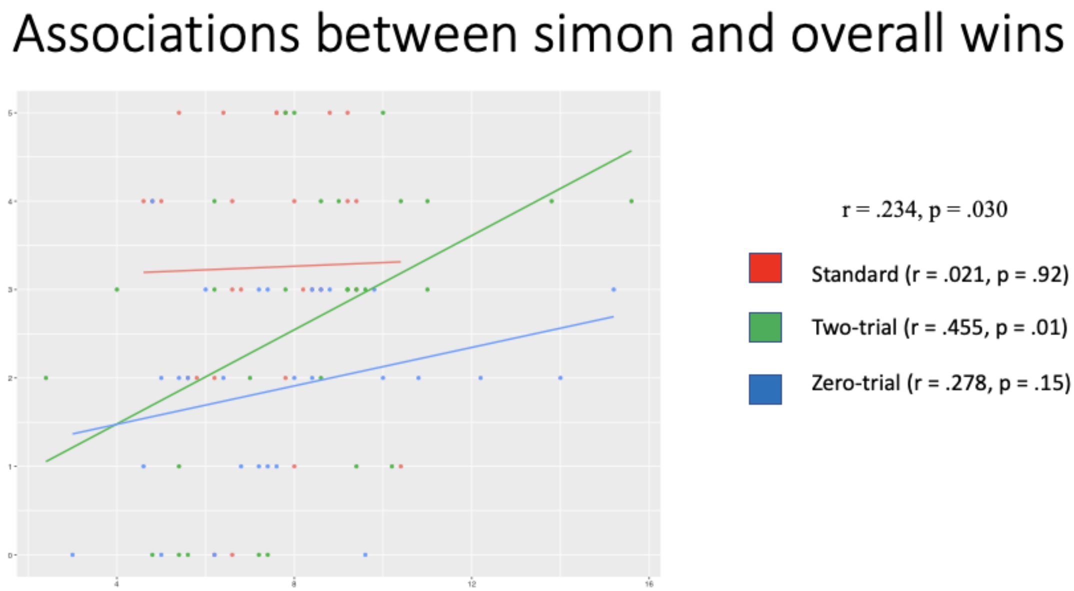

Figure S7: Scatter plot showing associations between Simon and efficiency of information gathering scores.

Figure S8: Bar graph of overall Dual N-Back scores between conditions. Bars indicate a significant difference between the standard and zero-trial conditions.

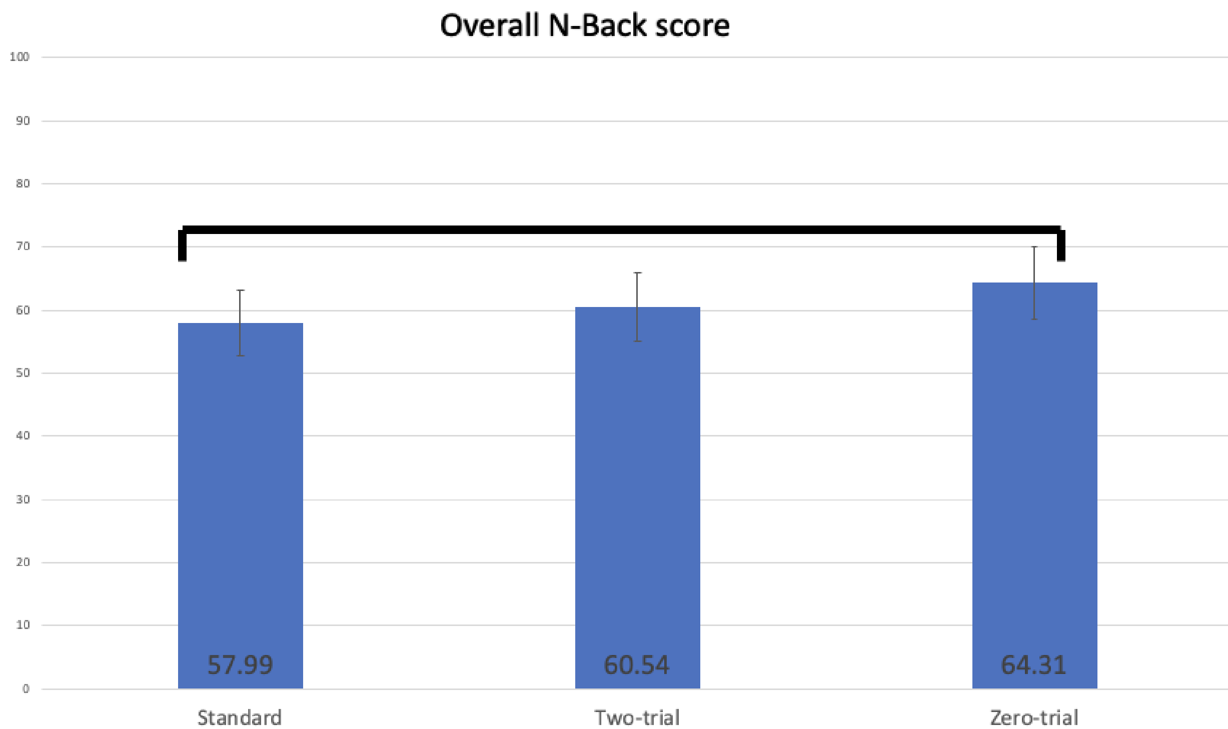

Figure S9: Bar graph of Dual N-Back scores in the N-1 portion of the task, between conditions.

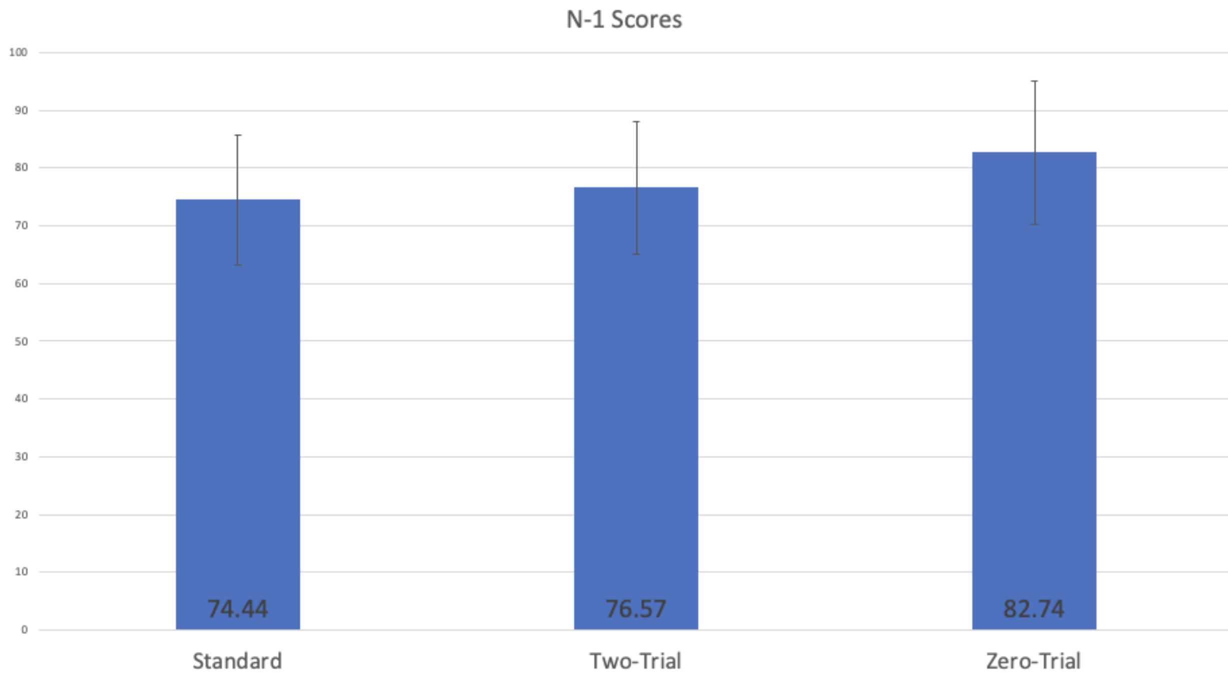

Figure S10: Bar graph of Dual N-Back scores in the N-2 portion of the task, between conditions. Bars indicate a significant difference between the standard and zero-trial conditions.

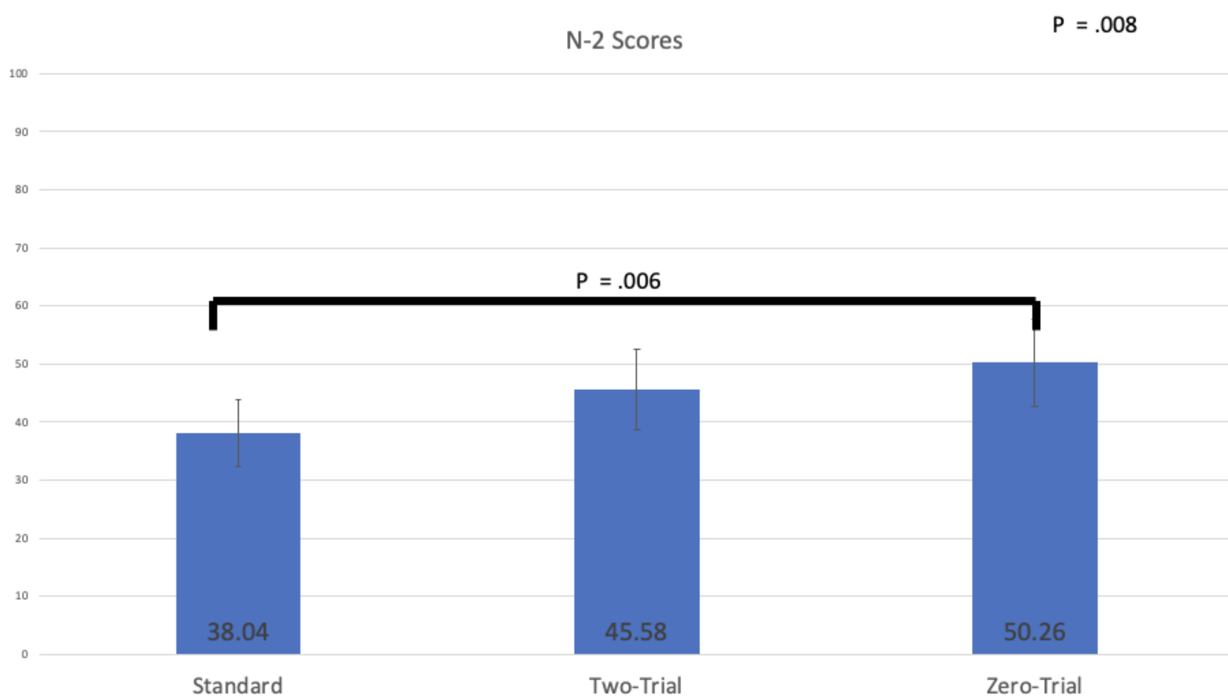

Figure S11: Bar graph of Dual N-Back scores in the N-3 portions of the task, between conditions.

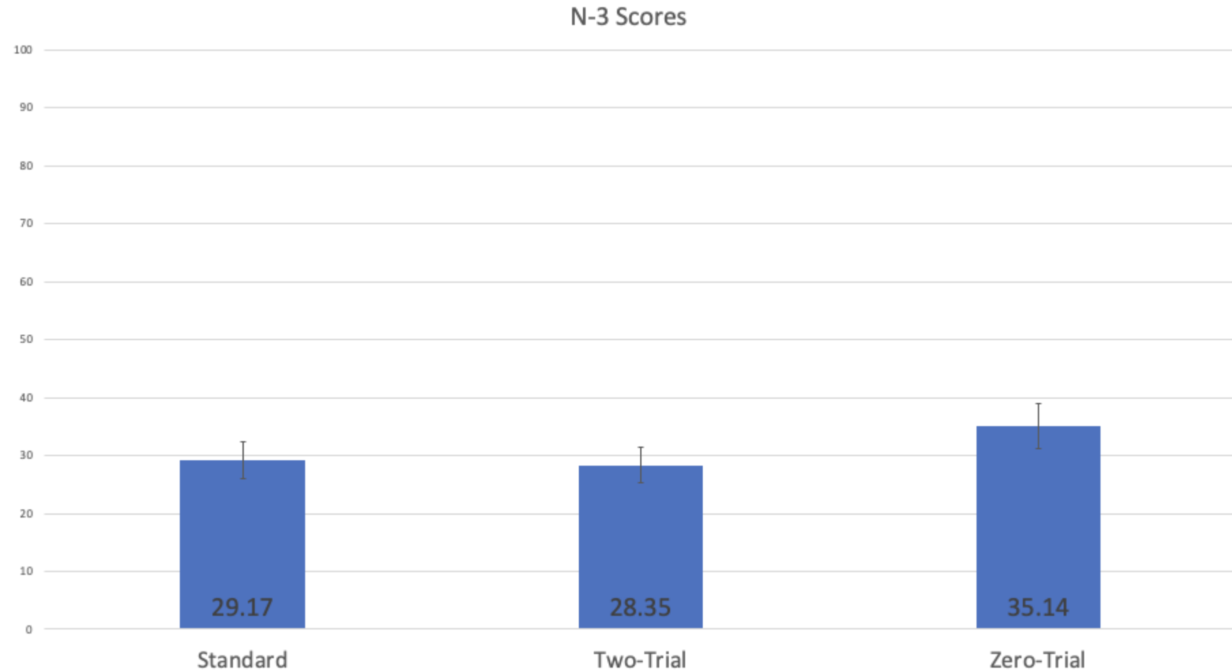

Figure S12: Scatter plot showing association between Dual N-Back scores and average wins in the code-breaking game

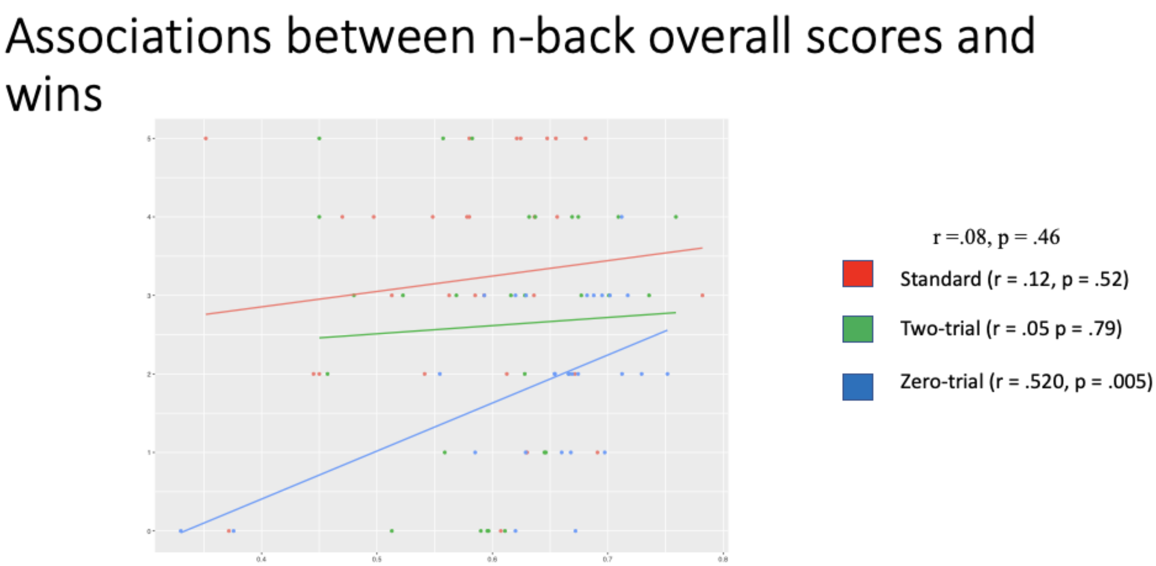

Figure S13: Scatter plot showing association between Dual N-Back scores and average attempts in the code-breaking game

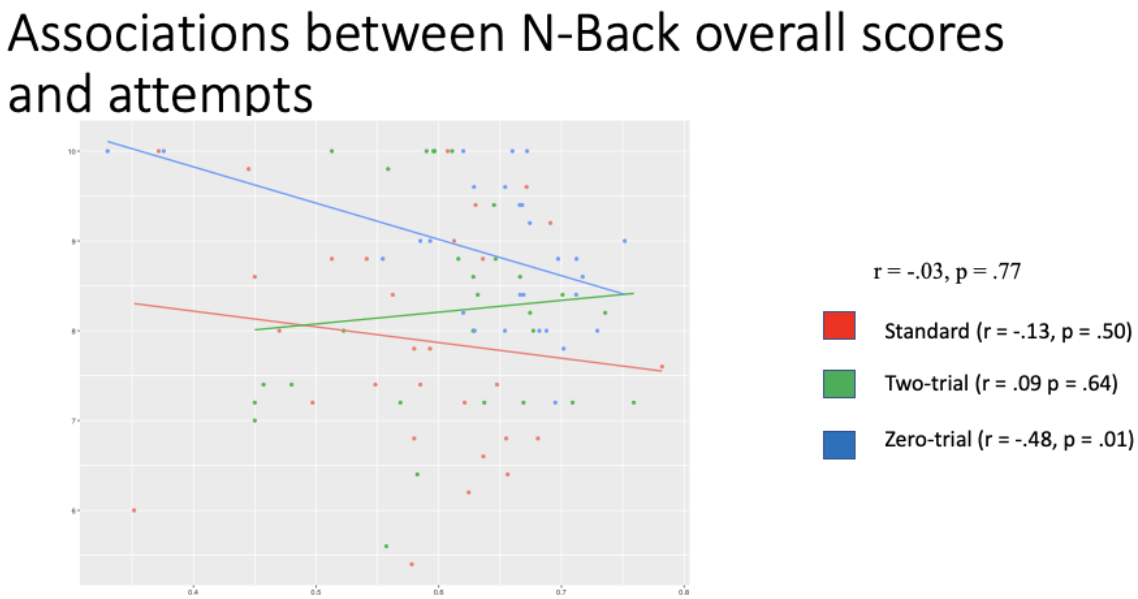

Figure S14: Scatter plot showing association between Dual N-Back scores and average efficiency scores in the code-breaking game