RUNNING HEAD: A MICROWORLD SIMULATION OF DYNAMIC COGNITION

# A microworld simulation of dynamic cognition as a test of executive function

Helen M. Hodgetts[1], Sonia Packwood[2], Francois Vachon[2], & Sébastien Tremblay[2]

[1] Cardiff Metropolitan University, UK

[2] Université Laval, Québec, Canada

Word Count (main text): 9276

## Author note

Sonia Packwood is now at Integrated Neuro, Vancouver, Canada.

Correspondence to: Helen Hodgetts, Department of Applied Psychology, Cardiff Metropolitan University, Cardiff, UK, CF5 2YB, or Sébastien Tremblay, École de psychologie, Université Laval, Québec (QC), G1V 0A6, Canada. E-mails: hhodgetts@cardiffmet.ac.uk or sebastien.tremblay@psy.ulaval.ca

## Abstract

Introduction: The lack of consensus regarding the nature or composition of executive functioning (EF) has led to a proliferation of executive tasks to assess the concept. Many do agree however that the theoretical concept of EF is a holistic one, leading us to consider whether it would be beneficial to assess EF in a more holistic manner. We explore how well a computerized simulation of dynamic cognition – that reproduces the context of real-world complex decision-making – can predict performance on nine classical neuropsychological tasks of EF. Method: A sample of 121 participants completed all tasks, and canonical correlations were used to assess the nine tasks as predictors of the three simulation performance metrics to evaluate the multivariate-shared relationship between the two variable sets: executive functions and dynamic cognition. Results: Results show that a substantial amount of variance in two indices of dynamic cognition can be explained by a linear combination of three key types of neuropsychological tasks (planning, inhibition, working memory), with a larger contribution from the planning tasks. Conclusion: Our findings suggest that tasks of dynamic cognition could augment traditional, separate tests of EF, offering benefits in terms of parsimony, ecological validity, sensitivity, and computerized delivery.

*Keywords*: executive function, neuropsychological tests, dynamic cognition

How to define and how to measure Executive Functions (EFs) is a contentious issue: Over the last few decades, research has deployed considerable effort to determining the organization of EFs, with opinion divided between a unitary model and the existence of multiple functions. Amongst those proponents of a unitary model, there is no agreement regarding the nature of this core underlying process; and for those subscribing to the view of multiple EFs, there is little consensus regarding the nature or even the number of functions. Furthermore, the proliferation of so-called functions has led to the development of a multitude of tasks with which to assess them (Baggetta & Alexander, 2016; Karr et al., 2011), making it increasingly difficult to directly compare between EF studies or diagnoses by different clinicians. Previously we have argued for greater coherence and uniformity regarding executive constructs (Packwood et al., 2011), and in the current article we extend this idea to consider whether greater parsimony could be achieved in terms of assessment. A recent systematic review of 106 papers found 109 different tasks used to test EF (Baggetta & Alexander, 2016); of these, 56 were used only once suggesting that researchers frequently develop their own tasks for their own specific study, thus adding to the circular problem of an escalating number of tasks and functions. Here, we consider whether it may be possible for a single task to predict performance on a range of tests of EF. In so doing, we look to tasks used outside of neuropsychology and consider single-concept views of cognition in other domains. Specifically, we turn to the idea of dynamic cognition, a concept used in cognitive psychology and human factors by which new behaviors and strategies emerge from the attempt to adapt to constant changes in the environment (Brehmer, 1992; Gonzalez, 2005a, b; Gonzalez et al., 2005). We suggest that dynamic cognition – or other comparable unitary concepts – could provide a parsimonious and ecologically valid indication of higher-order functioning in individuals, and we test whether performance on a dynamic cognition task is predictive of that on traditional tasks associated with three EFs: Planning, Inhibition, and Working Memory

(WM). We suggest that if a single task could predict performance on multiple aspects of EF, then there is the potential that such a task could be developed and standardized for clinical use in order to augment the traditional tests available to neuropsychologists, as well as to aid in research, education, and personnel selection.

There exists a lack of clarity around the concept of EF and the way in which it should be assessed. The proposal of multiple independent EFs adds to the abundance of tasks used to measure them, and to an extent, the continued development of new assessment tasks can influence how many separate EFs are identified (Packwood et al., 2011). Models that promote the existence of multiple EFs lack consensus regarding the number and nature of functions; indeed, there seem to be as many different organizations of EFs as there are studies addressing the issue (see Fournier-Vicente et al., 2008; Hull et al., 2008; Miyake et al., 2000). A task-specific approach is generally used in fractionating the executive system, by which the specificity of an EF is based upon the relationship between one EF measure and other executive tasks. According to this approach, any uncorrelated measures of EF would thus represent separate functions (see Miyake & Shah, 1999). For example, Miyake et al. (2000) used nine experimental tasks to validate their model of three separate EFs (shifting of mental sets; monitoring and updating of WM representations; inhibition of prepotent responses), and found that the tasks considered to predominantly tap each of the three target EFs were moderately correlated but clearly separable, thus supporting the hypothesis of three independent EFs. Baddeley (1996) used a multifaceted approach to clarify the central executive of his model: Based upon the observation of patient impairment and individual differences on four specific tasks (dual task, random generation task, selective attention task, and immediate recall task), he suggested the existence of four different EF components: i) the capacity to coordinate performance on two separate tasks; ii) to switch retrieval strategies; iii) to attend selectively to one stimulus and inhibit the disrupting effect of others; and iv) to hold and manipulate information in short-term memory, respectively. The task-specific approach gives rise to several models that support the idea of independent functions, with the validation generally of between three and six EFs. However, it seems that the more different tasks one uses to validate a model, the more different models one could potentially generate. Furthermore, there is still no common agreement on even one single subcomponent across all models (see Fournier-Vicente et al., 2008; Hull et al., 2008; Miyake et al., 2000). It could perhaps be argued that such efforts to fractionate the executive system contribute more to the continued ambiguity of the concept and measurement of EF rather than a resolution.

The idea that there may be a general dimension to EF is recognized by some researchers, with the suggestion that those not exploring this possibility will potentially overestimate the diversity of EF components (Karr et al., 2018). A recent systematic review revealed that nearly a quarter of models involved a global executive function factor (Karr et al., 2018), but even amongst those who advocate a more unitary approach there is still disagreement regarding the nature of the general mechanism that would underlie performance on any given executive task. One suggestion is that a weakening of associative links in WM could account for the diversity of executive deficits, such as perseveration or the selection of a prepotent action (Kimberg & Farah, 1993). Associative links in WM allow for the selection of appropriate actions based on contextual factors, but without such links, the integration of contextual information needed to perform the task would be prevented.

Another proposal is that of an executive-attention framework that unifies the role of EFs, WM and fluid intelligence (Kane & Engle, 2002). The framework is based on findings that high-WM participants perform better not only on span tasks, but also on tasks requiring controlled attention and the maintenance of task goals in the face of interference (i.e., executive attention). In contrast, there is no difference between low- and high-span participants on automatic tasks which require neither WM capacity nor EF (Kane et al., 2001).

The authors suggest that WM capacity underlies executive task performance in terms of the ability to actively maintain access to stimulus representations and goals in interference-rich contexts. A considerable amount of other research supports their view by demonstrating better performance on executive-attention tasks for high- than low-WM capacity individuals (Long & Prat, 2002), and an important correlation between WM tasks and measures of fluid intelligence that are acknowledged to be especially executive (Conway et al., 2002; Duncan et al., 1996; Frias et al., 2006). Furthermore, it has been proposed that the link between WM and intelligence is underpinned by variations in executive processes, with attentional control as the unique process that influences performance in more specific abilities (Kovacs & Conway, 2016; although see Schubert & Rey-Mermet, 2019). Another concept of interest is *placekeeping* – the ability to perform sequences in order without repetition or omission – with the suggestion that this executive construct could underlie differences in fluid intelligence to a greater extent than WM capacity (Hambrick et al., 2020). The close links between EF, WM, and general intelligence provide strong evidence for a global factor that can account for associations between these higher-order processes.

Despite suggestions of a general process, there is no consensus on a single task best placed to capture this global factor. In recent moves towards greater parsimony, efforts have looked at developing a composite score to understand aggregate performance (Iverson et al., 2021). While this would simplify the issue of multiple scores (for both clinicians and researchers), it would still require the time-consuming administration of multiple tests. An extensive test battery of course gives an in-depth analysis of patient deficits across a range of EFs, but there may be occasion when a quicker and more general assessment would be beneficial – for example, as an initial screening tool to determine whether or not further and more detailed investigation is necessary. We suggest that if a single task could be developed that taps into a more global executive factor, this could be a useful addition to the neuropsychologist’s toolkit. Furthermore, it may be useful to researchers who wish to observe individual differences in EF within the general population, for whom many neuropsychological tests suffer from ceiling effects, a narrow range, and do not exhibit a normal distribution within a non-clinical sample (Suchy, 2009).

In order to capture a global EF factor, the task used must be adequately complex and tap into higher-order functions rather than task-specific behaviors. Despite debate regarding the unity vs. diversity of EF, the majority of authors agree that the theoretical concept of EF is a holistic one; it involves a set of cognitive abilities and processes that guide human behavior in a controlled and adaptive way in complex and novel situations (Banich, 2009; Hull et al., 2008; Jurado & Rosselli, 2007). Suchy (2009) distinguishes between complex skills contributing to goal-directed behavior (reasoning, planning, problem solving), and elemental processes that form the basis for EF skills (such as sequencing, inhibition, initiation, response selection). While individuals may appear to have adequate EF in routine or well-practiced circumstances (their elemental skills are sufficient), deficits may emerge when they need to apply these skills in a novel situation. As such, tasks that encapsulate more goal-directed behavior could be deemed a better test of EF than those that focus on elemental skills in isolation. The Tower of Hanoi and Wisconsin Card Sorting task, e.g., are thought to be more ‘complex’ (e.g., Hull et al, 2008), perhaps encompassing several EFs, and therefore being more likely to tap into the purported global factor that correlates with intelligence. However, despite the proliferation of EF tests available to neuropsychologists, we suggest that it may be beneficial to take an interdisciplinary approach and consider whether concepts and methods from beyond neuropsychology can inform and augment those traditionally used. Indeed, there have been recent calls for greater integration and collaboration between psychometricians and cognitive psychologists to overcome disciplinary boundaries in order to make advances in theories and measurement (Ratcliff & McKoon, 2022; Schubert & Rey-Mermet, 2019).

The idea of an integrated and unitary view of cognition that encompasses multiple processes and functions is echoed in other domains such as intelligence (‘*g*’ factor; Spearman, 1904), education (complex problem solving; Funke et al., 2018), and cognitive psychology/human factors (dynamic decision-making or dynamic cognition; Brehmer, 1992; Gonzalez, 2005a, b; Gonzalez et al., 2005). These concepts eschew the idea of multiple independent and separately assessed cognitive processes, and instead fit well with our previous assertion that many proposed ‘processes’ are actually task-specific manifestations of behaviors deriving from a common function (Packwood et al., 2011). At the core of these concepts is the ability to ‘operate’, to be competent and functional, and to make appropriate decisions in complex (real-life) environments. This too, we believe, is the essence of EF, and we suggest that looking more broadly to the more parsimonious and ecologically aligned concepts in other disciplines could help with methodological and theoretical advances (e.g., Schubert & Rey-Mermet, 2019). In the current paper we borrow from the domain of cognitive psychology and human factors to propose that the concept of dynamic cognition – assessed using simulations of complex, novel, and dynamic real-world tasks – may capture the essence of EF and be a valuable predictor of performance on any executive task.

Dynamic cognition is used in the field of cognitive psychology to describe human complex decision-making processes and may be advantageous in the quest for a testable universal mechanism to indicate general executive ability. It is required in many everyday situations and can be described as a series of processes that interact dynamically to reach a goal; that is, to achieve and maintain control is a continuous activity requiring many decisions, each of which can only be understood in the context of the others. These decisions are not independent: Later ones are constrained by those made earlier, which in turn constrain those that come after them. The state of the problem changes, both autonomously and as a consequence of the decision maker’s actions (Edwards, 1962), and all these decisions are to be made in real time (Busemeyer & Townsend, 1993; Rapoport, 1975). Dynamic cognition is not a single process but a holistic view on cognition that goal-oriented behavior requires the dynamic interaction of different processes. It shares similarities with the concepts of complex problem solving (Funke et al., 2018) or dynamic decision-making (Gonzales, 2005a), which have been linked with higher-order global processes such as general intelligence (e.g., Hambrick et al., 2020), and are predictive of performance in real-life pursuits (Danner et al., 2011). We use this as just one example of a unitary and dynamic view on cognition beyond the boundaries of neuropsychology and psychometrics that offers parallels with EF.

The dynamic nature of cognition is evident in a whole range of real-world situations, particularly in critical work environments such as emergency response management, patient care in hospitals, and military operations. All these situations involve many abilities that are close to what we would describe as EFs: (a) the need for a good comprehension of the context, (b) to consider many different elements and subgoals while working on the problem as a whole (with the need to attend to some and inhibit others), (c) to form and test hypotheses about how the environment could possibly evolve, and (d) to develop an adaptive strategy in real time. These dynamic situations not only depend upon the subject’s own behavior, but also independent changes in the environment; as such, those capable of promptly comprehending their environment might be able to exercise greater control with effective and adaptive strategies. Thus, dynamic cognition seems to encompass many EFs at the same time (initiating, inhibition, WM, planning, changing strategies in the face of errors or environmental feedback) and is also close to the well-accepted general definition of EF: Cognitive control mechanisms that direct and coordinate human behavior in an adaptive way in a new and/or complex situation (see Lezak, 1995; Shallice, 1988). Therefore, it seems reasonable to believe that dynamic cognition would represent a source of commonality between all EFs, and encompass the essence of the concept in a more holistic manner without the need for fractionated subcomponents and the use of multiple executive tests.

A key benefit of looking to the human factors/cognitive psychology literature and using tasks of dynamic cognition is the potential to bridge the gap between ecological validity and control in neuropsychological assessment. Most neuropsychological tests tend to be performed in a structured environment, are generally static rather than dynamic, are not performed in real time, and involve a segmented and rigid structure with each trial predetermined and independent of actions/decisions taken in previous trials. This is not reflective of everyday EF, and also the fairly restricted nature of traditional tests means that practice or prior exposure to that or a similar task (and thus loss of novelty) can easily compromise validity. In contrast, scenarios tapping into the dynamic nature of cognition – although having some structure – are more flexible with no ‘right’ or ‘wrong’ answers (as in real life people have autonomy over different strategies), but they assess whether the task is approached in an optimal or adaptive way. They impose the need to ‘think on one’s feet’ in response to a novel and evolving situation. It is this which we believe is key to general executive abilities, but which may not be captured by traditional tests. Indeed, one could be able to carry out traditional executive tests as well as an average, non-clinical individual, but experience daily functional difficulties when dealing with non-practiced procedures and unstructured situations (Andrés, 2003; Andrés & van der Linden, 2002). That is, a neuropsychological test does not necessarily allow for the identification of difficulties that may occur in a more global, complex and dynamic context, and thus a task that assesses the dynamic nature of cognition may be a welcome addition to the traditional tasks available.

While a number of authors have assessed individuals’ task performance in real environments which are more representative of everyday life (Alderman et al., 2003; Norris & Tate, 2000), these are more challenging to organize and conduct than a laboratory-based task, and are also problematic in terms of control over extraneous variables. In the field, some internal validity is lost as external factors may intervene during the course of the task (e.g., the Multiple errands test; Shallice & Burgess, 1991), thus making it difficult to interpret results, compare to norms, and establish any deficits. We propose that a balance between internal and external validity can be achieved through the use of computerized simulations, or microworlds, that simulate the essential features of a dynamic system within a controlled environment (e.g., Funke et al., 2018; Hagemann & Kluge, 2017). Such a task could be run simply within a practitioner's setting to gain an indication of deficits of a more global nature and in a manner that is minimally demanding of time and resources. This could be used as a more ecological measure alongside traditional tests, or perhaps as an initial screening tool to determine whether a more time consuming and in-depth assessment of EF is required.

Microworlds retain the essence of real-world decision-making while importantly allowing a level of precision, assessment sensitivity, and control not afforded by field studies. There are many examples of microworld simulations (e.g., route finding, air traffic control, stock market investment, security surveillance, society management, company management; Danner, 2011; Dorner & Guss, 2022; Lafond et al., 2012; Rigas et al., 2002), each of which are simplified versions of actual tasks but which still embody the essential abstract cognitive characteristics of real-world problems. The complexity, dynamics, and opaqueness typical of a microworld task (e.g., Gonzalez, 2005a, b; Gonzalez et al., 2005) necessarily engage a broad combination of executive processes, akin to those required outside the laboratory for everyday functioning. Microworlds are complex in the sense that several problems must be solved in parallel and feedback for which is rare and/or delayed; this activates the interaction of several EFs such as the planning and scheduling of future actions, the monitoring and updating of relevant information in WM, and the inhibition of actions interfering with immediate problem solving. As with real life, microworlds are dynamic: changes occur as a consequence of decisions made by the participant, but also autonomous changes in the environment can place them in unpredictable situations. There is some structure of course, but the evolution of the situation is not predetermined; there are more ‘optimal’ ways of approaching the MW task but is less rigid than traditional tests for which there are ‘right’ or ‘wrong’ answers. Finally, the tasks are opaque in that all the information necessary to solve the problem is not immediately apparent; the participant has to make deductions and test predictions according to their own comprehension of the environment. The microworld therefore reproduces the complexity of real-life tasks in combination with everyday life stresses such as temporal pressure, uncertainty and limited resources (Granlund & Johansson, 2004), which is something that is missing from the classic neuropsychological tests. Furthermore, in contrast to field studies, the high degree of control allows us to infer the general underlying processes employed, which can be applied to a broad range of dynamic decision-making tasks.

A step in the right direction in recent years has been the use of virtual reality (VR) environments (see Borgnis et al., 2022) that offer a balance between internal and external validity, and include multiple EFs within a single scenario rather than using separate tests. While at first glance such tasks may seem comparable to microworlds, they still lack some critical elements to be a fully comprehensive assessment of EF. Example tasks include VR versions of multiple errands (Nir-Hadad et al., 2017; Rand et al., 2009), cooking (Chicchi Giglioli et al., 2019), office tasks (Jansari et al., 2014), and a virtual library (Renison et al., 2012), and are typically characterized by a ‘to-do’ list, that may incorporate some temporal overlap between subtasks. These tasks incorporate several EFs that are assessed in traditional neuropsychological tasks like inhibition, planning/organizing, and problem solving, but they do not quite capture any of the complex problem properties or the dynamic aspects of cognition. For example, although subtasks may be conducted in parallel, decisions are still fairly independent and later choices are not affected by earlier decisions; environmental change only occurs when effected by the participant themselves (e.g., moving an item from one room to another); and participants have feedback on their actions instantly, rather than needing to plan by projecting the potential impact of their earlier decisions to some evolved future state. While the VR tasks incorporate a number of separate executive elements, dealing with the dynamic evolution of a microworld within compressed time invokes processes that tap into EF to its full extent. Microworlds allow the study of how constructs interact in a more complex environment (Gray, 2002), and we believe it is this dynamic *interaction* between EFs – rather than simply the inclusion of multiple EFs within a single test – that can elicit a more general executive ability. The idea that this type of dynamic decision making provides unique access to a comprehensive executive factor is supported by other research associating complex problem-solving processes with other higher-order unitary constructs such as fluid intelligence (e.g., Funke, 2021; Hambrick et al., 2020).

## The present study

In the present study, we set out to assess whether a more holistic concept of dynamic cognition can capture the essence of three chosen EFs. The three selected were the most common according to a recent literature review of 60 of the most frequently cited papers on EF (Packwood et al., 2011): planning (48% of studies), inhibition (42%), and WM (42%). Set shifting (32% of studies) was also commonly assessed, probably because – along with inhibition and WM – it makes up the three-factors of one of the most highly-cited models of EF (Miyake et al., 2000). However, we chose to assess planning in favor of shifting, given its prevalence in nearly half of all the reviewed studies, and also the fact that it has been the subject of relatively little empirical work. Besides, planning is often viewed as a more complex factor, perhaps subsuming inhibition, WM, and shifting under it, and so we were interested to see whether such a complex and multidimensional executive construct would correlate more highly with our measures of dynamic cognition.

Taking into account the holistic conceptualization of EF that is well accepted in the literature – that is, cognitive control mechanisms that direct and coordinate human behavior in an adaptative way when no pre-established schema of action is available (see Lezak, 1995; Shallice, 1988) – we test a computer simulation of a complex problem-solving situation that, from our point of view, better represents this definition than some of the existing static executive tasks. Most neuropsychological tasks measure isolated mechanisms that do not necessarily represent complex and multi-component behavior typical of new and complex situations, and even realistic VR tasks lack the complexity, connectivity, dynamics, intransparency, and polytely (Funke, 2021) that underlie a microworld system. This simulation requires a series of cognitive operations that together achieve a final goal, and thus appears to be a good measure of a more general factor of executive control.

The microworld we have chosen is $C^3$Fire (see Figure 1), a complex and dynamic firefighting simulation whereby the aim is to save as many houses as possible (preventing them from catching alight or extinguishing them if they do). This is just one example of a microworld that we could have chosen, and the aim of the paper is to illustrate the usefulness of the microworld approach to assessing EF rather than validating any specific microworld per se. The context or surface characteristics of the chosen task were unimportant; the critical factor is that microworlds are high in construct validity and possess the same underlying characteristics that foster the use of higher-order executive processes (e.g., they are complex, opaque, dynamic; Brehmer & Dorner, 1993).

In C3Fire, participants must deploy firefighting units, whilst also controlling tank trucks to ensure that the units are provided with water. This simulation was chosen primarily because of its characteristics that are close to those features of real-world complex situations: a) it requires a series of decisions, b) these decisions are not independent, c) the state of the task changes, both autonomously and as a consequence of the actions taken by human, and d) the human has to act in real time (Brehmer, 1996; Gonzalez et al., 2005; Granlund, Johansson, Trnka, & Granlund, 2009). We extract three metrics of performance from the microworld, and assess whether these are predictive of performance on a selection of nine classical neuropsychological tasks (three tasks representing each of the three EFs of planning, WM and inhibition). If a single task is a good predictor of EF performance then this could prove a useful tool. For example, it could provide an initial assessment to indicate whether testing on the full battery of tasks would be necessary, or it could be used alongside traditional tests to augment the assessment process. Furthermore, if such a task is sensitive to individual differences in EF, it could be used more widely for applied issues (e.g., personnel selection) or to address theoretical questions (e.g., the correspondence between EF, intelligence, complex problem solving).

# Method

## Participants

The participants were 121 students from Université Laval in Quebec City, aged from 18 to 55 years (*n* = 77 women and *n* = 44 men; mean age = 25.03 years, *SD* = 7.75). They were paid nominal fees for their participation and were originally recruited through advertisements on the campus. All participants gave informed consent, and the research was approved by the Human Research Ethics Committee of Laval University (Comité d’Éthique et de la recherche de l’Université Laval, CÉRUL; approval number 2005-022/29-01-2007).

## Tasks

All tasks were computer-based and run on a desktop computer in a pre-randomized order.

### *Planning*

**TOL (based on Shallice, 1982).** The participant is presented on the screen with a starting arrangement of three different-colored balls on three pegs of decreasing heights. They are also shown a goal state arrangement that they must try to achieve by moving the balls from peg to peg using the mouse in the fewest moves possible. Two rules must be respected: (a) to move only one ball at a time and (b) not to put more balls on the peg than it is able to hold (i.e., one, two or three balls). The assessment of performance was based upon the speed of execution. This activity involved completing a series of 13 problems (3 practice trials and 10 experimental trials) and lasted about 15 minutes.

**Labyrinths (based on Porteus Maze Test; Porteus, 1965).** Ten labyrinths of increasing difficulty were presented to the participant, which were to be completed as quickly as possible without starting anew. Participants had to navigate the maze using the mouse from a starting point to the exit, and speed of execution was used as the dependent variable. This activity lasted about 10 minutes.

**WCST (Heaton et al., 1993).** Four reference cards showing colored shapes were displayed on the screen, below which a series of target cards were presented one at a time. The participant was required to categorize each target card by dragging and dropping them with the mouse to below the correct reference card, in accordance with one of the three following classification criteria: (a) color of the card stimuli (red, green, blue or yellow), (b) the number of stimuli on the card (1, 2, 3 or 4), or (c) the shape of the stimuli on the card (circle, cross, star or square). The participant was made aware that only one criterion would be used at a time to categorize each target card, and received feedback after each response. The classification criterion (e.g., color) remained the same until the participant gave the correct answer to eight consecutive classifications. The criterion was then changed (e.g., from color to number). The participant knew that the classification criterion of the cards changed during the activity, but did not know when. The dependent variable used in this task was the number of perseverations (i.e., the number of times the participant kept categorizing the cards according to the previous classification criterion).

### *Inhibition*

**Stroop (Golden, 1978; Stroop, 1935).** Participants were presented with a series of stimuli – either color-words or colored rectangles – to which they had to respond as quickly as possible, according to one of three conditions: a) color-naming condition (name the color of the rectangle presented), b) words condition (read the word presented), or c) interference condition (name the color of the ink in which an incongruent color-word was written, such as the word *YELLOW* written in red ink). The stimuli were presented in the center of the screen at a rate of one per second, and all related to one of four colors: blue, red, yellow or green. Time to read each list was recorded, and the dependent variable was the total time difference between interference and word conditions.

**CPT-II (Conners, 2000).** In this task, the participant had to press the space bar every time a letter appeared on the screen, except for the letter X. There were six blocks of 20 trials each, lasting 14 minutes in total. The dependent variable analyzed was the number of commissions (i.e., number of times the participant pressed the space bar following the appearance of the letter X).

**Stop signal (Logan, 1994).** This task consisted of two blocks of trials. In the first block, participants had to categorize presented words (by pressing one of two keys) as quickly and accurately as possible according to whether they were animals or non-animals (e.g., duck, gun). The length and frequency of words was controlled. In the second block, the participant completed the same categorization task but was instructed not to give a categorization response on trials when a signal was heard. Participants were told not to increase response time in order to wait for a potential signal, and were reminded of this if response time became too slow. The dependent variable of this activity was the number of times participants categorized the word at the same time that a signal occured (i.e., the number of times they responded when they should not have done).

### *Working Memory*

**Digit span (based on Wechsler Adult Intelligence Scale-III [WAIS-III], Digit span subtest; Wechsler, 1997).** This activity required the participant to recall backwards the serial order of digits presented. The level of difficulty increased, and the task ended when the participant made three consecutive mistakes. The dependent variable was backward digit span (i.e., the number of digits the participant succeeded in memorizing during two consecutive series).

**Spatial WM task (based on Owen et al., 1990).** In this task, the participant had to find the location of hidden tokens in an increasing number of boxes (i.e., 3, 4, 6 and 8 boxes). There was only one token in one of the boxes at any one time. Once a token was found, the box was emptied and the token concealed in a different box. The participant then started a new search – remembering which box(es) had already previously contained a token – until all boxes had yielded a token for that trial. An error was made if the participant returned to the same box to look for a token within the same search (item error), or returned to click on an already empty box within the same trial (attempt error). A result from one (perfect performance) to 37 (poor performance) was calculated to reflect the efficiency of the search strategy.

**Self-ordered-pointing task (SOPT; e.g., Petrides & Milner, 1982).** In this task, participants were shown a random arrangement of stimuli on the screen, and as many ‘pages’ (subsequent screens) as there were items on the screen (6, 8, 10 or 12). On each page, the items would be in a different arrangement and the participant had to select a different item on each page, therefore remembering those items already selected during that trial. The dependent variable is the number of items selected more than once; therefore, a perfect score of zero would indicate that the participant had selected a different item on each page.

### *Microworld*

The computer simulation chosen was *C³Fire* (Granlund et al., 2009; Granlund & Johansson, 2004; Hagemann & Kluge, 2017; Tremblay, Lafond et al., 2010; Tremblay, Vachon et al., 2010; Tremblay et al., 2012), which simulated a complex and dynamic firefighting activity (see Figure 1). The task was to save as many houses as possible according to three priorities: (a) to avoid houses catching alight, (b) to limit the spread of the fire, (c) to extinguish the burning houses. The participant had simultaneous control over two types of units: (a) four firefighter units used to extinguish fires and (b) four units of tank trucks used to provide the firefighters with water. Four types of distinct elements interacted in the course of the activity: the fire status, the geographical items, the meteorological conditions and the intervention units. The appearance of a new fire, and meteorological changes that alter the speed with which the fire spreads, were critical changes that occurred during the simulation. The secondary activities (e.g., providing the firefighters with water, detecting and extinguishing the fires) were numerous, non-routine activities that had to be completed in a simultaneous way. The simulation took place on a 40×40 matrix, as a scene viewed from above.

The task involves a number of higher-order processes necessary for successful performance, including managing multiple resources, initiating actions, planning/anticipation of future eventualities, co-ordination/prioritization of multiple activities, and the ability to achieve task relevant goals. Reflecting these, three dependent measures were taken: a) Time spent without water (time when a firefighting action could not be executed); b) monitoring effectiveness (total time spent when an action could have been initiated but was not, e.g., time the firefighter units had enough water supply to extinguish a fire but were inactive); and c) an overall performance measure. For this, actual performance on each scenario was compared to a hypothetical worst-case outcome (i.e., if no intervention had taken place at all). The proportion of saved cells (i.e., the number of saved cells divided by the number of cells that would have burned without intervention) and the proportion of saved houses (i.e., the number of saved houses divided by the number of houses that would have burnt with no intervention) were used to calculate the overall performance measure as follows:

$$\frac{\text{Proportion of saved cells} + \text{Proportion of saved houses}}{2}$$

## General Procedure

Testing took place in two 1.5-hour sessions, administered individually within a 2-week period. All tasks were adapted for computer administration and the order of task administration was pre-randomized. Participants were seated at a table in front of a computer screen, and an experimenter was seated to one side. Participants were allowed short breaks between tasks if they so wished.

# Results

## Transformations and Outlier Analysis

On the basis of overall between-subjects distributions of each task, we first identified any extreme outliers that were more than three standard deviations (*SDs*) from the mean and replaced them with a value that was three *SDs* from the mean. Thus, the scores for the outlying cases were still deviant, but not as extreme as they had been (Tabachnick & Fidell, 2007). No more than 2.02% of the observations were affected by these changes. As for the computer simulation ($C^3$Fire), three variables were generated by the program: time spent without water, monitoring effectiveness, and overall performance. From these three variables, six observations (representing 1.65% of all the observations) were outliers that were transformed to a value that was three *SDs* from the mean. If outliers were removed from the data set instead of being transformed, analyses yielded the same pattern of results despite the loss of four participants. Two randomly missing data points were also identified in the entire set of distributions. These were estimated using Expectation Maximization methods (EM).

EM forms a missing data correlation matrix by assuming the shape of a distribution (normal) for the partially missing data and basing inferences about missing values on the likelihood under that distribution. Such an operation did not affect the pattern of results. Finally, for 4 out of 12 distributions, we applied a square root transformation because these four distributions (i.e., TOL task, WCST, Stop-Signal task, and Spatial WM task) differed moderately from normal (see Tabachnick & Fidell, 2007). All of the RT and proportion correct measures achieved a satisfactory level of normality after this transformation process (see Table 1 for the skewness and kurtosis statistics). Unless stated otherwise, distribution transformations did not affect the obtained results. For ease of interpretation, all the dependent measures were standardized and adjusted so that larger numbers always indicated better performance. Item level data for the tasks were not collected and therefore we do not report reliabilities of the measures.

A summary of descriptive statistics for the 12 measures used in the analysis is presented in Table 1, and correlations among these measures are shown in Table 2. The correlations were generally low, consistent with the results from previous individual differences studies of EF. It is important to note, however, that the correlations among the 12 measures were not uniformly low; rather, measures associated with the computer simulation C$^3$Fire – that we believe taps a more general concept of EF – tended to show significant correlations with the other executive measures.

A canonical correlation analysis was carried out using the nine neuropsychological tasks as predictors of the three simulation performance metrics to evaluate the multivariate-shared relationship between the two variable sets: EFs and dynamic cognition. The analysis revealed three functions with canonical correlations of .776, .290, and .099 for each successive function. Collectively, the full model across all functions was statistically significant using the Wilks's $\lambda = .361$ criterion, $F(27, 318.98) = 4.927$, $p < .001$. Because

Wilks's $\lambda$ represents the variance unexplained by the model, $1 - \lambda$ yields the full model effect size in an $r^2$ metric. Thus, for the set of three canonical functions, the $r^2$ type effect size was .639, which indicates that the full model explained a substantial portion, about 64%, of the variance shared between the variable sets.

The dimension reduction analysis allows us to test the hierarchal arrangement of functions—i.e. pairs of canonical variates—for statistical significance. As noted, the full model (Functions 1 to 3) was statistically significant. However, Functions 2 to 3 and Function 3 failed to explain a statistically significant amount of shared variance between the variable sets ($F$s < 1, $p$s > .800). Given the effects for each function, only the first function was considered noteworthy in the context of this study (60.15% of shared variance). The two other functions only explained 8.43% and 0.99%, respectively, of the remaining variance in the variable sets after the extraction of the prior functions.

Table 3 presents the standardized canonical function coefficients, the structure coefficients ($r_s$), the squared structure coefficients ($r_s^2$), and the square canonical correlation ($R_c^2$) for the three functions. Following a convention in many factor analyses, structure coefficients above .45 are considered as the highest level of usefulness in the model (see Sherry & Henson, 2005). An examination of these structure coefficients for Function 1 revealed that relevant criterion variables were performance and monitoring effectiveness. This conclusion was supported by the squared structure coefficients. These two C$^3$Fire metrics were positively related to the function. Regarding the predictor variable set in Function 1, two of the planning tasks, namely TOL and Labyrinth, were the primary contributors to the predictor synthetic variable, with a secondary contribution by the Stroop and spatial WM tasks. Note that when the analysis was conducted on non-transformed data, there was a slight decrease in the usefulness of spatial WM in the model (structure coefficient went from .531 with transformation to .433 with no transformation). These neuropsychological tasks also tended to have the larger canonical function coefficients. A slight exception involved the Stroop task, which had modest function coefficients but relatively large structure coefficients. This result was due to the multicollinearity that this variable had with the three other best predictor variables. The four relevant neuropsychological tasks were positively related to the function.

These results, summarized in Figure 2, indicate that a substantial amount of variance in two dynamic cognition indices (performance and monitoring effectiveness) can be explained by a linear combination of the three main types of neuropsychological tasks, i.e. planning (TOL and Labyrinth), inhibition (Stroop), and WM (Spatial WM), with a larger contribution from the planning tasks.

## Discussion

The current paper reports an individual differences study in which we compared performance on a single computer simulation of complex problem-solving (C$^3$Fire microworld) with performance on nine neuropsychological tasks that are widely accepted within the EF literature. The microworld task, representing the single higher-order process of dynamic cognition, was a strong predictor of – i.e., accounting for over 60% of variation in – performance on the traditional EF measures (together representing three separate functions: Planning, Inhibition, and WM). The findings we present question the idea of fractionated EFs and multiple separate assessments, and instead provide support for the idea of a common higher-order function that underpins the notion of EF. We also make the novel suggestion that – if standardized for clinical use – the parsimony, ecological validity, and control afforded by a complex decision-making task such as C3Fire could make it a useful option for clinicians to consider. It could benefit neuropsychological testing either as 1) a single task; to save time and resources by administering an initial screening tool that taps into global aspects of EF, to determine whether a more extensive follow-up battery is necessary, or 2) as a complementary task; alongside a standard task battery to capture the dynamic nature of cognition and perhaps uncover deficits which may not come to light with traditional tests (e.g., a patient who exhibits problems with everyday functioning but presents with a normal profile on separate tests). Furthermore, the sensitivity of the microworld and the ability to capture global aspects of EF make it well placed as a task to help understand more about individual differences in EF, and relationships with other unitary concepts such as intelligence and complex problem-solving.

Our results are in keeping with prior studies that suggest a single underlying process may unite the disparate EFs that are currently proposed (Frias et al., 2006; Salthouse et al., 2003). Previous speculation has put forward that either the maintenance of goal and context information in WM, or some sort of inhibitory processes could represent the source of the commonality between EFs (Miyake et al., 2000); however, these explanations are more speculative and have not been empirically tested. Other work has suggested that the common higher-order executive process is closely linked to the *g* factor (e.g., see Duncan, 2010), a concept used to describe a broad range of abilities to control multi-step behavior in complex situations. Indeed, the *g* factor concept was first developed because many different psychometric tests of fluid intelligence all converge to the same general factor, and it seems reasonable that this logic could be applied to the concept of EF (see Boone et al., 1998; Engle et al., 1999). Frias et al. (2006) found poor executive functioning to relate to poor performance on an indicator of fluid intelligence, leading them to claim that low scores on executive tasks might be a proxy for low fluid intelligence and not a specific executive deficit per se. A direct comparison of both fluid intelligence and EF would improve the validity of each construct and allow one to evaluate whether or not the two are separate. In the current work, our novel proposal is that a higher-order integrative concept such as dynamic cognition (see also dynamic decision-making, Gonzalez et al., 2005; complex problem solving; Funke et al., 2018) – that assesses decision making in novel, complex, and dynamic environments – might capture the overlap between the two constructs of fluid intelligence and EF. Learning in real time dynamic decision-making tasks has been linked with general intelligence (Gonzalez, 2005b). Furthermore, complex problem solving has been suggested as the essence of intelligence (Hambrick et al., 2020), or even an ability more than intelligence (Funke, 2021). If intelligence alone is not sufficient to account for complex and dynamic problem-solving performance, perhaps a unique contribution is made by a global EF factor. The current work demonstrates the usefulness of dynamic cognition to encompass the essence of real-world decision making, bringing together several of the fractionated EF subcomponents into a more holistic concept, but further work is needed to understand the nature of higher-order unitary mechanisms and how they may be related (or separate) to one another.

The measures of dynamic cognition showed common variance with the nine neuropsychological tests, but some correlations were stronger than others. TOL, for example, was strongly related to the dynamic cognition measures. It was included here as a measure of planning, but has also been used to assess inhibition, problem solving, sustained attention, organization, impulsivity (Packwood et al., 2011), and so its multifaceted nature likely explains the stronger relationship with the holistic concept of dynamic cognition than that of simpler tasks assessing more isolated functions (e.g., CPT). Overall our results show that the abilities to manage multiple resources (as reflected by monitoring effectiveness) and to achieve task-relevant goals (as reflected by the performance index) in the simulation were positively related to capabilities in planning (as measured by TOL and Labyrinth tasks), inhibition (as indexed by the Stroop task), and WM (especially with Spatial WM). Although measures of dynamic cognition could predict a significant amount of variance of each on the three assessed EFs, they were particularly good at predicting individual differences in planning, which is often considered a more complex and multidimensional executive construct. While planning/TOL might be regarded as a good indicator of higher-level functioning due to its complexity and subsumption of other EFs, we argue that the microworld task is more complex and more holistic in nature, therefore making it even better placed to tap into a global executive factor.

The results support recent theoretical attempts to present a more holistic view of EFs (see Kimberg & Farah, 1993; Koechlin & Summerfield, 2007; Shallice & Burgess, 1996; Uttal, 2001). With at least 18 proposed EFs to be found in the literature (see Packwood et al., 2011), the numerous and different existing models demonstrate how difficult it is to incorporate all these potential functions into a meaningful and parsimonious model. So, in light of the current study, how might EF be conceptualized? We suggest that what have come to be referred to as different EFs are perhaps not separate functions at all, but could simply reflect task-specific manifestations of behaviors that all rely upon the same higher-order process. Such a process would be responsible for the acquisition of task context and the implementation of rules used to guide behavior, no matter what these behaviors are (e.g., inhibition of a prepotent response, planning, set-shifting, etc.) or what response is required by a specific task. Our cognitive system is economical and solves everyday problems using binding and heuristics. For instance, when facing a complex problem for which we have incomplete information, we can use simple and efficient rules that allow for a “good enough” solution in a shorter period of time. EF is about adaptation based on basic cognitive processes such as the capacity to retain information online, to focus attention, and to recollect information from long-term memory. When all these processes interact, different strategies emerge: that is the essence of EF. In fact, the capacity to develop an efficient strategy to be used in a complex and novel situation might be the common point between the multitude of EF components that have been generated from the concept. A microworld, with all the features that it involves, might be a task that can capture this very capacity.

The C$^3$Fire microworld (and other functional simulations of real-world tasks) possesses several notable features that make it likely to capture the global essence of everyday executive functioning in a way that classical neuropsychological tasks fail to do. Microworlds involve interrelated decision making and problem solving whereby, for example, a strategy or decision chosen at the beginning of the task will have a direct impact on the state of the problem later on and the subsequent decisions that must be made. This feature is strongly related to everyday decision making, yet most of the classic neuropsychological tasks involve only a series of isolated and predetermined trials that are not related to each other. For instance, the nature and complexity of trials in the TOL are predetermined and will not change according to the participant's performance/strategy in earlier trials. Microworlds also evolve without the participant's intervention, just as situations continue to change around us in daily life. If we decide not to respond to a problem we face at Time A, then this same problem at Time B may have worsened, e.g., a student who has not completed their homework on day one, may have double the amount of homework to do on day two. Finally, microworlds evolve in real time, unlike some classic neuropsychological tasks that do not take into account the time scale of problem solving (e.g., WCST). Classic neuropsychological tasks that do not include all of these features are less likely to provide an accurate profile of EF, and would explain why some participants perform well on isolated executive tasks yet are impaired when it comes to general functioning in everyday life that necessitates this kind of adaptive behavior (Andrés, 2003; Andrés & van der Linden, 2002). Developing a measure that includes all these features in a single task is not easy, but the results of the current study suggest that a microworld task could be very promising in capturing the complexity of our executive system. While we are of course not suggesting that classic methods of assessment be replaced, we propose that a relatively quick microworld task could provide an initial snapshot of functioning which could supplement well-established tasks and aid a more global assessment.

While C$^3$Fire correlated well with the traditional neuropsychological tasks, we would expect a similar pattern of results had any other microworld been used (e.g., using the task context of any other social, political, or economic system). Rather than the surface characteristics, it is the underlying structure of the task that gives rise to complex problem-solving behaviors that correlate well with the various individual measures of EF. What we demonstrate in this study is the benefit of this *type* of task, and a ‘wish list’ of characteristics that such a task should incorporate in order to tap into a global EF component. Neuropsychologists can use this set of microworld characteristics to develop and validate a task for clinical testing, the surface context of which could be any scenario deemed to be most ecologically appealing.

Another key strength of pursuing the idea of dynamic cognition (and other comparable holistic concepts) is the expansion beyond neuropsychology to the fields of human factors and cognitive psychology, and the innovation in methodology that cross-disciplinary collaboration affords (Schubert & Rey-Mermet, 2019). This is particularly the case given the ‘technology crisis’ in neuropsychology, whereby it is recognized that methods of assessment are becoming outdated and less efficient than they could be (Miller & Barr, 2017; Schmand, 2019). In recent years, there has been a growing trend to update the traditional neuropsychological desktop tasks for computerized versions, but sometimes these tasks are still using stimuli developed 100 or so years ago (e.g., list learning, picture completion) and could appear outdated. Furthermore, even computerized versions of traditional tests do not offer the degree of task engagement or ‘fun factor’ afforded by more realistic microworld tasks, which can increase willingness to engage in testing and ensure that boredom/fatigue is less of an issue than with a lengthy test battery. Moreover, the immersive nature of scenarios enhances ecological validity and provides a better assessment of naturalistic problem-solving skills than tasks that appear artificial and are far removed from everyday problems. Computerized testing allows for more uniform delivery, greater precision and ease of scoring, as well as remote and portable testing. For researchers wishing to assess EF in a non-clinical population, the microworld task offers even greater advantages for testing in terms of the flexibility of scripting scenarios. While practice effects can be a problem for some traditional tests, microworlds can use many different scenarios, allowing for multiple assessments with no learning of ‘correct’ responses. The flexibility also means that different levels of difficulty can be implemented, and difficulty can be adapted to the level of the individual. There is the possibility to inject events that change the course of action and, in turn, require an adaptation and/or a change of strategy. Such changes to the scenarios could be personalized so that one could stress a particular capability (e.g., the capability to adapt or to plan). Like other experimental tasks, dynamic cognition measures are normally distributed and likely more sensitive to assessing differences in EF within a general population. In short, measuring EF through a complex, dynamic, and immersive functional simulation of real-world decision-making offers many practical benefits to testing for both clinicians and researchers, in addition to the conceptual advantages of a holistic approach.

## Summary

Our study suggests that in adopting a more holistic concept it is possible to propose a single complex task that is predictive of performance on diverse neuropsychological tasks frequently used to measure EFs. This lends support to studies advocating the idea that EF performance is supported by a single-factor, or at least by a few very strongly correlated factors. Quite what that factor is – *g*, WM capacity, dynamic cognition – or the relationship between them, is a matter for further investigation. In this regard, neuroimaging would be invaluable to localizing shared neural networks activated when performing tests of fluid intelligence, WM, and dynamic cognition, and could help to identify the source of the commonality between the constructs. Nevertheless, the novel application of human factors/cognitive psychology to neuropsychological testing can broaden our conceptual thinking, as well as suggesting innovative and parsimonious methods that could complement traditional test batteries. Beyond the conceptual debate, we present a single task that captures what a multitude of tests measure. A microworld such as $C^3$Fire could provide a relatively quick indication of a potential executive deficit to decide whether or not further time and resources are needed to investigate a more detailed EF profile. Furthermore, the flexibility to develop such a task and personalize scenarios to the individual could have benefits for both clinicians looking to pinpoint specific deficits, and researchers assessing individual differences within a general population.

## Acknowledgements

This research was supported by an operating grant to Sébastien Tremblay from the Natural Sciences and Engineering Research Council of Canada (NSERC).

## Disclosure of interest

The authors report there are no competing interests to declare.

# References

Alderman, N., Burgess, P. W., Knight, C., & Henman, C. (2003). Ecological validity of a simplified version of the multiple errands shopping test. *Journal of the International Neuropsychological Society, 9,* 32-44.

Andrés, P. (2003). Frontal cortex as the central executive of working memory: time to revise our view. *Cortex, 39*, 871-895.

Andrés, P. & Van der Linden, M. (2002). Are central executive functions working in patients with focal frontal lesions? *Neuropsychologia, 40*, 835-845.

Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology, 49A*, 5-28.

Baggetta, P., & Alexander, P. A. (2016). Conceptualization and operationalization of executive function. Mind, Brain, and Education, 10, 10–33. http://dx.doi.org/10.1111/mbe.12100

Banich, M. T. (2009). Executive function: The search for an integrated account. *Current Directions in Psychological Science, 18,* 89-94.

Bentler, P. M. (2006). *EQS 6 Structural equations program manual.* Multivariate Software, Inc., Encino, CA.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Boone, K. B., Ponton, M. O., Gorsuch, R. L., Gonzalez, J. J., & Miller, B. L. (1998). Factor Analysis of four measures of Prefrontal Lobe functioning. *Archives of Clinical Neuropsychology, 13,* 585-595.

Borgnis, F., Baglio, F., Pedroli, E., Rossetto, F., Uccellatore, L., Oliveira, J.A.G., Riva, G., & Cipresso, P. (2022). Available Virtual Reality-Based Tools for Executive Functions: A Systematic Review. *Frontiers in Psychology, 13,* 833136.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica*, 81, 211-241.

Brehmer, B. (1996). Man as stabilizer of systems: From static snapshots of judgement processes to dynamic decision making. *Thinking and Reasoning, 2,* 225-238.

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior, 9,* 171–184. https://doi.org/10.1016/0747-5632(93)90005-D

Busemeyer, J. & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making. *Psychological Review, 100,* 432-459.

Chicchi Giglioli, I. A., Bermejo Vidal, C., and Alcañiz Raya, M. (2019). A virtual versus an augmented reality cooking task based-tools: a behavioral and physiological study on the assessment of executive functions. *Frontiers in Psychology, 10,* 2529. doi:10.3389/fpsyg.2019.02529

Conners, C. K. (2000). *Conners Continuous Performance Test*. Canada : Multi-Health Systems Inc.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D., & Minkoff, S. (2002). A latent variable analysis of working memory capacity, short term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163-183.

Danner, D., Hagemann, D., Holt, D. V., Hager, M., Schankin, A., Wüstenberg, S., et al. (2011). Measuring performance in dynamic decision making. Reliability and validity of the Tailorshop simulation. *Journal of Individual Differences, 32,* 225–233. doi: 10.1027/1614-0001/a000055

Dörner, D., & Güss, C. D. (2022). Human error in complex problem solving and dynamic decision making: A taxonomy of 24 errors and a theory. *Computers in Human Behavior Reports, 100222*.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences, 14*, 172-179.

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology, 30*, 257-303.

Duncan, J., Johnson, R., Swales, M., & Freer, C. (1997). Frontal lobe deficits after head injury: Unity and diversity of function. *Cognitive Neuropsychology, 14*, 713-741.

Duncan, J. & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences, 23*, 475-483.

Edwards, W. (1962). Dynamic decision theory and probabilistic information processing. *Human Factors*, *4*, 59-73.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102-134). New York: Cambridge University Press.

Fortin-Côté, A. et al (2018). Predicting Video Game Players' Fun from Physiological and Behavioural Data. In *Future of Information and Communication Conference*.

Fournier-Vicente, S., Larigauderie, P. & Gaonac'h, D. (2008). More dissociations and interactions within central executive functioning : A comprehensive latent-variable analysis. *Acta Psychologica*, *129*, 32–48.

Frias, C. M., Dixon, R. A., & Strauss, E. (2006). Structure of four executive functioning tests in healthy older adults. *Neuropsychology, 20,* 206-214.

Friedman, D., Nessler, D., Johnson, R. Jr., Ritter, W., & Bersick, M. (2008). Age-related changes in executive function: an event-related potential (ERP) investigation of task-switching. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology, and Cognition, 15*, 95-128.

Funke, J., Fischer, A., & Holt, D. V. (2018). Competencies for complexity: Problem solving in the 21st century. In E. Care, P. Griffin, & M. Wilson (Eds.), Assessment and teaching of 21st century skills: Research and applications (pp. 41–53). Dordrecht, the Netherlands: Springer.

Funke, J. (2021). It Requires More Than Intelligence to Solve Consequential World Problems. *Journal of Intelligence, 9,* 38.

Golden, C. J. (1978). *Stroop Color and Word Test.* Chicago: Stoelting.

Goldman-Rakic, P. (1987). Circuitry of primate prefrontal cortex and regulation of behaviour by representational knowledge. In F. Plum & V. Mountcastle (Eds.), *Handbook of Physiology* (pp. 373-417). Bethesda, MD: American Physiological Society.

Gonzalez, C. (2005a). Decision support for real-time, dynamic decision-making tasks. *Organizational Behavior and Human Decision Processes, 96,* 142–154.

Gonzalez, C. (2005b). Task workload and cognitive abilities in dynamic decision making. *Human Factors, 47,* 92-101.

Gonzalez, C., Vanyukov, P., & Martin, M.K. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior, 21,* 273–286.

Granlund, R., Johansson, B., Trnka, J., & Granlund, H. (2009). Investigating the impact of geographical information system in a municipality crisis management organization.

*Proceedings of the 9th International Conference on Naturalistic Decision Making, London, UK.*

Granlund, R., & Johansson, B. (2004). Monitoring distributed collaboration in the $C^3$ Fire microworld. Dans S. G. Schifflet, L. R. Elliott, E. Salas, & M. D. Coovert (Eds.), *Scaled Worlds: Development, Validation and Application* (pp.37-48). Cornwall, GB: Ashgate.

Gray, W.D. (2002). Simulated task environments: the role of high-fidelity simulations, scaled worlds, synthetic environments, and laboratory tasks in basic and applied cognitive research. *Cognitive Science Quarterly, 2,* 205–227.

Hagemann, V., & Kluge, A. (2017). Complex Problem Solving in Teams: The Impact of Collective Orientation on Team Process Demands. Frontiers in Psychology, 8, 1730. doi: 10.3389/fpsyg.2017.01730

Hambrick, D., Burgoyne, A., & Altmann, E. (2020). Problem-solving and intelligence. In R. J. Sternberg (Ed.), *Cambridge handbook of intelligence* (2nd ed.), pp. 553–579. Cambridge University Press.

Heaton, R. K., Chelune, G. K., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual: Revised and expanded*. Odessa, FL: Psychological Assessment Resources Inc.

Hu, L. T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76-99). Thousand Oaks, CA: Sage.

Hu, L. T., Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424-453.

Hull, R., Martin, R. C., Beier, M. E., Lane, D., & Hamilton , A. C. (2008). Executive Function in Older Adults: A Structural Equation Modeling Approach. *Neuropsychology, 22,* 508–522.

Iverson, G.L., Karr, J. E., Terry, D.P., Garcia-Barrera, M. A., Holdnack, J.A., Ivins, B.J., & Silverberg, N.D. (2020). Developing an executive functioning composite score for research and clinical trials. *Archives of Clinical Neuropsychology, 35(3),* 312–325. doi: 10.1093/arclin/acz070.

Jansari A., Devlin A., Agnew R., Akesson K., Murphy L., Leadbetter T. (2014). Ecological assessment of executive functions: A new virtual reality paradigm. *Brain Impairment 15,* 71–87. Doi: 10.1017/BrImp.2014.14

Jobidon, M., Turcotte, I., Aube, C., Labrecque, A., Kelsey, S. and Tremblay, S. (2017). Role variability in self-organising teams working in crisis management. *Small Group Research, 48,* 62-92.

Jurado, M. B. & Rosselli, M. (2007). The elusive nature of executive functions: a review of our current understanding. *Neuropsychological Review*, *17,* 213-233.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working memory capacity. *Journal of Experimental Psychology: General*, *130*, 169-183.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychonomic Bulletin & Review*, *9*, 637-671.

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and reanalysis of latent variable studies. *Psychological Bulletin*, *144,* 1147–1185.

Kimberg, D. Y. & Farah, M. J. (1993). A unified account of cognitive impairments following frontal lobe damage: The role of working memory in complex organized behaviour. *Journal of Experimental Psychology: General*, *122*, 411-428.

Koechlin, E. & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences, 11*, 229-235.

Kovacs, K., & Conway, A. R. A. (2019). A unified cognitive/differential approach to human intelligence: Implications for IQ testing. *Journal of Applied Research on Memory and Cognition, 8,* 255–272.

Lafond, D., DuCharme, M., Gagnon, J., & Tremblay, S. (2012). Support requirements for cognitive readiness in complex operations. *Journal of Cognitive Engineering and Decision Making, 6,* 393–426.

Lezak, M. D. (1995). *Neuropsychological Assessment* (3rd ed.). New York, N.Y. Oxford University Press.

Logan, G. D. (1994). On the ability to inhibit thought and action: A user's guide to the stop signal paradigm. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 189-239). San Diego, CA: Academic Press.

Long, D. L., & Prat, C. S. (2002). Working memory and Stroop interference: An individual differences investigation. *Memory & Cognition, 30*, 294-301.

McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology, 24,* 222–243. doi:10.1037/ a0017619

Miller, J. B., & Barr, W. B. (2017). The technology crisis in neuropsychology. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists, 32(5),* 541–554. doi:10.1093/arclin/acx050

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex frontal lobes tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49-100.

Miyake, A., & Shah, P. (1999). Toward unified theories of working memory: Emerging general consensus, unresolved theoretical issues, and future research directions. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 442-481). New York: Cambridge University Press.

Nir-Hadad, S. Y., Weiss, P. L., Waizman, A., Schwartz, N., and Kizony, R. (2017). A virtual shopping task for the assessment of executive functions: Validity for people with stroke. *Neuropsychology Rehabilitation, 27,* 808–833. doi: 10.1080/09602011.2015.1109523

Norris, G. & Tate, R. L. (2000). The Behavioral Assessment of the Dysexecutive Syndrome (BADS): ecological, concurrent and construct validity. *Neuropsychological Rehabilitation, 10,* 33-45.

Owen, A.M., Downes, J.J., Sahakian, B.J., Polkey, C.E., & Robbins, T.W. (1990): Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia 28,*1021–1034.

Packwood, S., Hodgetts, H. M., & Tremblay, S. (2011). A multiperspective approach to the conceptualization of executive functions. *Journal of Clinical and Experimental Neuropsychology, 24,* 1-15.

Petrides, M., & Milner, B. (1982). Deficits on subject-ordered tasks after frontal and temporal-lobe lesions in man. *Neuropsychologia, 20,* 249-262.

Porteus, S. D. (1965). *Porteus Maze Test: Fifty years Application*. Palo Alto, CA. Pacific Books.

Rabbitt, P. (Ed.). (1997). *Methodology of frontal and executive function*. Hove, UK: Psychology Press.

Rapoport, A. (1975). Research paradigms for the study of dynamic decision behavior. In D. Wendt and C. Vlek (Eds.), *Utility, probability and human decision making.* Dordrecht: Reidel.

Rand,D., Rukan, S. B. A.,Weiss, P. L., and Katz,N. (2009). Validation of the virtual MET as an assessment tool for executive functions. *Neuropsychology Rehabilitation, 19,* 583–602. doi: 10.1080/09602010802469074

Ratcliff, R., & McKoon, G. (2022). Can neuropsychological testing be improved with model-based approaches? *Trends in Cognitive Sciences, 26(11),* 899–901. https://doi.org/10.1016/j.tics.2022.08.015

Renison B., Ponsford J., Testa R., Richardson B., & Brownfield K. (2012). The ecological and construct validity of a newly developed measure of executive function: The virtual library task. *Journal of the International Neuropsycholical Society, 18,* 440–450. 10.1017/S1355617711001883

Rigas, G., Carling, E., & Brehmer, B. (2002). Reliability and validity of performance measures in microworlds. *Intelligence, 30,* 463–480.

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., Lawrence, A. D., McInnes, L., & Rabbitt, P. M. A. (1998). A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers : Implications for theories of executive functioning and cognitive aging. *Journal of the international neuropsychological society, 4,* 474-490.

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of experimental psychology-General, 132*, 566-594.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of a structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*, 23-74.Schmand, B. (2019)

Why are neuropsychologists so reluctant to embrace modern assessment techniques? *The Clinical Neuropsychologist, 33:2,* 209-219. DOI:10.1080/13854046.2018.1523468

Schubert, A.-L., & Rey-Mermet, A. (2019). Does process overlap theory replace the issues of general intelligence with the issues of attentional control? *Journal of Applied Research in Memory and Cognition, 8(3),* 277-283. DOI: https://doi.org/10.1016/j.jarmac.2019.06.004

Sergeant, J. A., Geurts, H., & Oosterlaan, J. (2002). How specific is a deficit if executive functioning for attention-deficit/hyperactivity disorder? *Behavioural brain research, 130,* 3-28.

Shallice, T. (1982). Specific impairment of planning. *Philophical Transaction of the Royal Society of London, B* (298), 199-209.

Shallice, T. (1988). *From Neuropsychology to Mental Structure*. New York: Cambridge University Press.

Shallice, T. & Burgess, P.W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain, 114,* 727– 741.

Shallice, T. & Burgess, P. (1996). The domain of supervisory processes and temporal organization of behaviour. *Philosophical Transactions of the Royal Society of London, 351*, 1405-1411.

Sherry, A., & Henson, R. K. (2005). Conducting and interpreting canonical correlation analysis in personality research: A user-friendly primer. *Journal of Personality Assessment, 84,* 37–48.

Slomine, B. S., Gerring, J. P., Grados, M. A., Vasa, R., Brady, K. D., Christensen, J. R., & Denckla, M. B. (2002). Performance on measures of "executive function" following pediatric traumatic brain injury. *Brain Injury, 16,* 759-772.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643-662.

Stuss, D. T., Floden, D., Alexander, M. P., Levine, B., & Katz, D. (2001). Stroop performance in focal lesion patients: Dissociation of processes and frontal lobe lesion location. *Neuropsychologia, 39*, 771-786.

Suchy, Y. (2009). Executive functioning: Overview, assessment, and research issues for non-neuropsychologists. *Annals of Behavioral Medicine, 37,* 106 –116. doi:10.1007/s12160-009-9097

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Mutivariate Statistics* (5$^{e}$ éd.). Boston, MA: Pearson Education, Inc.

Tremblay, S., Lafond, D., Gagnon, J.-F., Rousseau, V., & Granlund, R. (2010). Extending the capabilities of the C3Fire microworld as a testing platform for emergency response management. *Proceedings of the 7th International ISCRAM Conference*, Seattle.

Tremblay, S., Vachon, F., Lafond, D., & Hodgetts, H. M. (2010). Does teaming up make you less vulnerable to task interruption? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 54*, 1605-1609.

Tremblay, S., Vachon, F., Lafond, D., & Kramer, C. (2012). Dealing with task interruptions in complex dynamic environments: Are two heads better than one? *Human Factors, 54,* 70-83.

Uttal, W. R. (2001). *The new Phrenology: The limits of Localizing Cognitive Processes in the Brain.* Massachusetts: The MIT Press.

Wechsler, D. (1997). *WAIS-III administration and scoring manual.* The Psychological Corporation, San Antonio, Tx.

Young, S., Morris, R., Toone, B., & Tyson, C. (2007). Planning ability in adults with Attention-Deficit/Hyperactivity Disorder. *Neuropsychology, 21*, 581-589.

**Table 1**

*Descriptive Statistics for the Dependent Measures (N = 121)*

| Task | Mean (*SD*) | Range | Skewness | Kurtosis |
|---|---|---|---|---|
| TOL | 20.73 (4.02) | 13.31 to 31.35 | .69 | .22 |
| Labyrinth | 207.52 s (66.47) | 93.16 to 406.53 | .87 | .69 |
| WCST | 2.41 (.47) | 1.41 to 3.61 | .73 | .15 |
| Stroop | 51.45 s (15.06) | 23 to 96 | .70 | .49 |
| CPT-II | 13.64 (7.14) | 1 to 34 | .58 | .08 |
| Stop-Signal | 1.63 (.94) | 0 to 3.74 | .31 | .08 |
| Digit Span | 4.26 (1.2) | 2 to 7 | .58 | -.23 |
| Spatial WM | 2.75 (1.97) | 0 to 7 | .29 | -.60 |
| SOPT | 1.36 (0.79) | .33 to 3.33 | .56 | -.39 |
| C$^3$Fire - Performance | 0.49 (0.13) | 0.13 to 0.73 | -.51 | -.05 |
| C$^3$Fire – Time spent empty | 1134.70 (141.64) | 711 to 1544 | -.13 | .27 |
| C$^3$Fire – Monitoring effectiveness | 657.85 | 276 to 1356 | 1.19 | 2.05 |

*Note*. The data analysis used square root transformation for TOL, WCST, Stop-Signal, and Spatial WM tasks. Total of errors was used for Stop-Signal, Spatial WM, and SOPT tasks. Number of perseverative responses was used for WCST. Total time was used for TOL and Labyrinth tasks. Number of commissions was used for CPT-II. Backward digit span was used for Digit Span task. The total time difference between interference and word conditions was used for Stroop task.

**Table 2**

*Pearson Correlation Coefficients for the ten measures (N = 121)*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. TOL | _ | | | | | | | | | | |
| 2. Labyrinth | .469* | _ | | | | | | | | | |
| 3. WCST | .101 | .215* | _ | | | | | | | | |
| 4. Stroop | .397* | .356* | .097 | _ | | | | | | | |
| 5. CPT-II | -.199* | -.142 | -.124 | .075 | _ | | | | | | |
| 6. Stop-Signal | .098 | .078 | -.018 | .195* | -.023 | _ | | | | | |
| 7. Digit Span | .120 | .162 | .134 | .168 | .035 | -.023 | _ | | | | |
| 8. Spatial WM | .255* | .249* | .164 | .268* | .057 | .202* | .073 | _ | | | |
| 9. SOPT | .210* | .174 | .157 | .155 | .177 | .161 | .250* | .114 | _ | | |
| 10. C$^3$Fire Perf. | .554* | .484* | .307* | .337* | -.234* | .145 | .193* | .362* | .220* | __ | |
| 11. C$^3$Fire TSE | .124 | .003 | .091 | -.013 | -.124 | .023 | .016 | -.012 | -.092 | .137 | _ |
| 12. C$^3$Fire ME | .470* | .573* | .199* | .385* | -.121 | .105 | .158 | .368* | .284* | .569* | -.346* |

*Note*. TOL, Tower of London; WCST, Wisconsin Card Sorting Test; CPT-II, Continuous Performance Test-II; Spatial WM, Spatial Working Memory; SOPT, Self-Ordered Pointing Task; MW C$^3$Fire, Microworld C$^3$Fire.

*$p < .05$.

**Table 3**

*Canonical solution for executive functions predicting dynamic cognition for each of the three functions*

| | Function 1 | | | Function 2 | | | Function 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | Coef | $r_s$ | $r_s^2$ (%) | Coef | $r_s$ | $r_s^2$ (%) | Coef | $r_s$ | $r_s^2$ (%) |
| C$^3$Fire set | | | | | | | | | |
| Performance index | 0.447 | .875 | 76.56 | -0.833 | -.369 | 13.51 | 0.961 | .313 | 9.80 |
| Time spent empty | 0.265 | .088 | 0.08 | -0.311 | -.732 | 53.58 | -1.109 | -.676 | 45.70 |
| Monitoring effectiveness | 0.688 | .851 | 72.40 | 0.889 | .522 | 27.25 | -0.874 | .057 | 0.32 |
| $R_c^2$ | | | 60.15 | | | 8.43 | | | 0.99 |
| Executive functions set | | | | | | | | | |
| *Planning* – TOL | 0.380 | .778 | 61.34 | -0.685 | -.283 | 8.00 | -0.200 | -.155 | 2.40 |
| *Planning* – Labyrinth | 0.406 | .789 | 62.21 | 0.675 | .361 | 13.03 | -0.618 | -.388 | 15.05 |
| *Planning* – WCST | 0.151 | .385 | 14.79 | -0.507 | -.368 | 13.54 | 0.137 | .206 | 4.24 |
| *Inhibition* – Stroop | 0.131 | .531 | 28.23 | 0.278 | .226 | 11.80 | 0.024 | .014 | 0.02 |
| *Inhibition* – CPT-II | -0.182 | -.283 | 8.04 | 0.228 | .436 | 19.01 | -0.085 | .180 | 3.24 |
| *Inhibition* – Stop signal | 0.016 | .185 | 3.43 | -0.264 | -.119 | 1.42 | 0.088 | .224 | 5.02 |
| *WM* – Digit span | 0.056 | .256 | 6.58 | -0.228 | -.088 | 0.77 | 0.200 | .299 | 8.94 |
| *WM* – Spatial WM | 0.261 | .531 | 28.25 | 0.122 | .100 | 1.00 | 0.468 | .397 | 15.76 |
| *WM* – SOPT | 0.139 | .348 | 12.10 | 0.445 | .337 | 11.36 | 0.682 | .661 | 43.69 |

*Note.* Structure coefficient ($r_s$) greater than |.45| are underlined only for significant functions. Coef = standardized canonical function coefficient; $r_s$ = structure coefficient; $r_s^2$ = squared structure coefficient; $R_c^2$ = square canonical correlation.

**Figure 1**

02:25:00

Coordonateur de l'informatio...

Unit Info

| ID | Pos | GoTo | Activity | Wat | Fuel |
|---|---|---|---|---|---|
| F1 | F, 8 | | Refill Fuel | 30 | 80,00 |
| F2 | S, 14 | | Inactive | 30 | ,00 |
| F3 | V, 14 | | Inactive | 18 | 90,00 |
| B4 | I, 24 | | Inactive | 0 | 1,50 |
| B5 | R, 7 | | Inactive | 0 | ,00 |
| B6 | Q, 3 | | Inactive | 0 | 97,00 |
| S7 | Q, 18 | | Inactive | 0 | 1 |
| S8 | U, 9 | | Inactive | 0 | 1 |
| S9 | S, 19 | S, 22 | Moving | 0 | 1 |
| T10 | W, 18 | | Inactive | 0 | ,00 |
| T11 | M, 18 | | Inactive | 0 | ,00 |
| T12 | AM, 5 | Z, 9 | Moving | 0 | 100,00 |
| W13 | K, 16 | | Inactive | 100 | ,00 |
| W14 | Q, 15 | | Inactive | 100 | 15,00 |
| W15 | H, 16 | | Inactive | 100 | ,00 |
| O16 | F, 7 | | Tapping F... | 0 | 10,00 |
| O17 | X, 5 | AN, 6 | Moving | 0 | ,00 |
| O18 | AN, 5 | | Inactive | 0 | ,00 |

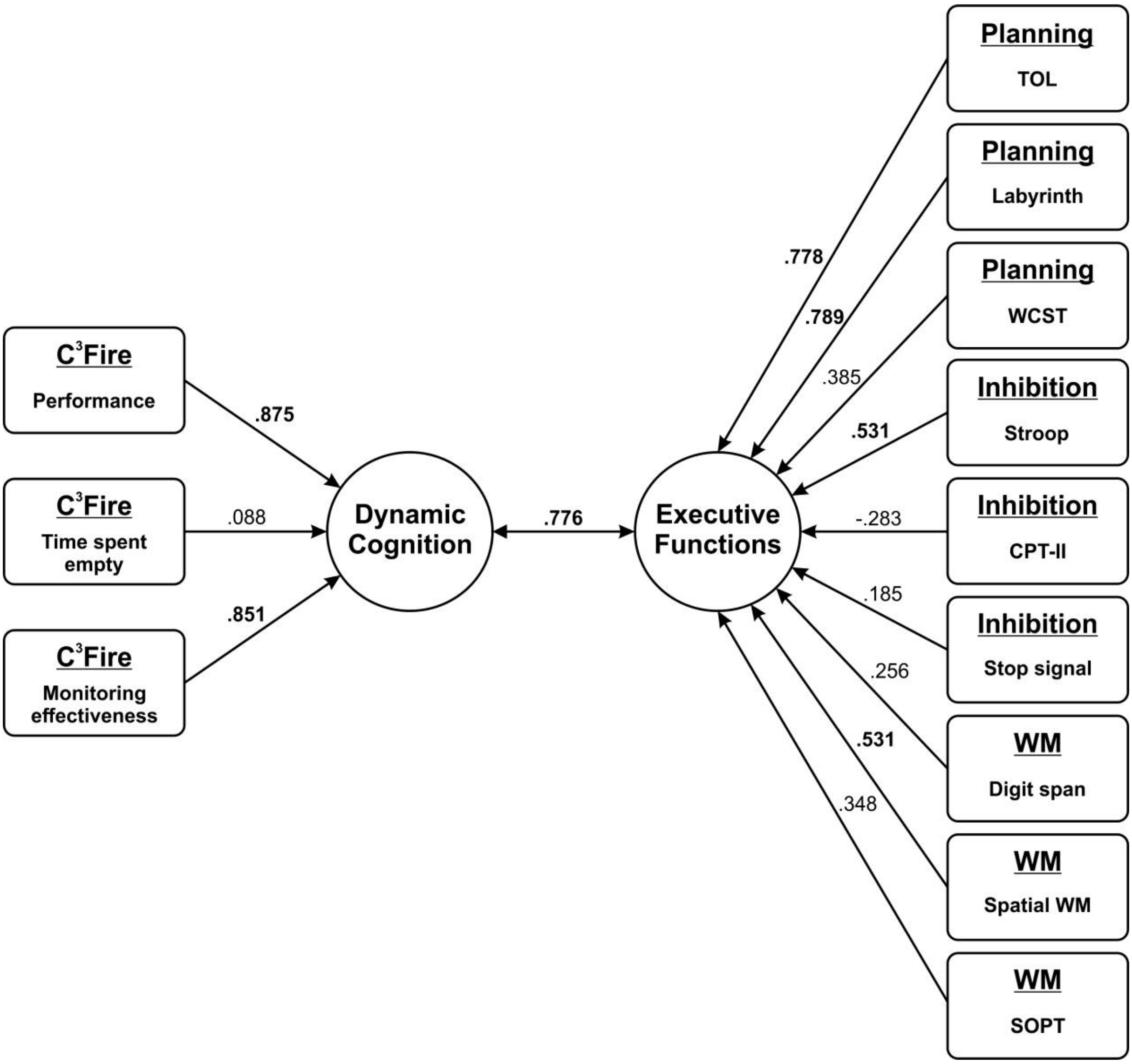
Viewer

Time: 02:16:08 From: S9 (7)

```
------------------------------------
Time: 02:16:08  From: S9
L4 - 3 personne(s).
```

Send To: XP

**Figure 2**


C$^3$Fire
Performance
C$^3$Fire
Time spent empty
C$^3$Fire
Monitoring effectiveness
.875
.088
.851
Dynamic Cognition
.776
Executive Functions
Planning
TOL
Planning
Labyrinth
Planning
WCST
Inhibition
Stroop
Inhibition
CPT-II
Inhibition
Stop signal
WM
Digit span
WM
Spatial WM
WM
SOPT
.778
.789
.385
.531
-.283
.185
.256
.531
.348

# Figure captions

**Figure 1.** Screen capture of Microworld C$^3$Fire.

**Figure 2.** Relationships among predictors and predictor synthetic variables for the first canonical function. Correlations greater than |.45| are in bold.