Journal of Experimental Psychology:
Learning, Memory, and Cognition
2018, Vol. 44, No. 5, 764–778

0278-7393/18/$12.00 http://dx.doi.org/10.1037/xlm0000487

How do Task Characteristics Affect Learning and Performance? The Roles of Variably Mapped and Dynamic Tasks

Brooke N. Macnamara and David J. Frank
Case Western Reserve University

For well over a century, scientists have investigated individual differences in performance. The majority of studies have focused on either differences in practice, or differences in cognitive resources. However, the predictive ability of either practice or cognitive resources varies considerably across tasks. We are the first to examine task characteristics' impact on learning and performance in a complex task while controlling for other task characteristics. In 2 experiments we test key theoretical task characteristic thought to moderate the relationship between practice, cognitive resources, and performance. We devised a task where each of several key task characteristics can be manipulated independently. Participants played 5 rounds of a game similar to the popular tower defense videogame Plants vs. Zombies where both cognitive load and game characteristics were manipulated. In Experiment 1, participants either played a consistently mapped version—the stimuli and the associated meaning of their properties were constant across the 5 rounds—or played a variably mapped version—the stimuli and the associated meaning of their properties changed every few minutes. In Experiment 2, participants either played a static version—that is, turn taking with no time pressure—or played a dynamic version—that is, the stimuli moved regardless of participants' response rates. In Experiment 1, participants' accuracy and efficiency were substantially hindered in the variably mapped conditions. In Experiment 2, learning and performance accuracy were hindered in the dynamic conditions, especially when under cognitive load. Our results suggest that task characteristics impact the relative importance of cognitive resources and practice on predicting learning and performance.

Keywords: task characteristics, learning, skill acquisition, practice, expertise

Supplemental materials: http://dx.doi.org/10.1037/xlm0000487.supp

Why do some people perform better than others when completing the same real-world task? Your answer to this question likely reflects your theoretical background. For example, psychologist who label themselves "expertise researchers," tend focus on differences in the amount of accumulated practice to explain variation in performance (e.g., Ericsson, Krampe, & Tesch-Römer, 1993). By contrast, those who label themselves "individual differences researchers," "intelligence researchers," or "working memory researchers" tend to focus on individual differences in cognitive resources as explanations for individual variation (e.g., Hunter & Hunter, 1984; McVay & Kane, 2012; Wai, 2014). Yet both practice-performance relationships (average $r^2 = .12$; Macnamara, Hambrick, & Oswald, 2014) and abilities-performance relationships (average $r^2 = .28$; Hunter & Hunter, 1984) for real-world tasks tend to be relatively small. That is, both the expertise and individual differences literatures leave the majority of interindividual variance in real-world performance unexplained.

Even more importantly, there is considerable heterogeneity in the predictive power of both practice and cognitive ability across studies (e.g., Hunter & Hunter, 1984; Macnamara et al., 2014; Salgado et al., 2003). We propose that the general lack of predictive power in both expertise and individual differences research, as well as the gross heterogeneity of findings, stem in part from a common problem: the failure to account for moderating task characteristics.

We first describe the evidence available from meta-analyses and skill acquisition research that supports the importance of task characteristics for understanding the interplay between practice and cognitive resources in predicting performance. We next describe a set of theoretical task characteristics that may additionally moderate the relationship between practice, ability, and performance. Lastly, we present results from two experiments that empirically test two of these task characteristics, while controlling for others.

Task Characteristics

Task characteristics have long been of interest in problem solving research. For example, problem difficulty is influenced by the problem space—the number of possible states and number of options at each stage of the solution (Newell & Simon, 1972). As the problem space increases, it can quickly exceed one's limited cognitive resources. As a result, people can only represent a subset of the problem space at any given time. However, even when the

This article was published Online First April 2, 2018.

Brooke N. Macnamara and David J. Frank, Department of Psychological Sciences, Case Western Reserve University.

The two authors contributed equally to this work.

Correspondence concerning this article should be addressed to Brooke N. Macnamara, Department of Psychological Sciences, Case Western Reserve University, 11220 Bellflower Road, Cleveland, OH 44106. E-mail: bnm24@case.edu

problem space is relatively small, other features can still produce difficulties (Kotovsky, Hayes, & Simon, 1985). For example, when a problem involves *moving* an object, such as moving disks in the Tower of Hanoi problem, it tends to be easier than when a problem involves *changing* the size or color of an object, such as changing the size of the globes in the Monster and Globes problem (Clément & Richard, 1997; Kotovsky et al., 1985). This is because the rules for change problems tend to be more linguistically complex, and are often novel and cannot be retrieved from long-term memory (Kotovsky et al., 1985). The seemingly obvious implication is that the larger the problem space and more complex or counterintuitive the rules, the more performance differences should be predicted by basic cognitive resources.

As with problem solving, task characteristics may play an important role in skill acquisition and expertise. In a meta-analysis, Macnamara, Hambrick, and Oswald (2014) found *performance domain* to be a significant moderator of the relationship between practice and performance: deliberate practice accounted for 26% of the variance in games such as chess, 21% in music, 18% in sports, 4% in education, and <1% in other professions. Additionally, Macnamara et al. (2014) examined *predictability of the task environment*—a variable that took into account how dynamic the task was and how consistent the task was—and found that this was also a significant moderator. Deliberate practice explained 26% of the variance in performance when tasks were highly predictable (e.g., running), 12% when they were moderately predictable (e.g., the sport of fencing), and 4% when they were low in predictability (e.g., handling an aviation emergency). Similarly, meta-analyses examining the relationship between general mental ability and job performance (e.g., Hunter & Hunter, 1984; Salgado et al., 2003) find that the relationship between general fluid intelligence and job performance differs based on job domain. Despite this, few studies have sought to determine the underlying features that dictate the extent to which a task relies on basic cognitive resources in addition to task-specific practice.

Although task characteristics have largely been ignored in experimental studies of individual differences in task performance, there is one notable exception. Ackerman's (1986) performance-ability relations theory suggests that task characteristics can be thought of in terms of having consistent (consistently mapped) components or inconsistent (variably mapped) components—which determine how practice and cognitive resources influence performance. Ackerman (1986) proposed that tasks with consistent components allow automatic processes to develop with practice (see also Shiffrin & Schneider, 1977). Once automatic processes are in place, individual differences in available cognitive resources are less associated with differences in task performance. In contrast, tasks with inconsistent (variably mapped) components continuously require controlled processing to perform the task despite training, thus recruiting cognitive resources even after accumulating task-specific practice.

As a test of the performance-ability relations theory, Ackerman and Woltz (1994) developed the noun-pair lookup task, where participants decided whether a centrally located pair of words (e.g., ivy-bird) matched any of several word pairs in a table. This task could be either consistently mapped, where the pairs in the table were the same from trial-to-trial, or variably mapped, where the table differed trial-to-trial. While accuracy was generally stable across the consistently and variably mapped conditions, efficiency on the task differed substantially. Specifically, response times in the consistently mapped version of the task decreased considerably with practice before reaching asymptote. By contrast, response times in the variably mapped version of the task showed only minimal improvement before reaching asymptote early in practice and were substantially slower than in the consistently mapped version. Additionally, Ackerman and Woltz (1994) found that, with practice, correlations between individual differences in perceptual speed and performance diminished in the consistently mapped condition, but remained high in the variably mapped condition. These results support Ackerman's (1986) theory that task characteristics—specifically consistently versus variably mapped attributes—differentially demand cognitive resources, which affect learning and performance.

As another example, Ackerman and Cianciolo (2002) asked participants to complete an air traffic control task where the mappings could be consistent—100% pilot compliance—or variably mapped—65% pilot compliance (e.g., pilot asking for the instructions to be repeated, failing to reply, a different pilot responding). In the variably mapped condition in this dynamic task, participants needed to adapt their instructions based on changes (e.g., positions of the aircrafts) that had taken place from the initial command issuance to the present moment. We therefore do not know how much accuracy was affected by the interaction between the dynamic state of the stimuli—requiring the contents of working memory and the action plan to be updated—and the variably mapped features. Perhaps unsurprisingly, Ackerman and Cianciolo (2002) found that basic cognitive resources were similarly predictive in their consistent and variably mapped versions of the air traffic control task. We suspect that this is because the dynamic nature of the air traffic control task placed strong cognitive demands on the performer, even under consistently mapped conditions. Other task characteristics in addition to the consistent versus variably mapped dimension may systematically interact to impact learning and performance.

Dimensions of Difficulty

Until recently, a set of additional task characteristics thought to impact learning and performance had not been formally proposed. In a recent book titled *Accelerated Expertise: Training for High Proficiency in a Complex World*, Hoffman et al. (2014) put forth eight *dimensions of difficulty* hypothesized to increase task difficulty across domains via their reliance on limited cognitive resources.[1] The eight dimensions are listed and described below. For simplicity, we will describe each in terms of the categorical extremes of the dimension and reference the characteristics as the task (e.g., "static tasks" rather than the more accurate but cumbersome "tasks with static characteristics"). In each case, the easier characteristic is presented first and the more cognitively demanding (i.e., difficult) characteristic is presented second:

1. Static versus dynamic: Important aspects of static tasks can be captured in "snapshots," whereas dynamic tasks are continuously changing. For example, chess is a static

[1] The dimensions of difficulty are based on research by Feltovich, Spiro, and Coulson (1989, 1993; see also Dawson-Saunders, Feltovich, Coulson, & Steward, 1990) who surveyed medical school instructors about difficult-to-learn biomedical concepts (e.g., mechanisms of hypoxia).

task where players can take their time to make decisions. By contrast Pacman (Namco, 1980) is a dynamic task where the information is constantly changing.

2. Discrete versus continuous: Attributes of discrete tasks are characterized by a small number of categories, whereas attributes of continuous tasks are characterized by a continua of features or a large number of categorical distinctions. For example, chess is a discrete task, with discrete categories for each piece and discrete positions on the board. By contrast, distance and angles in the game pool fall on a continuum.

3. Separable versus interactive: Processes in separable tasks occur independently or with weak interactions, whereas processes in interactive tasks are strongly interdependent. For example, games like Whac-A-Mole and Duck Hunt (Nintendo, 1984) are largely separable; whenever a stimulus appears, the player responds to it independently of the other stimuli—there is little to no need to coordinate her actions between stimuli. By contrast, air traffic control is a highly interactive task; each process, directing an aircraft to take off, land, or change altitude must be coordinated with the flight paths of all other aircraft.

4. Sequential versus simultaneous: Processes in sequential tasks occur one at a time, whereas processes in simultaneous tasks occur at the same time. For example, baseball is a sequential task; fielding and batting occur sequentially and a player is never engaged at both at once. By contrast, flying an aircraft involves a great deal of simultaneity; one must monitor gauges, speed, and heading, simultaneously.

5. Linear versus nonlinear: Relationships among features in linear tasks are proportional and can be conveyed with a single line of explanation, whereas relationships among features in nonlinear tasks are nonproportional and require multiple lines of explanation. For example, in baseball, the faster a pitcher moves his arm, the faster the pitch. By contrast, in basketball the relationship between the angle that a ball should be thrown and distance from the net is nonlinear.

6. Single versus multiple representations: Elements in single-representation tasks have one or very few interpretations or uses, whereas elements in multiple-representation tasks have multiple interpretations, and uses, based on context. For example, in chess, each piece affords a single representation; it can only behave in one manner. By contrast, in sports video games, the same button may have different functions depending on the context (offense or defense).

7. Mechanistic versus organic: Attributes in mechanistic tasks can be understood in terms of their parts. Effects in mechanistic tasks have direct causal agents, whereas organic tasks must be understood as a whole and effects in these tasks are due to system-wide functions. For example, the game Pong (Atari, 1972) is highly mechanistic; the behavior of the ball and paddle are the result of direct causal agents and can easily be understood in isolation. By contrast, chess is highly organic; only by understanding the relationship between all of the pieces on the board can one make optimal decisions.

8. Homogenous versus heterogeneous: Components and conceptual representations in homogenous tasks are uniform across a system (e.g., there is a single explanation), whereas components and conceptual representations in heterogeneous tasks are diverse.[2] For example, first-person shooter videogames are largely homogenous; if a player's health decreases it is always because he has taken damage from an enemy. By contrast, automotive diagnostics are heterogeneous; there may be multiple possible explanations for a single problem. For example, a car overheating could be due to insufficient coolant levels, poor airflow, or a malfunctioning fan.

The dimensions of task difficulty are intuitively appealing. However, they have not yet been empirically tested.

The Present Studies

The current studies are the first to examine task characteristics' impact on learning and performance in a complex task while controlling for other task characteristics. Using a novel test paradigm, we experimentally examine two dimensions: consistently versus variably mapped and static versus dynamic. These two dimensions are most similar to the components of the *predictability of the task environment* moderator examined in Macnamara et al.'s (2014) meta-analysis on deliberate practice. Additionally, these were two of the dimensions at play—but not examined independent of one another—in Ackerman and Cianciolo's (2002) air traffic control study.

To isolate the effect of a given dimension of difficulty, the task paradigm must hold all other dimensions of difficulty constant. Furthermore, to test whether a given characteristic thought to be more difficult (e.g., variably mapped, dynamic) places increased demands on basic cognitive resources, all other dimensions of difficulty must be kept at their "easiest" difficulty level (i.e., discrete, separable, sequential, etc.) so that potential decrements in performance are observable. To this end, we designed a tower defense game similar to the popular videogame Plants vs. Zombies (PopCap Games, 2009) that allows the manipulation of each task dimension independent of the others.

We first developed a *baseline* version of a task in which all task characteristics are on the easy side of each dimension of difficulty. That is, the characteristics of the baseline version of the task are consistently mapped, static, discrete, separable, sequential, linear, singly represented, mechanistic, and homogenous. We then created two manipulated versions of the task such that only one dimension is more difficult and all other dimensions are the same as the baseline version. In Experiment 1, the characteristics of the manipulated version of the task are variably mapped (while the

[2] Note that there is some degree of overlap between Ackerman's (1986) consistently mapped characteristic and Hoffman et al.'s (2014) single representation and homogenous task characteristics. Specifically, to achieve consistent mapping, a task would likely have to be homogenous and contain only single representations.

baseline version is consistently mapped) and both tasks are static, discrete, separable, sequential, linear, singly represented, mechanistic, and homogenous. In Experiment 2, the characteristics of the manipulated version of the task are dynamic (while the baseline version is static) and both are consistently mapped, discrete, separable, sequential, linear, single-represented, mechanistic, and homogenous.

Ackerman (1986) and Hoffman et al. (2014) both suggested that certain task characteristics are more difficult because they demand additional cognitive resources. We designed our task versions so that any differences in performance are due to additional cognitive demands. That is, any observed differences in task performance are *not* due to an artificial ceiling in the manipulated version or any kind of task mechanics that limits performance. Identical strategies will produce identical scores on each version of the task. This is true for both suboptimal rudimentary strategies as well as optimal strategies (see supplemental materials for data on experts trained on multiple task versions). Likewise, a computer would perform equally well on both task versions. Thus, any differences in the manipulated versions and the baseline version are due to cognitive demands of the task characteristics. We further manipulate cognitive load to test its effects on performance on the baseline and manipulated versions. If the manipulated version is more cognitively demanding, performance should suffer the most when experiencing the manipulated version under cognitive load.

Our goal is not only to test whether differences in performance emerge between the baseline and manipulated versions of the tasks, but to examine *how* various dimensions of difficulty differentially affect learning and/or performance. Ackerman and Woltz (1994) found that accuracy was similar between consistently mapped and variably mapped conditions in the noun-pair lookup task. However, they found that participants in the variably mapped condition had substantially slower response times. Based on Ackerman and Woltz's (1994) findings, we should expect no difference in accuracy between the baseline version and the variably mapped version of the task. However, participants in the variably mapped conditions should take longer to complete the task than participants in the baseline version, especially when under cognitive load. We provide the first test of Ackerman's (1986) theory in a complex task that also controls for other dimensions of difficulty.

Tests of Hoffman et al.'s (2014) dimension of difficulty are novel. Hoffman et al. (2014) did not detail how additional cognitive resources are recruited or how the dimensions of difficulty will impact task learning and performance. We assume that dynamic tasks, relative to static tasks, require additional cognitive resources because the contents of working memory must be continuously updated. Additionally, in order to respond to the incoming stimuli in constant flux, the inherent time pressure of a dynamic task means that the learner cannot pause to generate or implement a new strategy without the task environment changing. For these reasons, dynamic versions of the same task should hinder learning and performance scores. (Due to the nature of a dynamic task, efficiency cannot be measured because the dynamic version stimuli move on a timer regardless of participant response.) We expect that learning and performance accuracy will be hindered in the dynamic version relative to the baseline version, especially under cognitive load. This finding would support Hoffman et al.'s (2014) inclusion of static versus dynamic as a dimension of difficulty and would provide insight into the effects' underlying mechanisms.

Experiment 1

Method

Participants and design. A priori power analyses require knowledge of an expected effect size and assume linear analyses. Due to the novelty of our task paradigm and the nonlinearity of our planned analyses, estimates of power are unavailable. Our stopping rule was to collect 100 participants (25 per condition) or until the end of the semester's data collection period, whichever occurred second. One-hundred and 28 students enrolled in either General Psychology I or Research Design and Analysis at Case Western Reserve University participated in exchange for partial course credit or extra credit. To participate in our experiment, participants could not have experience with the commercial version of the Plants vs. Zombies videogame. This study was approved by Case Western Reserve University's Institutional Review Board. Participants were first informed about the procedures of the experiment and provided written consent if they agreed to participate. The study used a 2 (Task Version: baseline, variably mapped) $\times$ 2 (Cognitive Load: load, no load) between subjects design with roughly 32 participants per cell. Participants randomly assigned to the baseline version and those randomly assigned to the dynamic version did not differ on overall videogame experience, $t(126) = 0.97$, $p = .331$. Only one participant indicated any previous experience with tower defense videogames and thus a *t*-test is not possible to conduct.

Materials. The program was created with and administered via E-Prime 2 (Schneider, Eschman, & Zuccolotto, 2002) on Dell Optiplex 9030 computers at a resolution of 1,920 $\times$ 1,080.

Plants vs. Zombies task: Baseline version. Our version of Plants vs. Zombies is visually similar to the commercial videogame. However, the inherent structure and rules are different (videos of the task can be downloaded from https://osf.io/f4auv/).

Task overview. The participant is provided with the following game context: A zombie apocalypse has occurred and it is up to her avatar (named Crazy Dave) to collect energy for the town and fight off the zombies to keep them from escaping into the town. The task consists of these two alternating "missions," each of which is repeated five times. In the *energy collection missions*, participants plant sunflowers to collect energy from suns that move across the screen. In the *zombie fighting missions*, the participant plants pea plants to shoot and kill zombies that move across the screen. For each mission, the game environment comprises a 6 $\times$ 14 grid (Figures 1 and 2). The participant uses the arrow keys on a computer keyboard to move the avatar around the screen and the "z" key to plant a sunflower or pea plant at the avatar's current grid position. Only one sunflower/pea plant can occupy a space at a time.

The participant can take as long as she chooses to move her avatar or plant. Only after making two moves (planting also counts as a move), do the elements of the game move (the suns, energy from the suns, zombies, and pea shots). The number of moves remaining before the games' "turn" appears in an information bar at the top of the screen.

Designing a task with two distinct missions was done for two main reasons: (1) to test whether results can generalize across

Figure 1. Screenshot of an energy collection mission (baseline version). The small "starbursts" are sunlight energy (lumens) that travel from a sun to a sunflower on its right. See the online article for the color version of this figure.

different subtasks, and (2) to establish a usable paradigm for future studies. While the superficial design is similar across the two missions, they differ drastically in a number of important ways. First, the missions differ in their goals: The goal of the energy collection missions is to maximize beneficial opportunities, (i.e., to collect as many resources as possible), while the goal of the zombie fighting missions is to minimize negative consequences (i.e., to stop as many zombies from escaping as possible). Second, the availability of resources differs between the two. For the energy collection missions, flowers are only useful for one sun at a time and the suns will continue to travel across the screen regardless of planting. Thus, it is near impossible to run out of flowers because only a few have utility at a time. In contrast, for the zombie fighting missions, all plants in a row with a zombie are useful and zombies can be stopped from traveling across the screen and escaping. Thus, it is common to run out of plants because they can all have utility at once. Finally, the missions differ in their reactivity. The energy collection missions solicit a reactive strategy for the best performance: Participants will earn more points by focusing their efforts on large, slow suns as they appear and ignoring small, fast suns. Thus, participants should be reactive to the incoming stimuli, choosing their actions based on the location of certain suns as they emerge. In contrast, the zombie fighting missions demand a more proactive approach for the best perfor-

Figure 2. Screenshot of a zombie fighting mission (baseline version). The circles are the pea shots. See the online article for the color version of this figure.

mance: Participants will earn the most points by planning for and concentrating on large fast zombies while distributing enough resources for other zombies. Thus, participants should be proactive in order to have plants already set up for the best chance of killing the large, fast zombies. A similar pattern across both missions would provide robust evidence to the effect of the task characteristic being manipulated.

The other reason for having two missions is to test other dimensions of difficulty in the future. Specifically, when manipulating the sequential versus simultaneous dimension, two subtasks are needed: In the sequential condition the missions will remain separate for participants to manage one at a time; in the simultaneous condition the missions will be combined for participants to manage at the same time.

Energy collection missions. In the energy collection missions, the goal is to collect as much energy as possible by planting sunflowers to collect sunlight energy (lumens) from suns that move from left to right across the screen. A sunflower pulls a lumen—a small starburst of energy—from the first sun to its immediate left in the same row. The lumen moves from the sun to the right until it reaches a sunflower, at which point the energy is absorbed and added to the participant's energy score, which appears at the top of the screen. Each sun can expend only one lumen at a time. This single lumen will be absorbed by the first sunflower it reaches. No new lumens can be pulled from a sun until the current lumen reaches a sunflower and is absorbed. The closer a sunflower is to the right of a sun, the sooner the lumen reaches the sunflower and another lumen can be pulled. Thus, planting sunflowers close to a sun increases energy collection faster than planting farther away. However, once a sun reaches a sunflower, the flower wilts and can no longer collect energy. Wilted flowers are returned to the participant's resources.

Suns vary in color and size. The larger a sun, the more energy its lumens contain. Thus, more energy can be collected by focusing on collecting lumens from larger suns, planting new flowers in their paths to replace the wilted flowers. Suns come in three sizes: small, medium, and large. These suns produce lumens with four, eight, and 12 units of energy, respectively.

The color of each sun indicates how quickly it moves across the screen. Suns come in three colors: yellow, orange, and red. These suns move at rates of 15, 39, and 62 pixels per turn, respectively. Once a sun moves across the screen, no more lumens can be collected from it. Thus, more energy can be collected from slower suns because each sunflower will have time to absorb more lumens before being wilted by the sun passing over. The rules for the suns are displayed visually in the lower left part of the screen.

A participant can have up to 20 sunflowers on the screen at once. If a participant uses up all of her sunflowers (a near-impossible occurrence in this mission), she is unable to plant another sunflower until an existing sunflower wilts and is removed from the screen. After planting a sunflower, the participant has to wait three moves before planting again. Hence, the participant should strategically allocate her sunflower resources to the suns that will provide the most energy per sunflower plant. (Large yellow suns can provide the most energy.) Each energy collection mission contains a total of 27 suns (three of each possible Size × Color combination) with no more than 16 suns on the screen at once. When the final sun makes its way across the screen the mission ends.

The position and time at which each of the 27 suns emerges from the left side of the screen varies across rounds. Five arrangements are used and arrangement order is approximately counterbalanced across participants. Importantly, extensive piloting verified that comparable scores are possible across each arrangement both when rudimentary and optimal strategies are used. Extensive piloting also suggests that maximum performance on an energy collection mission is around 4,800 units of energy. Because we were concerned that the larger numbers for energy collection scores might be harder to track, we display energy collection scores to the participant as days and hours of accumulated energy that has been collected for the town.

Zombie fighting missions. In the zombie fighting missions, the goal is to plant pea plants that shoot peas to kill as many zombies as possible before they make their way across the screen and escape. Unlike sunflowers, where only the closest flower in a row interacts with a sun, *all* pea plants in the same row as a zombie will fire peas at the closest zombie to its right. Each zombie takes multiple hits before being killed. Thus, planting multiple peashooters in the same row will increase the rate at which zombies are killed. Each pea plant can have only one pea in the air at a time. The closer a pea plant is to the left of a zombie, the sooner the pea will hit the zombie and another pea can be fired from that plant. Thus, planting pea plants closer to zombies increases how rapidly the pea plant fires.

Zombies vary in color and size. The larger a zombie, the tougher it is and the more hits it can sustain before it dies. Thus, more pea shots are required to kill larger zombies. Zombies come in three sizes: small, medium, and large. These zombies can withstand 10, 26, and 42 hits, respectively. After sustaining enough hits, a zombie falls over and dies. For each zombie death, points are added to the zombie fighting score (displayed at the top of the screen). Small, medium, and large zombies are worth 10, 25, and 40 points, respectively.

The color of each zombie indicates how quickly it moves across the screen. Zombies wear one of three colored suits: blue, purple, or red. These zombies move at rates of 15, 39, and 62 pixels per turn, respectively. When a zombie reaches a space occupied by a pea plant, the plant is trampled and then disappears. Trampled pea plants are returned to the participant's resources. Faster zombies provide less time for the peashooters to kill them before they trample the pea plants and make it across the screen. The rules for the zombies are displayed visually in the lower right part of the screen.

A participant can have up to 20 pea plants on the screen at once. If a participant uses up all of her pea plants (a common occurrence in this mission), she is unable to plant another pea plant until either a plant is trampled and returned to her resources or she uproots a currently planted pea plant, which returns the plant to her resources for use. To do this, the participant positions her avatar over an existing pea plant and presses the "z" (plant) key. After planting or uprooting a pea plant, the participant has to wait three moves before planting or uprooting again. Participants need to strategically allocate pea plant resources to the zombies based on their speed and toughness. (Large red zombies are the most difficult to kill.) Each zombie fighting mission contains 27 zombies (three of each possible Size × Color combination). When the final zombie is killed or makes its way across the screen, the mission ends.

The position and time at which each of the 27 zombies emerges from the right side of the screen varies across rounds. Five arrangements are used and arrangement order is approximately counterbalanced across participants. Extensive piloting verified that comparable scores are possible across each arrangement both when rudimentary and optimal strategies are used. The maximum possible score for each zombie fighting mission (all zombies killed) is 675.

Additional feedback and instructions. Prior to each mission, participants are shown a reminder of the game rules for that mission (see supplemental materials). Starting with Round 2, the participant's scores for all previous missions appear so that participants can chart their progress across rounds. At the end of each mission, the participant's final score for that mission is presented in white at the center of a green screen for two seconds.

Baseline task characteristics. The baseline version's task characteristics are all on the easy side of each dimension of difficulty (see Table 1).

Plants vs. Zombies task: Variably mapped version. The variably mapped version is identical to the baseline version with the exception that the mapping of color and size to sun/zombie speed and energy amount/zombie toughness changes several times throughout each mission. Color can represent speed or amount of energy, with size representing the other property. The linear relationship within each feature—small, medium large; red, orange, yellow; red, purple, blue—is maintained but the direction can change. That is, small to large suns could represent slowest to fastest or fastest to slowest (or most to least energy or least to most energy). For example, a participant could start with red suns as the fastest (yellow as the slowest), and large suns producing the most energy (small producing the least amount). After 50 moves, the mapping could change to small suns as the fastest (large suns as the slowest), and yellow suns producing the most energy (red suns producing the least amount). Every 50 moves (six times per game) the words "RULE CHANGE" appear in large blue font on the screen for 0.5 s. At this point the rule summary at the bottom of the screen changes as does the size and color of the suns/zombies to conform to the new rules. Note that the behavior of the stimuli—the speed and energy/toughness of each sun and zombie—does not change. Rather, their physical appearance on the screen changes to fit the new mappings. The fastest and toughest zombie remains so, even though his appearance changes. This is important so that good strategies for performance remain good strategies regardless of physical changes. The eight possible rule combinations occur in the same fixed order for all participants (see supplemental materials). In summary, the baseline version uses a consistent set of rules for the appearance and property of stimuli. By contrast, in the variably mapped version, the relationship between stimulus features and their behavior changes periodically, but identical planting strategies in each version yields identical results. Thus, any differences in task performance can only be due to cognitive and not mechanical differences in task difficulty (videos of the task can be downloaded from https://osf.io/f4auv/).

Cognitive load manipulation. All participants wore professional-grade noise cancelling headphones. Participants in the load conditions were instructed to mentally rehearse and update letters of the alphabet each time they heard a beep in their headphones. That is, they were instructed to mentally rehearse the letter "A" the first time they heard a beep while performing the Plants vs. Zombies task, then update the mentally rehearsed letter to "B" when the heard the second beep, "C" when they heard the third beep, and so on. If they reached Z, they were instructed to start again at "A." The beeps occurred at random intervals with the exception that beeps could not occur fewer than 2 s apart or more than 6 s apart. At the end of each mission, participants were asked to type the last letter they had mentally rehearsed during that mission.

Open strategy reports. Following the Plants vs. Zombies task, participants were asked to "Please describe the strategies that

Table 1
Baseline Version Characteristics

Characteristic	Description
Consistently mapped	The rules for the stimuli are maintained throughout the task. For example, yellow suns are always the slowest, small zombies are always the weakest.
Static	The task does not advance until the participant makes her move. Thus, the relevant information for making decisions does not change during the decision process. The participant can pause to think before executing a decision without the relevant features changing.
Discrete	The suns and zombies each fall into three discrete categories of size and color. Likewise, the sunflowers and pea plants are discrete categories.
Separable	Each process (moving the avatar and planting plants) occurs in isolation. Although the participant must coordinate her moving and planting (i.e., taking the shortest route to the desired location before planting), this interaction is relatively weak.
Sequential	The energy collection missions and the zombie fighting missions occur in separate stages. The participant is never engaged in both missions at once.
Linear	The relationship between size and sun energy/zombie toughness is simple and mathematically linear. The relationship between color and speed are simple and mathematically linear.
Single representations	Each task stimulus has only one meaning throughout the baseline version. That is, a sun of a given size and color represents the same opportunity for energy collection throughout the task. The sunflower and pea plants function the same in all task situations.
Mechanistic	The task can generally be understood in terms of the individual parts. Causal agents (how plants interact with suns and zombies) are direct. The participant does not have to consider multiple aspects of the task to predict how a plant, sun, or zombie will behave.
Homogeneous	The task components (movement and planting counters, scores, task display, button mapping, etc.) are similar for both missions.

you used when collecting sunlight. In your description, please identify any changes in your strategy over time, or any insights that you had while performing the task." Participants typed their responses into a Microsoft Word document. Participants were next asked the same question regarding the zombie fighting mission.

Strategy questionnaire. A computerized multiple choice strategy questionnaire listed 24 strategies for the energy collection missions and 27 strategies for the zombie fighting missions (see supplemental materials). For each strategy, participants indicated whether they (a) did not think of the strategy, (b) thought of it but did not use it, (c) used it rarely, (d) used it sometimes, or (e) used it often (see supplemental materials for strategy questionnaire screenshots and results).

Procedures. Participants first completed a survey of videogame experience (see supplemental materials). Next participants read the Plants vs. Zombies task instructions and completed the five rounds of the task. Immediately after the task, participants rated their task experience in terms of task interest, fun, tediousness, difficulty, frustration, engagement, and fatigue. Participants in the load conditions also indicated how they responded on the letter counting task if they forgot which letter they were on (see supplemental materials).[3] Lastly, participants completed the open strategy reports and strategy questionnaire. Participants read the instructions, completed the task, filled out the open strategy report, and responded to the questionnaires at their own pace. Sessions lasted between 59 and 183 min with most participants taking approximately 90 min.

Results

Plants vs. Zombies scores. We first computed means and standard deviations for Round 1 of each mission. We then removed any outliers that were three standard deviations below the mean. This resulted in <1% of the data being removed: one observation from an energy collection mission and seven observations from zombie fighting missions.[4] Because the scores for each mission are on different scales, we next standardized those scores by subtracting the mean score for Round 1 in the baseline no load group, and then dividing by the standard deviation for that group. Thus, each score is the number of standard deviations above or below the mean score for the first round for the theoretically easiest condition.

Changes in performance are often nonlinear (Fitts, 1964). To account for this, we conducted mixed models including both round and round squared as predictors of performance scores using SAS Proc MIXED (Littell, Milliken, Stroup, & Wolfinger, 2000). Additionally, we included task version (baseline, variably mapped), cognitive load (load, no load), mission (energy collection, zombie fighting), and all possible interaction terms except for Round × Round Squared. The baseline, no load condition served as the reference group. Regression weights and fixed effects can be found in Tables 2 and 3, respectively. Figure 3 shows mean performance data over rounds.

Main effects of round, $F(1, 1124) = 96.29$, $p < .001$, and round squared, $F(1, 1124) = 32.76$, $p < .001$, indicated better performance over rounds with a negative exponential curve following typical learning curves (Fitts, 1964). The effect of task version was negative and significant, $F(1, 1124) = 4.86$, $p = .029$, indicating worse performance on the variably mapped version than the baseline version. We observed significant Task Version × Round, $F(1, 1124) = 4.03$, $p = .045$, and Task Version × Round Squared, $F(1, 1124) = 4.71$, $p = .030$, interactions. Together, these main effects and interactions with task version indicate that the performance of the variably mapped groups was initially lower than the baseline groups, improved rapidly from Round 1 to Round 2, but ultimately plateaued at a lower level than the baseline groups.[5]

A positive main effect of mission, $F(1, 123) = 29.15$, $p < .001$, indicated better performance on the zombie fighting mission relative to the energy collection mission. However, this was qualified by a positive Mission × Load interaction, $F(1, 123) = 10.11$, $p = .002$. This interaction resulted from worse performance on the energy collection mission in the presence of cognitive load, $t(126) = -2.38$, $p = .019$, $d = -0.42$ (significant after Bonferroni's correction, $\alpha = .025$), but similar performance for zombie fighting missions regardless of load, $t(125) = -1.70$, $p = .093$. A negative Mission × Round interaction, $F(1, 1124) = 4.52$, $p = .034$, and a positive Mission × Round squared interaction, $F(1, 1124) = 4.19$, $p = .041$, resulted from more linear performance gains on the zombie fighting relative to energy collection missions. Note, however, that mission did not interact with task version, indicating similar effects of consistent and variable mapping across both missions.

Completion times. We used round completion times as a measure of performance efficiency. Due to a programming error, Round 1 start times were not available for the energy collection mission. Thus, we focus our statistical analyses of performance efficiency on the zombie fighting missions (see Figure 4). For the variably mapped version, we subtracted from the total completion times the amount of time that the game was paused while the words "RULE CHANGE" appeared on the screen (exactly 30 s total). We removed any completion times that were more than three standard deviations above the mean for the corresponding round. This resulted in the removal of two observations in each of the baseline load, variably mapped load, and variably mapped no load groups.[6]

As with performance scores, we used SAS Proc MIXED (Littell et al., 2000) to analyze performance efficiency with the following variables as predictors: round, round squared, task version (baseline, variably mapped), cognitive load (load, no load), and all possible interaction terms except for Round × Round Squared. Again, the baseline no load group served as the reference group. Regression weights and fixed effects can be found in Tables 4 and 5, respectively.

A main effect of task version resulted from slower completion times for the variably mapped groups compared with the baseline groups, $F(1, 124) = 6.23$, $p = .014$. The main effects of round,

[3] Participants who indicated that they occasionally or always quit counting if they forgot which letter they were on did not differ from others in terms of overall task performance. Hence, we retained these participants for all analyses.

[4] This criterion was set only after observing non-normality. After removing outliers the distribution was approximately normal (skew and kurtosis between 1 and −1).

[5] When analyzed without removing outliers, the pattern of results remains unchanged with the exception that we additionally observe a main effect of phase, $F(1, 124) = 8.35$, $p = .005$, resulting from higher performance in the zombie fighting relative to the energy collection mission.

[6] When analyzed without removing outliers, the pattern of significance was unchanged.

Table 2
Experiment 1 Performance Scores Mixed Modeling Results

Effect	Estimate	*SE*	DF	*t*-value	*p*-value	Description of estimate
Intercept	−.01	.18	115	−.08	.939	Mean score estimate for Round 1 (coded as Round 0) of energy mission in the baseline no load group.
Version	−1.50	.26	115	−5.84	<.001	Difference in Round 1 score between the variable no load and baseline no load groups (for energy mission).
Load	−.44	.26	115	−1.74	.085	Difference in Round 1 score between the baseline load and baseline no load groups (for energy mission).
Mission	−.03	.20	115	−.14	.887	Difference in Round 1 score between zombie and energy missions in the baseline no load group.
Round	.32	.17	1,040	1.90	.058	Linear effect of round in the baseline no load group (for energy missions). Rounds coded 0–4.
$Round^2$	−.04	.04	1,040	−1.09	.274	The exponential effect of round in the baseline no load group (for energy missions).
Version × Load	.12	.37	115	.33	.745	Difference in Round 1 score between the variable load and variable no load groups (for energy mission).
Version × Mission	−.17	.29	115	−.56	.574	Difference in Round 1 score between zombie and energy missions for the variable no load group.
Load × Mission	.12	.29	115	.41	.679	Difference in Round 1 score between zombie and energy missions for the baseline load group.
Round × Version	.72	.24	1,040	2.96	.003	Difference in the linear effect of round between the variable no load and baseline no load groups (for energy missions).
Round × Load	.35	.24	1,040	1.43	.153	Difference in the linear effect of round between the baseline load and baseline no load groups (for energy missions).
Round × Mission	.12	.24	1,040	.52	.602	Difference in the linear effect of round between zombie and energy missions for the baseline no load group.
$Round^2$ × Version	−.13	.06	1,040	−2.20	.028	Difference in the exponential effect of round between the variable no load and baseline no load groups (for energy missions).
$Round^2$ × Load	−.06	.06	1,040	−1.07	.285	Difference in the exponential effect of round between the baseline load and baseline no load groups (for energy missions).
$Round^2$ × Mission	.00	.06	1,040	.07	.947	Difference in the exponential effect of round between zombie and energy missions for the baseline no load group.
Version × Load × Mission	−.03	.42	115	−.08	.938	Difference in Round 1 score between zombie and energy missions for the variable load group.
Round × Version × Load	−.72	.35	1,040	−2.06	.040	Difference in the linear effect of round between the variable load and variable no load groups (for energy missions).
Round × Version × Mission	−.29	.35	1,040	−.83	.408	Difference in the linear effect of round between zombie and energy missions in the variable no load group.
Round × Load × Mission	.01	.35	1,040	.02	.986	Difference in the linear effect of round between zombie and energy missions in the baseline load group.
$Round^2$ × Version × Load	.16	.08	1,040	1.90	.057	Difference in the exponential effect of round between the variable load and variable no load groups (for energy missions).
$Round^2$ × Version × Mission	.07	.08	1,040	.83	.407	Difference in the exponential effect of round between zombie and energy missions in the variable no load group.
$Round^2$ × Load × Mission	−.01	.08	1,040	−.12	.908	Difference in the exponential effect of round between zombie and energy missions in the baseline load group.
Round × Version × Load × Mission	−.09	.50	1,040	−.19	.849	Difference in the linear effect of Round between zombie and energy missions in the variable load group.
$Round^2$ × Version × Load × Mission	.02	.12	1,040	.18	.855	Difference in the exponential effect of round between zombie and energy missions in the variable load group.

Note. Version = task version (baseline, variably-mapped); Load = cognitive load (load, no load); Round = linear improvement over rounds; $Round^2$ = nonlinear (quadratic) improvement over rounds; Energy = energy collection; Zombie = zombie fighting; Variable = variably mapped.

$F(1, 498) = 109.52$, $p < .001$, and round squared, $F(1, 498) = 22.56$, $p < .001$, indicate faster performance over rounds with a positive exponential curve.

Discussion

As was the case in Ackerman and Woltz's (1994) noun-pair lookup task, we expected performance scores to be similar in the variably mapped and baseline versions of the task, but that performance efficiency would be reduced when the task was variably mapped. Additionally, we expected that cognitive load would further slow performance in the variably mapped version of the task. Our hypotheses were partially supported. Variable mapping both decreased performance scores and efficiency. However, these decrements were fairly small and were not influenced by cognitive load—indicating that the cognitive resources consumed by our cognitive load task (likely working memory capacity) were not those necessary for dealing with variable mapping. These results suggest that variable mapping increases the difficulty of the task in such a way that cannot be compensated simply with a speed–accuracy trade-off—participants in the variably mapped conditions were both slower and had worse performance.

Table 3
Experiment 1 Performance Scores Type 3 Fixed Effects

Effect	Num DF	Den DF	*F* value	*p*-value
Version	1	124	4.86	.029
Load	1	124	2.49	.117
Mission	1	123	29.15	<.001
Round	1	1,124	96.29	<.001
Round^2	1	1,124	32.76	<.001
Version × Load	1	124	.21	.647
Version × Mission	1	123	.10	.752
Load × Mission	1	123	10.11	.002
Round × Version	1	1,124	4.03	.045
Round × Load	1	1,124	.02	.892
Round × Mission	1	1,124	4.52	.034
Round^2 × Version	1	1,124	4.71	.030
Round^2 × Load	1	1,124	.15	.700
Round^2 × Mission	1	1,124	4.19	.041
Version × Load × Mission	1	123	.28	.597
Round × Version × Load	1	1,124	.17	.681
Round × Version × Mission	1	1,124	1.73	.188
Round × Load × Mission	1	1,124	1.06	.303
Round^2 × Version × Load	1	1,124	.19	.664
Round^2 × Version × Mission	1	1,124	.98	.324
Round^2 × Load × Mission	1	1,124	.17	.678
Round × Version × Load × Mission	1	1,124	.12	.725
Round^2 × Version × Load × Mission	1	1,124	.10	.754

Note. Type III fixed effects are interpreted as they would be in an ANOVA. For example, the main effect of version, test the mean difference between the two task versions after collapsing across all rounds and each level of load and mission. By contrast the effect estimates and their corresponding *t*- and *p*-values reported in Table 2, are conditional to all other effects, and are explained in the estimate description column. For theoretical testing, we focus on the Type III fixed effects and use the regression estimates to determine the direction of main effects and interactions. Version = task version (baseline, variably-mapped); Load = cognitive load (load, no load); Round = linear improvement over rounds; Round^2 = nonlinear (quadratic) improvement over rounds.

Given Ackerman and Cianciolo's (2002) findings on the air traffic control task, we might expect that accuracy would decrease under variably mapped *and* dynamic conditions, especially under

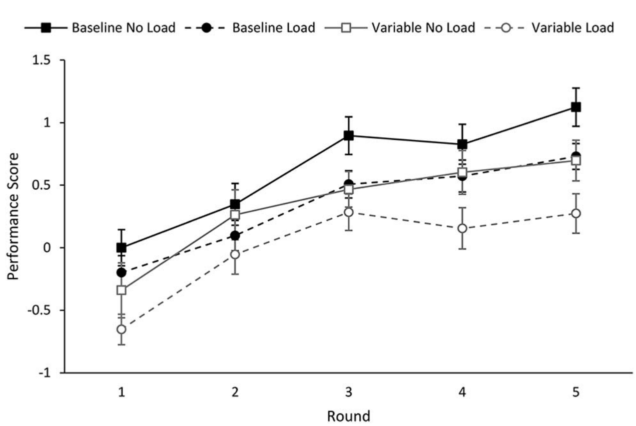

Figure 3. Experiment 1 mean performance on the Plants vs. Zombies task. Performance data are standardized around Round 1 performance in the baseline no load group, which has a *z*-score of 0. A *z*-score of 1 indicates performance one standard deviation above the Round 1 baseline no load group mean. Error bars indicate standard errors of the mean.

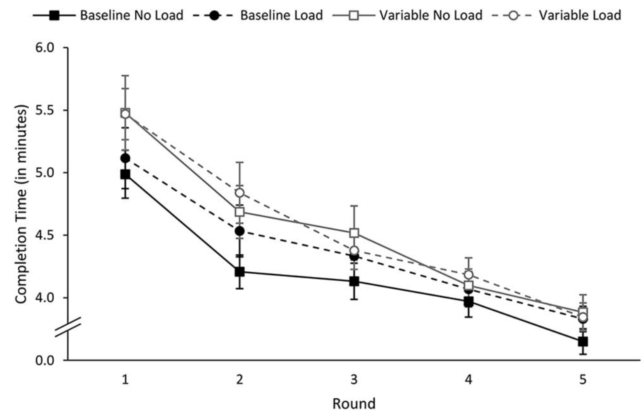

Figure 4. Experiment 1 mean task completion times in minutes for the zombie fighting mission. Error bars represent standard errors of the mean.

cognitive load. However, before task characteristic interactions can be investigated, the effect of task characteristics in isolation should be tested. Controlling for other dimensions of difficulty, we next investigate learning and performance when the task is static versus dynamic.

We expect that the demands of a dynamic task—constantly updating the contents of working memory and not being able to pause to strategize—will hinder learning and performance relative to a static task where learners have time to consider multiple strategies, especially when under cognitive load.

Experiment 2

Method

Participants and design. As in Experimenter 1, the novelty of our task paradigm and the nonlinearity of our planned analyses made estimates of power unavailable. Our stopping rule was to collect 100 participants (25 per condition) or until the end of the semester's data collection period, whichever occurred second. To participate in our experiment, participants could not have experience with the commercial Plants vs. Zombies videogame. One-hundred and 19 students enrolled in either General Psychology I or Research Methods and Design at Case Western Reserve University participated in exchange for partial course credit or extra credit. As in Experiment 1, this study was approved by Case Western Reserve University's Institutional Review Board. Participants were first informed about the procedures of the experiment and provided written consent if they agreed to participate. The study used a 2 (Task Version: baseline, dynamic) × 2 (Cognitive Load: load, no load) between subjects design with roughly 30 participants per cell. Participants randomly assigned to the baseline version and those randomly assigned to the dynamic version did not differ on overall videogame experience, $t(117) = 0.01$, $p = .994$, or tower defense videogame game experience, $t(117) = 0.89$, $p = .376$.

Materials. Materials and procedures were identical to those in Experiment 1 except where indicated.

Plants vs. Zombies task: Baseline version. Minor changes to the timing of the animation were made to make the baseline version in Experiment 2 comparable with the dynamic version.

This resulted in the lumens moving slightly faster in the energy collection missions relative to Experiment 1. As a result, energy collection scores were somewhat higher in Experiment 2, with a maximum possible score of roughly 5,000 (compared with 4,800 in Experiment 1) for the baseline and manipulated task versions (videos of the task can be downloaded from https://osf.io/f4auv/).

Plants vs. Zombies task: Dynamic version. The dynamic version is identical to the baseline version with two exceptions. First, the suns and zombies advance not after the participant makes two moves, but rather, every 500 ms. This timing allows the participant to move her avatar zero to three times before the suns and zombies move again. Given that participants can move their avatar up to one move more before the game's "turn" (relative to the baseline version where the game's turn is after the participant makes two moves) there is a slight mechanical *advantage* in the dynamic condition for moving the avatar into position to place or uproot a plant. Second, rather than allowing the participant to plant on the fourth move after the last time she placed or uprooted a plant, a planting timer allows her to plant exactly 1,332 ms after her previous planting or uprooting. This was equivalent to the baseline version. In summary, the baseline version functions such that the participant can pause and make decisions at her own pace before moving and advancing the game. By contrast, the game progresses in the dynamic version regardless of the participant's behavior (videos of the task can be downloaded from https://osf.io/f4auv/).

Cognitive load manipulation. The cognitive load manipulation was identical to Experiment 1.

Procedures. As in Experiment 1, participants completed a survey of videogame experience, read the Plants vs. Zombies task instructions and completed the five rounds of the task, rated their task experience, completed the open strategy reports, and completed the strategy questionnaire (see supplemental materials for strategy questionnaire screenshots and results). By virtue of the dynamic version being pace-controlled, participants in dynamic conditions completed the session in approximately 45 min. Participants in the baseline version completed the task at their own pace and took between 66 and 123 min with most participants taking approximately 90 min.

Results

Plants vs. Zombies scores. As in Experiment 1, we first computed means and standard deviations for Round 1 of each mission. We then removed any outliers that were three standard deviations below the mean. This resulted in $<1\%$ of the total data being deleted (three observations from energy collection missions and eight observations from zombie fighting missions).[7] We standardized performance scores as was done in Experiment 1.

As in Experiment 1, we used SAS Proc MIXED (Littell et al., 2000) to analyze performance scores with the following variables as predictors: round, round squared, task version (baseline, dynamic), cognitive load (no load, load), mission (energy collection, zombie fighting) and all possible interaction terms except for Round × Round Squared. The baseline no load group served as the reference group. Regression weights can be found in Tables 6 and 7. Figure 5 shows mean performance data over rounds.

A main effect of task version, $F(1, 115) = 103.16$, $p < .001$, resulted from poorer performance scores in the dynamic groups compared to the baseline groups. A main effect of cognitive load, $F(1, 115) = 4.87$, $p = .029$, resulted from poorer performance under cognitive load. The main effects of round, $F(1, 1040) = 111.19$, $p < .001$, and round squared, $F(1, 1040) = 29.54$, $p < .001$, indicate better performance over rounds with a negative exponential curve. However, these main effects for round and round squared were qualified by significant three-way interactions with task version and cognitive load, for round, $F(1, 1040) = 9.57$, $p = .002$, and round squared, $F(1, 1040) = 8.22$, $p = .004$. Specifically, the learning slope for the dynamic load group was significantly different from the other three groups. This resulted from delayed learning in the dynamic load group, which improved little from Round 1 to Round 2, whereas the other three groups improved markedly from Round 1 to Round 2. There were no main effects or interactions with mission indicating that task version and cognitive load affected the energy collection and zombie fighting missions similarly.[8]

Discussion

We expected performance scores to be worse when the task was dynamic than when it was static. Additionally, we expected that cognitive load would further reduce performance in the dynamic version of the task. Our hypotheses were supported. All participants improved with practice, but this was hindered in the dynamic conditions, where participants never "caught up" to participants in the baseline conditions. Cognitive load had no effect in the baseline version, but substantially impacted learning and performance in the dynamic version of the task. One potential argument is that the results are simply a speed–accuracy trade-off—that cognitive load increases difficulty, but that the effect does not emerge in the baseline conditions because participants could slow down to compensate for the additional difficulty. The data do not support this argument. Participants in the baseline load condition ($M = 4.38$, $SD = 1.11$) were no slower than participants in the baseline no load condition ($M = 4.19$, $SD = 0.91$), $t(117) = 0.26$, $p = .796$. Instead, our findings are consistent with Hoffman et al.'s (2014) suggestion that dynamic tasks hinder learning and performance because they require additional cognitive resources.

General Discussion

We investigated how task characteristics impact the relative importance of both cognitive resources and practice on learning and performance. In Experiment 1, we found that variable mapping resulted in slower improvements with practice and lower overall performance. Additionally, we found that variable mapping significantly slowed performance speed. Participants were unable to perform the task as efficiently, taking approximately 13% longer than participants in the baseline versions.

When examining the static versus dynamic dimension—where efficiency is not a viable metric—we found that performance

[7] This criterion was set only after observing non-normality. After removing outliers the distribution was approximately normal (skew and kurtosis between 1 and −1).

[8] When analyzed without removing outliers, the main effect of mission was significant, $F(1, 115) = 10.02$, $p = .002$, resulting from higher performance in the zombie fighting relative to the energy collection mission.

Table 4
Experiment 2 Completion Times Mixed Modeling Results

Effect	Estimate	*SE*	DF	*t*-value	*p*-value	Description of estimate
Intercept	4.89	.19	124	26.20	<.001	Mean Round 1 (coded as Round 0) completion time estimate in minutes for the baseline no load group.
Version	.52	.27	124	1.95	.054	Difference in Round 1 completion times between the variable no load and baseline no load groups.
Load	.36	.26	124	1.37	.173	Difference in Round 1 completion times between the baseline load and baseline no load groups.
Round	−.53	.12	504	−4.40	<.001	Linear effect of round in the baseline no load group. Rounds coded 0–4.
Round^2	.06	.03	504	2.06	.040	The exponential effect of round in the baseline no load group.
Version × Load	−.21	.37	124	−.56	.576	Difference in Round 1 score between the variable load and variable no load groups.
Round × Version	−.10	.17	504	−.56	.576	Difference in the linear effect of round between the variable no load and baseline no load groups.
Round × Load	−.16	.17	504	−.92	.359	Difference in the linear effect of round between the baseline load and baseline no load groups.
Round^2 × Version	.01	.04	504	.24	.813	Difference in the exponential effect of round between the variable no load and baseline no load groups.
Round^2 × Load	.03	.04	504	.84	.399	Difference in the exponential effect of round between the baseline load and baseline no load groups.
Round × Version × Load	.07	.24	504	.28	.779	Difference in the linear effect of round between the variable load and variable no load groups.
Round^2 × Version × Load	−.03	.06	504	−.49	.627	Difference in the exponential effect of round between the variable load and variable no load groups.

Note. Version = task version (baseline, dynamic); Load = cognitive load (load, no load); Round = linear improvement over rounds; Round^2 = nonlinear (quadratic) improvement over rounds; Variable = variably mapped.

scores were substantially lower in the dynamic version relative to the baseline version, especially when under cognitive load. Those in the dynamic versions improved with practice, but did not reach the performance scores of those in the baseline versions. The finding that cognitive load did not affect learning or performance in either mission in the baseline version but hindered learning and performance in the dynamic version supports Hoffman et al.'s (2014) claim that the more difficult end of the dimension (in this case, dynamic) recruits additional cognitive resources.

Table 5
Experiment 1 Completion Times Type 3 Fixed Effects

Effect	Num DF	Den DF	*F* value	*p*-value
Version	1	12	4.90	.020
Load	1	12	1.90	.170
Round	1	50	112.79	<.001
Round^2	1	50	26.77	<.001
Version × Load	1	12	.31	.570
Round × Version	1	50	.27	.600
Round × Load	1	50	1.03	.310
Round^2 × Version	1	50	.02	.880
Round^2 × Load	1	50	.50	.470
Round × Version × Load	1	50	.08	.770
Round^2 × Version × Load	1	50	.24	.620

Note. Type III fixed effects are interpreted as they would be in an ANOVA. For example, the main effect of version, test the mean difference between the two task versions after collapsing across all rounds and each level of load. By contrast the effect estimates and their corresponding *t*- and *p*-values reported in Table 4, are conditional to all other effects, and are explained in the estimate description column. For theoretical testing, we focus on the Type III fixed effects and use the regression estimates to determine the direction of main effects and interactions. Version = task version (baseline, dynamic); Load = cognitive load (load, no load); Round = linear improvement over rounds; Round^2 = nonlinear (quadratic) improvement over rounds; Variable = variably mapped.

Our results suggest that task characteristics are an important component of any model or theory of skill acquisition or expertise. Specifically, task characteristics affected the impact of cognitive load and practice amounts on learning and performance. This finding suggests that the predictive power of cognitive resources on expertise and the predictive power of practice on expertise are not set amounts that can be applied to any task or any performance domain. Rather, the influence of these factors is systematically heightened or reduced depending on the characteristics of the task.

The task characteristics we investigated are general properties that are applicable across very different kinds of tasks. Although we cannot be certain that our paradigm will generalize to every task paradigm, our results suggest some degree of generalizability. Namely, although the shape of learning curves differed somewhat across our two missions in Experiment 1, the effect of task characteristics was similar across missions in both experiments. One could argue that this provides only limited evidence of generalizability given the similarity between our two missions (the keyboard controls, the screen layout). Indeed performance on each mission was correlated in Experiments 1, $r(125) = .60$, $p < .001$, and Experiment 2, $r(117) = .68$, $p < .001$. Although the two missions share a number of surface features, the strategies that produce superior performance in each mission are actually opposites. For energy collection, it is best to focus all of one's resources on just a few suns with the best features (slow, high energy suns). By contrast, for zombie fighting it is best to divide one's resources evenly among each row, then deploy additional resources to fight the more difficult zombies (tougher or faster zombies). Thus, high correlations between scores on the two missions cannot be the

Table 6
Experiment 2 Performance Scores Mixed Modeling Results

Effect	Estimate	*SE*	DF	*t*-value	*p*-value	Description of estimate
Intercept	−.01	.18	115	−.08	.939	Mean score estimate for Round 1 (coded as Round 0) of energy mission in the baseline No Load group.
Version	−1.50	.26	115	−5.84	<.001	Difference in Round 1 score between the dynamic no load and baseline no load groups (for energy mission).
Load	−.44	.26	115	−1.74	.085	Difference in Round 1 score between the baseline load and baseline no load groups (for energy mission).
Mission	−.03	.20	115	−.14	.887	Difference in Round 1 score between zombie and energy missions in the baseline no load group.
Round	.32	.17	1,040	1.90	.058	Linear effect of round in the baseline no load group (for energy missions). Rounds coded 0–4.
$Round^2$	−.04	.04	1,040	−1.09	.274	The exponential effect of round in the baseline no load group (for energy missions).
Version × Load	.12	.37	115	.33	.745	Difference in Round 1 score between the dynamic load and dynamic no load groups (for energy mission).
Version × Mission	−.17	.29	115	−.56	.574	Difference in Round 1 score between zombie and energy missions for the dynamic no load group.
Load × Mission	.12	.29	115	.41	.679	Difference in Round 1 score between zombie and energy missions for the baseline load group.
Round × Version	.72	.24	1,040	2.96	.003	Difference in the linear effect of round between the dynamic no load and baseline no load groups (for energy missions).
Round × Load	.35	.24	1,040	1.43	.153	Difference in the linear effect of round between the baseline load and baseline no load groups (for energy missions).
Round × Mission	.12	.24	1,040	.52	.602	Difference in the linear effect of Round between zombie and energy missions for the baseline no load group.
$Round^2$ × Version	−.13	.06	1,040	−2.20	.028	Difference in the exponential effect of round between the dynamic no load and baseline no load groups (for energy missions).
$Round^2$ × Load	−.06	.06	1,040	−1.07	.285	Difference in the exponential effect of round between the baseline load and baseline no load groups (for energy missions).
$Round^2$ × Mission	.00	.06	1,040	.07	.947	Difference in the exponential effect of round between zombie and energy missions for the baseline no load group.
Version × Load × Mission	−.03	.42	115	−.08	.938	Difference in Round 1 score between zombie and energy missions for the dynamic load group.
Round × Version × Load	−.72	.35	1,040	−2.06	.040	Difference in the linear effect of round between the dynamic load and dynamic no load groups (for energy missions).
Round × Version × Mission	−.29	.35	1,040	−.83	.408	Difference in the linear effect of Round between zombie and energy missions in the dynamic no load group.
Round × Load × Mission	.01	.35	1,040	.02	.986	Difference in the linear effect of round between zombie and energy missions in the baseline load group.
$Round^2$ × Version × Load	.16	.08	1,040	1.90	.057	Difference in the exponential effect of round between the dynamic load and dynamic no load groups (for energy missions).
$Round^2$ × Version × Mission	.07	.08	1,040	.83	.407	Difference in the exponential effect of round between zombie and energy missions in the dynamic no load group.
$Round^2$ × Load × Mission	−.01	.08	1,040	−.12	.908	Difference in the exponential effect of round between zombie and energy missions in the baseline load group.
Round × Version × Load × Mission	−.09	.50	1,040	−.19	.849	Difference in the linear effect of round between zombie and energy missions in the dynamic load group.
$Round^2$ × Version × Load × Mission	.02	.12	1,040	.18	.855	Difference in the exponential effect of round between zombie and energy missions in the dynamic load group.

Note. Version = task version (baseline, dynamic); Load = cognitive load (load, no load); Round = linear improvement over rounds; $Round^2$ = nonlinear (quadratic) improvement over rounds; Energy = energy collection; Zombie = zombie fighting.

result of transfer of specific strategies across missions. Rather, performance should, theoretically, be highly correlated because the missions shared critical features. That is, missions shared the key task features—consistent mapping, variable mapping, static features, or dynamic features—which should require the same cognitive resources across missions. Additionally, our results for Experiment 1 mimic Ackerman and colleagues (Ackerman, 1986; Ackerman & Woltz, 1994) findings using the noun pair look-up task.

By examining task characteristics' impact on learning and performance we can better understand how task demands affect cognitive resources and processes. This knowledge can contribute to current skill acquisition and expertise theories. These findings also have the potential to influence other research areas such as industrial/organizational psychology. Currently, industrial/organizational psychologists are aware that general cognitive resources best predict occupational level and performance (even better than job experience), as well as rate of learning when receiving job training (Schmidt & Hunter, 2004). Additionally, the predictive power of general cognitive ability increases as complexity—information processing requirements—of the job increases (Hunter & Hunter, 1984). However, better understanding of how task characteristics impact information processing requirements can refine job classifications and enhance systems for personnel recruitment, work placement, and training.

Future Directions and Conclusion

We propose that the heterogeneity of the predictive power of practice and of cognitive abilities stems in part from a failure to account for moderating task characteristics. Before correlational research examining individual differences in practice and cognitive abilities on real-world performance takes place, task characteristics should be experimentally tested. We found evidence that task characteristics interact with practice and cognitive resource availability.

The present studies build the first empirical step in a framework for multiple future directions. First, our paradigm is designed to continue investigations with other task characteristics (e.g., sequential vs. simultaneous, discrete vs. continuous) that allow cross-dimension comparisons. Second, our paradigm is designed to enable future investigation of dimensions' interactions. For example, comparing how tasks that are both dynamic and variably mapped impact learning and performance relative to tasks that are only dynamic or only variably mapped. Third, our paradigm is designed to examine individual differences in both practice and cognitive abilities on performance while manipulating task characteristics. These investigations will set the stage for carefully examining how task characteristics moderate the predictive power of both practice and cognitive resources on real-world task performance.

We believe task characteristics are important considerations in skill acquisition and expertise research. Specifically, examining a set of factors that potentially moderates both the predictive power of practice and cognitive resources on performance brings expertise research and individual differences research into the same theoretical model. By incorporating multiple factors into a theoretical framework we are more likely to increase our understanding of complex human performance.

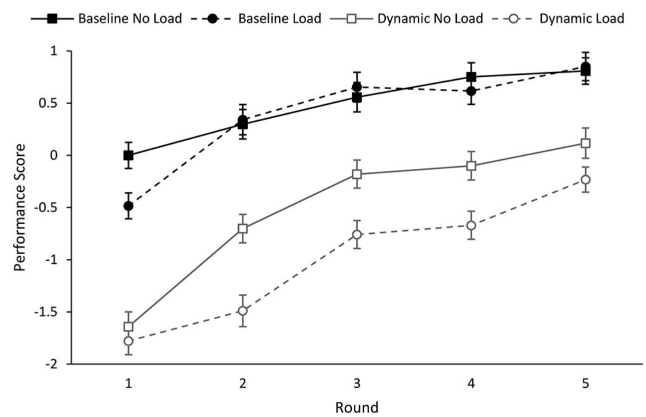

Figure 5. Experiment 2 mean performance on the Plants vs. Zombies task. Performance data are standardized around Round 1 performance in the baseline no load group, which has a score of 0. A score of 1 indicates performance one standard deviation above the Round 1 baseline no load group mean. Error bars indicate standard errors of the mean.

Table 7
Experiment 2 Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	*F* value	*p*-value
Version	1	115	103.16	<.001
Load	1	115	4.87	.029
Mission	1	115	.31	.577
Round	1	1,040	111.19	<.001
Round^2	1	1,040	29.54	<.001
Version × Load	1	115	.12	.734
Version × Mission	1	115	.75	.390
Load × Mission	1	115	.25	.620
Round × Version	1	1,040	2.52	.113
Round × Load	1	1,040	.06	.806
Round × Mission	1	1,040	.10	.747
Round^2 × Version	1	1,040	.10	.748
Round^2 × Load	1	1,040	.34	.559
Round^2 × Mission	1	1,040	1.71	.191
Version × Load × Mission	1	115	.01	.938
Round × Version × Load	1	1,040	9.57	.002
Round × Version × Mission	1	1,040	1.81	.179
Round × Load × Mission	1	1,040	.03	.868
Round^2 × Version × Load	1	1,040	8.22	.004
Round^2 × Version × Mission	1	1,040	1.80	.180
Round^2 × Load × Mission	1	1,040	.00	.983
Round × Version × Load × Mission	1	1,040	.04	.849
Round^2 × Version × Load × Mission	1	1,040	.03	.855

Note. Type III fixed effects are interpreted as they would be in an ANOVA. For example, the main effect of version, test the mean difference between the two task versions after collapsing across all rounds and each level of load and mission. By contrast the effect estimates and their corresponding *t*- and *p*-values reported in Table 6, are conditional to all other effects, and are explained in the estimate description column. For theoretical testing, we focus on the Type III fixed effects and use the regression estimates to determine the direction of main effects and interactions. Version = task version (baseline, dynamic); Load = cognitive load (load, no load); Round = linear improvement over rounds; Round^2 = nonlinear (quadratic) improvement over rounds.

References

Ackerman, P. L. (1986). Individual differences in information processing: An investigation of intellectual abilities and task performance during practice. *Intelligence, 10,* 101–139. http://dx.doi.org/10.1016/0160-2896(86)90010-3

Ackerman, P. L., & Cianciolo, A. T. (2002). Ability and task constraint determinants of complex task performance. *Journal of Experimental Psychology: Applied, 8,* 194–208. http://dx.doi.org/10.1037/1076-898X.8.3.194

Ackerman, P. L., & Woltz, D. J. (1994). Determinants of learning and performance in an associative memory/substitution task: Task constraints, individual differences, volition, and motivation. *Journal of Educational Psychology, 86,* 487–515. http://dx.doi.org/10.1037/0022-0663.86.4.487

Atari. (1972). *Pong*. Sunnyvale, CA: Atari Inc.

Clément, E., & Richard, J. (1997). Knowledge of domain effects in problem representation: The case of Tower of Hanoi. *Thinking & Reasoning, 3,* 133–157. http://dx.doi.org/10.1080/135467897394392

Dawson-Saunders, B., Feltovich, P. J., Coulson, R. L., & Steward, D. E. (1990). A survey of medical school teachers to identify basic biomedical concepts medical students should understand. *Academic Medicine, 65,* 448–454. http://dx.doi.org/10.1097/00001888-199007000-00008

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100,* 363–406. http://dx.doi.org/10.1037/0033-295X.100.3.363

Feltovich, P. J., Spiro, R. J., & Coulson, R. L. (1989). The nature of conceptual understanding in biomedicine: The deep structure of complex ideas and the development of misconceptions. In D. A. Evans & D. A. Patel (Eds.), *Cognitive science in medicine: Biomedical modeling* (pp. 113–172). Cambridge, MA: The MIT Press.

Feltovich, P. J., Spiro, R. J., & Coulson, R. L. (1993). Learning, teaching, and testing for complex conceptual understanding. In N. Frederiksen, R. Mislevy, & I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 181–217). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fitts, P. M. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 243–285). San Diego, CA: Academic Press.

Hoffman, R. R., Ward, P., Feltovich, P. J., Dibello, L., Fiore, S. M., & Andrews, D. H. (2014). *Accelerated expertise: Training for high proficiency in a complex world*. New York, NY: Psychology Press.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96,* 72–98. http://dx.doi.org/10.1037/0033-2909.96.1.72

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology, 294,* 248–294. http://dx.doi.org/10.1016/0010-0285(85)90009-X

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (2000). *SAS system for mixed models* (4th ed.). Cary, NC: SAS Institute.

Macnamara, B. N., Hambrick, D. Z., & Oswald, F. L. (2014). Deliberate practice and performance in music, games, sports, education, and professions: A meta-analysis. *Psychological Science, 25,* 1608–1618. http://dx.doi.org/10.1177/0956797614535810

McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General, 141,* 302–320. http://dx.doi.org/10.1037/a0025250

Namco. (1980). *Pacman*. Tokyo, Japan: Author.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Nintendo. (1984). *Duck hunt*. Kyoto, Japan: Author.

PopCap Games. (2009). *Plants vs. zombies*. Redwood City, CA: Electronic Arts.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., de Fruyt, F., & Rolland, J. P. (2003). A meta-analytic study of general mental ability validity for different occupations in the European community. *Journal of Applied Psychology, 88,* 1068–1081. http://dx.doi.org/10.1037/0021-9010.88.6.1068

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology, 86,* 162–173. http://dx.doi.org/10.1037/0022-3514.86.1.162

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools, Inc.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84,* 127–190. http://dx.doi.org/10.1037/0033-295X.84.2.127

Wai, J. (2014). Experts are born, then made: Combining prospective and retrospective longitudinal data shows that cognitive ability matters. *Intelligence, 45,* 74–80. http://dx.doi.org/10.1016/j.intell.2013.08.009

Received February 3, 2017
Revision received July 25, 2017
Accepted July 27, 2017 ■