THE EFFECTS OF INCUBATION TASK CHARACTERISTICS ON IDEA GENERATION

by

ADRIAN ABELLANOZA

DISSERTATION

Submitted in partial fulfillment of the requirements

for the degree of Doctor of Philosophy at

The University of Texas at Arlington

December, 2021

Arlington, Texas

Supervising Committee:

Jared Kenworthy, Supervising Professor

Paul Paulus

Laura Marusich

Lauren Coursey

Hunter Ball

Abstract

The effect of short breaks on creativity and idea generation, referred to as “incubation”, has had a considerable body of work devoted to its causal mechanisms. Despite this, many issues persist in the research literature, including difficulty in testing between similar competing theoretical frameworks, mixed findings regarding incubation effects, as well as measurement and practicality issues related to the use of non-expert human raters. Therefore, the present study combines methods used in recent research to evaluate which proposed mechanisms are supported. Additionally, to explore potential solutions to rater issues, a novel research methodology is explored using text mining methods and a lexical database called WordNet (Fellbaum, 1998). Results indicate a slight incubation effect for novelty of ideas during a standard divergent thinking task but failed to find support for the mind-wandering account of incubation effects. Instead, the pattern of results suggests limited support for the unconscious work theory of incubation. Additionally, results support the notion that WordNet-based measures of category flexibility are highly correlated with human-judged categories, which may prompt future research into similar methods as an alternative to the use of human raters.

Keywords: creativity, incubation, breaks, WordNet, fixation, unconscious work, mind-wandering

Acknowledgements

I would like to thank my dissertation committee members Drs. Jared Kenworthy, Paul Paulus, Lauren Coursey, Laura Marusich, and Hunter Ball each for their unique contribution to my dissertation, as well as the impact their instruction has had on my academic career. I would also like to thank Rebecca Roten, Jacob Lopez, Emily Harding, Rachel Cole, Tabitha Griffin, and Michael Byrne for serving as human raters for this project, for without their efforts the study would not have been possible.

Dedication

I'd like to thank all my teachers, mentors, friends and family for their support throughout my graduate training, as they motivate me to work harder, think more deeply, and constantly improve myself as both a scientist and as a person. I would especially like to thank my parents, Abner and Evangeline Abellanoza, as well as my sister, Cheryl Abellanoza, for their unending love, support, and guidance throughout my life. I would also like to dedicate this work to my girlfriend, Kimberly Bruner, whose devotion and encouragement has taught me more about life than I could ever hope to learn on my own. All my accomplishments—past, present, and future—are the product of all your support.

Table of Contents

The Effects of Incubation Task Characteristics on Idea Generation

While the history of creativity research is long and storied, creativity and innovation have become increasingly important in recent decades. Organizations rely on innovation to maintain a competitive advantage in their respective markets or fields (Alves et al., 2007). For example, technology companies have pushed to create supportive “innovation climates” in which employee innovation is championed (Büschgens et al., 2013). The US military has long invested in the study of team dynamics, including the creativity of teams, to maximize their performance on missions (Goodwin et al., 2018). Thus, various ways of enhancing creativity have been explored by researchers. A variety of factors at different levels of analysis have been identified and explored. Among these are individual-level factors such as verbal fluency and personality of individuals, team-level factors such as diversity and leadership, as well as organizational-level factors like creativity culture and resource management within organizations. These factors influence how often individuals and teams engage in creative behavior, share ideas, and communicate them across the organization. One area of research is concerned with the effect of breaks on creativity. This project will test hypotheses derived from theories about the role of breaks in creativity, as well as explore the use of newer tools in creativity research.

One aspect of creativity that has developed significantly over the past few decades is idea generation, or “brainstorming”. Researchers have attempted to explain the link between verbal ability, associative memory, and idea output. The context in which idea generation occurs also seems to influence the quality and quantity of ideas generated. Brainstorming is affected depending on whether ideas were generated via electronic, written, or verbal means. Additionally, individuals tend to outperform groups, with social factors such as production blocking, evaluation apprehension and social comparison assumed to be the cause of the relatively poor performance of groups, especially during verbal brainstorming (Diehl & Stroebe, 1987; Brown & Paulus, 1996; Paulus et al., 2002). However, brief interruptions or breaks in idea generation have not been as extensively studied. Most studies have participants complete tasks without interruption. The continuous nature of most laboratory creativity research is not realistic from the perspective of industry and business; breaks are likely to occur, and creative problems may take years to solve (Dijksterhuis & Meurs, 2006). While the effects of short breaks have been discussed in previous literature, much is still unknown about the process driving their potential benefit, referred to as “incubation effects” (Smith & Blankenship, 1989).

A variety of measures for factors related to creativity, such as personality factors and final product quality, have been devised (Feldhusen & Goh, 1995; Zhou & Shalley, 2011). However, objectively measuring creativity has proven to be difficult for researchers. Researchers have been skeptical about the subjective nature of creativity and the field’s overreliance on non-expert judgments, questioning the validity of creativity research. Measures of performance in idea generation tasks typically involve a set of raters judging participants’ written or typed responses to a task prompt. Idea quantity is considered a more objective measure of performance and is usually measured by counting ideas generated during a task. Other measurements, like idea quality, feasibility, or novelty, require human judgment and usually take the form of Likert-type items or similar measurements that attempt to quantify creative dimensions. These raters are typically trained for each specific performance measure necessary for a given study (i.e., how to rate novelty, feasibility, etc.). After ratings are recorded, reliability is assessed among raters. While the use of human raters is a flexible and convenient way to obtain holistic appraisals of creative performance, their use may also present methodological issues, including the representativeness of raters and the degree to which findings are affected by rater effects. Other methods that limit or prevent these issues should also be explored.

Considering these gaps in the current research literature, this dissertation project aims to examine the effects of incubation task characteristics on performance in an idea generation task. One main contribution of this study is an attempt to replicate how the type and difficulty of an intermittent task influences the incubation process (e.g., Sio & Ormerod, 2009; Gilhooly et al., 2013), and in turn their effects on idea generation and category flexibility. Additionally, the attempt to replicate findings related to task-unrelated thoughts during incubation, also known as “mind-wandering” (Baird et al. 2012; Smeekens & Kane, 2016; Steindorf et al., 2020), will help address some of the theoretical debate surrounding the proposed mechanisms underlying incubation. Finally, the present study investigates the use of computerized text analyses as a complement to traditional research using human raters.

Literature Review

Creativity, Divergent Thinking, and Measurement

Though the exact definition has varied considerably over the history of creativity research, creativity can be defined as the complex process in which individuals create novel solutions or products in response to a problem or goal (Feldhusen & Goh, 1995). Many factors play a role in this process. These include individual factors, such as intelligence and personality, as well as contextual and social factors (Feist, 2010; Feldhusen & Goh, 1995). Empirical creativity research has mainly focused on personality and cognitive factors. Amabile’s (1983, 2011) theory of componential creativity was among the first to describe social and environmental factors in creativity, alongside individual differences such as personality, domain expertise, and task motivation. Experimental methods in studies of creativity usually include some open-ended task which allows participants to respond in novel ways, versus a straightforward task that has a clear solution (Zhou & Shalley, 2011). Types of tasks used may include artistic tasks, such as the creation of a painting, or problem-solving tasks that reflect specific contexts, such as the design of a product or solving management-related issues (Zhou & Shalley, 2011). The type of task is typically chosen by researchers to match the facet of the creative process that is of interest. For example, Lu and Luh (2012) compared different product assessment methods using a Taiwanese design competition.

Of course, a design competition is not the only method of assessing creativity. One important aspect of creativity is divergent thinking, or the ability to generate different potential solutions or ideas. For example, in the Unusual Uses task (Guilford et al., 1978; Silvia, 2011), also known as the Alternative Uses Task (Gilhooly et al., 2007), participants are asked to come up with a variety of uses for various objects, such as a tire, brick, or chair. Tasks like these are among the most widely used and validated for studying creativity (Gilhooly et al., 2007; Silvia et al., 2013). An important aspect of this type of research is that higher rates of idea generation (i.e., *fluency*) should result in an increased probability of generating unique ideas merely as a function of the total number of ideas generated. Consistent, strong correlations between fluency and overall creativity have been demonstrated (Silvia et al., 2013). However, Silvia and colleagues (2013) recommend that researchers take steps to avoid conflating verbal fluency (i.e., merely the rate of word or idea generation) with divergent thinking. These include prompting participants specifically to think of creative uses, as well as having raters judge idea quality in addition to idea output (Silvia et al., 2013).

The quality of ideas in divergent thinking tasks is often assessed through a panel of judges, also referred to as raters or coders. Raters undergo training, sometimes referred to as frame-of-reference training, in which raters are familiarized with the definitions of different aspects of idea quality, such as novelty, usefulness, and feasibility (Zhou & Shalley, 2011). Interrater reliability is then assessed, usually with an intraclass correlation coefficient (Koo & Li, 2016). If this value is sufficiently high, with a typical cutoff being a value of .70, scores are averaged across raters to produce a final score for each aspect of creativity (Zhou & Shalley, 2011). The use of raters is an important tool in creativity research, as it gives researchers a way to examine the subjective aspects of idea generation. For example, Coursey and colleagues (2018) used raters to assess the novelty of ideas generated asynchronously in an online discussion board. The use of raters is not necessarily limited to rating the final product(s) of a creativity task. Firestien and McCowan (1988) used raters to count the number of instances of various behaviors during group idea generation, including indications of criticism, support, and humor. Some of our preliminary work on group creativity in "hackathon"-like settings (Paulus et al., 2021) uses raters to assess behaviors such as designation of roles and displays of positive emotion.

However, the procedure does not come without its flaws. Rater effects pose threats to the validity of creativity measurement and are not often examined in creativity research (Hung et al., 2012). For example, a "halo effect" can occur in which raters are biased towards rating one item or attribute similarly to others that they have rated (Hung et al., 2012). Rater judgments can also be affected by severity or leniency, in which raters consistently tend to rate items more (or less) harshly than others (Hung et al., 2012). While reliability analyses account for the consistency between raters, they do not guarantee that the measure is valid. Hung and colleagues (2012) investigated potential rater effects in a study of constrained creativity using a many-facets Rasch model, a psychometric technique which allows researchers to examine effects like item difficulty and rater severity alongside creative performance. In their study, rater effects were not found, but the researchers suggest that rater effects be examined more carefully in future research (Hung et al., 2012). Additionally, the choice of who to use as raters is an important methodological decision. Some researchers have argued that only those with a certain level of expertise can serve as proper judges. For example, to judge products of a drawing task, art experts or professional artists would be most appropriate (Kaufman & Baer, 2012). Considering the issues that may arise using human raters, other forms of assessing creative performance should be explored.

Breaks, Incubation, and Creativity

While divergent thinking tasks serve as an approximation of how some creative ideas are generated, the context in which they are performed is not totally representative of the creative process. People may spend multiple sessions generating or refining ideas, across long periods of time. Almost any task requiring mental or physical effort will eventually require a break or rest period. The beneficial effects of breaks have been studied in industrial psychology research since the early 20th century (Mayo, 1924). Smith and Blankenship (1989) were some of the first to describe cognitive effects from breaks as "incubation", or increased performance after a delay in working on a problem. Since then, various theories explaining the benefit of breaks to creative tasks have been proposed. This section will summarize some of the notable findings in recent creativity research, delineate differences between types of incubation theories, and discuss areas of research that remain underexplored.

Considering the myriad of potential factors in creative problem-solving and idea generation, researchers have sought to develop models that describe effects at multiple levels of analysis (e.g., micro to macro). Nijstad and Stroebe (2006) developed a cognitive model that describes the retrieval of ideas from long-term memory. Dubbed the "search for ideas in associative memory", or SIAM model, the authors describe how different cues lead to a deep exploration of specific domains in semantic space. Similarly, Doboli and Brown (2010) simulated cognitive, social, and motivational factors using a dynamic model, combining previous models by describing the search and evaluation of ideas while accounting for social factors such as production blocking and social comparison.

In a review of effective brainstorming, Paulus and Kenworthy (2019) discuss these different cognitive models of idea generation, present their own model, and review various factors in both individual and group brainstorming. In this review, the authors outline their cognitive-social-motivational model of group creativity. Like the previous researchers (Doboli & Brown, 2010; Nijstad & Stroebe, 2006), they discuss how cognitive aspects of idea generation, represented by Brown and Paulus's (2002) associative memory model (AMM), is influenced by social and motivational factors. The authors also describe methods that enhance brainstorming performance, such as providing instructions (e.g., Osborn, 1958), using facilitators, using cues to tap different semantic categories, and brief breaks. While discussing the potential benefit of breaks, the authors discuss fixation as a potential cause of the slowdown of idea generation. Additionally, the authors indicate that more research is needed on the length and number of breaks.

Since these models generally describe creative performance as resulting from factors across all these levels, research into incubation effects has also explored a variety of different contexts, individual differences, types of problems, and different incubation tasks. Additionally, various theories about why incubation might influence creative performance have been presented. Beeftink et al. (2008) examined the effect of interruptions on performance in solving word problems. In one condition, participants were not allowed to switch between puzzles; in two other conditions, participants were either forced to switch or allowed to switch at their discretion. The authors found that those who switched at their discretion outperformed the other conditions, attributing the positive effects of switching-at-will to limiting negative emotions, such as frustration or confusion (Beeftink et al., 2008). Madjar and Shalley (2008) also examined the effects of having the discretion to switch tasks, as well as goal setting, on incubation and creative performance. In contrast to Beeftink and colleagues' work (2008), the authors found that having the discretion to switch did not affect creativity. However, an interaction between goal setting and task switching discretion was found. The condition that allowed participants to switch, while having goals for both creative tasks as well as the incubation task, yielded the highest creative performance. The authors explain the effects in terms of increased cognitive stimulation, pointing to increased demand during the incubation task as the root cause of incubation effects (Madjar & Shalley, 2008).

When individuals are asked to generate ideas collaboratively, social factors also influence whether participants share ideas, as well as the quality of their ideas. In a laboratory setting, Harinck and De Dreu (2008) showed a benefit of brief breaks in buyer-seller negotiations, especially when participants were distracted by another task during the break, or when cooperative thinking was induced during the break. Breslin (2019) examined the effect of off-task breaks on idea generation in groups. All groups completed the Unusual Uses Task (UUT) as the main creativity task but differed based on an interpolated task. One condition had participants work together to build a tower using model bricks, which was thought to induce unequal involvement among participants. Another condition had participants sort model bricks into different colors, in which equal participation among group members was induced. Finally, a no-task break condition was also included, as well as a no-break condition in which participants did not stop working on the UUT. Overall, more original ideas were generated post-break only for the equal-participation, color-sorting task. Breslin (2019) frames the results as attributable to the limiting of negative emotions and competitive thinking (much like Beeftink et al., 2008), as well as limiting phenomena such as production blocking and free riding compared to the unequal participation condition. Zhou and colleagues (2019) examined the effect of evaluation apprehension and incubation on group idea generation. The authors found that when participants were exposed to others' ideas, evaluation apprehension reduced the number of ideas, as well as the number of categories of ideas one generated (i.e., flexibility). This effect was reduced when participants were not exposed to others' ideas. Incubation periods also reduced this effect, as well as increased the number of ideas one generated (Zhou et al., 2019).

Explaining Incubation Effects

Theories about incubation can be split generally into two types: conscious and unconscious (Williams et al., 2018). Conscious theories of incubation state that top-down cognitive processes are the primary driver of incubation effects (Williams et al., 2018). Recovery-from-fatigue theory (Gilhooly, 2016; Williams et al., 2018) attributes the benefits of an incubation period to rest and recovery from mental fatigue. For example, Paulus and Yang (2000) examined sharing ideas in a brainwriting context, finding that idea sharing in a group-writing paradigm outperformed nominal groups, and attributing this effect to enhanced cognitive stimulation. Additionally, Paulus and colleagues (2006) demonstrated a benefit of brief breaks during brainwriting sessions in which groups of participants wrote ideas down on paper, but this benefit was not present in electronic brainstorming sessions. The authors explain the effects of breaks as having limited fatigue in the more demanding writing conditions, as opposed to the electronic condition. Intermittent thought theory (e.g., Brown & Paulus, 1996; Gilhooly, 2016; Paulus & Nijstad, 2003; Williams et al., 2018) posits that individuals benefit from continuing to work and think about a problem during break periods, even during laboratory experiments that explicitly instruct participants not to do so.

Unconscious theories of incubation effects focus on unconscious thought alone, or both associative and attention-based mechanisms (Williams et al., 2018). Smith and Blankenship's (1989) explanation for improvement on a creative task after incubation was that more time away from a problem increases the chance that irrelevant or unhelpful information is discarded. The authors organized their explanation of the effects as the "forgetting-fixation hypothesis", also known as "beneficial forgetting". The association between fixation and the environment it was encountered in was further postulated as the "context-dependent fixation hypothesis" (Smith & Beda, 2020). Two other closely related theories of incubation, mind-wandering (Baird et al., 2012) and unconscious work theory (Dijksterhuis & Nordgren, 2006), also attempt to explain possible unconscious mechanism driving incubation effects. The mind-wandering theory of incubation posits that unrelated cognition (e.g., off-topic thinking or behavior) allows one to draw potentially useful associations from seemingly unrelated concepts (Baird et al., 2012; Barron et al., 2011). Like intermittent thought theory, unconscious work theory (Dijksterhuis & Nordgren, 2006) argues that an individual continues to work on a problem during an incubation period. However, a key aspect of this theory is that the individual is unaware that they are still devoting effort to finding a solution (Dijksterhuis & Nordgren, 2006). This lack of awareness is the main distinction between the two theories. In both unconscious work theory (Dijksterhuis & Nordgren, 2006) and forgetting-fixation theory (Smith & Blankenship, 1989), associative memory plays a key role. Each theory proposes different incubation mechanisms that increase leverage of long-term memory in finding high-quality solutions. In forgetting fixation theory, either memory decay of unhelpful categories, or reduction of interference during the incubation period, results in a performance increase (Gilhooly, 2016). In unconscious work theory, incubation effects are thought to occur due to unconscious and automatic spreading activation across an associative memory network (Gilhooly, 2016). In either case, incubation effects should drive a search for potential solutions in different, more distally related categories, especially when a particular category of ideas is exhausted (Iyer et al., 2009).

Some investigations (Ellwood et al., 2009; Baird et al., 2012; Gilhooly et al., 2012; Hao et al., 2014; Sio & Ormerod, 2009; Williams et al., 2018) have failed to find empirical support for conscious theories of incubation effects. A meta-analysis of incubation studies (Sio & Ormerod, 2009) found that participants given an undemanding cognitive task during the break (versus demanding tasks or no-incubation controls) had better performance on linguistic creativity tasks, contradictory to both intermittent thought theory and recovery-from-fatigue theory. Baird and colleagues (2012) found that participants' intermittent, explicit thoughts about a creativity task were not found to be related to quality of performance, and that participants who had a rest period between creative task sessions did not benefit from an incubation effect. Gilhooly and colleagues (2012) found that performance on interpolated tasks was shown to be unrelated to main task performance, contrary to both theories' predictions. In a review of the literature, Gilhooly (2016) characterized these findings as fatal flaws in conscious theories of incubation effects, particularly regarding the intermittent work hypothesis. Recovery-from-fatigue is also unlikely based on these results (e.g., Baird et al., 2012; Ellwood et al., 2009), as participants should perform optimally if allowed to rest during incubation periods and should not see increased performance relative to controls by performing interpolated tasks. Incubation effects found in immediate incubation studies (e.g., Gilhooly et al., 2012) in which participants are given an incubation task before starting the main creative task, also limit conscious theories' explanatory power, as there should be no effect of incubation if fatigue is not induced first. However, Ellwood et al. (2009)

discussed possible interpretations of their results under both fatigue recovery and unconscious work theories. Earlier work by Brown and Paulus (1996), as well as more recent work by Williams and colleagues (2018) state that both attentional and associative processes are likely to play a role in the creativity process. As a recent trend of support (e.g., Baird et al., 2012; Ellwood et al., 2009; Gilhooly et al., 2012; Williams et al., 2018) has been found for unconscious processes as a primary driver of incubation effects, the present study focuses on unconscious theories of incubation, with implications at the individual level. To highlight some of the differences between unconscious theories of incubation, the following sections describe a brief selection of relevant findings, according to which theory they appear to support.

Support for Unconscious Work Theory

In a review article covering incubation findings in creativity research, Ritter and Dijksterhuis (2014) argued that then-recent findings (e.g., Ellwood et al., 2009; Dijksterhuis & Meurs, 2006; Sio & Omerod, 2009) pointed to unconscious processes as having an important role in creative thinking. The authors also discussed the importance of neuroscientific research to confirm hypotheses related to unconscious thought, citing Creswell et al.'s (2013) findings that unconscious processes are associated with prefrontal cortex activity. Personality may also play a role; Gallate and colleagues (2012) examined incubation effects as a function of individual differences in creative ability, with results suggesting both that incubation effects arise due to unconscious processing, and that individuals higher in creative ability benefit more from incubation.

Instead of idea generation, Ritter et al. (2012) also examined the effect of unconscious thought on idea selection. Using an immediate incubation paradigm, participants generated ideas either immediately, after an incubation period where they were instructed to consciously think about the problem, or after a distraction task meant to induce unconscious thought. Afterwards, all participants were instructed to select their most creative idea, which was then compared against independent creativity ratings from trained raters. The authors found that while idea generation was similar between conscious and unconscious conditions, those in the unconscious condition more consistently selected their best ideas as judged by raters.

In an investigation of immediate incubation effects, Nam and Lee (2015) probed potential unconscious processing using a lexical decision task (LDT), in which participants judge whether stimuli presented on-screen are words or non-words. The authors posit that novel solutions are first accessed unconsciously, which is then transferred to a conscious level of processing. Response times on the LDT were used to assess unconscious processing but did not correlate with creative performance on the remote association task (RAT). Additionally, no immediate incubation effect was found (Nam & Lee, 2015).

Support for Forgetting-Fixation Theory

Using puzzles consisting of pictures and words dubbed "rebus" problems, Smith and Blankenship (1989) found support for forgetting-fixation theory with a series of four experiments. The results showed higher incubation effects for groups with higher forgetting of misleading initial cues (Smith & Blankenship, 1989). In a series of experiments on unconscious work theory, Dijksterhuis and Meurs (2006) found that participants who were asked to complete a distraction task before a creative task performed better on the latter, versus those who were told to consciously work on the task beforehand.

In comparing the forgetting fixation and conscious work hypotheses, Kohn and Smith (2009) found that incubation only affected performance on a Remote Associates Test (RAT) when fixation cues were given before the task, supporting the forgetting fixation interpretation of incubation. It should be noted that the RAT, first used by Mednick and colleagues in early experiments on incubation (e.g., Mednick et al., 1964) is more of a convergent thinking task (see Nam & Lee, 2015), as participants must come up with a solution that "fits" the three words presented to them.

Kohn and Smith (2011) tested the idea that fixation occurs in group settings due to members sharing ideas with each other, causing group members to conform to others' idea domains. Results indicated that fixation through group conformity did occur, restricting the breadth of ideas through exposure to others' ideas and making group members' ideas more like each other. This fixation, however, did not influence the total number of ideas generated by groups. However, taking a break stymied the natural decline of quantity and variety of ideas.

In a series of four experiments, Smith and colleagues (2017) tested the forgetting fixation hypothesis by having participants generate ideas in specific categories (e.g., birds, fattening foods, heavy things). In one condition, participants completed each category before moving on to the next in a sequence. In another condition, participants switched frequently between categories. The authors found that participants who switched between categories generated more ideas in total, as well as more novel ideas; this finding also held for participants who switched between idea generation and an unrelated maze-solving task (Smith et al., 2017).

Chiang and Chen (2017) examined whether fixation during divergent thinking tasks influenced creative outcomes. In a series of two experiments using both delayed and immediate incubation, the authors found that providing weak fixation cues, as opposed to robust or no fixation, produced more divergent ideas. The authors also found that immediate incubation, as opposed to delayed incubation, reduced the influence of both robust and weak fixation cues. The authors explain their results as being partially supportive of forgetting-fixation as opposed to unconscious work theory, as in the second experiment participants had unconscious work induced via a distraction task, but still generated a similar number of convergent ideas as an immediate generation group (Chiang & Chen, 2017). This runs contrast to predictions that unconscious thought is more likely to lead to divergent thinking (Dijksterhuis & Nordgren, 2006).

In a study on group creativity, Choi et al. (2019) examined the effect of value orientation and self-representation on group creativity. Value orientation refers to a focus on collective (collectivistic) versus personal (individualistic) goals, whereas self-representation concerns a perception of being separate (independent) versus connected (interdependent) with others (Choi et al., 2019). By priming each dimension, the authors found that more original ideas were generated when members held a collectivistic value orientation and independent self-representation, and that this relationship was mediated by a reduction in idea fixation. The authors suggest that social identity factors serve as a potential cause of fixation, which in turn influences creative performance (Choi et al., 2019).

To test the importance of context in fixation forgetting, Smith and Beda (2020) had participants complete a series of word problems known as the Remote Associates Test (RAT), which are overlayed on scenic images representing context. Greater performance on the RAT problems were found for those who were given a new context photo as opposed to the original “fixation” photo (Smith & Beda, 2020).

Support for Mind-Wandering

In a review on the effects of mind-wandering, Mooneyham and Schooler (2013) discuss several negative effects of mind-wandering, such as reductions in reading comprehension, and lower scores on tests of working memory and general intelligence. However, the authors discuss several potential benefits of mind-wandering, including the ability to cycle between streams of information, dishabituation of learning through a “refreshing” of attentional processes, and benefits to creative thinking.

Tan and colleagues (2015) used a number reduction task (NRT) to test the effect of mind-wandering on insight (i.e., a convergent thinking task). Participants had to respond to a series of number stimuli using a set of rules. The researchers “hid” a pattern in the responses, such that the correct answer to participants’ second response would always be the same as the seventh response. Mind-wandering was assessed with a retrospective Likert-type item. The authors found that those who discovered this hidden rule (“solvers”) reported significantly more mind-wandering than those who did not (“non-solvers”), suggesting that mind-wandering was important in solving this task (Tan et al., 2015).

In another investigation on insight, Zedelius & Schooler (2015) examined whether the effect of trait mind-wandering (operationalized as lack of mindfulness) differed based on different approaches to a creative task. The authors compared an approach in which participants gaining sudden awareness of a solution (defined here as insight) and an approach where participants explicitly search for an idea or solution through active reasoning (termed “analytic strategy”). In the first of two studies, participants self-reported which approach they took. In the second, they attempted to manipulate which strategy participants used through instruction. The authors found that mindfulness had a negative association with problem-solving on the RAT overall but found that mindfulness was positively associated with problem-solving when participants either self-reported or were instructed to approach problems analytically (Zedelius & Schooler, 2015).

Additionally, Gable and colleagues (2019) completed a series of two diary studies on physicists and writers. Participants were asked to record their most creative ideas and to describe their state of mind when the idea occurred. One-fifth of participants reported their most significant ideas during mind-wandering, defined by the authors as engaging in an activity other than work and while thinking about something unrelated to the generated idea. Participants were also more likely to report significant ideas as overcoming an impasse on a problem and representing an “aha” moment, versus ideas generated while on-task. The authors describe this as evidence of mind-wandering being crucial to the creative process.

Amongst all mind-wandering paradigms, one prominent study has been published by Baird and colleagues (2012), who found that those who had an undemanding task between creative sessions saw an increase in task performance, versus those who had a more demanding task, a rest period, or no break at all. The authors framed their results as support for the mind-wandering interpretation of incubation effects, as self-reported measures of mind-wandering were also found to be related to creative performance. In a similar fashion to Baird et al. (2012), Yamaoka and Yukawa (2016; 2019) completed two separate studies on mind-wandering. The authors performed an incubation study in which participants completed the Unusual Uses Task (UUT). In-between sessions of the UUT, participants completed an n-back task, then asked to rate their level of mind-wandering retrospectively. Results showed that participants with higher reported mind-wandering also scored more highly on flexibility and originality on the UUT (Yamaoka & Yukawa, 2016). In a follow-up study, the authors expanded on their previous work by including mood and working memory capacity as additional factors (Yamaoka & Yukawa, 2019). They found that those with higher mind-wandering did not differ on any creative outcome, though groups were significantly different on idea flexibility before adjusting for multiple comparisons. The authors also found an inverse relationship between mind-wandering and affect, suggesting that higher mind-wandering may worsen mood (Yamaoka & Yukawa, 2019). They discuss the potential impact of directing mind-wandering to other domains, like future- or self-related thinking, that have shown a relationship with more positive mood (Ruby et al., 2013; Yamaoka & Yukawa, 2019).

Disentangling Unconscious Theories of Incubation

Smith and Beda (2020) argue that, as Williams and colleagues (2018) note in the conscious vs. unconscious processes debate, unconscious incubation theories (i.e., mind-wandering, unconscious work, and forgetting fixation) need not be in direct contention to each other. For example, mind-wandering behavior can contribute to forgetting of unhelpful solutions through the exploration of alternative ones (Smith & Beda, 2020). Ritter and Dijksterhuis (2014) describe creative thinking as alternating between the two types of processes. This is echoed in additional commentary on Gilhooly's review (2016). Yuan and Shen (2016), as well as Gilhooly himself (2017), discussed the potential of dual-process theory to solve tasks, as well as the inherent differences between convergent and divergent idea generation tasks. These considerations make the various unconscious theories surrounding incubation difficult to tease apart. However, key theoretical differences between these theories exist. Combining existing research paradigms introduced to test their predictions offers ways of comparing theories of incubation more cleanly.

Unconscious work on a creative problem can be distinguished from mind-wandering, such that mind-wandering involves undirected internal thought that a person is aware of, but is not consciously initiated (Barron et al., 2011; Williams et al., 2018). In contrast, unconscious work is likely to be performed during a distracting task and is carried out without the person's awareness (Dijksterhuis & Nordgren, 2006). Mind-wandering is likely to occur during an incubation period with an undemanding task (Baird et al., 2012), whereas unconscious work is likely to occur during demanding or distracting tasks (Dijksterhuis & Meurs, 2006). One of the elements of unconscious work theory is that as the task becomes more complex, unconscious thinking is more likely to generate a quality solution (Dijksterhuis & Nordgren, 2006). Conscious, focused attention is of limited capacity and is thought to increase schema use. This was supported through empirical investigation on stereotyping (Dijksterhuis & Nordgren, 2006). In contrast, unconscious thought is more representative of a weighted summary of experiences and is by nature a more divergent process. Using performance on decision-making tasks as the target variable (i.e., selecting a roommate, selecting an apartment, etc.), Dijksterhuis (2004) experimentally manipulated conscious and unconscious thought before a creative task. The author induced conscious or unconscious thought by either allowing participants to think about the task beforehand, or distracting them (Dijksterhuis, 2004). Participants were asked to rate fictional roommates and apartments, some of which were constructed by the researchers to have more positive attributes than the others. Those in the unconscious thought condition had better performance on decision-making tasks versus those in the conscious condition, as they tended to correctly rate desirable roommates and apartments higher than less desirable ones (Dijksterhuis, 2004; Dijksterhuis & Nordgren, 2006).

Other researchers have also explored the effects of demand of interpolated tasks on creative problem solving, as well as their relation to mind-wandering and unconscious work theory. Stronger incubation effects when the incubation is spent performing an undemanding task were previously supported by systematic review (Sio & Ormerod, 2009). After finding similar support for the effects of undemanding tasks, Baird and colleagues (2012) interpreted their findings as supportive of the mind-wandering theory of incubation, as a self-report measure of mind-wandering was implemented as a manipulation check. While unconscious work is, by definition, unable to be assessed by self-report, one method of ascertaining participants' level of mind-wandering is to retrospectively survey them after performing a task. Researchers have used self-report measures of mind-wandering to explore the neural correlates of mind-wandering (Barron et al., 2011). Results supported the "decoupling" hypothesis of mind-wandering, which posits that mind-wandering occurs when attention is diverted away from external percepts and toward internal thoughts.

However, retrospective surveying has fallen under criticism in recent years (Smeekens & Kane, 2016; Weinstein, 2018) due to its vulnerability to memory decay. Instead, cognition researchers have argued for the use of time-sensitive "thought-probes" that are displayed in real-time during the task (Smeekens & Kane, 2016; Weinstein, 2018). Indeed, attempts to replicate Baird et al.'s (2012) findings failed in several additional investigations by Smeekens and Kane (2016), Steindorf et al. (2020), and Murray et al. (2021). One key addition of these studies is the use of thought probes. Smeekens and Kane (2016) attempted to replicate Baird et al.'s (2012) findings using an updated research paradigm. In a series of three experiments, the authors failed to find a relationship between task-unrelated thought

(TUT) measured using thought probes and performance on a divergent thinking task, with the third experiment being a close replication of Baird et al. (2012). In contrast to Baird et al. (2012), the authors report the results of both retrospective mind-wandering self-report measures as well as thought probes, with both failing to support the mind-wandering account of incubation. The authors also failed to find an effect of working memory capacity on creativity. Additionally, a similar replication by Steindorf et al. (2020) failed to find incubation effects across four different incubation conditions: one with thought probes, one with trivia question probes, one with no probes, and a no-incubation control.

In yet another attempted replication of Baird et al.'s (2012) findings, Murray and colleagues (2021) were unable to replicate the original finding that mind-wandering enhanced performance on the AUT, even after closely matching their experimental design to that of Baird et al. (2012). Additionally, the authors performed exploratory analyses using automated semantic distance scores from Beaty and Johnson's (2021) "semDis" tool, discussed further in the next section. The authors found similar (null) results for the effect of mind-wandering on both human-rated novelty and semantic distance calculated using semDis. One exception to this pattern is the work by Hao et al. (2015). Using the thought-probe method of recording mind-wandering, the authors found that higher mind-wandering reduced idea fluency and originality on the Alternative Uses Task, when probed during the task itself as opposed to during incubation.

Additionally, the type of task performed during the break has been experimentally manipulated to examine differences in the prediction between unconscious work theory and forgetting-fixation theory. Gilhooly and colleagues (2013) compared the predictions of unconscious work theory and forgetting-fixation by varying the type of task used. The interference-based mechanism of forgetting-fixation implies that a task of similar modality is more likely to cause forgetting of the unhelpful category or mindset (Gilhooly et al., 2013). In contrast, unconscious work would be impeded by a task that is too similar to the original, as the interpolated task will occupy the same underlying substrates (e.g., verbal, spatial) that continued unconscious work would otherwise be reserved for (Gilhooly et al., 2013). The researchers varied whether participants completed a spatial task or a verbal task. This was done for both the creative task as well as the interpolated task (Gilhooly et al., 2013). Support for the unconscious work interpretation of incubation was found, as significant incubation effects on idea fluency and overall creativity ratings were only found when the creative task and interpolated task were dissimilar, i.e. when the creative task was verbal and the incubation task was spatial, as well as vice versa (Gilhooly et al., 2013).

Therefore, this study is designed to help address the inconsistent findings regarding incubation effects, task complexity, and the demand of tasks during incubation periods. Specifically, the present study aims to examine how these factors affect individual performance on a divergent thinking task. The results of the study may have implications for why incubation effects occur in divergent thinking tasks, and under which conditions incubation effects are enhanced or attenuated. Specifically, the results could point to either forgetting-fixation, mind-wandering, or unconscious work as the most likely mechanism for incubation effects.

Text Analysis Software and WordNet

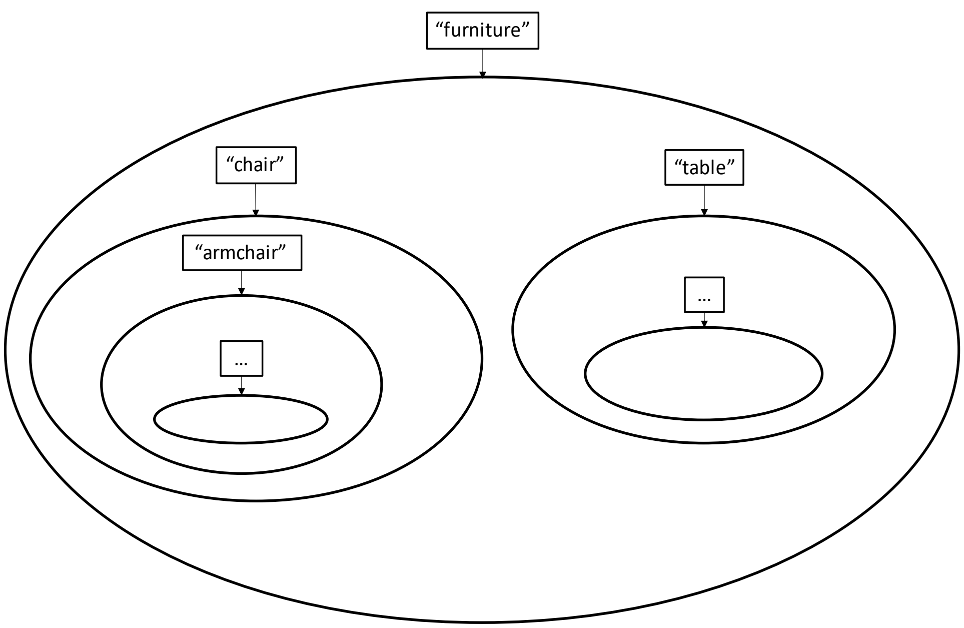
Methodological issues, in addition to theoretical issues, have also been considered in previous creativity research. The way that data is obtained and analyzed from idea generation tasks is also of interest. As alluded to earlier, issues exist with using human raters in quantifying creative performance. Rater variation in judgments/perception (Hung et al., 2012), as well as the expertise and/or authority of the raters' in making judgments (Kaufman & Baer, 2012), are threats to the validity of studies that use human raters. Exploring analysis methods that do not rely on human judgment is important not only for academic research, but also for businesses that rely on innovation or other creative processes to maintain a competitive advantage over their rivals (Jafari et al., 2020). Over the past few decades, technological advancements in personal computing have empowered researchers and data professionals to make use of computationally intense methods and/or large volumes of data. One area that has seen major advancement is the processing and analyzing of text. Tools to process and extract information from bodies of text are now standard within both industry and academic research. Business intelligence professionals regularly leverage methods such as sentiment analysis, in which customer attitudes toward products or services can be inferred by parsing through customer comments. This project will explore the utility of text mining to obtain measures of creative performance, taking advantage of the benefits that automation provides.

Recently, computational creativity research has applied text mining and machine learning techniques to analyze creativity data (Christensen et al., 2017; Toubia & Netzer, 2017). However, these techniques are usually not applied to generate understanding behind theories of creativity, but to either generate or detect creative ideas in an automated fashion. For example, Christensen and colleagues developed a classification model for detecting ideas in large amounts of online text (Christensen et al., 2017). Toivonen and colleagues (2015) also published work on a machine learning model that detects ideas from Internet sources. The increasing usage of the Internet ensures that researchers will have large swaths of text data available for decades to come. A particularly promising tool for text analysis is Beaty and Johnson's (2021) "semDis" tool. The automated tool combines several modeling approaches to latent semantic analysis (LSA), which is a natural language processing technique used to quantify the relatedness of text. Groups of ideas that are more "remote" will therefore have a greater semantic distance. In a series of five studies, the authors demonstrate strong correlations between human-rated novelty and semDis scores, as well as evidence of convergent validity with other measures of creativity.

While more advanced measures are needed to capture a multifaceted construct like novelty, a more simplistic measure can be used to assess the number of categories ideas cover (i.e., flexibility). Different kinds of machine-readable dictionaries organize words into different senses. One of the most popular of these dictionaries is WordNet (Fellbaum, 1998), a database of English words organized into synonym sets, or “synsets”. Each synset has a common meaning and has a unique identifier. A visual example of WordNet synsets is provided in Figure 1 below. Using WordNet, one can parse through a body of text and match individual words to synset classifications. This has made it popular in research domains involving text classification and quantifying similarity between text documents. Since associative memory and accessibility is such a key preceding factor in idea generation, the utility of a linguistic database that neatly organizes terms to related concepts may be impactful to future creativity research, especially as richer text sources become available.

Figure 1

Illustration of WordNet Synsets



Note. The above illustration shows the organization of nouns in WordNet. Nouns are organized into a hierarchy based on sets of synonyms, called “synsets”. Synsets may have superordinate or subordinate synsets, which represent more general categories or specific types/instances (Fellbaum, 1998).

Alternatively, we may consider WordNet synsets to be too specific. Refer to Figure 2 for an example. The words “shirt”, “pants”, and “garment” could all reasonably fall under the category of “clothing”. However, because “shirt” and “pants” are not direct synonyms for each other, they will not be counted as having the same synset. WordNet also organizes words and their corresponding synsets into distinct files, called lexicographer files. For nouns, there are 25 distinct categories of nouns that are relatively broad---one lexicographer file, for example, contains all words and synsets referring to man-made objects. Lexicographer file numbers can be identified and counted in the same manner as synset IDs. This will allow for a broader measure of category accessibility. Refer to Table 1 below for a full list of the noun categories within WordNet. This study used WordNet synsets and lexicographer file numbers as a way of objectively counting concepts in the text generated from participants' idea generation.

Figure 2

Comparison of Synset IDs and Lexicographer File Numbers

"hat" Synset ID: **103459694** Lexicographer File #:**6** (man-made object)	*"stone"* Synset ID: **109281147** Lexicographer File #:**17** (natural object)
"cap" Synset ID: **102925972** Lexicographer File #:**6** (man-made object)	*"rock"* Synset ID: **109281147** Lexicographer File #:**17** (natural object)

Note. The above diagram displays the differences in specificity between synset IDs and lexicographer file numbers. Here, "hat" and "cap" share the same lexicographer file number (denoting the general category "man-made object") but different synsets. This is despite their being closely related, though not interchangeable, terms. Conversely, "stone" and "rock" are true synonyms of each other, and therefore share the same synset ID and lexicographer file number---they are the same concept.

Table 1

WordNet Lexicographer Files – Nouns

File Number	Name	Contents
3	noun.Tops	unique beginner for nouns
4	noun.act	nouns denoting acts or actions
5	noun.animal	nouns denoting animals
6	noun.artifact	nouns denoting man-made objects
7	noun.attribute	nouns denoting attributes of people and objects
8	noun.body	nouns denoting body parts
9	noun.cognition	nouns denoting cognitive processes and contents
10	noun.communication	nouns denoting communicative processes and contents
11	noun.event	nouns denoting natural events
12	noun.feeling	nouns denoting feelings and emotions
13	noun.food	nouns denoting foods and drinks
14	noun.group	nouns denoting groupings of people or objects
15	noun.location	nouns denoting spatial position
16	noun.motive	nouns denoting goals
17	noun.object	nouns denoting natural objects (not man-made)
18	noun.person	nouns denoting people
19	noun.phenomenon	nouns denoting natural phenomena
20	noun.plant	nouns denoting plants
21	noun.possession	nouns denoting possession and transfer of possession
22	noun.process	nouns denoting natural processes
23	noun.quantity	nouns denoting quantities and units of measure
24	noun.relation	nouns denoting relations between people or things or ideas
25	noun.shape	nouns denoting two and three dimensional shapes
26	noun.state	nouns denoting stable states of affairs
27	noun.substance	nouns denoting substances
28	noun.time	nouns denoting time and temporal relations

Note. File numbers, names and content descriptions for WordNet lexicographer files pertaining to nouns (Fellbaum, 1998).

Therefore, a secondary aim of the study is to explore the utility of using WordNet categories as measures of flexibility. The primary feature of WordNet is its organization of words into categories based on similarity. The count of these categories neatly provides a reliable measure of flexibility, as the categories are simply based on the text submitted by the participant—specific words will always share the same identification number. This contrasts with the rater method typically used by researchers (e.g., Lu et al., 2017). This study explores its convergence with rater judgments and its potential utility in creativity research.

Summary and Study Hypotheses

The effect of incubation periods on individual idea generation and its underlying mechanism is still unclear, despite a growing amount of research around the topic. To understand more about the conditions and mechanisms underlying incubation effects, the present study attempts to expand on previous findings, specifically borrowing from the designs of Baird et al. (2012) and Smeekens and Kane (2016). The present study attempted to replicate findings from these and other studies and to disentangle incubation theories through path modeling. Participants underwent two sessions of a divergent thinking task, with one of four versions of an incubation task, which varied by difficulty and modality (i.e., verbal or spatial). Mind-wandering was assessed using thought probes during the incubation task. Generally, the hypotheses are framed under mind-wandering theory, with the goal being to test any mediation of mind-wandering between the experimental manipulations and outcomes on the idea generation task. This was also done because of the intent to replicate the studies mentioned above. However, assessing the path model should also shed light on whether mechanisms proposed under the other theories (unconscious work and forgetting fixation) are supported. As a secondary aim, the study also examines the utility of text classification using WordNet within creativity research.

Hypotheses 1a through 1e concern the demand of the task during the incubation period. Previous research (Baird et al., 2012; Sio & Ormerod, 2009) has found a beneficial effect of undemanding incubation tasks on creative performance, as opposed to a demanding task or rest period. Evidence in favor of mind-wandering theory has also been found using retrospective self-report measures of mind-wandering (Baird et al., 2012), but the more robust method of thought probing used by Smeekens and Kane (2016) and later Steindorf et al. (2020) failed to replicate these results.

Therefore, Hypotheses 2a through 2d are positioned as an additional attempt to replicate Baird et al.'s (2012) results, which support a mind-wandering interpretation of incubation effects. In contrast, unconscious work theory posits that distraction away from conscious attention on the main task is the main driver of incubation effects (Dijksterhuis & Meurs, 2006; Dijksterhuis & Nordgren, 2006); thus, a higher-demand task would yield greater distraction and therefore greater creative performance. Furthermore, forgetting-fixation theory posits that fixation should be reduced regardless of the demand of the interpolated task (Gilhooly, 2016; Segal, 2004).

Hypothesis 1a: Those in undemanding break task conditions will generate more ideas than those in demanding conditions.

Hypothesis 1b: Those in undemanding break task conditions will have higher human-rated flexibility than those in demanding conditions.

Hypothesis 1c: Those in undemanding break task conditions will have higher WordNet synset counts than those in demanding conditions.

Hypothesis 1d: Those in undemanding break task conditions will have higher WordNet lexicographer file counts than those in demanding conditions.

Hypothesis 1e: Those in undemanding break task conditions will have higher novelty than those in demanding conditions.

Additionally, the effect of similarity of breaks (i.e., verbal vs. spatial) is differentially predicted by unconscious work theory and forgetting fixation theory. Under unconscious work theory, a task of similar modality would interfere with unconscious work on the main task (Gilhooly et al., 2013). However, under forgetting-fixation theory, a task of similar modality would be more effective at inducing forgetting (Gilhooly et al., 2013). While there are no direct predictions for demanding vs.

undemanding tasks under mind-wandering theory, the verbal nature of mind-wandering (Bastion et al., 2017) suggests a similar prediction as unconscious work theory, as performing a verbal task would hamper mind-wandering. Therefore, a spatial interpolated task was expected to produce more pronounced incubation effects.

Hypothesis 2a: Those in spatial break task conditions will generate more ideas than those in verbal conditions.

Hypothesis 2b: Those in spatial break task conditions will have higher human-rated flexibility than those in verbal conditions.

Hypothesis 2c: Those in spatial break task conditions will have higher WordNet synset counts than those in verbal conditions.

Hypothesis 2d: Those in spatial break task conditions will have higher WordNet lexicographer file counts than those in verbal conditions.

Hypothesis 2e: Those in spatial break task conditions will have higher novelty than those in verbal conditions.

The next set of hypotheses concern how mind-wandering relates to both the incubation conditions and the creativity outcomes. Based on previous literature, demand should influence the amount of mind-wandering, such that more difficult tasks should coincide with higher mind-wandering. No previous work has examined the effect of task similarity on mind-wandering. As stated previously, due to the role of inner speech and language in mind-wandering (Bastian et al., 2017), as well as the findings for a beneficial effect of switching modalities predicted by unconscious work theory (Gilhooly et al., 2013), a higher level of mind-wandering is hypothesized.

Hypothesis 3a: Those in the undemanding break task conditions will have higher levels of mind-wandering than those in demanding break task conditions.

Hypothesis 3b: Those in the spatial break task conditions will have higher levels of mind-wandering than those in verbal break task conditions.

Next, for mind-wandering theory to be fully supported, a mediation effect is implied wherein mind-wandering should mediate the relationship between the task conditions and creative performance. Therefore, Hypotheses 4a through 4e posit that mind-wandering will mediate the relationships between incubation task condition and each of the creativity outcomes.

Hypothesis 4a: Mind-wandering will mediate the relationship between the break task conditions and idea generation, such that changes in mind-wandering due to condition will then result in higher idea generation.

Hypothesis 4b: Mind-wandering will mediate the relationship between the break task conditions and human-judged categories, that changes in mind-wandering due to condition will then result in a higher number of categories.

Hypothesis 4c: Mind-wandering will mediate the relationship between the break task conditions and WordNet synsets, such that changes in mind-wandering due to condition will then result in higher counts of WordNet synsets.

Hypothesis 4d: Mind-wandering will mediate the relationship between the break task conditions and WordNet lexicographer files, such that changes in mind-wandering due to condition will then result in higher counts of WordNet lexicographer files.

Hypothesis 4e: Mind-wandering will mediate the relationship between the break task conditions and novelty, such that changes in mind-wandering due to condition will then result in higher counts of novelty.

Finally, Hypothesis 5 concerns the relationship between the three measures of category flexibility. Since WordNet synset and lexicographer file counts provide a way of identifying unique nouns used in each idea, the number of unique WordNet synsets and lexicographer files should be strongly, positively correlated to the number of unique categories as judged by human raters.

Hypothesis 5: The number of unique WordNet synsets and lexicographer files, as well as human-judged categories, will be positively correlated with each other.

Methods

Participants

A total of 200 participants were recruited for the present study. Participants were recruited using Prolific, an online platform in which volunteers can participate in studies for compensation. Each participant was paid $7.25 per hour for participating. Participants could enter the study provided they were at least 18 years old, had access to a desktop or laptop computer, and had no significant visual impairment (e.g., color blindness). All participants remained anonymous, identifiable only through Prolific ID numbers. Among the recruited 200 participants, 40 were removed from analysis. These included participants whose session data were not saved ($n = 7$) or were incomplete ($n = 2$), those whose work was rejected ($n = 1$), those who did not pass an attention check (i.e., had multiple non-responses to thought probes; $n = 9$), and those who spent too much time in between tasks or timed out of the session ($n = 21$). This resulted in a final sample of $N = 160$ participants. Demographic information for the final sample can be found in Table 2 below.

Table 2

Demographics and Descriptive Statistics

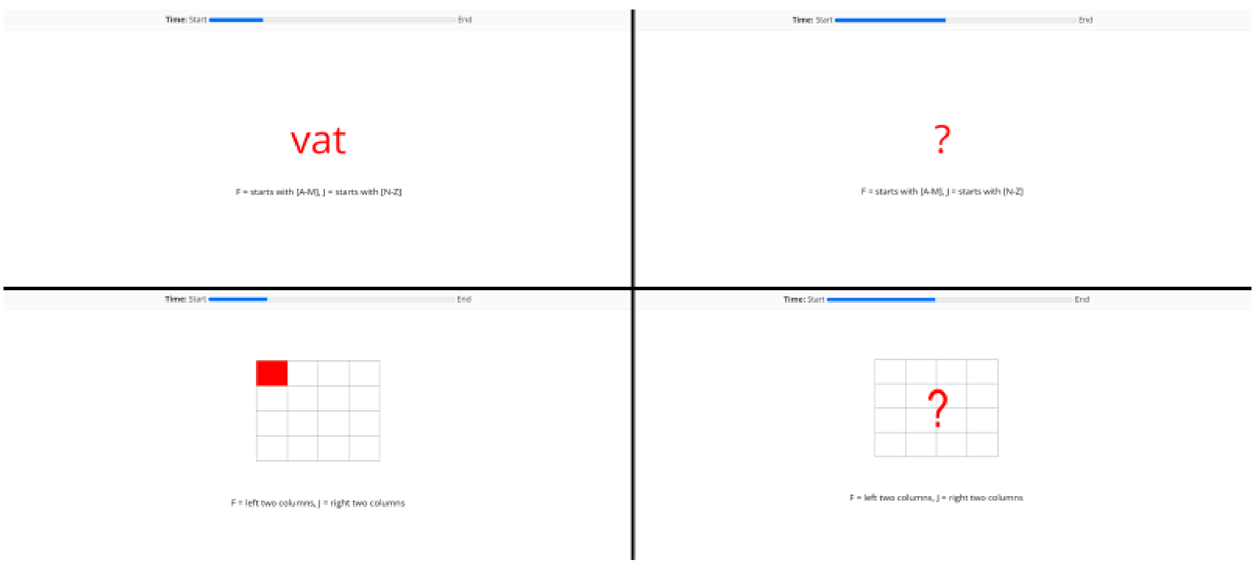
Variable	Group	*N*
Gender	Female	105
	Male	50
	Nonbinary/Other	5
Race	African American/Black	13
	Alaskan Native/Native American	2
	Asian	16
	Caucasian/White	104
	Hispanic/Latino	19
	Middle Eastern/North African/Arab American	2
	Other	4
Condition	Verbal/Easy	37
	Verbal/Hard	49
	Spatial/Easy	38
	Spatial/Hard	36
Variable	*M*	*SD*
Age	28.82	10.40
Pre-Task Fluency	12.02	5.02
Ideas – Block 1	33.58	15.12
Ideas – Block 2	34.28	20.65
Synsets – Block 1	58.72	27.00
Synsets – Block 2	57.50	31.00
Lexicographer Files – Block 1	15.49	4.05
Lexicographer Files – Block 2	14.88	4.51
Categories – Block 1	10.48	3.35
Categories – Block 2	10.26	4.21
Novelty – Block 1	2.05	0.33
Novelty – Block 2	2.14	0.43
On-Task Thought Count	8.49	3.92
Shoelace Thought Count	0.46	1.18
Mind-Wandering Count	5.05	3.80
Proportion Correct – N-Back	0.87	0.19
Idea Difference	0.69	12.01
Synset Difference	-1.22	17.22
Lexicographer File Difference	-0.62	3.05
Category Difference	-0.21	2.98
Novelty Difference	0.09	0.35
SDMWS - Spontaneous Subscale	4.80	0.35
SDMWS - Deliberate Subscale	5.03	0.35

Experimental Design

Participants were randomly assigned to one of four conditions of a 2 [modality: verbal or spatial] x 2 [difficulty: easy or hard] between-subjects design. Random assignment was done after participants agreed to the informed consent document, with participants assigned an internal variable denoting which n-back task was displayed. All participants completed the incubation task between the two idea generation sessions. An example of each of the conditions is shown in Figure 3 below.

Figure 3

N-Back Stimuli by Condition



Note. Top left = verbal/easy, top right = verbal/hard, bottom left = spatial/easy, bottom right = spatial/hard. The “easy” condition prompts are identical to the non-response stimuli, save for the coloring (red instead of black) and the instructions on which key to press. In these conditions, participants make a judgment on the current stimuli presented (i.e., a 0-back task). The “hard” conditions replaced the stimuli with a question mark, and task the participant with judging the previous stimuli (i.e., a 1-back task).

Materials

Study Interface. The study was administered completely online using a website hosted on UT Arlington servers. The website was coded in JavaScript using the jsPsych library (de Leeuw, 2015), a toolkit for creating cognitive experiments. The library allows for a large degree of flexibility in creating behavioral and cognitive experiments, including options for obtaining survey responses as well as presentation and timing of various stimuli. Experiments coded using jsPsych also show response times comparable to standard tools like E-Prime (de Leeuw, 2015).

Tendency to Mind-Wander. The Spontaneous and Deliberate Mind-Wandering Scale, or SDMWS (Marcusson-Clavertz & Kjell, 2018), was used to assess propensity to mind-wandering. This trait-level measure of mind-wandering was used to control for individual differences in mind-wandering. Example items include "I find my thoughts wandering spontaneously" and "I find mind wandering is a good way to cope with boredom". Items were rated from "1" = rarely/not at all true to "7" = a lot/very true. Subscales for both spontaneous and deliberate mind-wandering were used as variables in analysis. A full list of items can be seen in Appendix A.

Fluency Task. A word fluency task was used to account for individual differences in the ability to generate ideas. Participants generated words and short phrases according to three prompts: "science", "business", and "technology". Participants were given one minute for each prompt to generate ideas, and the average submission count of these three one-minute sessions was taken as a pre-task fluency measure.

Alternative Uses Task. The Alternative Uses Task, or AUT (Gilhooly et al., 2007; Guilford et al., 1978; Silvia, 2011), was used to measure creative performance. Participants were asked to generate alternative uses for a "shoelace" in 15 minutes. Each participant finished two sessions of the task: one right before the incubation period (i.e., n-back task), and one immediately afterward. Idea submissions were then checked by research assistants for spelling and grammatical errors, which were corrected and cleaned for text analysis.

Novelty Ratings. Three research assistants provided novelty ratings (1 = not novel at all, 5 = very novel) for the ideas that participants generated in the AUT. First, to check reliability, raters assessed 25% of the data. Because of the large corpus (n = 12,815) of ideas, each rater only rated a subset (one-third) of the remaining ideas. Considering the method of obtaining the majority of ratings (i.e., single raters that do not rate all items), ICC(1, 1) was selected as the appropriate reliability measure (Koo & Li, 2016). Results indicated moderate reliability, ICC(1, 1) = .57, $p < .001$.

Rater Categories. Three additional research assistants joined the previous raters in order to place ideas into categories. Each participant again only categorized a subset of ideas (one-sixth), since the number of ideas was so large. However, unlike novelty ratings, minimal rater training was used, and raters independently derived their own set of categories to use while rating their assigned ideas. The researcher then curated the final set of categories by merging existing categories, and then relabeling each idea using the new set. A final list of categories used in this process can be seen in Table 3 below.

Table 3

List of Categories from Human Raters

#	Category	#	Category	#	Category
1	accessibility	21	entertainment	41	organization
2	action	22	exchange	42	planting
3	analogy	23	exercise	43	prank
4	animal	24	experiment	44	protection
5	appearance	25	extension	45	reminder
6	aroma	26	fame	46	repair
7	art	27	fishing	47	restraint
8	bedding	28	food	48	retrieval
9	blockade	29	gifts	49	rope
10	camouflage	30	hanging	50	ruler
11	camping	31	hunting	51	securing
12	cleaning	32	junk	52	sports

13	clothing	33	magic	53	starter
14	communication	34	marker	54	survival
15	companionship	35	measurement	55	television
16	currency	36	medical	56	tool
17	decoration	37	movement	57	toy
18	divider	38	music	58	trading
19	drugs	39	nature	59	weapon
20	education	40	occupational		

WordNet Categories. Ideas were also categorized using WordNet, a lexical database that provides a hierarchical structure of English based on the similarity of words. For each part of speech (i.e., noun, verb, adjective, adverb), words are given superordinate and subordinate groupings based on their similarity. The more fundamental form of categorization is the "synonym set", or "synset". Similar words may share the same synset or may share a superordinate synset that their immediate synsets have in common. An illustration of these groupings is shown above in Figure 1. A larger, less granular categorization is referred to as the "lexicographer file". These group words more broadly, e.g. "nouns describing man-made objects". A visual comparison of synsets and lexicographer files is shown in Figure 2 above.

The "wordnet" library in the R programming language (Feinerer & Hornik, 2020; Wallace, 2007) was used to obtain both lexicographer file IDs and synset IDs. The text from each submitted idea was cleaned by removing "stop" words unrelated to the content of the idea (e.g., prepositions, pronouns, conjunctions, etc.). Then, the remaining words were matched with lexicographer files and synsets in WordNet. Matches were done with respect to the noun form of each word. Nouns were chosen for the present study, since nouns carry a significant amount of subject matter information in an idea, and since there are many more possible IDs with which the words could be matched. Words were matched with the first definition/listing in WordNet, and IDs were pulled according to that definition.

N-Back Task. The n-back task (e.g., Dijkstergyus & Meurs, 2006; Smeekens & Kane, 2016) is a working memory task that involves a participant remembering aspects of stimuli that are presented sequentially. The “n” in n-back refers to the previous number of stimuli one must recall. For example, a 3-back task would involve a participant given a sequence of stimuli, and (when prompted) having to recall the stimuli that was presented to them three presentations ago.

Participants in the two verbal conditions were given a series of three-letter words; a full list of word stimuli are shown in Appendix B. Difficulty was manipulated by the number of previous stimuli one had to hold in working memory. In the easy condition, the prompt was merely the current stimuli on the screen, colored red. Participants made their judgments based on the stimuli that was on-screen (i.e., a “0-back” task). In the difficult condition, the prompt was a red question mark (“?”). Participants had to make a judgment about the previous stimuli shown before the prompt (i.e., a “1-back” task).

Thought Probes. A thought probe procedure like the one used by Smeekens and Kane (2016) was adapted for the present study. Throughout the n-back task, participants were randomly asked what they were thinking about in the moment before the prompt appeared. Participants were given ten seconds to select one of three options. Pressing the “1” key indicated that they were thinking about the present task (the n-back task, referred to as a reaction or memory game). Pressing “2” indicated that they were thinking about the previous task (the AUT, referred to as the shoelace task). Finally, the third “3” option indicated that the participant was thinking about neither of these two tasks, i.e. mind-wandering. Participants were “allowed” a single non-response among the 14 total thought probes, with additional non-responses considered to be an attention check failure. If participants did not respond to one of the 14 probes, this single instance was added to their mind-wandering count. If participants missed more than one thought probe they were excluded from analysis. Therefore, the variables “on-task thought”, “shoelace thought”, and “mind-wandering” represent counts of each category of responses (“1”, “2”, and “3”, respectively). Higher scores on each of these variables represent a higher number of responses.

Procedure

Participants first read a brief description of the study through an advertisement on Prolific, then were taken to the study website to read and agree to the informed consent document. After agreeing to the informed consent document, participants entered their Prolific ID and filled out a series of items to indicate their race, age, and gender. Participants then completed the SDMWS, and then were given directions on how to complete the three sessions of the fluency task. Whenever directions were displayed to participants, they confirmed their reading of the directions and began the task either by pressing “Enter” or clicking “Next” on the page. Directions for each task can be seen in Appendix C.

Once the participants completed the set of fluency tasks, they were shown directions on how to complete the first session of the AUT, including Osborn's (1958) rules of creativity (also see Paulus et al., 2002). After reading these directions, participants proceeded to complete the first AUT, timed at 15 minutes. Then, participants were shown directions for the n-back task. After the 15-minute n-back task, directions for the second AUT were shown that were nearly identical to the first session. After participants completed the second 15-minute AUT, the participants were thanked for their contribution to the study and given instructions to navigate back to Prolific to receive their payment.

Results

Manipulation Check

First, in order to establish that the difficult conditions were indeed more difficult than their easy counterparts, a 2 x 2 ANOVA was performed comparing the proportion of correct n-back answers. To assess whether correct n-back responses differed based on incubation task modality (verbal/spatial) and difficulty (easy/hard), a 2 x 2 ANOVA was performed. There was a main effect of modality, $F(1, 156) = 12.29$, $p = .001$, $\eta^2 = .07$. There was also a main effect of difficulty, $F(1, 156) = 5.69$, $p = .02$, $\eta^2 = .03$. However, the interaction effect was not significant, $F(1, 156) = .73$, $p = .39$, $\eta^2 = .004$. The easy task conditions (*M* = .92, SE = .02) show higher correct n-back responses than those who performed hard tasks (M = .85, SE = .02), which was intended by the manipulation. However, the result may also mean that the spatial tasks (*M* = .93, *SE* = .02) were inherently easier than the verbal tasks (M = .83, SE = .02). Implications of these results for the other findings in this study are addressed in the Discussion section.

Intercorrelations between study variables can be found in Table 4 below. Descriptive statistics for creativity outcomes per condition can be seen in Table 5 below. Path analysis was used to test the study hypotheses. Due to potential issues with non-normality among endogenous variables (i.e., variables which are predicted by other variables in the model), robust maximum likelihood estimation using the Yuan-Bentler method was used (Rosseel, 2012). See Appendix D for Q-Q plots of each endogenous variable; deviations of data points from the trendline indicate deviations from normality. In order to compare the main effects of each n-back condition on the endogenous variables in the model, two binary variables representing n-back similarity (0 = “verbal”, 1 = “spatial”) and n-back difficulty (0 = “easy”, 1 = “hard) were entered into the model as predictors. Mind-wandering count was used as a mediator. Creativity outcomes were assessed using difference scores (i.e., the increase or decrease between the first and second sessions of the AUT). The model is shown graphically in Figure 4 below.

Table 4

Intercorrelations Between Study Variables

Variable	1	2	3	4	5	6	7	8	9	10	11
1. Pre-Task Fluency	-										
2. On-Task Thought Count	-.16	-									
3. Shoelace Thought Count	-.17	-.25	-								
4. Mind-Wandering Count	.22	-.95***	-.05	-							
5. Proportion Correct - N-Back	.12	.07	-.18	-.02	-						
6. Ideas - Block 1	.49***	-.11	-.08	.14	.08	-					
7. Synsets - Block 1	.42***	-.09	.00	.10	.06	.82***	-				
8. Lexicographer Files - Block 1	.36***	-.09	.05	.07	.09	.69***	.84***	-			
9. Categories - Block 1	.36***	-.12	-.02	.13	.11	.73***	.65***	.57***	-		
10. Novelty - Block 1	.09	-.05	.09	.02	.04	.36***	.51***	.47***	.43***	-	
11. Ideas - Block 2	.35**	-.05	-.02	.06	.04	.82***	.64***	.52***	.61***	.25	-
12. Synsets - Block 2	.38***	-.07	.05	.05	.03	.74***	.83***	.69***	.59***	.39***	.83***
13. Lexicographer Files - Block 2	.33**	-.02	.05	.01	.03	.64***	.77***	.75***	.55***	.41***	.67***
14. Categories - Block 2	.35**	-.03	-.02	.04	.06	.78***	.71***	.65***	.71***	.39***	.86***
15. Novelty - Block 2	.19	-.06	.09	.03	.01	.34**	.49***	.42***	.34**	.60***	.15
16. Idea Difference	-.01	.05	.06	-.07	-.05	.15	.06	.03	.12	-.03	.69***
17. Synset Difference	.02	.03	.10	-.06	-.03	.04	-.07	-.09	.05	-.09	.49***
18. Lexicographer File Difference	.01	.08	.01	-.09	-.08	.04	.02	-.22	.06	-.02	.30*
19. Category Difference	.09	.09	-.01	-.09	-.05	.28	.27	.27	-.12	.08	.53***
20. Novelty Difference	.14	-.02	.02	.02	-.03	.07	.12	.06	.01	-.21	-.06
21. SDMWS - Spontaneous	.10	-.16	-.01	.17	-.03	.20	.12	.09	.16	.09	.15
22. SDMWS - Deliberate	.20	-.15	-.08	.18	-.08	.14	.13	.05	.05	.00	.13

Note. Adjustment of *p*-values for multiple comparisons were done using the Holm method.
*$p < .05$, **$p < .01$, ***$p < .001$

Table 4 (cont.)

Intercorrelations Between Study Variables

Variable	12	13	14	15	16	17	18	19	20	21	22
1. Pre-Task Fluency											
2. On-Task Thought Count											
3. Shoelace Thought Count											
4. Mind-Wandering Count											
5. Proportion Correct - N-Back											
6. Ideas - Block 1											
7. Synsets - Block 1											
8. Lexicographer Files - Block 1											
9. Categories - Block 1											
10. Novelty - Block 1											
11. Ideas - Block 2											
12. Synsets - Block 2	-										
13. Lexicographer Files - Block 2	.86***	-									
14. Categories - Block 2	.80***	.75***	-								
15. Novelty - Block 2	.40***	.49***	.31**	-							
16. Idea Difference	.49***	.34**	.50***	-.18	-						
17. Synset Difference	.49***	.33**	.34**	-.05	.79***	-					
18. Lexicographer File Difference	.36***	.48***	.25	.17	.47***	.61***	-				
19. Category Difference	.47***	.44***	.61***	.05	.56***	.42***	.29*	-			
20. Novelty Difference	.11	.21	.00	.65***	-.19	.02	.23	-.01	-		
21. SDMWS - Spontaneous	.10	.06	.11	.08	.00	-.01	-.02	-.03	.02	-	
22. SDMWS - Deliberate	.12	.07	.11	.10	.04	.03	.05	.09	.12	.38***	-

Note. Adjustment of *p*-values for multiple comparisons were done using the Holm method.
*$p < .05$, **$p < .01$, ***$p < .001$

Table 5

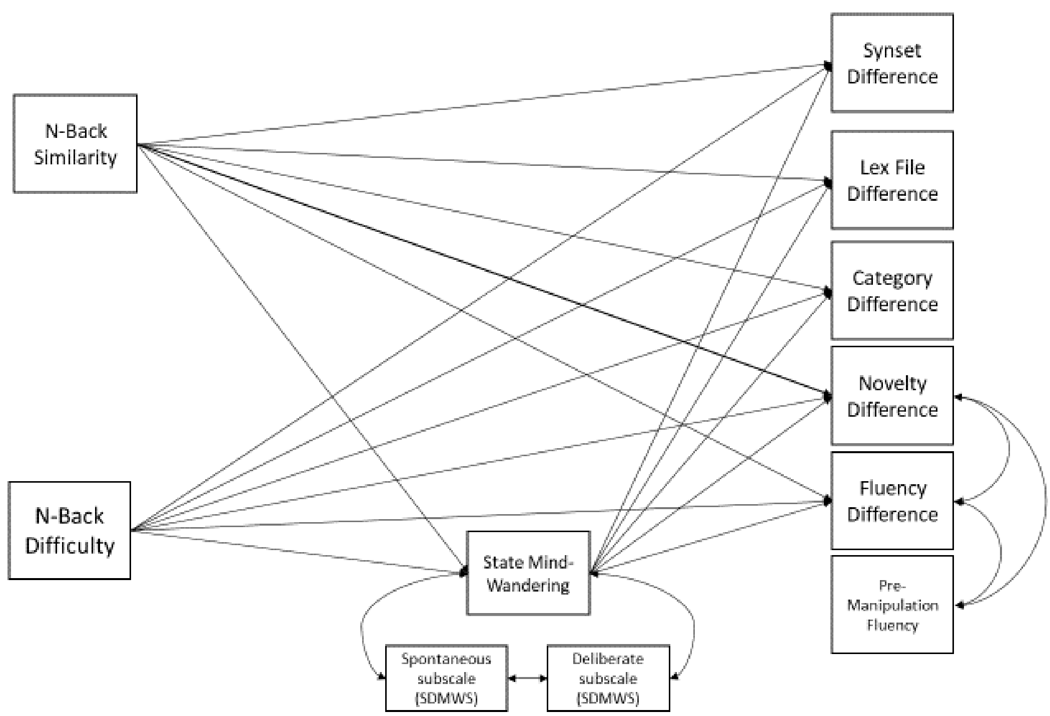
Summary of Creativity Outcomes by N-Back Condition

Verbal/Easy	Block 1	*M*	*SD*
	Ideas	34.10	17.40
	Synsets	63.20	28.70
	Lexicographer Files	16.50	3.75
	Categories	10.30	3.45
	Novelty	2.08	0.30
	Block 2	*M*	*SD*
	Ideas	35.70	24.70
	Synsets	63.90	36.30
	Lexicographer Files	15.60	4.73
	Categories	10.60	4.42
	Novelty	2.09	0.34
Verbal/Hard	Block 1	*M*	*SD*
	Ideas	33.60	15.40
	Synsets	60.10	30.20
	Lexicographer Files	15.30	4.66
	Categories	10.50	3.81
	Novelty	2.05	0.37
	Block 2	*M*	*SD*
	Ideas	33.20	18.60
	Synsets	57.80	31.80
	Lexicographer Files	15.10	4.48
	Categories	10.30	4.41
	Novelty	2.19	0.47
Spatial/Easy	Block 1	*M*	*SD*
	Ideas	33.20	12.30
	Synsets	53.60	21.70
	Lexicographer Files	14.60	3.80
	Categories	10.80	2.61
	Novelty	1.96	0.30
	Block 2	*M*	*SD*
	Ideas	33.60	17.90
	Synsets	52.20	25.60
	Lexicographer Files	13.70	4.52
	Categories	10.10	3.97
	Novelty	2.10	0.46
Spatial/Hard	Block 1	*M*	*SD*
	Ideas	33.40	15.50
	Synsets	57.60	25.80
	Lexicographer Files	15.70	3.59

Categories	10.20	3.37
Novelty	2.09	0.33
Block 2	*M*	*SD*
Ideas	35.10	22.10
Synsets	56.10	29.20
Lexicographer Files	15.00	4.29
Categories	10.10	4.11
Novelty	2.14	0.44

Figure 4

Diagram of Path Model



Hypotheses 1a- 2e

Model fit was found to be acceptable, $\chi^2(22) = 22.82$, $p = .41$, CFI = 1.00, RMSEA = .02. Refer to Table 6 below for path coefficients and associated statistics. In order to test Hypotheses 1a through 2e, coefficients associated with the effect of similarity and difficulty on the difference scores for each outcome (synsets, lexicographer files, human-judged categories, novelty, and idea count/fluency) were examined. None of the coefficients were significant. Therefore, Hypotheses 1a through 2e were not supported.

Table 6

Path Model Coefficients

path	estimate	*SE*	*z*	*p*	95% CI
MW ~ similarity (verbal to spatial)	1.67	0.57	2.94	< .001***	[0.56, 2.78]
MW ~ difficulty (easy to hard)	-0.06	0.57	-0.98	.33	[-1.67, 0.55]
synset difference ~ MW	-0.27	0.38	-0.71	.48	[-1.00, 0.47]
lex difference ~ MW	-0.06	0.06	-0.92	.36	[-0.19, 0.07]
category difference ~ MW	-0.06	0.06	-0.95	.34	[-0.19, 0.07]
novelty difference ~ MW	0.00	0.01	-0.24	.81	[-0.01, 0.01]
idea difference ~ MW	-0.22	0.27	-0.83	.40	[-0.75, 0.30]
synset difference ~ similarity	-0.13	2.80	-0.05	.96	[-5.62, 5.36]
lex difference ~ similarity	-0.20	0.50	-0.40	.69	[-1.18, 0.78]
category difference ~ similarity	-0.41	0.47	-0.87	.38	[-1.34, 0.51]
novelty difference ~ similarity	0.02	0.06	0.30	.77	[-0.09, 0.13]
idea difference ~ similarity	0.87	1.85	0.47	.64	[-2.76, 4.50]
synset difference ~ difficulty	-1.71	2.75	-0.62	.53	[-7.10, 3.69]
lex difference ~ difficulty	0.44	0.48	0.91	.36	[-0.51, 1.39]
category difference ~ difficulty	-0.04	0.48	-0.09	.93	[-0.98, 0.90]
novelty difference ~ difficulty	0.03	0.05	0.57	.57	[-0.07, 0.14]
idea difference ~ difficulty	-0.68	1.92	-0.35	.73	[-4.45, 3.09]

Note. MW = mind-wandering count, lex = lexicographer file, easy = undemanding, hard = demanding.
†$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$

Hypotheses 3a-4e

In order to test Hypotheses 3a and 3b, coefficients were assessed for the effect of similarity and difficulty on mind-wandering. Those in the easy conditions did not significantly differ on mind-wandering than those in the difficult conditions, which did not support Hypothesis 3a. However, those who were in the spatial conditions reported significantly higher levels of mind-wandering, supporting Hypothesis 3b. To test Hypotheses 4a through 4e, coefficients were assessed for the relationship between mind-wandering and the difference scores for each creativity outcome. None of the outcomes of the creativity task were significantly related to mind-wandering. Thus, Hypotheses 4a through 4e were not supported.

Hypothesis 5

In order to test Hypothesis 5, a set of factor analyses were performed using the “psych” (Revelle, 2021) and “lavaan” (Rosseel, 2012) libraries in R. First, intercorrelations between the three categorization variables (synsets, lexicographer files, and human-judged categories) were run. The three variables were found to be highly correlated with one another, supporting Hypothesis 5. Furthermore, in order to probe the potential factor structure of these variables, factor analytic methods were performed. Using the “psych” library, scree plots (Figures 5 and 6 below) were examined for each session of the AUT (i.e., each time point) and found that a one-factor solution fit both time points. Results of exploratory factor analyses show that the three items load onto a single factor. Cronbach’s alpha was used to test the reliability of the three variables, while maximum likelihood with an oblimin rotation was used to find the factor solution. Results are shown in Table 7 below. Reliability was found to be low, α_{AUT1} = .44, α_{AUT2} = .49. However, reliability increased if only lexicographer files and human-judged categories were included, α_{AUT1} = .72, α_{AUT2} = .86. This may be due to the similarity in the range of responses for each type of flexibility measure; there were 59 possible human-judged categories and 25 possible lexicographer files, while there are many more possible synsets (e.g., the maximum number of synsets reported was 144 for the first AUT session and 151 for the second). Thus, the similarity of lexicographer files and human-judged categories in terms of granularity may be driving this finding. Additionally, Table 8 contains correlations between each measure, and Table 9 shows regressions predicting human-judged categories from the WordNet measures. Interestingly, lexicographer files did not significantly predict human-judged categories in the first block but was a significant predictor in the second block. Additionally, both models accounted for a considerable amount of variance in human-judged categories (41% and 67%, respectively).

Figure 5

Scree Plot – AUT Session 1 Category Measures

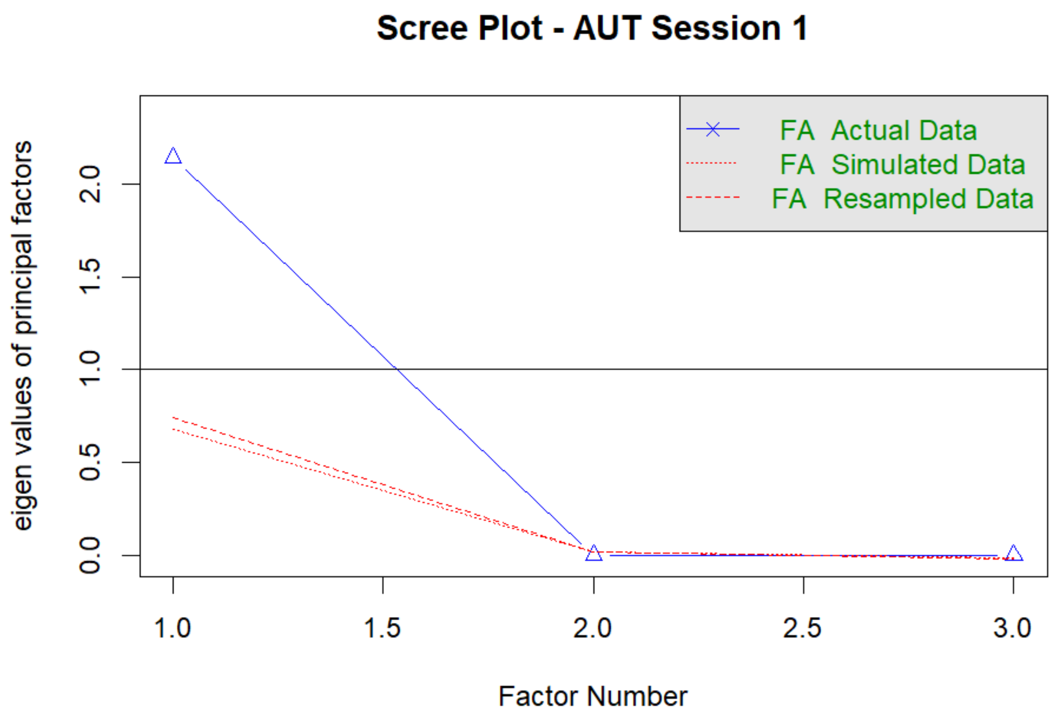


Figure 6

Scree Plot – AUT Session 2 Category Measures

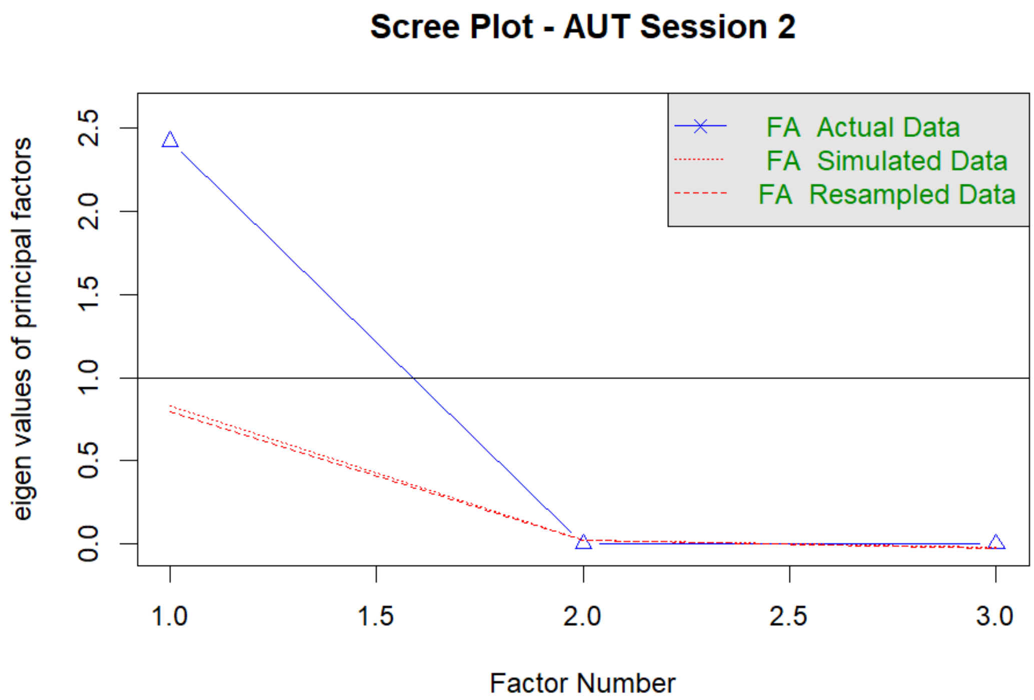


Table 7

Reliability and Factor Loadings of Flexibility Measures

Session	α		$α_{removed}$	Loading	h^2	u^2
Session 1	.44	Synset IDs	.72	.98	0.95	0.05
		Lex File IDs	.27	.86	0.86	0.25
		Categories	.40	.66	0.66	0.56
Session 2	.49	Synset IDs	.86	.96	0.96	0.08
		Lex File IDs	.35	.89	0.89	0.20
		Categories	.39	.84	0.84	0.30

Table 8

Correlations Among Flexibility Measures

Session 1				
		Synset IDs	Lex File IDs	Categories
	Synset IDs	.	.84***	.65***
	Lex File IDs		.	.57***
	Categories			.
Session 2				
		Synset IDs	Lex File IDs	Categories
	Synset IDs	.	.86***	.80***
	Lex File IDs		.	.75***
	Categories			.

***$p < .001$.

Table 9

Regression of Human-Judged Categories on WordNet Measures

Block 1	F	df_{effect}	$df_{residual}$	p	*adj.* R^2	Predictor	b	SE	t	p
	68.20	2	195	< .001***	.41	Synsets	0.07	0.01	4.84	< .001***
						Lex Files	0.12	0.09	1.30	.44
Block 2	F	df_{effect}	$df_{residual}$	p	*adj.* R^2	Predictor	b	SE	t	p
	198.20	2	195	< .001***	.67	Synsets	0.08	0.01	6.93	< .001***
						Lex Files	0.23	0.07	3.18	< .001***

***$p < .001$.

To see if the factor structure of the flexibility measures changed over time, measurement invariance tests were used to test whether the factor loadings and intercepts for a one-factor solution changed between the first and second sessions of the AUT. Factor models for configural, metric, and scalar invariance were compared using χ^2 likelihood ratio tests. These models respectively test whether the factor form (i.e., number of factors and factor-indicator patterns), factor loadings, and factor intercepts (i.e., item means) are equal across the two time points (Putnick & Bornstein, 2016). Results are provided in Table 10 below.

Table 10

χ^2 Difference Test of Measurement Invariance Models

Model	*df*	AIC	BIC	χ^2	$p_{difference}$
Configural	0	6010.2	6078		
Weak (factor loadings equivalent)	2	6015.7	6076	9.51	.01***
Strong (factor loadings and intercepts equivalent)	4	6014.9	6067.7	12.74	.18

While configural invariance seems to have been established by the previous analyses, metric and scalar invariance were not established based on these results. This means that factor loadings and intercepts likely changed between the first and second AUT sessions, limiting the interpretability of the proposed latent flexibility factor. Overall, the results of the factor analyses show partial support for Hypothesis 5 and lend support to the idea that WordNet IDs (lexicographer IDs in particular) could be a viable indicator of flexibility. However, the variability in factor structure across time raises concerns about the consistency of the WordNet measures, and their relation to human-judged categories.

Exploratory Analysis: Verbal/Easy vs. Other Conditions

Because of the previously mentioned issues with manipulation, i.e. the possibility that the n-back task difficulty was also confounded with task type, additional analyses were conducted to examine effects of both task similarity and difficulty on creative outcomes. Specifically, the path model was adjusted to compare the verbal/easy group to the other three conditions (verbal/hard, spatial/easy, and spatial/hard). The new model is shown graphically in Figure 7 below. Model fit was found to be acceptable, $\chi^2(25) = 23.47$, $p = .55$, CFI = 1.00, RMSEA < .001. Refer to Table 9 below for path coefficients and associated statistics. Moving from verbal/easy to verbal/hard was associated with a significant increase in novelty ($p = .04$), while the comparison between verbal/easy and spatial/easy on novelty yielded a small, but nonsignificant, p-value ($p = .06$). Additionally, results of 2 x 2 ANOVA models are also presented and illustrated in Appendix E; the slight increase of novelty for the verbal/hard and spatial/easy groups is shown visually. The pattern of results suggests support for unconscious work theory, but interpretation of these results is limited, as discussed in the following sections.

Figure 7

Exploratory Path Model Diagram

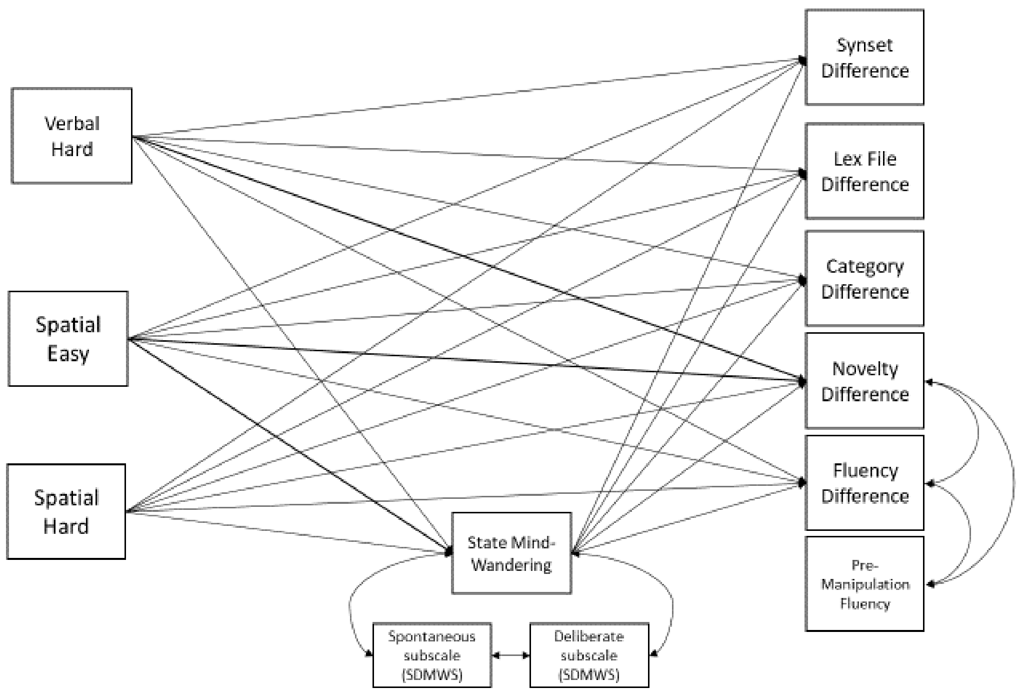


Table 11

Exploratory Path Model Coefficients

path	estimate	*SE*	*z*	*p*	95% CI
MW ~ verbal /hard	0.65	0.77	0.84	.40	[-0.87, 2.17]
MW ~ spatial/easy	3.02	0.82	3.69	< .001***	[1.42, 4.62]
MW ~ spatial/hard	1.10	0.81	1.36	.17	[-0.49, 2.69]
synset difference ~ MW	-0.24	0.39	-0.61	.54	[-1.01, 0.53]
lex difference ~ MW	-0.07	0.07	-0.99	.32	[-0.20, 0.06]
category difference ~ MW	-0.05	0.06	-0.78	.43	[-0.18, 0.08]
novelty difference ~ MW	0.00	0.01	-0.65	.52	[-0.02, 0.01]
idea difference ~ MW	-0.19	0.28	-0.69	.49	[-0.74, 0.36]
synset difference ~ verbal/hard	-2.77	4.06	-0.68	.50	[-10.72, 5.18]
lex difference ~ verbal/hard	0.75	0.67	1.12	.26	[-0.56, 2.05]

category difference ~ verbal/hard	-0.49	0.63	-0.79	.43	[-1.72, 0.74]
novelty difference ~ verbal/hard	0.14	0.07	2.06	.04*	[0.01, 0.27]
idea difference ~ verbal/hard	-1.96	2.44	-0.80	.42	[-6.75, 2.82]
synset difference ~ spatial/easy	-1.37	4.47	-0.31	.76	[-10.13, 7.39]
lex difference ~ spatial/easy	0.16	0.75	0.22	.83	[-1.30, 1.63]
category difference ~ spatial/easy	-0.94	0.68	-1.38	.17	[-2.27, 0.39]
novelty difference ~ spatial/easy	0.14	0.08	1.85	.06†	[-0.01, 0.29]
idea difference ~ spatial/easy	-0.65	2.64	-0.24	.81	[-5.81, 4.52]
synset difference ~ spatial/hard	-1.85	4.36	-0.42	.67	[-10.39, 6.69]
lex difference ~ spatial/hard	0.25	0.80	0.31	.76	[-1.33, 1.82]
category difference ~ spatial/hard	-0.46	0.72	-0.64	.52	[-1.88, 0.96]
novelty difference ~ spatial/hard	0.05	0.07	0.67	.50	[-0.09, 0.19]
idea difference ~ spatial/hard	0.18	2.87	0.06	.95	[-5.46, 5.81]

Note. MW = mind-wandering count, lex = lexicographer file, easy = undemanding, hard = demanding.
†$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$

Discussion

General Discussion

The present study investigated the role of mind-wandering in how incubation periods affect idea generation and creative performance. The study replicated previous findings by Smeekens and Kane (2016) and Steindorf et al. (2020), who failed to replicate effects found by Baird and colleagues (2012). The present study also introduced some novel changes to the methodology of the previous authors, including online data collection and text analysis methods using WordNet. Mind-wandering did not correlate with any of the creativity measures. However, WordNet measures did correlate with human-judged categories. The results, while limited, may inform incubation theory and creativity research methodology going forward. The following section firstly provides a general discussion of findings and their relevance to theories of incubation. The next section covers limitations of the study, then discusses the merit and contribution of the present results. Finally, directions for future research, including implications for creativity research, potential follow-up analyses, and suggestions for future studies, are presented.

The findings indicate some support for unconscious work theory, as the pattern of task demand and similarity in the exploratory analyses matches the predictions under unconscious work theory, i.e., that higher demand and higher difficulty will incur increases in unconscious spreading activation and therefore increase creative performance. This is partially a replication of studies finding similar effects (see Sio & Ormerod, 2009) and points to support that something other than fatigue reduction and/or continuing to consciously work on the creative problem is responsible for incubation effects. In terms of unconscious work theory, the finding that increasing difficulty or changing modality could be explained by reliance on more intuitive, automatic processes that are caused by the increased difficulty of the incubation task, or because of the freeing up of networks that are needed for the creative task. However, unlike studies such as Nam and Lee (2015), we did not include any indicators of unconscious work, making it difficult to attribute the effects to the often vaguely defined mechanisms of unconscious work theory. Future studies may improve on the work reported here by changing the manipulations used, including new manipulations, and adding conditions to task difficulty or modality.

Mind-wandering in this paradigm did not seem to influence creative performance, showing no relationship with any of the creativity outcomes. As shown in Table 2, participants reported being on-task more often than reporting mind-wandering, and instances of thinking about the creative (shoelace) task were not reported very often. This may be due to demand characteristics, since participants paid through Prolific must avoid being reported as a “bad” participant. Considering that this is now the latest in a line of similar studies that failed to find benefits of mind wandering (Murry et al., 2021; Smeekens & Kane, 2016; Steindorf et al., 2020). However, it was shown in the first path model that moving from a verbal to spatial task increased mind-wandering. Additionally, exploratory analyses that thought probes reporting mind-wandering were higher amongst those in the spatial/easy group than the verbal/easy group. While mind-wandering during an incubation period was not shown to be influential when it comes to creative performance, reliability inducing higher levels of mind-wandering through altering the task was demonstrated. This may be useful for researchers attempting to induce mind-wandering for other studies, regardless of whether creativity is of interest.

Follow-up studies and/or additional analyses may be designed to compare unconscious work theory in a more direct fashion. As the present study was conceptualized as a replication/extension of the line of studies starting with Baird et al. (2012), major deviations from that paradigm were not considered. However, the thought-probing methods used could be adapted to assess unconscious work more directly. For example, a Likert-type item could be presented similarly to the thought probes in this study that tap how aware (or unaware) individuals are about their thoughts, with awareness serving as an indicator of unconscious thought. Additionally, fixation may be induced or assessed through specific methods like the ones used by Smith and Beda (2020). Furthermore, the present data may be re-analyzed later for other, theoretically relevant effects. For example, the number of new unique categories may be assessed between session 1 and session 2 of the AUT, with new unique categories generated during the second session potentially representing a loss of fixation.

The continued discussion about these various mechanisms will likely continue. As previously mentioned, these theories may not necessarily contradict each other (Smith & Beda, 2020; Williams et al., 2018). For example, fixation can prevent unconscious spreading activation in potentially useful, distant semantic areas from occurring. Such an explanation combines elements forgetting fixation and unconscious work theory without necessarily invalidating either theory. Alternatively, mind-wandering could be a way of presenting new search cues in idea generation, such that cues that are no longer useful are forgotten (i.e., mind-wandering leads to forgetting fixation). Considering that the last major meta-analysis of the literature was conducted more than a decade ago (Sio & Ormerod, 2009), it may be of interest for researchers to collect effect sizes across all the studies that have been conducted since then, which happens to include Baird et al. (2012) and all the replications following its publication. A new systematic review of the incubation literature would ideally address the variety of creativity research that exists, including factors in group creativity, different incubation conditions, and different types of both convergent/divergent thinking tasks.

Limitations

Lack of Incubation Effects. While incubation periods have been reported to increase multiple facets of creative performance, such as fluency and novelty (Gilhooly, 2016), in the present study only novelty increased from the first AUT session (M = 2.05, SD = .33) to the second, (M = 2.14, SD = .43), t(299.1), p = .04. The lack of major changes in performance from the first AUT to the second makes attributing effects to mind-wandering or the incubation manipulations difficult. As shown in the exploratory analyses (Table 9), a significant p-value (p = .04) was associated with the comparison between the base group ("verbal/easy") and the "verbal/hard" group on novelty. A non-significant p-value (p = .06) was found for the comparison between the base group and spatial/easy on novelty, with the 95% CI lower limit close to zero. In both cases, the estimated coefficient was positive. While the pattern of results suggests alignment with unconscious work theory, as an increase in performance is associated with both changing modalities and increasing difficulty, the relatively weak statistical evidence for these effects, as well as a non-significant comparison of novelty between verbal/easy and spatial/hard, hinders comparison of incubation theories with each other. Additionally, there is no clear theoretical rationale to explain why only difficulty within the verbal modality (i.e., verbal/easy vs. verbal/hard) produced differences in novelty. Instead, the results of the exploratory analyses may have to do with a conflation of difficulty and modality, as discussed in more detail below.

Issues with Experimental Manipulation. As previously mentioned, the proportion of correct responses on the n-back task was not significantly different between the easy and hard conditions of the spatial task. However, when assessed with a multiple comparison procedure, the proportion of correct responses did significantly differ between the spatial/easy condition and both verbal conditions. The results indicate that the verbal tasks may have been inherently harder to perform than the spatial tasks, which may mean that task demand and task modality were confounded in this study. Therefore, it is difficult to parse out the meaning of moving from the verbal/easy to the verbal/hard group in terms of gains in novelty. On one hand, the results can be interpreted as partially supportive of unconscious work theory, as higher demand (verbal/easy to verbal/hard) yielded higher novelty in session 2 of the AUT. However, the verbal/easy and verbal/hard conditions were not significantly different in terms of correct responses ($p = .13$). The conflation of modality and demand may also explain the small, but non-significant, effect of moving from verbal/easy to spatial/easy on novelty as described above.

Sample Size, Power, and Representativeness. During the planning stages of the current study, an a-priori power analysis was conducted, with the original proposed model containing 34 parameters (see Appendix F for a visual representation of the model, and the original power analysis, respectively). The final model contained a total of 41 parameters, which likely impacted the true power of the present study. Therefore, it is uncertain whether the null findings are due to low power, or due to the true absence of an effect. Instead of using a post-hoc power analysis, the use of which has been widely criticized (e.g., Zhang et al., 2019), another a-priori power analysis was conducted for a study like the present study (RMSEA < .08, $\alpha < .05$, df = 22), with a minimum power of .80. Results indicated that the estimated sample size needed for such a study is $N = 156$. This may indicate that power may not be as much of a concern in interpreting the present results.

Limits of Experimental Control in Online Settings. Conducting a study over the Internet has several benefits to researchers, including the ability to recruit participants that are more representative than other convenient samples (e.g., college students). However, a tradeoff is that researchers have minimal control over the environment that the participant is performing tasks in. A typical laboratory experiment is usually devoid of distractions and other aspects of one's environment that could change results. In the present study, none of those aspects about the environment were controlled, and outside of timestamp information recorded by the website, no information about participants' environment was obtained. However, since the two AUT sessions and n-back task require a minimum of 45 minutes to complete, the final sample was limited only to those who spent no more than five additional minutes (50 minutes total) past the beginning of the first AUT session. This was an attempt to partially control for any additional incubation effects that may have occurred while participants were in-between tasks.

Rule-Based Text Analysis. As discussed previously, a simple matching procedure was used to link individual nouns within an idea to synset and lexicographer file IDs in WordNet. The first definition for each word was used to pull both types of WordNet IDs. This procedure ignores information from the idea that would help researchers model the context of ideas, such as part of speech, similarity of ideas, etc. However, this rule was implemented to give a straightforward procedure of assessing categories, without resorting to such techniques as topic modeling or other techniques that are more complex and harder to interpret. More advanced techniques in natural language processing should be considered for future studies where feasible (e.g., Christensen et al., 2017; Toubia & Netzer, 2017), as these methods attempt to infer and model this contextual information. One of the most promising and purpose-built tools for assessing creativity seems to be the "semDis" software package (Beaty & Johnson, 2021).

Interrater Reliability. As previously stated, interrater reliability for novelty was low, and agreement across categories was not examined. Ratings were split up due to the large number (n = 12,815) of ideas in the dataset. The lower reliability for novelty was considered a trade-off for the raters to be able to complete the task in a reasonable amount of time. Similarly, categories were curated by the main researcher, and not through typical conventions (e.g., multiple rounds of rater alignment discussion and revision). This introduces threats to the validity of the methods related to human judgment.

Focus on Individual Creativity. Finally, it should be repeated that the present study does not account for social and motivational factors that are introduced in group settings. Several previous investigations have found differential support for both conscious (e.g. Beeftink et al., 2008; Harinck & De Dreu, 2008; Paulus et al., 2006; Paulus & Yang, 2000) and unconscious (e.g., Choi et al. 2019; Kohn & Smith 2011) theories of incubation, focusing on factors that are more relevant in a social or collaborative context, such as evaluation apprehension. Since participants in this study only generated ideas alone, further research using group idea generation is needed to examine the explanatory power of each theory.

Conclusion

Based on the present study, there does not seem to be sufficient evidence that mind-wandering plays a role in incubation effects when examined in individuals performing a divergent thinking task. However, limited evidence was found for increased n-back task demand corresponding with an increase in novelty, with a similar (though non-significant) trend for n-back task modality, matching predictions made by unconscious work theory. Some minor support is therefore found for this theory, but this evidence is by no means definitive considering the limitations listed above. It could be a possibility that the mechanisms for forgetting fixation and/or mind-wandering are more relevant when generating ideas in group settings, when completing a convergent rather than a divergent task, or in various other situations that this study did not examine.

This study was also one of the first to explore the utility of WordNet in idea generation research. Despite the shortcomings of the method used in this study, measures derived from WordNet seemed to strongly correlate with human-judged categories, though this correlation changed between the first and second idea generation task. This provides limited evidence for WordNet as a potential tool for assessing flexibility, as an alternative to using human raters. However, due to the methods in this study being quite simplistic to the wide range of text analysis and natural language processing techniques available, future studies should consider using more complex modeling to capture more contextual information about each idea. Again, one of the most promising and purpose-built tools for assessing creativity seems to be the “semDis” software package (Beaty & Johnson, 2021). Similar analyses to the ones performed for this study, such as factor analysis and measurement invariance tests, can also be leveraged on data derived from such techniques.

In closing, the present study contributes to the existing literature by exploring the three main unconscious theories of incubation effects. Findings indicate waning support for the idea that mind-wandering plays a key role in incubation, but further research is needed to address the limitations of the present study. Furthermore, this study presents a novel method of assessing idea category flexibility, with suggestions on how this method (and similar techniques) could be leveraged to avoid the labor-intensive work that is typically required of creativity studies that use human judgment. Future research that explores the role of unconscious processes during incubation periods is needed to more clearly determine their contribution to creativity.

References

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, *16*(2), 270-301.

Alves, J., Marques, M. J., Saur, I., & Marques, P. (2007). Creativity and innovation through multidisciplinary and multisectoral cooperation. *Creativity and Innovation Management, 16*(1), 27-34.

Amabile, T. M. (1983). The social psychology of creativity: a componential conceptualization. *Journal of Personality and Social Psychology, 45*(2), 357.

Amabile, T.M. (2011). *Componential theory of creativity* (pp. 538-559). Harvard Business School.

Baird, B., Smallwood, J., Mrazek, M. D., Kam, J. W., Franklin, M. S., & Schooler, J. W. (2012). Inspired by distraction: mind wandering facilitates creative incubation. *Psychological Science, 23*(10), 1117 -1122.

Bakker, M., & Wicherts, J. M. (2014). Outlier removal, sum scores, and the inflation of the type I error rate in independent samples t tests: the power of alternatives and recommendations. *Psychological Methods*, *19*(3), 409.

Barron, E., Riby, L. M., Greer, J., & Smallwood, J. (2011). Absorbed in thought: the effect of mind wandering on the processing of relevant and irrelevant events. *Psychological Science, 22*(5), 596 -601.

Bastian, M., Lerique, S., Adam, V., Franklin, M. S., Schooler, J. W., & Sackur, J. (2017). Language facilitates introspection: verbal mind-wandering has privileged access to consciousness. *Consciousness and Cognition*, *49*, 86-97.

Beaty, R. E., & Johnson, D. R. (2021). Automating creativity assessment with SemDis: an open platform for computing semantic distance. *Behavior Research Methods*, *53*(2), 757-780.

Beeftink, F., Van Eerde, W., & Rutte, C. G. (2008). The effect of interruptions and breaks on insight and impasses: do you need a break right now? *Creativity Research Journal*, *20*(4), 358-364.

Breslin, D. (2019). Off-task social breaks and group creativity. *The Journal of Creative Behavior*, *53*(4), 496-507.

Brown, V. R., & Paulus, P. B. (2002). Making group brainstorming more effective: recommendations from an associative memory perspective. *Current Directions in Psychological Science*, *11*(6), 208 -212.

Brown, V., & Paulus, P. B. (1996). A simple dynamic model of social factors in group brainstorming. *Small Group Research*, *27*(1), 91-114.

Büschgens, T., Bausch, A., & Balkin, D. B. (2013). Organizational culture and innovation: a meta-analytic review. *Journal of Product Innovation Management, 30*(4), 763-781.

Champely, S. (2020). *pwr: Basic Functions for Power Analysis*. R package version 1.3-0. https://CRAN.R-project.org/package=pwr

Chandler, J., & Shapiro, D. (2016). Conducting clinical research using crowdsourced convenience samples. *Annual Review of Clinical Psychology*, 12.

Chiang, N. C., & Chen, M. L. (2017). Benefits of incubation on divergent thinking. *Creativity Research Journal*, *29*(3), 282-291.

Choi, H. S., Seo, J. G., Hyun, J., & Bechtoldt, M. (2019). Collectivistic independence promotes group creativity by reducing idea fixation. *Small Group Research*, *50*(3), 381-407.

Christensen, K., Nørskov, S., Frederiksen, L., & Scholderer, J. (2017). In search of new product ideas: identifying ideas in online communities by machine learning and text mining. *Creativity and Innovation Management, 26*(1), 17-30.

Creswell, J. D., Bursley, J. K., & Satpute, A. B. (2013). Neural reactivation links unconscious thought to decision-making performance. *Social Cognitive and Affective Neuroscience*, *8*(8), 863-869.

de Leeuw, J. R. (2015). jsPsych: a JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1-12. doi:10.3758/s13428-014-0458-y.

Dijksterhuis, A. (2004). Think different: the merits of unconscious thought in preference development and decision making. *Journal of Personality and Social Psychology, 87(*5), 586.

Dijksterhuis, A., & Meurs, T. (2006). Where creativity resides: the generative power of unconscious thought. *Consciousness and Cognition, 15*(1), 135-146.

Dijksterhuis, A., & Nordgren, L. F. (2006). A theory of unconscious thought. *Perspectives on Psychological Science, 1(*2), 95-109.

Doboli, S., & Brown, V. R. (2010, July). An emergent attractors model for idea generation process. *The 2010 International Joint Conference on Neural Networks (IJCNN),* 1-8.

Ellwood, S., Pallier, G., Snyder, A., & Gallate, J. (2009). The incubation effect: hatching a solution? *Creativity Research Journal, 21*(1), 6-14.

Feinerer, I., & Hornik, K. (2020). *wordnet: WordNet Interface*. R package version 0.1-15, https://CRAN.R-project.org/package=wordnet.

Feist, G. J. (2010). The function of personality in creativity: the nature and nurture of the creative personality. In J. C. Kaufman & R. J. Sternberg (Eds.), *The Cambridge Handbook of Creativity* (pp. 113–130). Cambridge University Press. https://doi.org/10.1017/CBO9780511763205.009

Feldhusen, J. F., & Goh, B. E. (1995). Assessing and accessing creativity: An integrative review of theory, research, and development. *Creativity Research Journal, 8*(3), 231-247.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Bradford Books.

Firestien, R. L., & McCowan, R. J. (1988). Creative problem solving and communication behavior in small groups. *Creativity Research Journal*, *1*(1), 106-114.

Gable, S. L., Hopper, E. A., & Schooler, J. W. (2019). When the muses strike: creative ideas of physicists and writers routinely occur during mind wandering. *Psychological Science*, *30*(3), 396-404.

Gallate, J., Wong, C., Ellwood, S., Roring, R. W., & Snyder, A. (2012). Creative people use nonconscious processes to their advantage. *Creativity Research Journal*, *24*(2-3), 146-151.

Gilhooly, K. J. (2016). Incubation and intuition in creative problem solving. *Frontiers in Psychology*, *7*, 1076.

Gilhooly, K. J. (2017). Response: Commentary: Incubation and intuition in creative problem solving. *Frontiers in Psychology*, *8*, 465.

Gilhooly, K. J., Fioratou, E., Anthony, S. H., & Wynn, V. (2007). Divergent thinking: strategies and executive involvement in generating novel uses for familiar objects. *British Journal of Psychology, 98*(4), 611-625.

Gilhooly, K. J., Georgiou, G., & Devery, U. (2013). Incubation and creativity: do something different. *Thinking & Reasoning*, *19*(2), 137-149.

Goodwin, G. F., Blacksmith, N., & Coats, M. R. (2018). The science of teams in the military: contributions from over 60 years of research. *American Psychologist*, *73*(4), 322.

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504-528.

Guilford, J. P., Christensen, P. R., Merrifield, P. R., & Wilson, R. C. (1978). *Alternate uses: manual of instructions and interpretations*. Orange, CA: Sheridan Psychological Services.

Hao, N., Ku, Y., Liu, M., Hu, Y., Grabner, R. H., & Fink, A. (2014). Enhancing verbal creativity via brief interventions during an incubation interval. *Creativity Research Journal, 26*(1), 30-38.

Hao, N., Wu, M., Runco, M. A., & Pina, J. (2015). More mind wandering, fewer original ideas: be not distracted during creative idea generation. *Acta Psychologica*, *161*, 110-116.

Harinck, F., & De Dreu, C. K. (2008). Take a break! Or not? The impact of mindsets during breaks on negotiation processes and outcomes. *Journal of Experimental Social Psychology*, *44*(2), 397-404.

Hung, S. P., Chen, P. H., & Chen, H. C. (2012). Improving creativity performance assessment: a rater effect examination with many facet Rasch model. *Creativity Research Journal*, *24*(4), 345-357.

Iyer, L. R., Doboli, S., Minai, A. A., Brown, V. R., Levine, D. S., & Paulus, P. B. (2009). Neural dynamics of idea generation and the effects of priming. *Neural Networks*, *22*(5-6), 674-686.

Jafari, A., Shariatpanahi, S. P., Zolfagharzadeh, M. M., Abdollahyar, M., & Sadeghipour, M. (2020). From creativity to innovation and the role of competition networks: a cancer inspired two-step evolutionary agent-based model. *Journal of Simulation*, 1-11.

Johnson, P. C., Barry, S. J., Ferguson, H. M., & Müller, P. (2015). Power analysis for generalized linear mixed models in ecology and evolution. *Methods in Ecology and Evolution*, *6*(2), 133-142.

Kaufman, J. C., & Baer, J. (2012). Beyond new and appropriate: who decides what is creative? *Creativity Research Journal*, *24*(1), 83-91.

Kenworthy, J. B., Marusich, L. R., Paulus, P. B., Abellanoza, A., & Bakdash, J. Z. (2020). The impact of top performers in creative groups. *Psychology of Aesthetics, Creativity, and the Arts*.

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

Kohn, N. W., & Smith, S. M. (2011). Collaborative fixation: effects of others' ideas on brainstorming. *Applied Cognitive Psychology*, *25*(3), 359-371.

Kohn, N., & Smith, S. M. (2009). Partly versus completely out of your mind: effects of incubation and distraction on resolving fixation. *The Journal of Creative Behavior*, *43*(2), 102-118.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155-163.

Lu, C. C., & Luh, D. B. (2012). A comparison of assessment methods and raters in product creativity. *Creativity Research Journal*, *24*(4), 331-337.

Lu, J. G., Akinola, M., & Mason, M. F. (2017). “Switching On” creativity: task switching can increase creativity by reducing cognitive fixation. *Organizational Behavior and Human Decision Processes*, *139*, 63-75.

Madjar, N., & Shalley, C. E. (2008). Multiple tasks' and multiple goals' effect on creativity: forced incubation or just a distraction? *Journal of Management*, *34*(4), 786-805.

Malycha, C. P., & Maier, G. W. (2017). The random-map technique: enhancing mind-mapping with a conceptual combination technique to foster creative potential. *Creativity Research Journal*, *29*(2), 114-124.

Mayo, E. (1924). Recovery and industrial fatigue. *Journal of Personnel Research, 3*, 273–281.

Mednick, M. T., Mednick, S. A., & Mednick, E. V. (1964). Incubation of creative performance and specific associative priming. *The Journal of Abnormal and Social Psychology*, *69*(1), 84.

Mooneyham, B. W., & Schooler, J. W. (2013). The costs and benefits of mind-wandering: a review. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *67*(1), 11.

Moshagen, M. (2020) semPower: Power Analyses for SEM. R package version 1.1.0. https://CRAN.R-project.org/package=semPower

Murray, S., Liang, N., Brosowsky, N., & Seli, P. (2021). What are the benefits of mind wandering to creativity? *Psychology of Aesthetics, Creativity, and the Arts*.

Nam, H., & Lee, C. H. (2015). The immediate incubation effect on creative problem solving: using the remote association task. *Psychologia*, *58*(2), 98-113.

Nijstad, B. A., & Stroebe, W. (2006). How the group affects the mind: a cognitive model of idea generation in groups. *Personality and Social Psychology Review*, *10*(3), 186-213.

Orr, J. M., Sackett, P. R., & Dubois, C. L. (1991). Outlier detection and treatment in I/O psychology: A survey of researcher beliefs and an empirical illustration. *Personnel Psychology*, *44*(3), 473-486.

Osborn, A. F. (1958). Creative thinking. *American Association of Industrial Nurses Journal*, *6*(9), 23-25.

Paulus, P. B., & Kenworthy, J. B. (2019). Effective brainstorming. *The Oxford Handbook of Group Creativity and Innovation*, 287-386.

Paulus, P. B., & Nijstad, B. A. (Eds.). (2003). *Group creativity: Innovation through collaboration*. Oxford University Press.

Paulus, P. B., & Yang, H. C. (2000). Idea generation in groups: a basis for creativity in organizations. *Organizational Behavior and Human Decision Processes*, *82*(1), 76-87.

Paulus, P. B., Nakui, T., Putman, V. L., & Brown, V. R. (2006). Effects of task instructions and brief breaks on brainstorming. *Group Dynamics: Theory, Research, and Practice*, *10*(3), 206.

Paulus, P. B., Putman, V. L., Dugosh, K. L., Dzindolet, M. T., & Coskun, H. (2002). Social and cognitive influences in group brainstorming: predicting production gains and losses. *European Review of Social Psychology*, *12*(1), 299-325.

Paulus, P.B, Kenworthy, J.B., Abellanoza, A. (2021). *The Influence of Group Dynamics in "Hackathon" Settings.* Unpublished manuscript.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods*. doi: 10.3758/s13428-018-01193-y

Reiter-Palmon, R., Forthmann, B., & Barbot, B. (2019). Scoring divergent thinking tests: a review and systematic framework. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 144.

Revelle, W. (2021). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 2.1.9, https://CRAN.R-project.org/package=psych.

Ritter, S. M., & Dijksterhuis, A. (2014). Creativity—the unconscious foundations of the incubation period. *Frontiers in Human Neuroscience*, *8*, 215.

Ritter, S. M., Van Baaren, R. B., & Dijksterhuis, A. (2012). Creativity: The role of unconscious processes in idea generation and idea selection. *Thinking Skills and Creativity*, *7*(1), 21-27.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1-36. http://www.jstatsoft.org/v48/i02/.

Ruby, F. J., Smallwood, J., Engen, H., & Singer, T. (2013). How self-generated thought shapes mood—the relation between mind-wandering and mood depends on the socio-temporal content of thoughts. *PloS One*, *8*(10), e77554.

Segal, E. (2004). Incubation in insight problem solving. *Creativity Research Journal*, *16*(1), 141-148.

Silvia, P. J. (2011). Subjective scoring of divergent thinking: examining the reliability of unusual uses, instances, and consequences tasks. *Thinking Skills and Creativity, 6*(1), 24-30.

Silvia, P. J., Beaty, R. E., & Nusbaum, E. C. (2013). Verbal fluency and creativity: general and specific contributions of broad retrieval ability (Gr) factors to divergent thinking. *Intelligence*, *41*(5), 328 -340.

Sio, U. N., & Ormerod, T. C. (2009). Does incubation enhance problem solving? A meta-analytic review. *Psychological Bulletin*, *135*(1), 94.

Slotter, E. B., Winger, L., & Soto, N. (2015). Lost without each other: the influence of group identity loss on the self-concept. *Group Dynamics: Theory, Research, and Practice*, *19*(1), 15.

Smeekens, B. A., & Kane, M. J. (2016). Working memory capacity, mind wandering, and creative cognition: an individual-differences investigation into the benefits of controlled versus spontaneous thought. *Psychology of Aesthetics, Creativity, and the Arts*, *10*(4), 389.

Smith, S. M., & Beda, Z. (2020). Old problems in new contexts: the context-dependent fixation hypothesis. *Journal of Experimental Psychology: General*, *149*(1), 192.

Smith, S. M., & Blankenship, S. E. (1989). Incubation effects. *Bulletin of the Psychonomic Society*, *27*(4), 311-314.

Smith, S. M., Gerkens, D. R., & Angello, G. (2017). Alternating incubation effects in the generation of category exemplars. *The Journal of Creative Behavior*, *51*(2), 95-106.

Steindorf, L., Hammerton, H. A., & Rummel, J. (2020). Mind wandering outside the box—about the role of off-task thoughts and their assessment during creative incubation. *Psychology of Aesthetics, Creativity, and the Arts*.

Tan, T., Zou, H., Chen, C., & Luo, J. (2015). Mind wandering and the incubation effect in insight problem solving. *Creativity Research Journal*, *27*(4), 375-382.

Torrance, E. P. (1966) *Torrance Tests of Creative Thinking: Norms-technical manual (Research Edition).* Personnel Press.

Toubia, O., & Netzer, O. (2017). Idea generation, creativity, and prototypicality. *Marketing Science*, *36*(1), 1-20.

Wallace, M. (2007). *Jawbone Java WordNet API*. https://sites.google.com/site/mfwallace/jawbone.

Weinstein, Y. (2018). Mind-wandering, how do I measure thee with probes? Let me count the ways. *Behavior Research Methods*, *50*(2), 642-661.

Williams, K. J., Lee, K. E., Hartig, T., Sargent, L. D., Williams, N. S., & Johnson, K. A. (2018). Conceptualising creativity benefits of nature experience: attention restoration and mind wandering as complementary processes. *Journal of Environmental Psychology*, *59*, 36-45.

Yamaoka, A., & Yukawa, S. (2016). Mind-wandering enhances creative problem solving. *Shinrigaku Kenkyu: The Japanese Journal of Psychology*, *87*(5), 506-512.

Yamaoka, A., & Yukawa, S. (2020). Does mind-wandering during the thought incubation period improve creativity and worsen mood? *Psychological Reports*, *123*(5), 1785-1800.

Yuan, Y., & Shen, W. (2016). Commentary: Incubation and intuition in creative problem solving. *Frontiers in Psychology*, *7*, 1807.

Zedelius, C. M., & Schooler, J. W. (2015). Mind wandering "Ahas" versus mindful reasoning: alternative routes to creative solutions. *Frontiers in Psychology*, *6*, 834.

Zhou, X., Zhai, H. K., Delidabieke, B., Zeng, H., Cui, Y. X., & Cao, X. (2019). Exposure to ideas, evaluation apprehension, and incubation intervals in collaborative idea generation. *Frontiers in Psychology*, *10*, 1459.

Appendix A

List of Items from the Spontaneous and Deliberate Mind-Wandering Scale (SDMWS)

Items from the SDMWS *(*Marcusson-Clavertz & Kjell, 2018):

1. "I find my thoughts wandering spontaneously." (1 = rarely, 7 = a lot)
2. "When I mind wander, my thoughts tend to be pulled from topic to topic." (1 = rarely, 7 = a lot)
3. "I mind wander even when I'm supposed to be doing something else." (1 = rarely, 7 = a lot)
4. "I allow my thoughts to wander on purpose." (1 = rarely, 7 = a lot)
5. "I enjoy mind wandering." (1 = rarely, 7 = a lot)
6. "I find mind wandering is a good way to cope with boredom." (1 = not at all true, 7 = very true)
7. "I allow myself to get absorbed in pleasant fantasy." (1 = rarely, 7 = a lot)

Appendix B

List of Word Stimuli used in Verbal N-Back Tasks

1. 'aim'
2. 'cat'
3. 'dip'
4. 'for'
5. 'hem'
6. 'jet'
7. 'lob'
8. 'map’
9. 'not'
10. 'orb'
11. 'pet'
12. 'ran'
13. 'sit'
14. 'try'
15. 'vat'
16. 'wet'

Appendix C

N-Back Instructions for Each Condition

Verbal/Easy:

*****PLEASE READ ALL INSTRUCTIONS ON THIS PAGE CAREFULLY BEFORE BEGINNING!*****

The next task is a "reaction game". You will perform this task for the next **15 minutes**.

In this task, words will appear every few seconds. Periodically, the words will be colored **red**.

If the red word **starts with the letters A-M**, press the letter "F" on the keyboard as fast as you can.

If the red word **starts with the letters N-Z**, press the letter "J" as fast as you can.

cat	try
Press the F key	**Press the J key**

You will also be periodically asked what you are thinking about as you complete the task, e.g. a "thought probe".

The order of each presentation is **RANDOM**, so you may get multiple probes or red prompts in a row.

Continue to work until the task ends automatically.

Press **Enter** to begin.

Appendix C (cont.)

Verbal/Hard:

*****PLEASE READ ALL INSTRUCTIONS ON THIS PAGE CAREFULLY BEFORE BEGINNING!*****

The next task is a "memory game". You will perform this task for the next **15 minutes**.

In this task, words will appear every few seconds. Periodically, a **red** question mark will appear.

Your goal is to remember the word that appeared before the question mark.

?

If the word **started with the letters A-M**, press the letter "F" on the keyboard as fast as you can.

If the word **started with the letters N-Z**, press the letter "J" as fast as you can.

cat try

Press the F key on ? **Press the J key on ?**

You will also be periodically asked what you are thinking about as you complete the task, e.g. a "thought probe".

The order of each presentation is **RANDOM**, so you may get multiple probes or red prompts in a row.

Continue to work until the task ends automatically.

Press **Enter** to begin.

Appendix C (cont.)

Spatial/Easy:

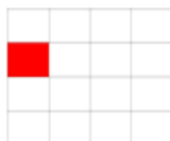
*****PLEASE READ ALL INSTRUCTIONS ON THIS PAGE CAREFULLY BEFORE BEGINNING!*****

The next task is a "reaction game". You will perform this task for the next **15 minutes**.

In this task, boxes will appear in a grid. Periodically, the box will be **red**.

If the red box is **in the left two columns**, press the letter "F" on the keyboard as fast as you can.

If the red box is **in the right two columns**, press the letter "J" as fast as you can.



Press the F key **Press the J key**

You will also be periodically asked what you are thinking about as you complete the task, e.g. a "thought probe".

The order of each presentation is **RANDOM**, so you may get multiple probes or red prompts in a row.

Continue to work until the task ends automatically.

Press **Enter** to begin.

Appendix C (cont.)

Spatial/Hard:

*****PLEASE READ ALL INSTRUCTIONS ON THIS PAGE CAREFULLY BEFORE BEGINNING!*****

The next task is a "memory game". You will perform this task for the next **15 minutes**.

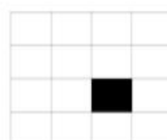
In this task, boxes will appear in a grid. Periodically, a **red** question mark will appear.

Your goal is to remember the position of the box that appeared before the question mark.

If the box was **in the left two columns**, press the letter "F" on the keyboard as fast as you can.

If the box was **in the right two columns**, press the letter "J" as fast as you can.

Press the F key on ?



Press the J key on ?

You will also be periodically asked what you are thinking about as you complete the task, e.g. a "thought probe".

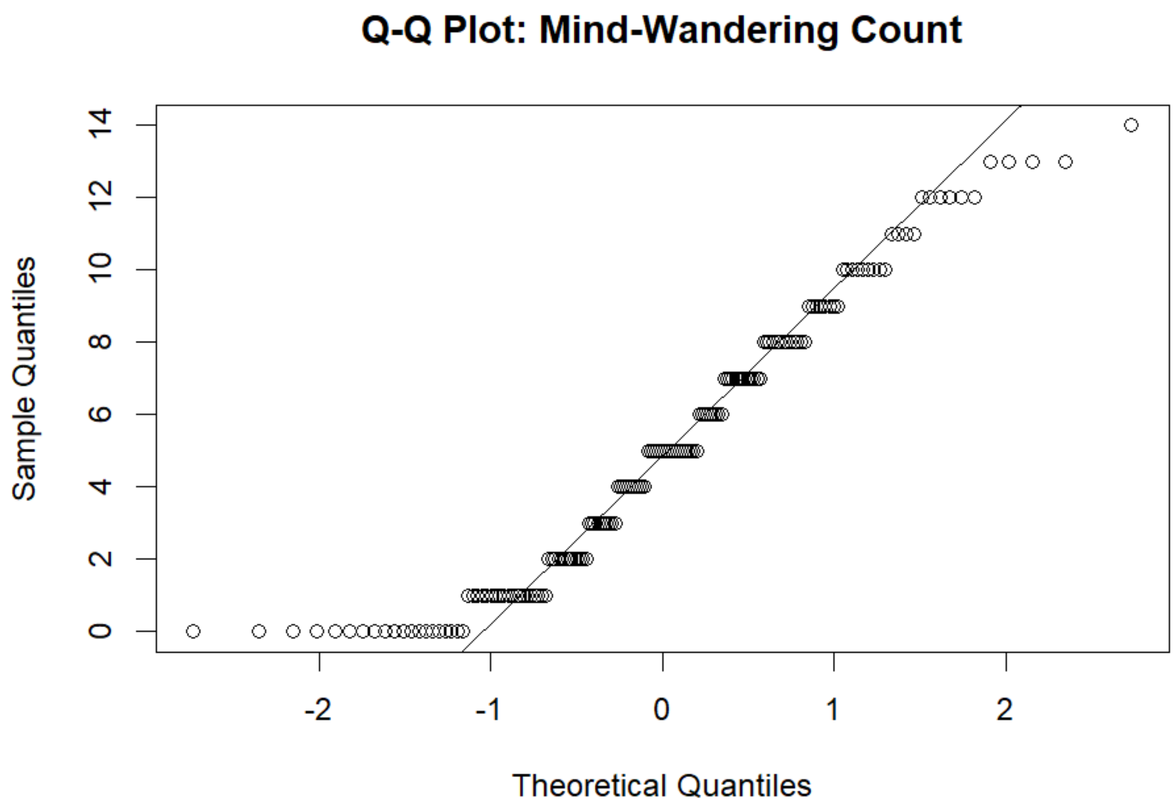
The order of each presentation is **RANDOM**, so you may get multiple probes or red prompts in a row.

Continue to work until the task ends automatically.

Press **Enter** to begin.

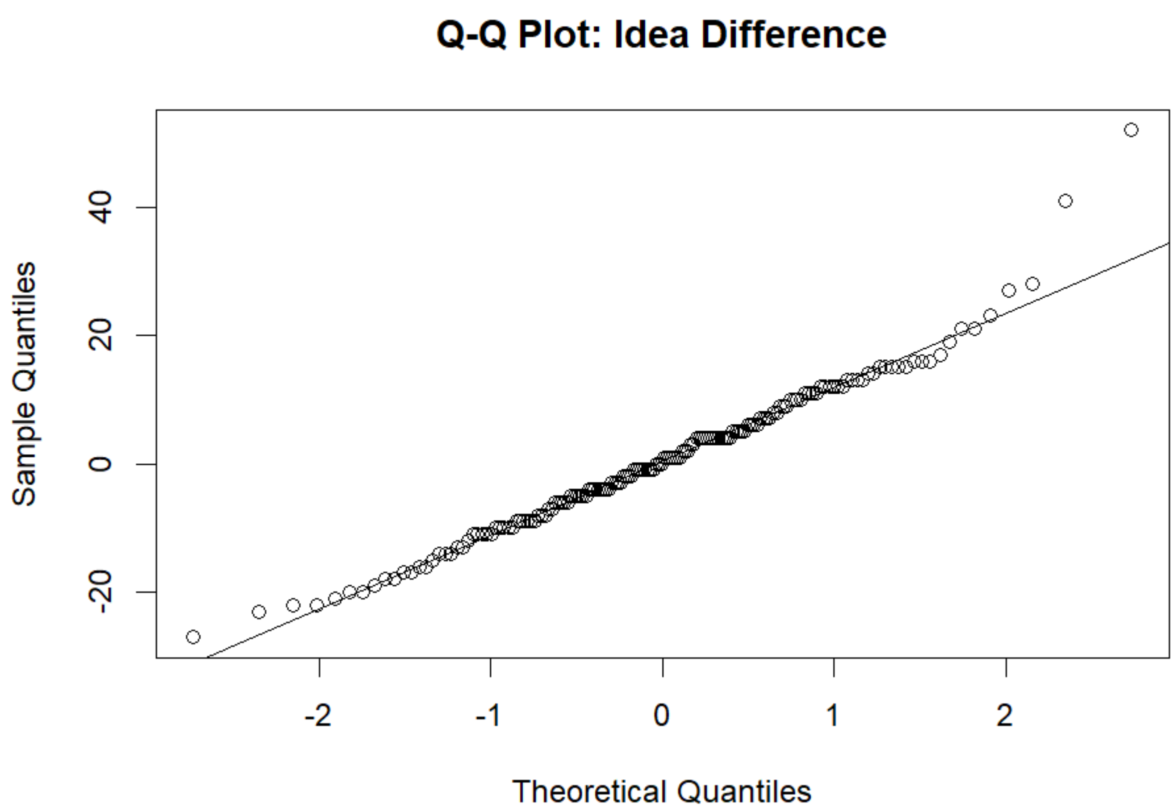
Appendix D

Q-Q Plots of Endogenous Variables in Path Model



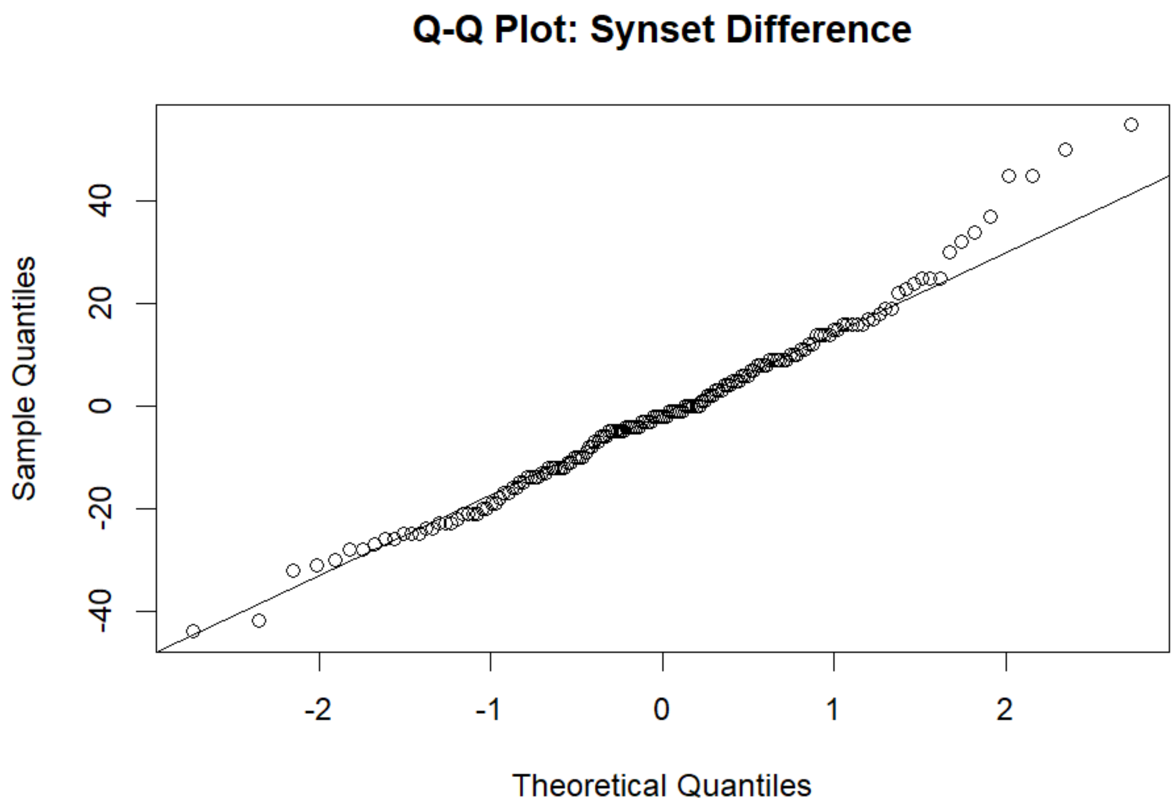
Note. Plot indicates significant deviation from normality.

Appendix D (cont.)



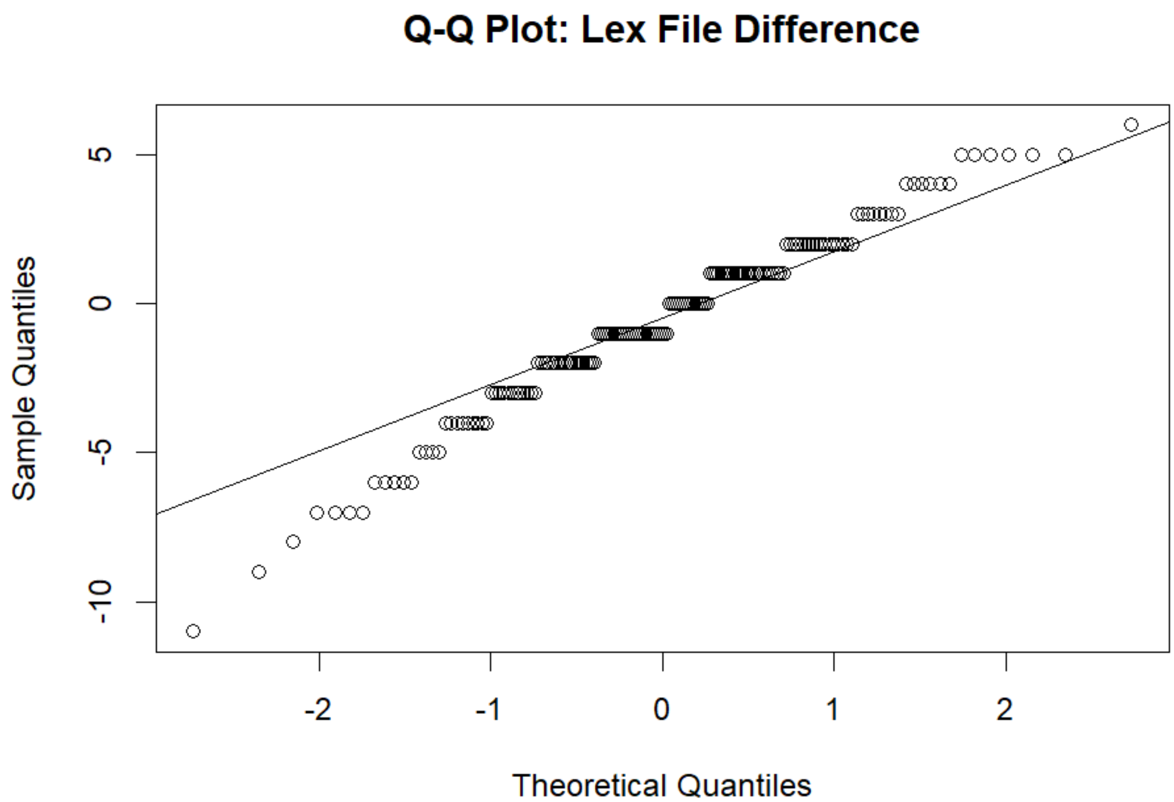
Note. Plot indicates significant deviation from normality.

Appendix D (cont.)



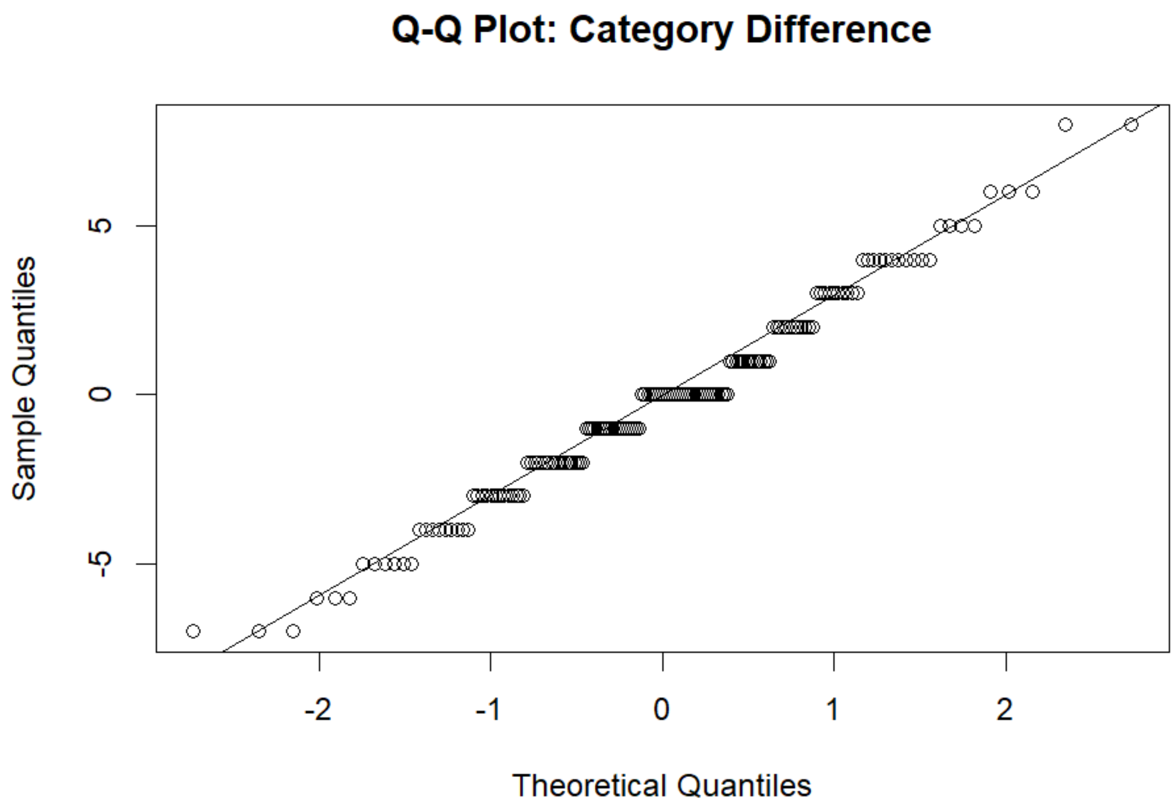
Note. Plot indicates significant deviation from normality.

Appendix D (cont.)



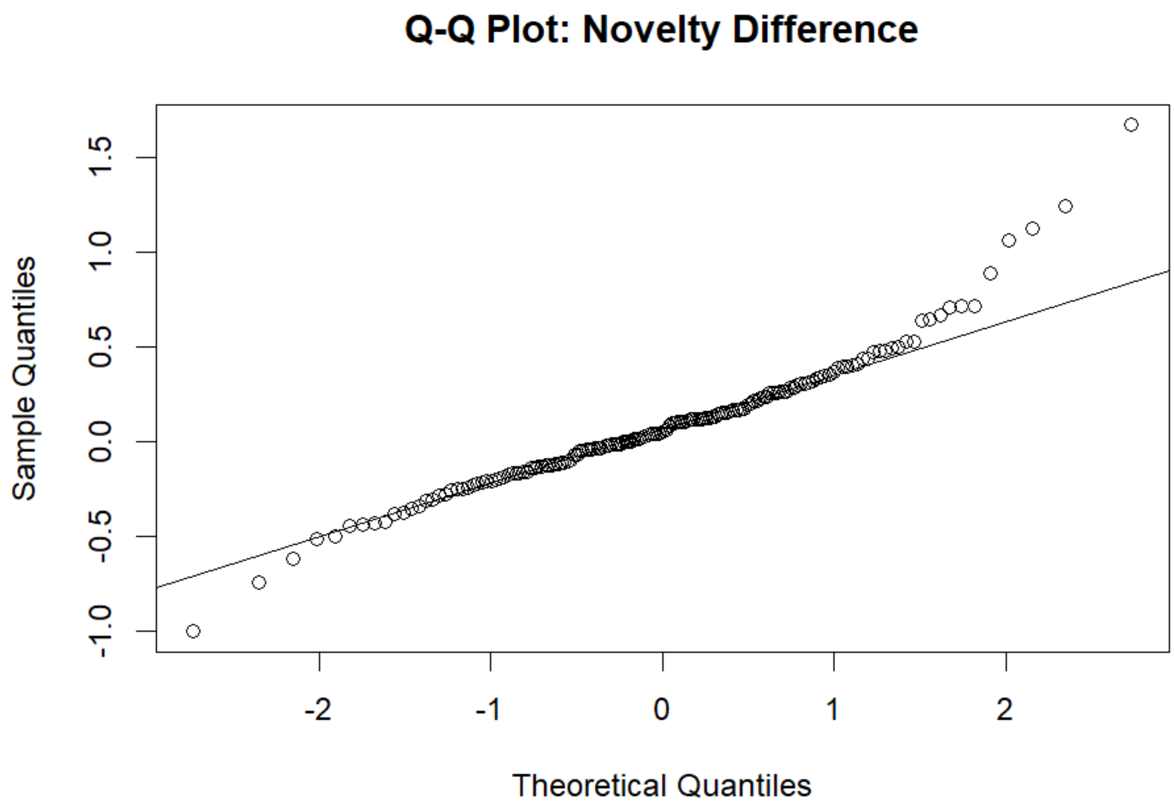
Note. Plot indicates significant deviation from normality.

Appendix D (cont.)



Note. Plot indicates slight deviation from normality.

Appendix D (cont.)



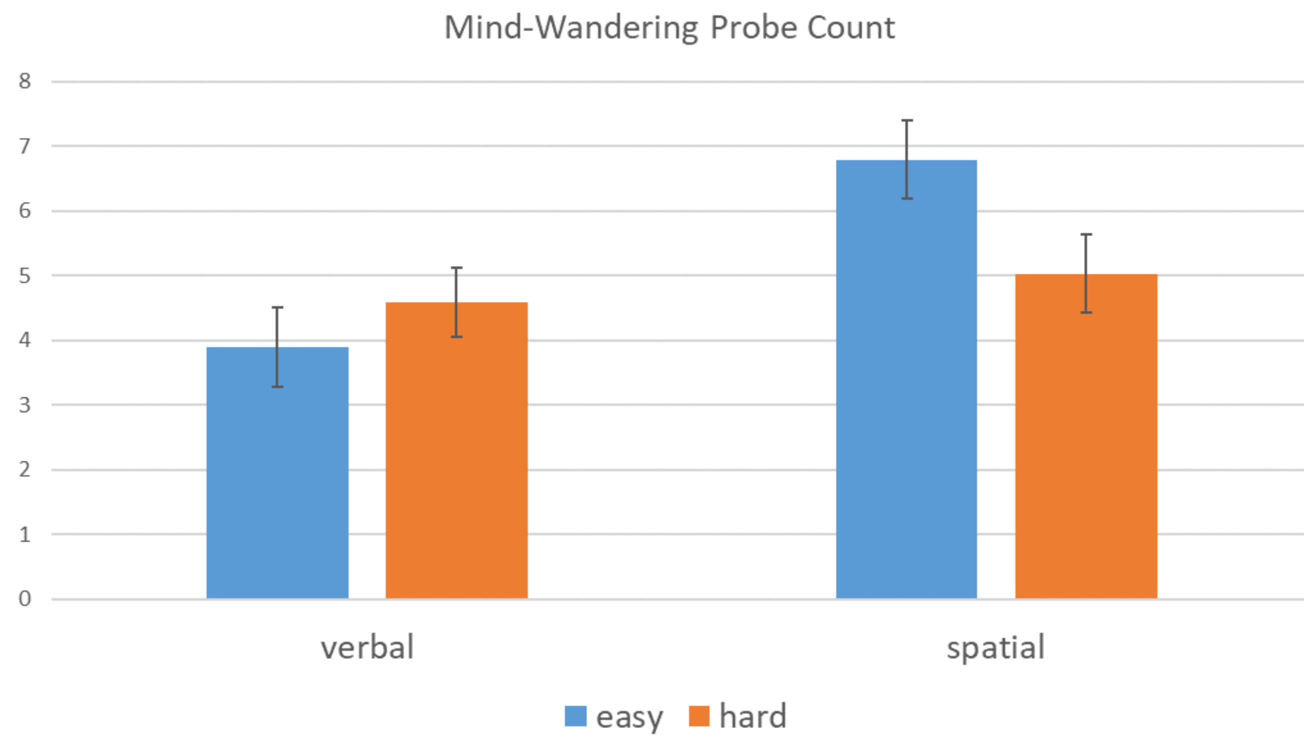
Note. Plot indicates significant deviation from normality.

Appendix E

ANOVAs: Creativity Measures & Mind-Wandering

Mind-Wandering Probe Count

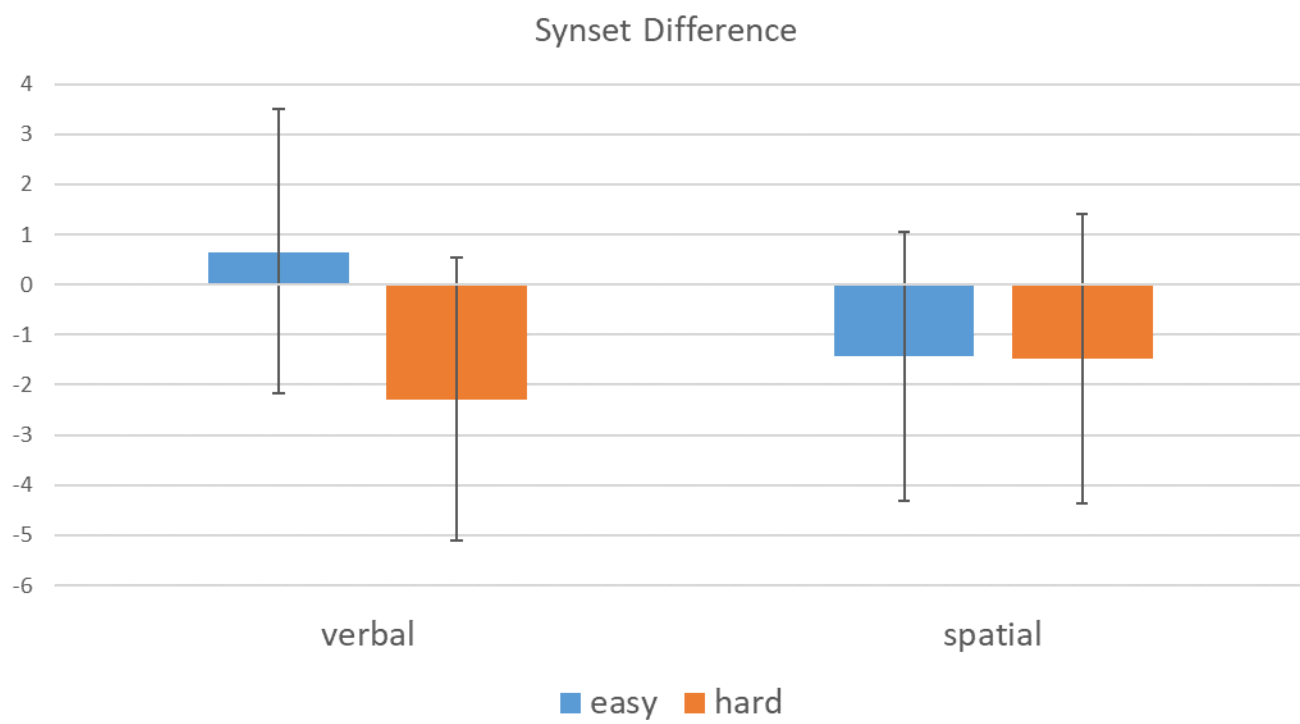
Effect	F	df_{effect}	$df_{residual}$	p	η^2
Similarity	7.89	1	156	.01*	.05
Difficulty	0.59	1	156	.44	.003
Similarity X Difficulty	4.4	1	156	.04*	.03



Appendix E (cont.)

Synset Difference Score

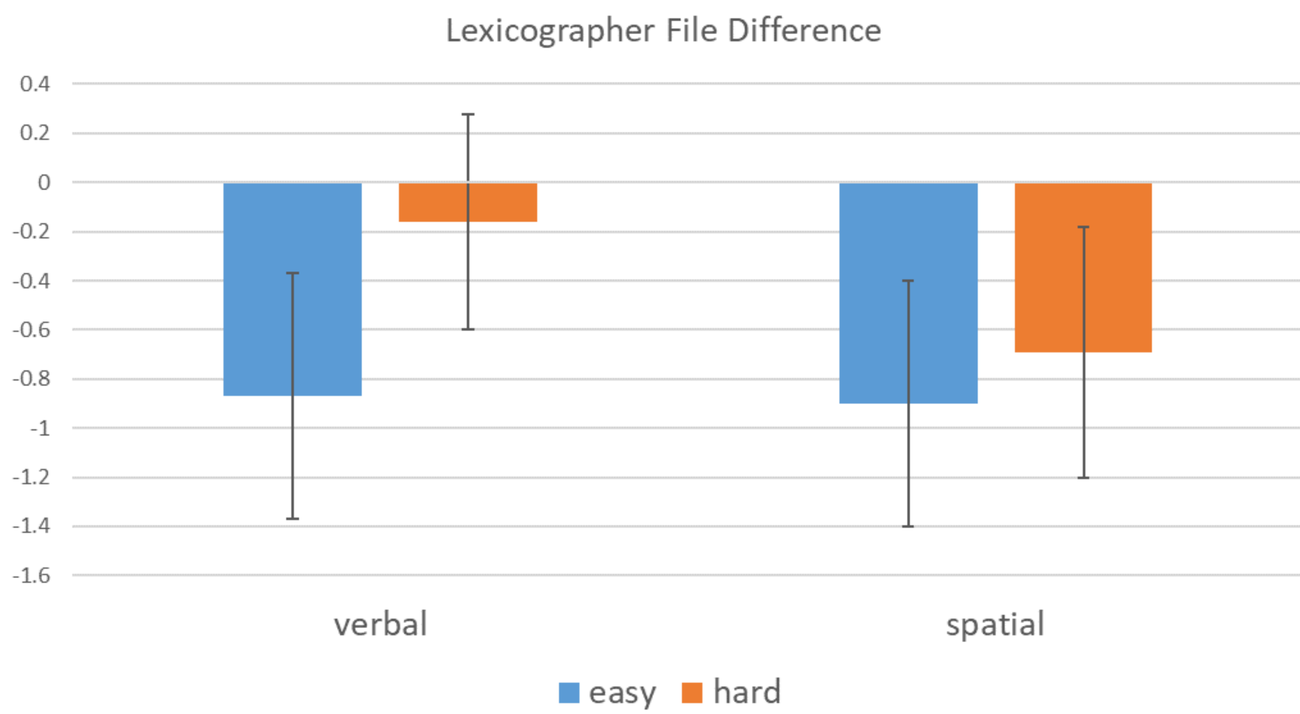
Effect	F	df_{effect}	$df_{residual}$	p	η^2
Similarity	0.02	1	156	.88	< .001
Difficulty	0.33	1	156	.57	.002
Similarity X Difficulty	0.27	1	156	.60	.002



Appendix E (cont.)

Lexicographer File Difference Score

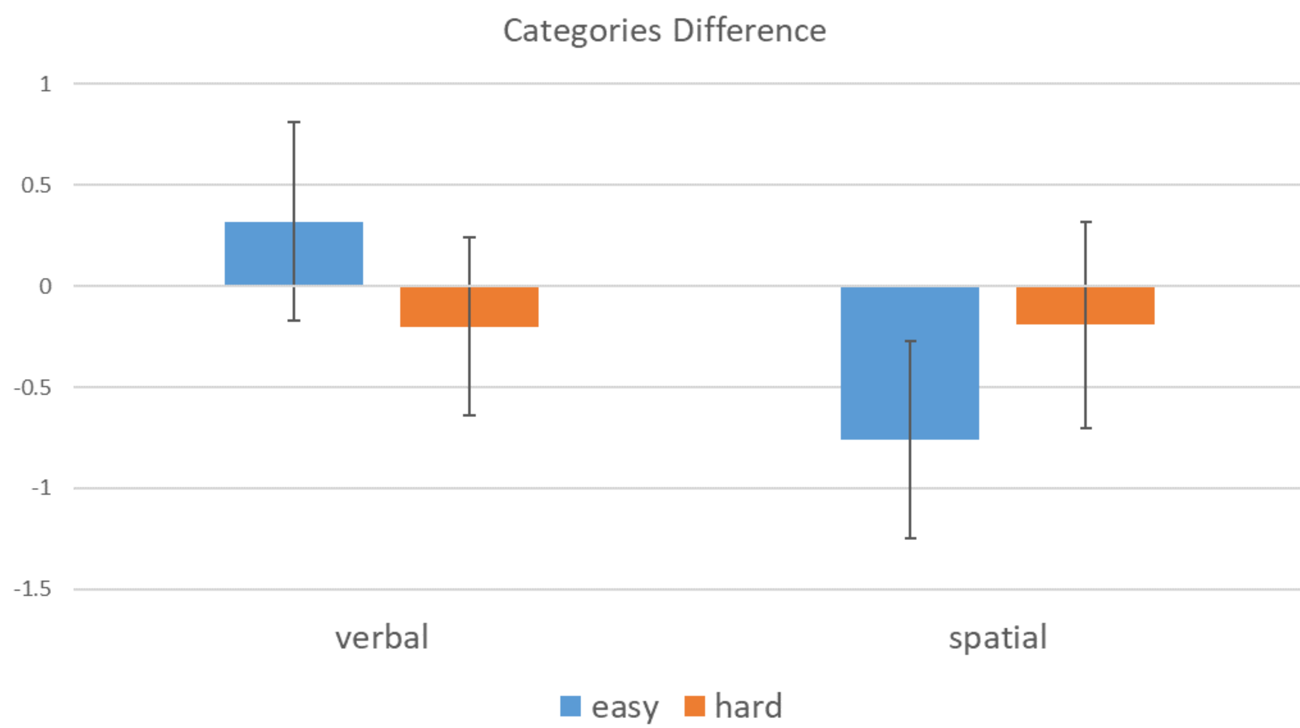
Effect	*F*	df_{effect}	$df_{residual}$	*p*	η^2
Similarity	0.47	1	156	.50	.003
Difficulty	0.92	1	156	.34	.01
Similarity X Difficulty	0.26	1	156	.61	.002



Appendix E (cont.)

Categories Difference Score

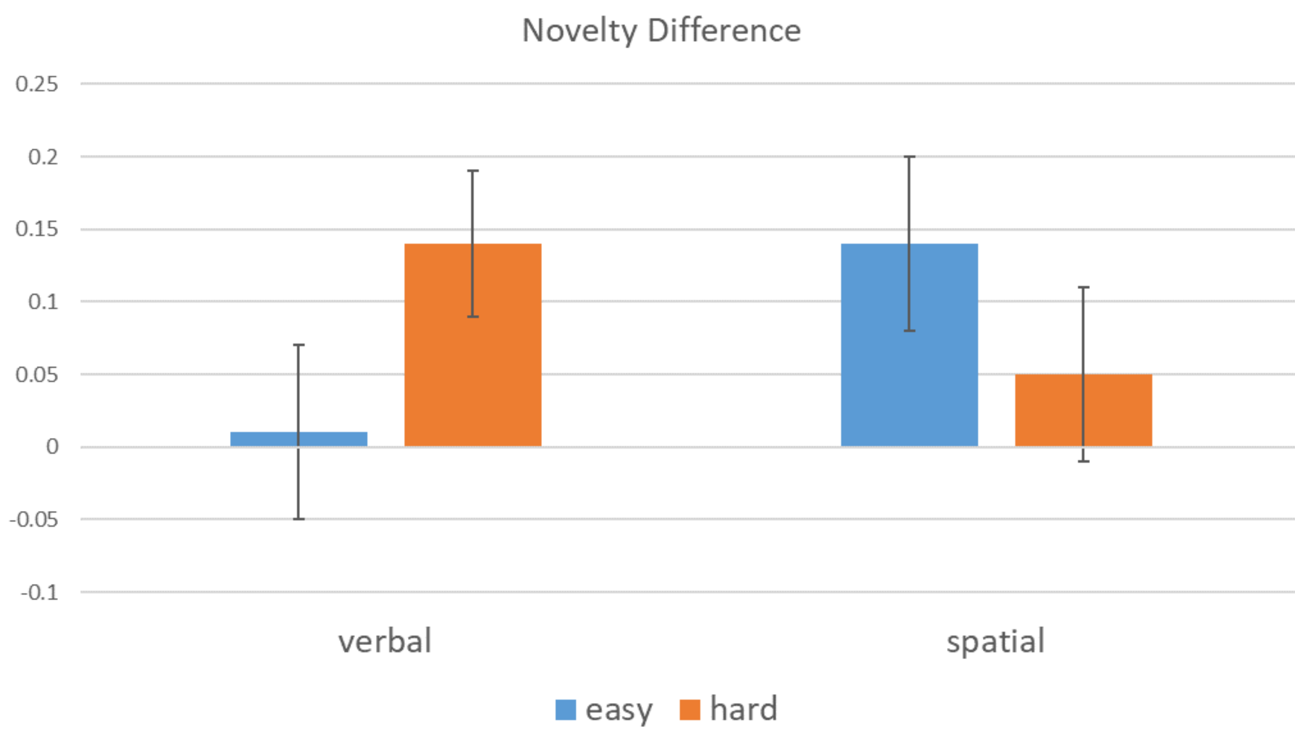
Effect	*F*	df_{effect}	$df_{residual}$	*p*	η^2
Similarity	1.16	1	156	.28	.01
Difficulty	0.001	1	156	.97	< .001
Similarity X Difficulty	1.33	1	156	.25	.01



Appendix E (cont.)

Novelty Difference Score

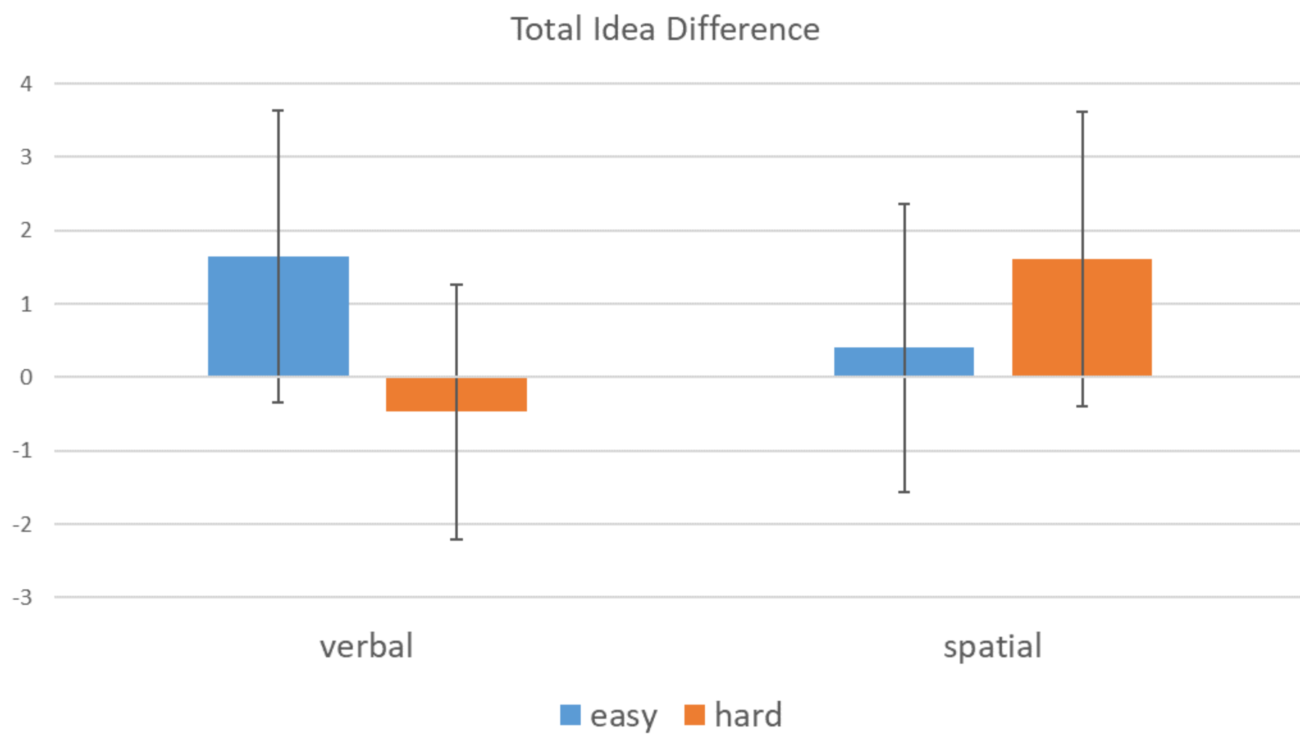
Effect	*F*	df_{effect}	$df_{residual}$	*p*	η^2
Similarity	0.05	1	156	.82	< .001
Difficulty	0.30	1	156	.59	.002
Similarity X Difficulty	4.15	1	156	.04*	.03



Appendix E (cont.)

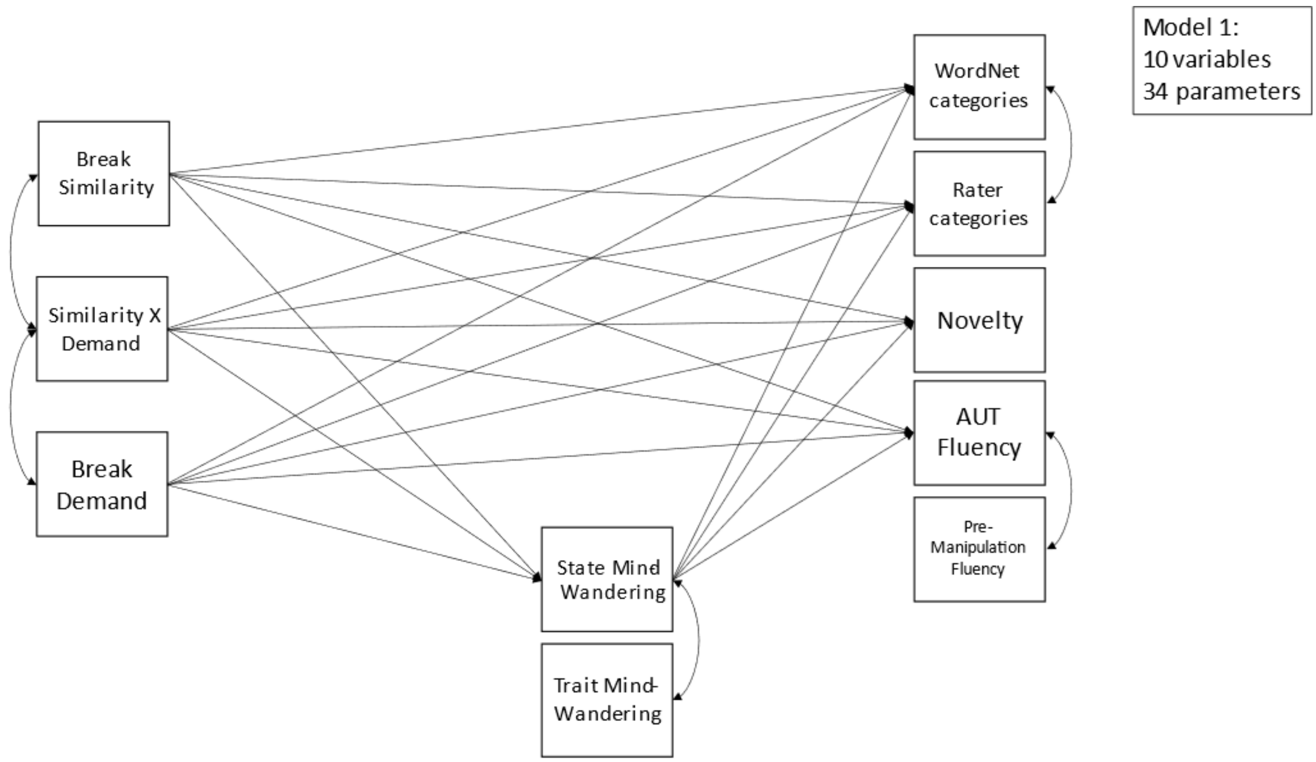
Total Idea Difference Score

Effect	F	df_{effect}	$df_{residual}$	p	η^2
Similarity	0.08	1	156	.78	.001
Difficulty	0.09	1	156	.77	.001
Similarity X Difficulty	0.75	1	156	.39	.01



Appendix F

Original Proposed Model and Power Analysis



Appendix F (cont.)

```
### formula:
# k = total number of variables
# i_cv = number of elements in covariance matrix
# n_param = number of parameters estimated
# i_cv = (k*(k+1)) / 2
# df = i_cv - n_param

### calculation using example diagram:
# k = 10, n_param = 34
# i_cv = (10*11) / 2  = 55
# df = 55 - 34 = 21

# a-priori power analysis
summary(semPower.aPriori(effect = .08, effect.measure = 'RMSEA',
                                  alpha = .05, power = .80, df = 21))

##
##  semPower: A-priori power analysis
##
##  F0                         0.134400
##  RMSEA                      0.080000
##  Mc                         0.935008
##
##  df                         21
##  Required Num Observations  160
##
##  Critical Chi-Square        32.67057
##  NCP                        21.36960
##  Alpha                      0.050000
##  Beta                       0.199664
##  Power (1-beta)             0.800336
##  Implied Alpha/Beta Ratio   0.250421
```