UNIVERSITY OF CALIFORNIA

Los Angeles

Creative Problem-Solving in Mathematics: Immersion, Impasse, Incubation, and Insight

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Psychology

by

Stacy Tamsen Shaw

2020

ABSTRACT OF THE DISSERTATION

Creative Problem Solving in Mathematics: Immersion, Impasse, Incubation, and Insight

By

Stacy Tamsen Shaw

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2020

Professor James Stigler, Chair

Although creativity research frequently borrows anecdotes from mathematicians, most research is conducted in a lab setting with abstract tasks known to be heavily confounded with verbal fluency (e.g. RATs, anagrams). This is unfortunate, as utilizing an area such as math would not only diversify creativity research, but allow exploration of how factors such as identity and affect can relate to creative processes. In the current dissertation, I employed a creative math puzzle to extend previous work on creativity and insight to the realm of mathematics, and explored how trait individual differences and state differences relate to solve rates both within the lab, and outside the lab (up to three days later). Study 1 recruited 231 undergraduate students, who were brought into the lab and randomly assigned to a condition—a low-demand incubation condition (LD), high-demand incubation condition (HD), or a control group. All groups had six minutes to work on the puzzle, but students in the LD and HD conditions took a break after three minutes to complete a signal detection task (LD condition) or complex- reading task (HD condition) for 2.5 minutes. If students were unable to solve the puzzle in lab, they were provided a follow-up survey link to fill out if they solved the math puzzle later, or if three days had passed and they had not solved. Results showed that there was no effect of incubation condition on problem solving within the lab, but it was significantly related to solving outside of the lab. Interestingly, control condition students had a greater probability of solving outside the lab compared to LD students. I also found that several factors significantly related with problem solving in the lab (i.e. math anxiety, trait emotions) were not related to solving the problem outside of the lab.

Study 2 attempted to replicate the findings of study 1 by adopting the same procedure (but limiting conditions to control and LD) and extending to trait individual difference measures such as openness, intellect, and different aspects of curiosity. Study 2 also evaluated an opportunistic assimilation account of the findings of study 1, which suggest that control participants may be out-solving their LD counterparts in the wild because they were reaching an impasse more than LD students, allowing them to pick up on cues from their environment that aid with solving the problem outside the lab. However, results from 252 students showed that incubation condition had no effect on solving in lab or the wild, failing to replicate the results from study 1. Further, impasse was not found to relate to solving in the wild, and while some students reported hints had helped them solve in the wild, this was only a small subset of the sample. Among trait individual difference measures, intellect and curious-I were found to positively relate to solving in the lab, but along with the other trait and individual difference measures, failed to predict whether students solved in the wild. Collectively, this dissertation highlights the complex nature of creative problem-solving in mathematics, and how different aspects of data collection (lab and wild) can contribute to a richer understanding of students' creative cognition.

The dissertation of Stacy Tamsen Shaw is approved.

Alan Castel

Keith Holyoak

Gerardo Ramirez

James Stigler, Committee Chair

University of California, Los Angeles

2020

This dissertation is dedicated to the memory of Lois Shaw, Bud Shaw, Caroline Krajcar, and John Krajcar, my loving grandparents.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Dissertations are hard, and there are many people to thank for aiding me and supporting me on this very long journey. First and foremost, I'd like to thank my partner Webster for not only an incredible 15 years full of adventures, love, and support, but also putting up with me in dissertation-mode while quarantined for over the last three months. There is absolutely no way I could have done this without you. I'd like to thank my parents who have always let me be my independent self, and believed in my ability even when it was not obvious to me. Huge thanks to my sister, Megan, who has gone above and beyond the call of sisterhood to ensure that I was well-fed, rested, encouraged, and even had some fun in graduate school through multiple fancy dinners and Disneyland trips (you win the best sister award). I also incredible grateful to my brother Jeff, who has always been a wonderful role model in my life but somehow, always managed to keep my spirits high through the most hilarious facetime calls. I'm also incredibly grateful to be a sister-in-law to Lindsey and Andrew, two fantastic people who have not only supported me through much of my schooling, but allowed me to be a proud aunt to Evan, Sydney, and Brynley. Last but not least I'd like to acknowledge the support over the years from Walt and Annette Webster, Rachel Sommers, as well my numerous cousins, aunts, and uncles, who have stepped up in many ways to support me in the last five years and continued to be my biggest cheerleaders.

I also wish to express how grateful I am to my academic family. I'd like to thank my two graduate advisors—Gerardo Ramirez and Jim Stigler, who are not only brilliant researchers and scientists, but two very kind and dedicated mentors which I have benefited instrumentally from. I could not have asked for two better advisors. I'd also like to thank Karen Givvin and Ji Son for their informal mentorship, and thank Alan Castel and Keith Holyoak for their invaluable feedback on this dissertation. Special thanks to Laura Fries, Mary Tucker, Marissa Ogren, Danny Rahal, Tawny Tsang, Vanessa Hilo, Sarah Tashjian, Dani Smith, Anahit Pogossian, Michelle Luna, Gwen Price, João F. Guassi Moreira, and the other graduate students in psychology at UCLA who have been my friends, collaborators, and fellow TAs. I am also incredibly grateful to Christina Chin-Newman and Jay Feldman, my mentors before graduate school who taught me so much about psychological research, education and equity, and helped me develop the skills I needed to become the researcher I am today.

The research presented in this dissertation was funded by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1650604. Any opinions, findings, and conclusions or recommendations pressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

STACY TAMSEN SHAW

EDUCATION

M.A. Psychology, University of California, Los Angeles — 2016
Co-advisors: Gerardo Ramirez and James Stigler

B. A. Human Development, California State University, East Bay — 2012
Minor in Psychology
Minor in Statistics

PROFESSTIONAL RESEARCH POSITIONS

Worcester Polytechnic Institute — 2020 - Present
Assistant Professor
Department of Psychological and Cognitive Sciences

University of California, Los Angeles — 2015 – 2020
Graduate Student Researcher
Department of Psychology

RTI International — 2012 – 2015
Education Analyst II (Research Analyst)
Center for the Evaluation and Study of Educational Equity

California State University, East Bay — 2010 – 2015
Research Associate
Arts, Creativity, and Education Lab

AWARDS, FELLOWSHIPS, AND SCHOLARSHIPS

- Dena Chertoff Graduate Student Service Award, UCLA Department of Psychology — 2020
- Society for Neuroscience of Creativity Annual Conference, NSF Travel Award — 2019
- Graduate Research Fellowship, National Science Foundation — 2017
- Graduate Summer Research Mentorship, UCLA Department of Psychology — 2017
- Graduate Summer Research Mentorship, UCLA Department of Psychology — 2016
- Distinguished University Fellowship, UCLA Department of Psychology — 2015
- Research Scholarship, California State University, East Bay — 2012

PUBLICATIONS

Peer-Reviewed Journal Articles

Stigler, J. W., Son, J. Y., Givvin, K. B., Blake, A., Fries, L., **Shaw, S. T.,** & Tucker, M. (in press). The *Better Book* model for education research and development. *Teacher College Record.*

Shaw, S. T., Pogossian, & Ramirez, G. (2020). The mathematical flexibility of college students: The role of affective and cognitive factors. *British Journal of Educational Psychology.* doi: 10.1111/bjep.12340

Shaw, S. T., Spink, K. S., & Chin-Newman, C. S. (2019). "Do I really belong here?": The stigma of being a community college transfer student at a four-year university. *Community College Journal of Research and Practice.* doi: 10.1080/10668926.2018.1528907

Ramirez, G., **Shaw, S. T.,** & Maloney, E. A. (2018). Math anxiety: Past research, promising interventions, and a new interpretation framework. *Educational Psychologist, 5*(3), 145-164. doi: 10.1080/00461520.2018.1447384

Shaw, S. T. & Chin-Newman, C. S. (2017). "You can do it!" Social support during the transition from community college to a four-year university. *Journal of the First-Year Experience & Students in Transition, 29*(2), 65-78.

Chin-Newman, C. S., & **Shaw, S. T.** (2013). The anxiety of change: How transfer students face challenges. *Journal of College Admission*, (221), 1. p. 15-21.

Technical Reports and Other Publications

Chin-Newman, C. S. & **Shaw, S. T.** (2015). Strengthening campus-based social support for transfer students. *E-SOURCE for College Transitions, 12*(2), 11-13.

Woo, J. & **Shaw, S.** (2015). *Trends in Graduate Student Financing: Selected Years, 1995–96 to 2011–12 (NCES 2015-026).* National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Shaw, S. & Radwin, D. (2014). *Comparison of Original and Revised Student Financial Aid Estimates for 2007-2008 (NCES 2014-179).* National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Ifill, N. & **Shaw, S.** (2013). *Undergraduate Financial Aid Estimates by Type of Institution in 2011-2012 (NCES 2013-169).* National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Radford, A., Lew, T., & **Shaw, S.** (2013). *Today's Baccalaureate: The Fields and Courses that Make Up Current Baccalaureate Recipients' Degrees in 2007–08 (NCES 2013-755).* National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Creative Problem Solving in Mathematics: Immersion, Impasse, Incubation and Insight

At its core, the essence of mathematical thinking is creative thinking (Mann, 2006). Creative problem solving in particular, which generally describes the process of making new and meaningful connections through the generation of unusual or original possibilities (Treffinger, Isaksen, & Stead-Dorval, 2005), is a valuable skill in math. At the forefront of mathematics, expert mathematicians rely on creative problem-solving to make new connections across different areas of math to solve problems and create scientific breakthroughs (see Mackenzie, 2006). In the classroom, the practice of creative problem-solving signals both an understanding of the problem and the ability to think flexibly about concepts in math (Ervynck, 2002; Liljedahl & Sriraman, 2006)—critical skills for the development of expertise (Hatano & Inagaki, 1984). Creative problem-solving is highly valued both in math classrooms (Mann, 2006) and as a general skill (Craft, Gardner, & Claxton, 2007; Feldman, Csikszentmihalyi, & Gardner, 1994). Despite immense interest, the cognitive processes underlying creative problem solving in mathematics are not fully understood.

The nature of creative cognition is complex, but one popular theory of creative problem-solving has inspired a simple four-stage model of creativity. Wallas (1926) argued that creative problem-solving begins by first, immersing oneself in the problem to better understand it and exhaust conventional ideas. After a period of immersion, one likely reaches an impasse, or a mental block. Once an impasse has been reached, people enter a period of incubation, where they temporarily shift their attention away from the problem and do something else. During this period of incubation, or upon return to the problem, people can experience an *aha!* moment of insight, where an idea surfaces with “brevity, suddenness, and immediate certainty” (Hardy, 1946, p. 54). After experiencing insight, the potential solution or breakthrough is evaluated.

Although proposed almost 100 years ago, the four stages of creative problem-solving – immersion, incubation, insight, and evaluation – has generally held up well in empirical research across the years and continues to be a popular framework. The primary interest of researchers, however, revolves around the first three phases: immersion, incubation, and insight (Mumford, Lonergan, & Scott, 2002; Silvia, 2008). Immersion, incubation, and insight are of particular interest for the field of mathematics, where problems are often ill-structured, and require overcoming fixations, making remote connections, and thinking both convergently and divergently about concepts to arrive at solutions—all characteristics of creative problem-solving (Runco, 2014).

We know from research and first-hand accounts that people experience immersion, incubation, and insight in mathematics (e.g. Hadamard, 1945; Poincaré, 1946; Savic, 2016). Through this dissertation, I hope to answer some important questions about these processes and how an individual experiences them. For instance, what are these experiences like for everyday students? How important is immersion in math for a student to reach a moment of insight? Do previous findings of incubation replicate when it comes to math? What are the experiences of students who have *aha!* moments of insight in their everyday life? And do individual differences in cognition and affect play a role in these creative processes? Below, I define and highlight research on each of the first three stages of creativity and discuss its potential role and relevance in mathematics. Afterwards, I summarize the limitations of the existing literature, then provide an introduction to the current set of studies.

Immersion

Famed physicist Hermann Helmholtz once said that it was almost impossible to reach an insight "without long preliminary work" (Cahan, 1995, p. 389). The period of immersion describes this preparation phase, and is characterized by intense periods of conscious effort to understand and solve a problem (Savic, 2016). Haylock (1987) argued that in mathematics, immersion produces a full and thorough investigation of the problem at hand, and a complete understanding of all aspects of the problem. Dorfman, Shames, and Kihlstrom (1996) note that in addition to developing a stronger representation of the problem and exhausting conventional ideas, immersion promotes a "definite problem attitude" for the problem solver that increases the awareness that there is a lingering problem to solve (p. 258). Work on open goals has found evidence that only when participants fail to solve problems, but are still motivated to solve them, do they show benefits of incubation periods (Bos, Dijksterhuis, & Van Baaren, 2008; see also Zeigarnik, 1938).

The immersion phase of creativity shares some commonalities with *productive struggle* (Hiebert & Grouws, 2007), *productive failure* (Kapur, 2014) and other constructivist approaches to teaching in math education in which a student expends great effort during an initial period to explore and try to solve a math problem before receiving direct instruction. Receiving an opportunity to fully immerse oneself in a problem before direct instruction has been found to lead to better conceptual knowledge and a greater ability to transfer new knowledge to other problems (Kapur, 2014) as students engage in more diverse strategies and develop a stronger understanding of the features of the problem (DeCaro & Rittle-Johnson, 2012). In sum, immersion practices show great benefits in critical aspects of learning and creative problem-solving. Some researchers have argued, however, that beyond developing an understanding of the problem, the most important process of immersion is that it temporarily exhausts the problem solver, prompting them to take a break from the problem (Savic, 2016). This break, then paves the way for incubation.

Incubation

After struggling to solve a problem during the immersion phase, problem solvers frequently feel the need to take breaks from the problem. Referred to as incubation, the process of walking away from a particularly a troubling problem and allowing oneself to "incubate" has garnered much attention from creativity researchers (see Sio & Ormerod, 2009), as incubation has been found to have great benefits to problem-solving. Such benefits include the reorganization of information, polarization of information, and reliance on more gist-based memory (Ritter & Dijksterhuis, 2014). Equally important, the processes of incubation can allow one to forget about elements that lead to fixation when solving problems (Smith & Blankenship, 1989). Upon return to the problem, the restructuring of information and forgetting of fixations can allow for false cues to become *less* accessible, leaving the cues needed to solve the problem *more* accessible (Schooler & Melcher, 1995). This can result in an *aha!* moment of insight.

One popular account of incubation is unconscious work theory (Ritter & Dijksterhuis, 2014; Zhong, Dijksterhuis, & Galinsky, 2008) which posits that the benefits of incubation are largely due to subconscious processes that continue to aid with problem solving during breaks. Whether this processing is mostly made up of further activation of relevant semantic information (a *semantic-activation* account; see Sio & Rudowicz, 2007), sensitivity to related environmental information (an *opportunistic-assimilation* account, see Seifert, Meyer, Davidson, Patalano, & Yaniv, 1994), inhibition of fixating cues or ideas (a *selective-forgetting* account; see Smith, 1995; Smith & Blankenship, 1991) or some sort of mix of these processes is unclear, as evidence has been found for each account. But regardless of the exact mechanism, incubation has been strongly linked to improved problem solving and creativity, both in the form of spontaneous insights during incubation (*aha!* moments that come without previous conscious thought of the problem) or reaching an insight shortly after a returning to working on the problem. A meta-analysis of incubation studies found an overall effect of incubation breaks on problem solving, with the strongest effects of incubation on divergent thinking tasks and for breaks when participants complete tasks with a low-level of cognitive demand (Sio & Ormerod, 2009).

Two of the most cited anecdotes of incubation come from mathematicians Poincare and Archimedes. Research papers frequently cite the story of Poincaré who reported making a major breakthrough in his Fuchsian function problem while stepping off a bus during a geological excavation (e.g. Benedek & Jauk, 2018; Gilhooly, 2016; Sadler-Smith, 2015), as well as the famed story of Archimedes who had a moment of sudden realization regarding how to calculate the volume of a crown using displaced water in the bathtub (e.g. Lawson, 2001; Simonton, 2018; Ward, Smith, & Finke, 1999). From the limited empirical research on incubation within math, we know that individuals do experience incubation and insight when solving math problems. For example, some research has found that expert mathematicians not only take incubation breaks, but even have more developed and diverse routines during incubation compared to more junior mathematicians (Savic, 2012, 2015). This indicates that expert mathematicians have established routines that can provide them with incubation opportunities. But how everyday students can use incubation to improve math learning and problem-solving is a different question, and worth exploration.

Insight

Synonymous with *aha!* moments, insights are described as “the sudden experience of comprehending something that you didn’t understand before, thinking about a familiar thing in a novel way, or combining familiar things to for something new.” (Kounios & Beeman, 2015, p.

5). Insights are also distinguished from other experiences of problem solving by a sense of suddenness (Gick & Lockhart, 1995; Metcalfe & Wiebe, 1987), a new way of looking at a problem (Csikszentmihalyi & Sawyer, 2014) and often, and accompanying feeling of elation or happiness (Shen, Yuan, Liu, & Luo, 2016). Spontaneous insights are a special type of insight, that describe a sense of suddenness, but reflect an experience of arriving at a solution without previous conscious thought (e.g. an answer "popping" into your head). For mathematicians, insight has been described as finding a remote connection, switching on a light, and suddenly developing a greater understanding for how concepts relate together (Burton, 1998). But it's not just expert mathematicians who experience insight in math. Research has also found that students experience insights when learning mathematical content (Barnes, 2000; Liljedahl, 2005). Student insights in math are extremely important, as they not only represent leaps of understanding on the part of the student, but insights can also spark interest and confidence in students who previously disliked mathematics (Liljedahl, 2004).

Insights are often achieved when students are in a relaxed mood and doing activities that require little cognitive effort (Sio & Ormerod, 2009) and are believed to be the product of unconscious processes (Metcalfe & Wiebe, 1987; Ritter & Dijksterhuis, 2014; Schooler & Melcher, 1995;). What separates insights from what's referred to as *analytical problem-solving* (e.g. conscious work on the problem) is a sense of suddenness and unexpectedness (Kounios & Beeman, 2015). For example, a student walking to campus one day who unexpectedly arrives at the solution to a problem would have solved via insight, whereas a student who continuously works on the problem and gradually arrives at the solution through reasoning and logic is said to have solved it analytically. There may be some positive feelings of solving for the student who solves via an analytical problem solving method, but the student who experiences an *aha!*

moment would experience much stronger and more positive feelings (Kounios & Beeman, 2015). Therefore a critical element that distinguishes how students arrive at solutions is how sudden and unexpected the solution feels.

Limitations to Prior Work

Our understanding of creative problem-solving has come a long way since Wallas first published his work on creativity in 1926. But there remain several limitations in studying immersion, incubation and insight—especially in math. For example, much of the research on immersion in mathematics is conducted in relation to performance (e.g. productive struggle) but how this immersion leads to creative problem-solving, not just retention of information or the ability to execute a series of procedures, is unknown. When immersion is studied in relation to incubation and insight, researchers frequently rely on abstract problems that do not reflect the types of problems that people will need to solve in real life. The Remote Associates Test (RAT), for example, provides participants with three words and ask the participant to generate a fourth word that connects the three words together (e.g. "*dust", "gold", "shooting"* are all related to the word *"star"*). Not only is this task heavily confounded with verbal fluency of the participant, but has little connection to real-world problems, or educational content. Further, classic RAT incubation studies give upwards of 20 RAT items to solve (showing participants 60 words in total), provide a break, and then test to see how many more can be solved after a break (e.g. Dodds, Smith, & Ward, 2002; Smith, & Blankenship, 1991) but some work suggests that students have difficulty remembering so many different sets of problems (Moss, Kotovsky, & Cagan, 2007), possibly overpowering processes involved in incubation, and obscuring results.

Perhaps the greatest limitation is that much of the research on immersion, incubation and insight take place in a controlled testing environment which makes it difficult to study incubation and insight that happens in everyday life. Of course laboratory research has provided vital information about these processes, but to build a more complete picture, research must be conducted outside the lab as well. Dunbar (2001a), who studied how scientific insight and reasoning arises utilized more naturalistic settings (such as lab meetings), was able to provide fruitful information about these processes— information that would otherwise be difficult to capture in a controlled testing environment. In addition to providing richer information, research from the analogy literature has found that participants tend to focus on superficial details in controlled testing environments, whereas outside the lab, they tend to focus on deeper structural information (Dunbar, 2001b). Might similar differences be found with incubation and insight, where participants engage in deeper forms of thinking outside the lab?

Capturing incubation and insight in student's everyday lives not only provides a richer account of creative problem-solving, but allows for the space, time, and activities that real-world incubation and insight requires. Research on incubation in natural settings suggests that incubation periods often take much longer than an hour—the standard time of a full lab experiment (Savic, 2012). The activities mathematicians engaged in are also diverse and tailored to the specific routine of the mathematician, something that could only be captured in a naturalistic setting. Further, the general creativity literature has found that incubation often occurs during the five "B's" – bed, bathroom, bus, bars, and boring meetings (Benedek & Jauk, 2018), making it difficult to recreate such conditions in a lab environment. By primarily relying on research that restricts the creative problem-solving experience to lab visits (and limits naturalistic research to experts), the field has missed out on rich and detailed information as to how students use immersion and incubation to reach mathematical insights in their everyday lives. Hence, one of the main goals of the dissertation is to address whether incubation and insight outside the lab reveals a more rich and representative picture of what creative problem-solving in math is like.

The Current Studies

My dissertation investigates how students come to solve a creative math puzzle both inside and outside the lab. In the first experiment, I examine how an immersion/incubation manipulation in lab leads students to experience insight for a novel math problem puzzle. Students are brought into the lab, complete a number of trait individual measures, attempt to solve a math puzzle, and then report on various state measures about their experiences and feelings in the lab. For students who are unable to solve the puzzle in the lab, they are instructed to go back to their lives as normal. I follow up with them up to three days later and survey them about their experiences solving in their everyday lives. For study 2, I attempt to replicate the findings of study 1 and expand upon study 1 by including new measures that capture different trait individual differences, as well as collect more detailed information from students who solve outside the lab.

Study 1

Studies of deep immersion and incubation often cite anecdotes of mathematicians to illustrate the power and usefulness of these practices to reach insight, but the majority of these studies use anagrams, RAT items, or other divergent thinking assessments—all of which rely on verbal ability to solve (Sio & Ormerod, 2009). In study 1, I extend a classic creative incubation and insight paradigm into a mathematical context by administering a novel creative math puzzle, and examine how a break in problem solving in the lab may boost students' problem-solving rates. Students who were unable to solve the math puzzle in lab were enrolled in a subsequent phase of the study where they were given three days to solve the puzzle outside of the lab. I also investigate how affective factors (such as math anxiety) and trait measures of cognitive engagement (such as Need for Cognition) may relate to processes that lead to incubation and insight.

Research Questions and General Design

There were three primary research questions in this first study, 1) Does a classic incubation manipulation boost solve rates in math? 2) Do individual differences measures of affect and cognitive engagement relate with solve rates of the math puzzle inside and outside the lab? 3) How do students arrived at creative insights in math, and what are their experiences? To answer these questions, undergraduate participants were brought into the lab and randomly assigned to one of three conditions— 1) a low-demand (LD) incubation condition that required students to complete a task that requires little cognitive effort halfway through the puzzle-solving attempt, 2) a high-demand (HD) incubation condition that required participants to complete a task that required a lot of cognitive effort half-way through, or 3) a control condition that had participants continuously work on the problem without interruption.

All three conditions received the same amount of total time spent working on the puzzle, the only difference between the conditions was whether they received a break half-way through, and what task they completed during this break. Note, this manipulation both serves to test the efficacy of different incubation conditions as well as the benefits of receiving a long period of uninterrupted immersion (as opposed to interjecting this period with an opportunity for incubation). Students who did not solve the problem during the lab portion of the experiment were enrolled in a second phase of the study, where they went about their everyday lives and were asked to take a follow-up survey either immediately after arriving at the solution, or if three days had passed and they had not solved. Subsequently, I refer to the first phase of the study as the “lab” portion of the study, and refer to the second phase of the study as the “wild” portion of the study. Students who solved in the lab will thus be referred to as “lab solvers” and those in the wild as “wild solvers,” the two of which are mutually exclusive (but not exhaustive) groups of participants.

Hypotheses

For the first research question – which asked whether an incubation manipulation in the lab predicts solve rates in the math puzzle – I predicted that students assigned to the conditions with the incubation break would show better solve rates in the lab compared to students in the control condition. Between the two incubation conditions (LD and HD), I predicted that students assigned to the LD condition would out-solve the HD condition. This hypothesis was informed by findings from a meta-analysis of incubation that found low-demand conditions show the highest rates of problem-solving (Sio & Ormerod, 2009).

For the second research question, I hypothesized that for students with a negative emotional disposition or negative reaction to math (i.e., high math anxious, negative PANAS scores) would generally show a lower solve rate compared to students with less of a negative emotional disposition. I surmised that the relationship between math anxiety and solve rates would be more pronounced for students assigned to either of the incubation conditions, given that rest periods have been found to lead to suppression/forgetting of information that is stressful (Ramirez, 2017). In addition to this, I predicted that students who were more curious to know the solution after failing to solve in the lab, and those generally higher in need for cognition would be more likely to solve in the wild. The third research question – what are students solve experiences like outside of the lab – was exploratory in nature and designed to examine whether the students’ solve experiences were consistent with literature in the insight problem solving literature, such as solving during periods of rest, while doing activities that are low in cognitive demand, or during the five b's (bed, bathroom, bus, bars, and boring meetings).

Method

Participants and Sample Size Justification

Participants consisted of 252 undergraduate students enrolled in an undergraduate psychology course who completed the study for course credit. Because the math puzzle used in this study had not been used in prior empirical research, the estimated solve rate was unknown, and thus, the effect size of such incubation conditions to conduct a power analysis was not available. Thus, my goal for the total sample size was to have at least 60 participants in each incubation condition (control, LD, HD conditions) totaling 180 participants. However, because there is also a wild component to this study, I ran as many participants past 180 as was logistically possible (within constraints) so that follow-up analyses could be as well-powered as possible. The resulting sample was comprised of 252 students, 180 which were females (71 males, 1 other), with an average age of 20 years old, representing race/ethnicity of Black/African American (2%), Asian (34%), Indian (4%), Middle Eastern (6%), Latino (15%), White (28%), and Biracial or Other (10%, with an additional participant preferring to not report).

Procedure

Participants were tested in groups of up to five students at a time. The experimenter began the lab session by providing participants with an overview of the experiment (including an overview of the wild phase of the study) and set up each participant at a station with a computer and scratch paper to work on the math puzzle. Participants began by filling out questionnaires that measures their math anxiety, math identity, need for cognition, and the PANAS (measures discussed in more detail in the subsequent measure section). Afterwards, they viewed a video with instructions to the puzzle, and were randomly assigned to one of three conditions. Those assigned to the LD incubation condition worked on the puzzle for three minutes, then completed a signal detection task for 2.5 minutes, before being given another three minutes to work on the math puzzle. Students randomly assigned to the HD condition had a similar procedure, except instead of completing a signal detection task during the incubation break, they read an advanced microbiology passage and answered difficult questions about the passage. Participants randomly assigned to the control condition worked on the problem for an uninterrupted 6 minutes, similar to other controls in incubation studies (e.g. Bos et al., 2008; Dijksterhuis & Meurs, 2006; Gilhooly, Georgiou, & Devery, 2013). The allotted times of work and the breaks associated with the conditions were informed by a meta-analysis of incubation that provided estimates of the optimal time of breaks (Sio & Ormerod, 2009) and by previous pilot testing.

If participants believed they had solved the math puzzle during the lab session, they raised their hand and the experimenter either confirmed the solution, or pointed out the rule they violated and instructed the participant to keep trying to solve. After the puzzle-solving part of the lab study, participants ended the session by reporting on their experience with the math puzzle (state measures) and filled out demographic information. Participants who did not solve the puzzle during the lab session were enrolled in the second phase of the study, and were provided with a survey link to the follow-up wild survey. Students were instructed to take the survey as soon as they had arrived at the solution to the problem, but if they had not found the solution after three days, I followed up with them three days later to take the follow-up survey. Participants were not instructed to try to come up with the solution, but rather live their lives as normal and take the survey if they happened to think of the solution. See Figure 1 for an illustration of the lab procedure, as well as when various measures were administered to the participant. A description of the measures is included below.

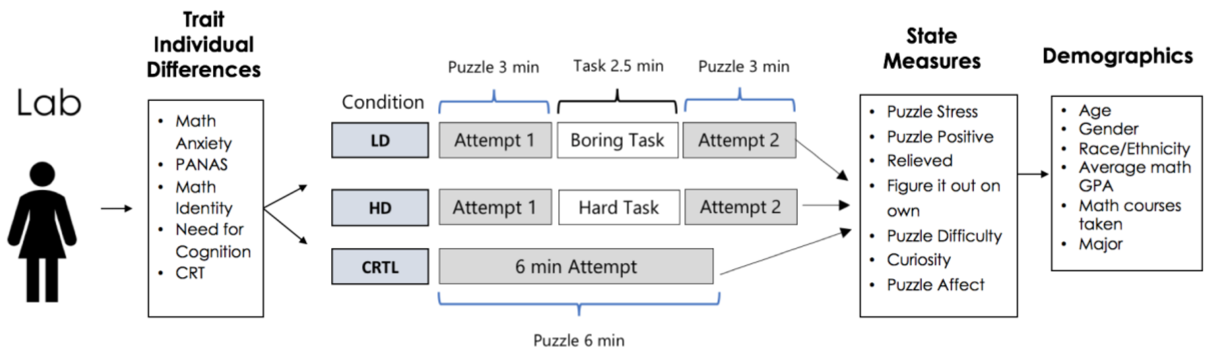

Figure 1. An overview of when trait and state measures appeared in study 1

Measures

Math puzzle. The math puzzle used in the current study presents participants with four digits (2 3 4 5) and two symbols (+ =) and asks them to create a balanced equation using each digit and symbol once and only once, without adding any digits or symbols (Miller, n.d.). Participants were encouraged to use any mathematical procedure they could think of, as long as it satisfied the rules. The solution to this problem is the equation $3^2 = 4 + 5$. To solve the puzzle, students must adopt a creative problem-solving approach that requires escaping conventional algorithmic procedures to overcome the common fixation of combining the numbers to create larger numbers (e.g. 4 + 5 = 23). This puzzle also requires inhibition of fixations, interference management, and can benefit from executive switching—processes found to play a large role in creative processes (e.g. Beaty & Silvia, 2012). Additionally, this puzzle can also be solved analytically through gradual steps and hypothesis testing which allowed me to measure variation in students' problem-solving attempts as well as how they came to solve the puzzle.

Trait individual differences.

Math anxiety. Math anxiety was measured using the Abbreviated Math Anxiety Scale (Hopko, Mahadevan, Bare, & Hunt, 2003). This scale presents nine items to the participant and asks them to rate how anxious each item would make them feel (e.g. opening a math textbook, going up to the front of the class to solve a problem on the board, etc.) on a scale from 1 (low anxiety) – 5 (high anxiety). These items were averaged across the nine items to create a composite score of math anxiety.

PANAS. I used the International Positive and Negative Affect Schedule Short Form (I-PANAS-SF; Thompson, 2007) to assess participants general tendency to feel negative and positive emotions. This version of the general PANAS presents five positive affect adjectives (*determined, attentive, alert, inspired, and active*) and five negative affect adjectives (*afraid, nervous, upset, ashamed, and hostile*) and asks the participant to rate how much they generally feel each of the adjectives, on a scale of 1 – 5. Averages for the positive and negative adjectives are then calculated to create a composite score for trait positive and negative emotion.

Math identity. To measure a positive disposition towards math, I administered a commonly used set of items to capture the degree to which math is an important part of students' self-concept (Ramirez, McDonough, & Jin, 2017). This scale consists of six items that participants rate on a scale from 1 – 7 based on how much they agree with each statement. Examples of these statements include, "It is important to me that I am good at mathematics" and "Compared to others, I feel I understand mathematics well."

Need for cognition. I also assessed how much participants generally like to engage in cognitively effortful activities using a modified version of the Need for Cognition scale (Cacioppo, Petty, & Kao, 1984). Participants were presented with 10 items (e.g. "I prefer complex to simple problems", "thinking is not my idea of fun") and were asked to rate on a scale of 1 to 5 how characteristic each statement was of them. Half of the items presented were statements reverse scored. Averages were taken from the ten items to make a composite score.

Cognitive reflection task. Students were asked to complete three items from CRT-2, an updated alternative cognitive reflection task (Thomson & Oppenheimer, 2016) designed to measure a participant's ability to override intuitive responses to reach correct solutions. This measure may help explain if students are able to reach a solution, as it requires students to overcome the impulse to try to combine numbers (e.g. making 2, 3 into 23) and extend their thinking to other mathematical properties not immediately apparent (specifically, an exponent).

State measures. After completing the problem solving attempt, students who did not solve in the lab were asked to report on a variety of state measures (listed below) and proceeded to answer demographic questions.

Puzzle stress and puzzle positive. Participants who did not solve in the lab were asked to rate how much they agreed with the following two prompts: "I thought that the math problem was overall really stressful and challenging," and, "I thought that the math problem was overall challenging, but in a positive way" on a scale from 1 (strongly disagree) – 7 (strongly agree). The first measure, that I will now refer to as Puzzle Stress, aims to capture the state-level maladaptive stress the student may experience with the math puzzle. This is distinctive from trait-level differences in negative affect as it is related to stress and the student's experience with math puzzle. The second measure, now referred to as Puzzle Positive, teases apart students who may have found the lab environment and math puzzle challenging, but found this challenge enjoyable.

Puzzle affect. Students were additionally asked to rate the extent to which the math puzzle made them feel frustrated, stupid/dumb, inferior, unsure, or doubtful on a scale of 1 (strongly disagree) to 7 (strongly agree). These measures were included to capture different state measures of affect, and were informed from pilot data where students reported the math puzzle made them feel these different emotions. A composite variable was then created by averaging across these five items.

Relieved lab attempt is over. Participants who did not solve in the lab were additionally asked to rate on a scale from 1 (not at all) to 7 (a great deal) how relieved they felt that the problem-solving attempt in the lab was over. This measure stands in contrast to the puzzle affect measure, as it captures positive feelings associated with a potentially negative experience in the lab, which may explain later solve rates in the wild.

Puzzle difficulty. Students also reported how difficult they found the problem on a scale from 1 (very easy) to 5 (very difficult). This measure was included to ascertain the role of ability perceptions in subsequent problem solving in the wild.

Figure out on own and curious to know solution. After answering the previous state measures, participants who did not solve in lab were asked to indicate how much they agreed, on a scale from 1 – 5, with the following statement: "I would rather figure out the math problem I was previously presented on my own instead of being told the answer." I also collected information as to how curious students were to know the solution before they left the lab on a scale of 1 (not curious at all) to 7 (very curious). Curiosity and desire for independent problem solving are important measures to capture as past research has suggested that motivation to solve is a necessary component for incubation periods and solve experiences (e.g. Bos et al., 2008;

Zeigarnik, 1938). I reasoned that these measures may hold some explanatory power as to whether a participant solves in the wild.

Demographic information. At the end of the lab section, participants reported demographic information such as age, gender, and race/ethnicity. They were also asked to report the average grade they received across previous math courses. This measure, referred to as average math grade, was quantified by assigning values to these grades on a scale from 1 (Below a C-) to 9 (A+). Participants additionally reported which math courses they had taken and passed from a list of courses (algebra, geometry, algebra 2, pre-calculus, calculus, or statistics) to potentially account for any differences in math background. Participants additionally completed the Content Knowledge for Teaching in Mathematics (CKTM), a measure of basic mathematical understanding of participants. The CKTM is comprised of 14 math questions that test student's knowledge of fractions, patterns and functions, as well as algebra. Students solve math problems on the CKTM, which are scored and summed to create a percent correct, representing a students' understanding of basic mathematics. This measure was included to characterize the sample in terms of their math knowledge beyond courses taken and passed.

Wild measures. As a reminder, students who were unable to solve the math puzzle in the lab session were subsequently dismissed and told to go about their day. These participants were provided a link to a follow-up wild survey with instructions to take the survey as soon as possible if they happened to solved the puzzle. If three days had passed since their lab visit and they had not reported solving, I contacted each participant and asked them to take the wild follow-up survey. The primary function of this survey was to ask students if they solved in the wild (and confirm their solution). However, I also included other measures, such as whether the participant googled information or solved it with someone else (used to filter out cheaters), a memory test of digits and rules (used to filter out students who did not remember the problem), as well as how often they thought about the problem outside the lab. Students who reported solving the puzzle in the wild were additionally asked to indicate if they were consciously thinking of the problem before solving, how the solution came, and were asked to describe what they were doing right before they solved (see below). After showing students the correct solution, they were asked why they did not think to use exponents (used for filtering out students who thought to use exponents, but believed it was against the rules because of the “^” symbol that is sometimes used). These measures are discussed in more detail below.

How often thought about puzzle. All students who took the wild survey were asked to rate how often they thought about the puzzle after the lab session on a Likert scale from 1 = never thought about it, 5 = thought about it a great deal. This information was collected to ensure that any potential differences between conditions in wild solve rates was not due to one condition simply thinking about the puzzle for longer outside the lab.

Consciously thinking about puzzle. Students who reported solving were asked to report if they were thinking of the puzzle right before they solved (Yes/No). This measure was included to investigate if students experienced a spontaneous insight when solving, which is characterized by the solution suddenly and unexpectedly coming to conscious awareness, without consciously thinking of the problem beforehand.

How Solution Came. Related to the previous item, participants were also asked to report which of their solve experiences best matches from a list of three different experiences.

- “I could feel myself slowly getting closer to the solution, until the solution came.”
- “I tried many different things until it seemed like there were no solutions. Then it came to me.”
- “The solution seemed to come out of nowhere.”

The first option was designed to illustrate an analytical problem-solving method that captures gradual progression without a sense of unexpectedness. The second statement was a designed to illustrate a mix of both analytical problem-solving and classic definitions of the *aha!* moment. The third statement was designed to illustrate a sudden and unexpected solve experience representing a classic insight moment.

Solve Experience (open-ended description). Wild solvers were asked to report, in as much detail as possible, what they were doing right before they solved through an open-ended response question. These descriptions are exploratory in nature, and were collected to shed light on the variety and diversity of students' solve experiences.

Solve Experience (open-ended description). Students who did not report solving in the wild were asked why they did not think to use exponents to solve, presented through a multiple-choice question. They chose a response from the following statements:

- "It just never came to mind"
- "I did think of using exponents, but I couldn't get the math to work out"
- "I did think of using exponents, but I did not think exponents were allowed"
- [Other, where a participant reported a description]

This question was used to filter out students who selected "I did think of using exponents, but I did not think exponents were allowed." This decision was made to ensure the integrity of the data, as these students may have been able to solve in lab but because of the variety of ways one can signal an exponent (such as "3^2" or 3^2) they may have discarded the idea (contributing to false negatives in lab solve).

Data Analysis Plan

The first research question of study 1 was whether students assigned to incubation conditions were more likely to solve the math puzzle compared to students randomly assigned to a control condition. To test for the presence of an incubation effect, a logistic regression was first run predicting lab solve rate from the participant's randomly assigned condition. If this model is significant, it would indicate that the condition each participant was assigned to helped explain variability in the lab solve rates. For the second research question – which asked whether negative dispositions towards math and other trait individual differences could explain solve rates – logistic regressions were run with the individual difference measures predicting solve rate in the lab. A significant model among any of these measures would suggest that the individual difference plays a significant role in whether a students solved the math puzzle. I additionally ran these analyses predicting wild solve rates as well, to see if condition and individual differences not only affected solve rates in lab, but also outside of the lab. Descriptive statistics for solvers were examined to better answer the final research question which asks about students' experience of solving in the wild.

Data exclusion rules. Participants were excluded from all analyses if 1) they reported seeing the puzzle before, and 2) if they reported thinking of exponents, but believed they were not allowed because of the carrot symbol. Participants were also excluded from wild analyses if they 1) reporting googling the problem, or reported solving it with someone else, 2) if they failed the memory test (defined as not reporting the correct digits and symbols). In addition to these noted data exclusion rules, one participant was removed from the total study, and another removed from the wild analyses because both failed to follow instructions (e.g. did not report solving in lab after solving).

Results

Math Experiences

Part of the demographic measures included in this study concerned information about participant's math backgrounds, which were included to characterize the sample. Results for the CKTM revealed a highly skewed distribution with an overwhelming majority of the sample performing very well on the measure (median score was 100%). Further, 88% of the sample reported having taken and passed calculus, suggesting students had the prior knowledge in mathematics to solve the math puzzle. To further ensure that solve rates were not simply a function of past experience in math, I ran logistic regressions predicting solving in lab and in the wild from successfully taking and passing different math courses (algebra, geometry, trigonometry, pre-calculus, calculus, and statistics). None of these models were significant, indicating that variation in past math experience could not predict solve rates in the math puzzle. Interestingly, student's self-reported general GPA for their average math class predicted wild solve rates ($p = .03$), such that for every one unit increase in reported GPA (e.g. B+ to A-, or A to A+), the expected odds of solving increase by 56%, though this did not predict solving in the lab ($p = .28$).

Division of the Sample and Solve Rate

Division of the sample can be found in Figure 2. In total, 252 participants were recruited and took part in this study. Data from 231 students were used in analyses of the lab phase of study, and 166 of the participants were followed up with outside of the lab for wild analyses. Of the total 231 participants in the study, 22% of students solved the puzzle in the lab. Of the 166 students who did not solve in the lab, 52% of students successfully solved the math puzzle in the wild.

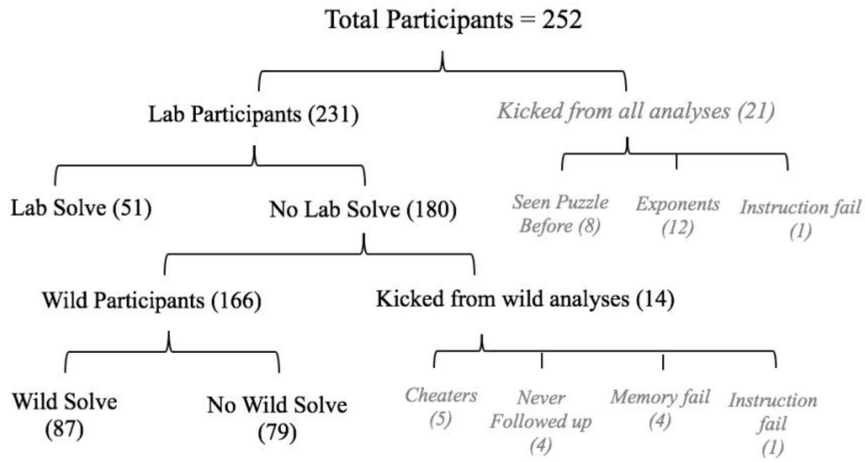

Figure 2: Division of total sample across study 1. Sections that appear in gray represent participants who had been kicked from analyses.

Effect of Incubation Break

To test the first research question – whether an incubation manipulation improved math puzzle solve rates – I ran a logistic regression predicting lab solve from condition. This model was not significant (LD vs. Control p = .574, HD vs. Control p = .282), indicating that there was no effect of incubation condition on solve rates in the lab. However, when I ran a model predicting wild solve rates from condition, there was a significant effect. Specifically, participants in the control condition had higher odds of solving in the wild compared to those in the low-demand incubation condition (p = .04), such that the expected odds of the control group solving was 31% higher than for those in the LD condition. Participants in the control condition (64%) solved the puzzle at a higher rate than those in the HD condition (48%) but this difference was not significant (p =.10). Figure 3 shows the solve rates by condition for both wild and lab solve.

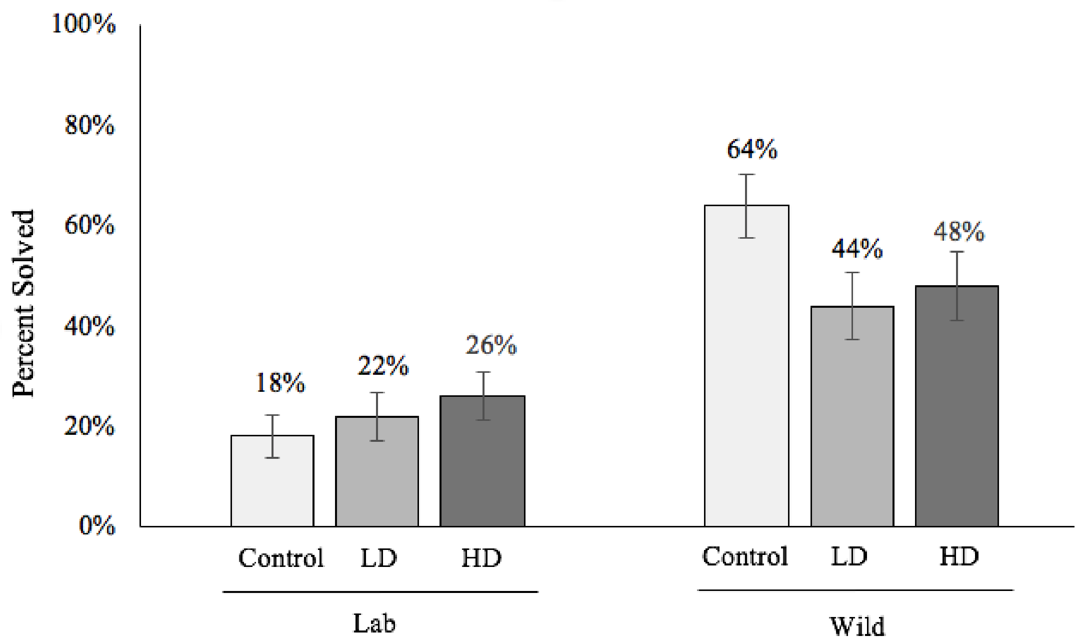

Figure 3: Solve rates in lab and in the wild by condition for study 1.

One initial explanation for why control participants showed higher rates of solve in the wild could be that participants in the control condition were more curious to know the solution, or thought about the problem longer after leaving the lab. This is not the case, however, as students in the control condition did not differ in their level of curiosity to know the solution compared to students in the low-demand condition (p = .08), nor did they differ in the amount they reported thinking about the problem in the wild (p = .30). In terms of the hypothesis that math anxiety might affect incubation students disproportionally, I tested the interaction between math anxiety and condition, and found no significant interaction predicting lab solve or wild solve rates (p = .24 - .80).

Trait Individual Differences

For the second research question – whether trait individual difference measures predict solve rates – I ran a series of logistic regressions predicting lab solve from trait measures, and predicting wild solve rates from trait individual difference measures. Table 1 presents the descriptive statistics of these measures, and Table 2 presents the outcomes of these measures predicting lab and wild solve rates. In the context of lab solve, two trait individual differences measures helped explain the odds of solving – math anxiety ("marginally" significant at p = .048) and PANAS positive (p = .02). These measures negatively predicted the odds of solving—meaning that the more math anxious a participant was, or the more positive emotion they reported on the PANAS, the less likely the odds of solving in the lab. None of the other measures predicted the odds of solving in the lab. In terms of solving in the wild, none of the trait individual difference measures were significant predictors.

State Measures

I next examined how state measures taken after a failed attempt at solving the puzzle in the lab predicted odds of solving in the wild (see Table 2). Of all state measures taken, only two significantly predicted the odds of solving in the wild. The first was the extent to which students reported being curious to know the solution to the puzzle (taken at the end of the lab session), such that for every 1 unit increase in curious to know the solution, there was an expected increase in wild solve rates by 60%. The second significant state measure was the extent to which participants reported feeling relieved the math puzzle was over in the lab. Unlike curious to know solution, feeling of relief *negatively* predicted wild solve rates, such that for every one unit increase in feelings of relief, there was a predicted decrease in solve rate by 45%. Other state measures that were not significant included puzzle stress, puzzle positive, puzzle affect, puzzle difficulty, and the extent to which participants wanted to figure out the solution on their own.

Context for Solving in the Wild

While only 22% of students could solve the problem in the lab, 52% of the remaining students solved the problem after leaving the lab. As a reminder, all students who were enrolled

Table 1

Descriptive statistics of trait individual difference measures used for lab and wild solve analyses in study 1, and state measures used in wild solve analyses.

	n	mean	*s*	median	min	max	range
Trait Individual Differences							
Math Anxiety	231	2.53	0.76	2.4	1	4.9	3.9
PANAS Positive	231	17.02	3.28	17	5	23	18
PANAS Negative	231	10.97	3.32	11	5	23	18
Math Identity	231	4.38	1.32	4.7	1	7	6
Need for Cognition	231	3.52	0.59	3.6	1.6	4.8	3.2
CRT Sum	231	1.54	0.70	2	0	3	3
State Measures							
Puzzle Stress	166	4.08	1.70	4	1	7	6
Puzzle Positive	166	4.95	1.65	5	1	7	6
Puzzle Affect	166	4.35	1.46	4.4	1	7	6
Relieved	166	3.05	1.82	3	1	7	6
Difficult for You	166	4.20	0.59	4	2	5	3
Figure Out Own	166	3.61	1.21	4	1	5	4
Curious	166	6.48	0.91	7	2	7	5

Note: Math Anxiety, Math Identity, Need for Cognition, scores are averaged across a series of items, whereas PANAS items are summed across 5 items, and CRT across 3 items.

Table 2

Outcomes of logistic regressions predicting lab solve and wild solve rates from individual difference measures in study 1.

	Lab Solve			Wild Solve		
	b estimate	SE	Sig.	*b* estimate	SE	Sig.
Trait Individual Differences						
Math Anxiety	**-0.45**	**0.29**	***p* = .048**	-0.20	0.20	*p* = .32
PANAS Positive	**-0.10**	**0.05**	***p* = .02**	0.01	0.05	*p* = .72
PANAS Negative	-0.04	0.05	*p* = .36	-0.05	0.05	*p* = .30
Math Identity	0.06	0.12	*p* = .58	0.15	0.12	*p* = .20
Need for Cognition	0.02	0.27	*p* = .95	0.16	0.27	*p* = .55
CRT Sum	-0.07	0.23	*p* = .75	0.21	0.22	*p* = . 34
State Measures						
Puzzle Stress	--	--	--	-0.05	0.09	*p* = *.53*
Puzzle Positive	--	--	--	0.02	0.09	*p* = .80
Puzzle Affect	--	--	--	0.05	0.11	*p* = .65
Relieved	--	--	--	**-0.18**	**0.09**	***p* = .04**
Difficult for You	--	--	--	-0.24	0.27	*p* = .38
Figure Out Own	--	--	--	0.22	0.13	*p* = .09
Curious	--	--	--	**0.39**	**0.18**	***p* = .04**

Note: Output in bold represent findings that were statistically significant.

in the wild portion of the study left the lab with the link to the follow-up wild survey. If students solved the puzzle, they were instructed to take the survey at their earliest convenience and report that they had solved the puzzle. If students did not solve, they were contacted after three days to take the follow-up survey, where they reported they had not solved and answered relevant survey items. It is important to note here that none of the students were explicitly instructed to continue solving the problem outside the lab. They were simply told to complete the follow-up wild survey should they arrive at the solution.

All students who were enrolled in the wild phase of the study reported how often they thought about the puzzle after leaving the lab in general, their responses were varied—11% reported "rarely or never", 27% reported "occasionally", 25% reported "a moderate amount", and 36% reported "a great deal." Students' responses on this measure did significantly predict solve rates in the wild, such that for every one unit increase on this measure, the expected odds of solving increased by 67%. This suggests, not surprisingly, that the more often students thought about the puzzle the greater likelihood they were to solve it.

While students who did not report solving were asked why they did not think to use exponents (used for filtering purposes), those who did solve were asked a series of questions about their wild experience. Surprisingly, 82% of students who solved the puzzle in the wild reported that they were consciously thinking about the problem right before solving, 11% reported they were not, and 6% could not remember. Wild solvers were then asked to choose a statement that best characterized how the solution came to them— one option described an insight moment ("The solution came out of nowhere"), one described an analytical problem-solving experience ("I could feel myself slowly getting closer to the solution, until the solution came"), and a third provided a mix of the two ("I tried until it seemed like there were no solutions. Then it came to me"). In total, 29% of participants selected the statement describing insight solving, 36% selected the statement describing analytical solving, and 35% chose the statement that represented a mix of the two.

Solve Experiences (open-ended)

As a reminder, students who solved in the wild were provided with an open-ended question asking what they were doing right before they solved the math puzzle. These open-ended responses showed a diverse range of experiences and activities. Some students reported consciously working on the problem outside of the lab, and it was this work that led to the solution. For example, one student wrote "I was sitting in my tutoring session for my psychology class. I wrote the numbers and symbols out repeatedly until I saw how I could arrange the numbers without adding or taking anything away." Another student, shared:

> "I walked out of lab and threw out that flyer [the debrief form] and like damn I can't solve it there's gotta be some trick with combining the numbers somehow. I tried thinking to combine digits like 23=45 but then the rules kinda reminded me I need to use the + So then I'm like oh man let me try exponents and I tried a bunch of combinations and found the answer."

Other students reported experiences of doing everyday activities when they thought of the solution. For example, students reported solving when they were walking to class, eating, driving, cleaning, and reading email. Many students listed activities that fit well with the finding that incubation and insight happens during periods of rest. For example, students provided evidence of four of the five B's of incubation— one student reported solving in the shower (bath), six were sitting in bed or about to fall asleep, and another student reported they solved on the bus while zoning out. Nine students reported solving during lectures (which likely falls under "boring meetings"), but for the "bar" category, no student reported drinking alcohol.

In addition to students solving through conscious work and during periods of rest, some students shared that relevant clues in their environment had helped them to solve. Overall, eight students had shared that something in their environment had helped them to solve or reported doing an activity related to math right before they solved. For instance, one student reported, "I was helping my friend do her math homework and I saw a problem that just helped me click the solution in my head." Another student stated,

> I was in my dorm's floor lounge playing a game on my laptop where I had to avoid obstacles and my floormates were working out a math problem on the board, where at one point they talked about squaring both sides of an equation and I thought of the problem.

These experiences illustrate how incidental encounters with relevant stimuli can help facilitate insight, even when the clues in the environment are not the exact solutions and require a connection (in contrast to a line of studies that find hints do not help students solve unless they are explicitly made aware of them; see Dodds et al., 2002). One student shared a particularly telling example of this:,

> I was doing my homework for English 4W and the main topics for this class are 'Form and Power.' Then that was when it hit me, that taking the power of a number doesn't require another symbol and I smacked myself in the face.

Discussion

Study 1 provides encouraging results for the use of the math puzzle in studying creative problem-solving by demonstrating good variability in the lab and in the wild, and by eliciting differences in how students arrive to the solution of this problem (insight, analytical, mix). The manipulation of incubation condition in the lab did not show differences in solve rate within the lab, rejecting the hypothesis that students in the incubation conditions would outperform students in the control condition in lab solve. One possible explanation is that using math puzzle instead of a task like RAT requires longer periods of an incubation break to show an effect, or longer periods of uninterrupted time for immersion and impasse (such as in the control condition) provided a benefit equal to the benefit of an incubation break. Another interpretation of this finding is that if incubation effects do exists, they might be weaker and unreliable in the lab.

Although all three conditions led participants to perform relatively the same within the lab, students in the control condition outperformed the other two conditions in terms of raw solve rates in the wild, and showed a statistically significantly greater solve rate in the wild compared to students in the low-demand condition. In his work on scientific insights and reasoning, Dunbar (2001a) noted the importance of capturing information from naturalistic settings, but also found that that scientific insights can often happen when control conditions counteract hypothesized results (Dunbar, 2001b; see also Buckner, Andrews-Hanna, & Schacter, 2008).

So what might explain the benefits of the control condition? Although all three conditions received the same amount of total time to work on the puzzle in lab (6 minutes), providing students with an *uninterrupted* amount of time (control condition) may provide added value in helping students to immerse themselves deeper, exhaust more ideas, and better reach an impasse (Beeftink, Van Eerde, & Rutte, 2008; Segal, 2004). Indeed, some research has found that only when students exhaust ideas and bring themselves to impasse are they able to solve, as this impasse sensitizes students to hints in the environment that help them solve (Moss, Kotovsky, & Cagan, 2011; Seifert et al., 1994).

Study 1 also revealed that students solved the math puzzle in the wild under a diverse array of conditions and activities. When students were asked to characterize their solve experience using a multiple-choice measure, students indicated solving the puzzle via descriptions that represent an insight, analytical thinking and or a combination of the two at a roughly equal rate. Another interesting finding from the open-ended description measure is that some students reported that chance encounters with math and other hints (e.g. "Form and Power") sparked an *aha!* moment for them. This is particularly interesting, as some lab studies of hints have found that hints do not improve solve rates unless participants are told that they may encounter hints (e.g. Dodds et al., 2002; Smith, Sifonis, & Angello, 2012).

However one theory, termed *opportunistic assimilation,* suggests that when an individual reaches an impasse, this sensitizes them to hints in their environment, and it is only through the diverse environment of everyday life that they encounter chance hints that trigger insight moments (Seifert et al., 1994). This may explain why students in the control condition only showed a benefit of solving when they were in a naturalistic setting, opposed to the lab—they may be more likely to exhaust ideas and reach an impasse (as they are not interrupted during the problem-solving period to take a break), which may help them benefit more from potential hints in their everyday lives compared to participants in the other two conditions.

Lastly, it was found that trait individual differences known to relate to problem solving in the lab did not relate with creative problem-solving in the wild. Specifically, math anxiety and PANAS positive was negatively predictive of solving the math puzzle in the lab, but these relationships did not hold for solving the math puzzle in the wild. These measures were administered again in study 2 in an effort to replicate findings.

Study 2

The primary goal of study 2 was to replicate and expand upon the findings of study 1. The decision to replicate study 1 was based on a few factors. Notably, the most interesting finding of study 1 – a difference between control and LD condition on rates of solving the math puzzle in the wild – was statistically significant at $p = .04$. Although this falls under conventional thresholds for significance ($p < \alpha$), such a value has been argued to represent weaker evidence of a finding, requiring replication (see Simonsohn, Nelson, & Simmons, 2019). In addition, the main theoretical account that may explain this finding, the *opportunistic assimilation* hypothesis, would argue that control participants show better solve rates in the wild because they were more likely to reach an impasse compared to the other groups, sensitizing them to information in the environment that helps them solve. But to properly evaluate this theory, measures of impasse and information about participants' wild environment are required.

Thus, it was decided that study 2 be a replication of study 1 to ensure the primary finding replicates, and collect new information to shed light about a potential mechanism (impasse) that would explain this finding. And of course, study 2 expands upon the (mostly null) findings of study 1 trait individual difference measures by including new trait individual difference measures that capture a participants' tendency to engage with abstract information and perceptual information, providing valuable information that may help explain whether participants pick up on cues from their environment.

Although the two studies have almost identical procedures, study 2 differs in a few key ways. Notably, in study 2 participants were randomly assigned to only two of the original three lab conditions—control and LD. The high-demand (HD) condition was removed from the second study as there was no difference between control and HD in the wild in study 1, and assigning students to two groups increase statistical power to detect differences between the control and LD condition. In addition to this change, I also provided students explicit instructions to report on their experience in the wild in greater detail, asked them to report information on their environment when they solved, as well as report any potential hints or cues that may have helped them to solve in the wild.

If deeper immersion periods sensitize students to incidental hints in the wild environment, and this is responsible for the wild solve rates in study 1, then it would be important to measure how well students attend to sensory and perceptual information in their environment. Therefore, I extend individual differences measures to ascertain potential variation in pattern recognition, attendance to sensory and perceptual information, and openness to new experiences. These additional measures can help predict why some students may see greater connections between the puzzle and pertinent information in the environment.

Research Questions and Hypotheses

The first research question of study 2 asks about the effect of providing participants longer periods of time working on the math puzzle (i.e., the original control condition) relative to requiring a break in between (LD Condition). I predict that the findings from study 1 will replicate, including no difference between conditions for lab solve, but a difference in wild solve rates. If the primary finding from study 1 replicates, the next step will be to evaluate one account of why individuals in the control condition may be outperforming students in the LD condition in the wild. Thus, the second research question asks if reaching an impasse is responsible for an effect of condition on wild solve rates. Drawing on the opportunistic assimilation hypothesis, I predict that participants in the control condition will reported reaching an impasse more often, and to a greater extent than LD participants, and that this impasse can explain the relationship between condition and wild solve rates. However, if the finding from study 1 does not replicate, I will evaluate a more general research question about the role of impasse on wild solve rates.

In terms of new trait individual difference measures, I ask whether openness to sensory and abstract information (openness / intellect), feelings of anxiety at having to think creatively (creativity anxiety), and two facets of curiosity (interest and deprivation) predict solve rates. I predict that openness and intellect will be positively associated with solving in the wild, as openness may benefit more from sensory information and intellect benefit from increased tendency to reason with abstract information, which may require more time than the lab session allowed. I also predict that greater creativity anxiety will be associated with worse solve rates in the lab and wild, as anxiety is known to impede performance. Moreover, I have no *a priori* predictions regarding measures of curiosity trait measures (discussed below), so these results should be considered exploratory.

My last research question remains the same from study 1: How do students arrived at creative insights in math, and what are their experiences like? This is an exploratory research question designed to evaluate whether this study's data align with traditional laboratory data, which suggest that creative insights happen after deep immersion periods, commonly following a break or periods of rest, and often come suddenly and unexpectedly (Dijksterhuis & Meurs, 2006; Ritter & Dijksterhuis, 2014; Smith & Blankenship, 1989). Based on the results of study 1, I predict about half of students who do not solve in the lab will end up solving in the wild, and their experiences will represent a diverse array of solve experiences. Further, to build off the results of study 1, I added a few new measures in study 2 to better characterize how the solution came to participants (see below).

Method

Participants and Sample Size Justification

A power analysis was conducted using Stata's powerlog function with an α of .05 and a P1 = .64 and a P2 = .44 to represent wild solve rates in the control condition and in the LD condition found in study 1. For 80% power, an *a priori* sample size was estimated at 57 students per condition. As the goal for study 2 is a replication, I aimed for 85 - 90% power which was estimated to require 65 - 76 participants per condition. As study 1 initially recruited 252 students and ended up with 76 - 78 participants per condition for wild solve analyses, I aimed to recruit 280 participants to ensure I met the necessary sample size.

The resulting sample began with 280 students, but 23 were removed from all analyses because they had either seen the puzzle before, or later reported they had thought to use exponents but believed it broke the rules. Thus, the total sample sized used in analysis consisted of 257 undergraduate students enrolled in an undergraduate psychology course and took place in the study for course credit. The sample was comprised of 195 females (58 males, 4 non-binary), with an average age of 20 years old, representing race/ethnicity of Black/African American (2%), Asian (43%), Indian (5%), Middle Eastern (4%), Latino (19%), White (21%), and Biracial or Other (7%, with an additional 4 participants preferring to not report).

Measures

Trait individual differences. In order to replicate some of the findings from study 1, I retained the two individual difference measures from study one that were significant in relation to lab solve: math anxiety and PANAS positive. Additional measures of individual and state differences were included to expand upon the findings of study 1. For example, measures of openness to experience (openness / intellect), curiosity (interest / deprivation) were included to explain differences in students' ability to pick up information from their environment. Moreover, a new measure that assessed apprehension towards creative thinking was also included (creativity anxiety). Below, I provide greater detail on the trait individual difference measures. See Figure 4 for an illustrative example of when these measures appeared during the procedure.

Openness and Intellect. Openness to experience is one of the five personality traits from the five factor model of personality (also known as the Big Five), and is primarily made up of two constructs— openness and intellect. Openness captures a person's tendency to engage with perceptual and sensory information as well as with fantasy and emotions (Kaufman et al., 2016; McCrae & Costa, 1997). In contrast, intellect describes a person's natural tendency to engage with abstract and semantic information through reasoning (Kaufman et al., 2016). Openness and intellect are distinct but related constructs (DeYoung, Quilty, & Peterson, 2007; Johnson, 1994; Saucier, 1992) that collectively capture cognitive exploration. To measure openness and intellect, the Big Five Aspect Scales (BFAS) for both constructs (DeYoung et al., 2007) will be administered. This involves presenting ten items for each construct which participants rate on a scale of 1 – 5 how much they agree with each of the ten statements (e.g. "I formulate ideas clearly" for intellect, "I get deeply immersed in music" for openness). Four of the ten items are reversed scored, and then averaged to create a composite variable.

Curiosity interest and deprivation scale. This 10-item Likert scale (Litman, 2008) measures two facets of *epistemic curiosity* (described as curiosity related to a drive to learn information, take on intellectual challenges, and eliminate gaps in understanding; Berlyne, 1954). The first is I-type, henceforth referred to as *curious-I*, and captures a person's curiosity to learn for the simple joy of learning something new. Example interest items include asking participants to a rate on a scale from 1 – 4 how often they generally "enjoy exploring new ideas" and find it "fascinating to learn new information." Those who score high on this subscale represent students who engage and learn for pleasure, engage in divergent information seeking, and feel enjoyment in these activities. In contrast, those who score high on deprivation type, to which I will now refer to as *curious-D,* represent those who are more focused on solving a specific problem in an effort to eliminate a gap of knowledge or feelings of uncertainty. Example curious-D items ask participants to rate how often they "work like a fiend at problems that I feel must be solved” and “feel frustrated if I can’t figure out the solution to a problem, so I work even harder to solve it.” Those who score high on curious-D are considered to be motivated by a desire to decrease negative emotions associated with not knowing, rather than for the joy of knowing (as is the case with interest). Researchers have made connections between curious-I and mastery-oriented learning, characterized by intrinsic motivation, and curious-D and performance-oriented learning, characterized by a focus on outcomes and persistence of studying (Elliot & McGregor, 1999).

Creativity anxiety scale. This eight-item scale measures trait-level anxiety that is specific to creative thinking (Daker, Cortes, Lyons, & Green, 2019). Participants are provided a series of statements and are asked to rate how much anxiety they would feel for a given situation on a scale from 1(none at all) – 5(very much). Example items include “Having to come up with a creative solution to a problem” and “Focusing on novelty over precision when doing something”. Scores are averaged across items to create a composite measure.

State measures. State measures that were previously listed and described in study 1 that were included again in study 2 are puzzle stress, puzzle positive, feelings of relief, difficulty of the puzzle, how much the participant wanted to figure out the puzzle on their own, as well as a measure of how curious the participant was to know the solution. I also added a new state measure (lab impasse; see below) that was design to assess the extent to which students have exhausted strategies and reached an impasse. See Figure 4 for an account of when these measures appeared.

Lab impasse. In addition to the above measures, study 2 included two measures of impasse in the lab. The first comes from a question that asks students who did not solve the puzzle in the lab if they reached an impasse and ran out of ideas (yes / no). Students who reported that they did not reach an impasse moved onto the next question, but for those who did report reaching one, they were additionally asked to report how great of an impasse on a scale from 1 (not stuck at all/ I still have lots of ideas) – 4 (very stuck/completely out of ideas). These two measures were included to test the hypothesis that reaching an impasse may explain the control condition's solve rate in the wild, and may generally be related to solving in the wild.

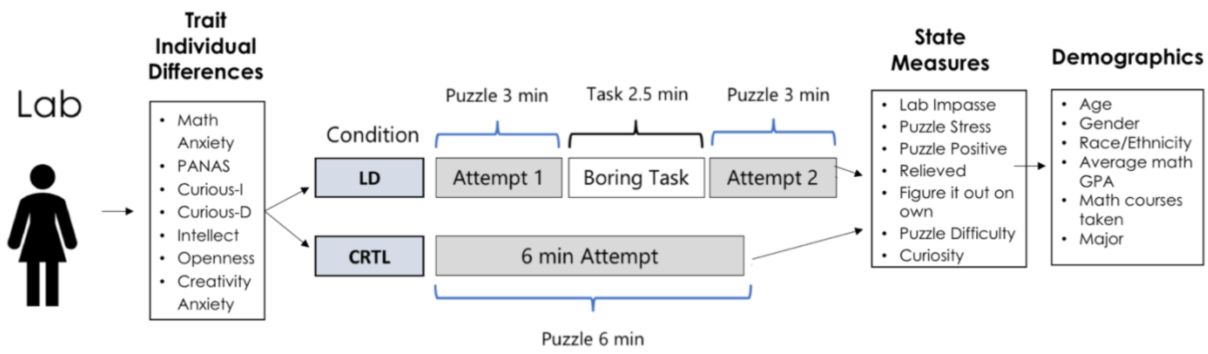

Figure 4. Overview of lab procedure in study 2, including when measures appeared for participants.

Wild measures. Many of the measures taken in the wild of study 1 were present in study 2. For instance, participants reported how often they thought about the puzzle after leaving the lab (1 = never thought about it, 5 = thought about it a great deal), and if they reported solving, whether they were consciously thinking about the problem (yes/no), how the solution came (analytical, insight, mix of both), and their solve experience (open-ended). New measures were added to study 2 (and slight modifications of measures from study 1) which described below in more detail.

Thinking episodes and total thinking time. In study 1, all wild participants reported how long they thought about the math puzzle. This measure was included again, but I also included an additional measure to capture how many *times* participants thought about the math puzzle (“thinking episodes”) and reported their best guess as to the total time they spent thinking about the problem (“total thinking time”: “*Please enter your best estimate of how long in total you thought about the math puzzle outside the lab*”). These estimates were meant to capture a more concrete measure of time spent thinking about the problem, compared with the subjective measure “How often thought about it” used in study 1.

Tried to solve in the wild. Through the open-ended descriptions of solve experiences from study 1, we learned that some participants reported actively trying to solve the problem in the wild, whereas others arrived at a solution without conscious effort, or sometimes, without even thinking about the problem (e.g. a spontaneous insight, or hint from the environment). To better quantify this in study 2, all participants who took the wild survey were asked if they tried to solve the math puzzle after leaving the lab. As attempts to solve may have occurred either by mentally reasoning about the puzzle, or writing the digits down and actively working the puzzle, this question was presented to participants as a multiple-choice question, in which they could choose from the following responses:

- No, I didn’t think about it or try to solve it.
- No, but I did think about it in general.
- Yes, I tried to solve it in my head.
- Yes, I even wrote down the digits and symbols to work on the math problem.
- Other (please specify)

Wild impasse. Participants who reported that they had tried to solve the problem outside the lab were additionally asked “When you were working on solving the math puzzle outside the lab, did you reach an impasse (i.e. felt stuck or ran out of ideas) at any point?” and responded with either yes or no. This measure was added to ensure that students who did not bring themselves to impasse in the lab, but may have outside the wild, were included in analyses that test the role of impasse on solve rates in the wild.

*How solution came (new modification and new additions)***.** In study 1, participants were provided with three statements, one designed to represent an analytical-problem solving approach, one an insight experience, and one a mix of both and participants were asked to select which statement best represented their solve experience. These measures were retained in study 2, but instead of asking participants to select the one that best characterized how the solution came to them, they were asked to *agree/disagree* with each statement as it pertained to how the solution came to them. This change was made to better understand the complex solve experiences students reported in their open-ended responses, which may encompass more than one of these experiences. The three statements that they were asked to agree/disagree to were:

- "I could feel myself slowly getting closer to the solution, until the solution came"
- "The solution seemed to come out of nowhere"
- "I tried many different things until it seemed like there were no solutions. Then it came to me"

In addition to the these three statements, two more statements were added to study 2 to tease apart nuance differences in solve experience. The first one, "The solution came suddenly and unexpectedly," was reported with the other measures and participants either agreed or disagreed that this characterized how the solution came to them. This statement aimed at characterizing insight, was derived from the perspective of Kounios and Beeman (2015) who argue that a distinguishing difference between analytical problem solving and an *aha!* moment is the sense of suddenness and unexpectedness. It may not be enough to capture this feeling as I originally did ("the solution seemed to come out of nowhere"), so this measure explicitly asked about suddenness and unexpectedness. The second new measure simply asked participants, "When you solved the math problem, would you say it felt like an 'aha' moment?" for which participants either reported yes or no. This measure comes from Gable, Hopper, and Schooler (2019) who captured the presence of an *aha!* moment by asking participants to simply report if they had one. I reasoned initially in study 1 that solutions that come suddenly would constitute an *aha!* moment, but this measure is more direct about their subjective experience and identification with an *aha!* moment.

Solve experiences (open-ended). Whereas in study 1 wild solvers were asked to report what they were doing right before solving through an open-ended response item, wild solvers in study 2 received more thorough instructions, and were asked to, "describe in at least four sentences what were you doing right before you solved the problem. Please provide as much detail as possible (even if it does not seem relevant)." The goal with these more thorough instructions and minimum required number of sentences was to encourage students to write more about what they were doing, and avoid terse responses (e.g. "eating").

Wild solve activities. Wild solvers in study 2 were additionally asked to select from a list of specific activities what they were doing right before solving (i.e. walking, showering/bathing, cleaning, listening to a lecture, etc.). These items were derived from the open-ended descriptions of solving from study 1, and help to better estimate the most common activities students are engaged with when solving.

Environmental description and potential hints. In addition to their solve experiences, wild solvers were asked if there was there anything in their environment that helped them solve the problem (e.g. someone having a conversation about math, seeing a formula written on a whiteboard, etc.) through a yes/no response item. Yet, one difficulty in capturing the influence of environmental hints with this method is that students are not always aware of environmental cues that might help them solve. Therefore, students were also asked to describe the environment around them when they solved in at least four sentences. These open-ended environmental descriptions were then coded for evidence of a potential hint (e.g. “I was helping a friend with their math homework”), creating a measure of hints that encompasses both students who reported a hint and students who reported information in their environment that could have acted as a hint to create a composite measure of potential hints.

Procedure

Participants began in the lab by filling out trait individual difference measures and then were randomly assigned to either a LD condition or the control condition. Participants assigned to the control condition worked on the math puzzle uninterrupted for 6 minutes. Participants assigned to the LD condition worked on the math puzzle for 3 minutes, then switched to a signal detection task for 2.5 minutes, before returning to the math puzzle for another 3 minutes. These conditions are identical to the control and LD condition in study 1. After the math puzzle attempt ended, participants who did not solve in the lab answered state measures (e.g. impasse, how curious they were to know the solution, etc). Participants then reported identical demographic information as in study 1 (e.g. age, gender, race/ethnicity, average math GPA, math courses taken, etc).

Before participants left the lab, non-solvers were provided a survey link and instructed to take the follow-up survey as soon as possible if they found a solution to the problem. Participants were then tracked up to three days after they left the lab. If a participant had not reported solving the problem after three days had passed, they were emailed instructions to take the follow-up survey (same as study 1 protocol).

Data Analysis Plan

Given that study 2 is a replication of study 1, the data analysis plan largely reflects the same analyses from study 1. The main research question looking at the effect of condition on lab solve rates (a binary outcome of solved, did not solve) and wild solve rates (also a binary outcome of solved, did not solve) will be tested using binary logistic regression, just as it was used in study 1. Similarly, trait individual difference measures will be used to predict lab solve rates and wild solve rates to examine whether they can significant predict solving in the lab or wild. State measures will then be used to predict solving in the wild (as they are collected after students' lab solve attempt), and student's experiences in the wild will be used in a descriptive and exploratory fashion to shed light on what these experiences are like for students. The same data exclusion rules from study 1 were applied in study 2.

To aid with clarity, I have divided the results section in two parts. The first part details the analyses of trait individual differences and state measures on lab and wild solve rates, with findings generally organized findings between those that are consistent with study 1, those inconsistent with study 1, and new analyses of variables not previously collected. In part 2, I focus on measures collected in the wild (e.g. how often thought about it) as well as the experience of students who solve in the wild.

Results

Part I

Division of the sample and solve rate. Division of the sample can be found in Figure 5. A total of 280 students were recruited for this study, but 23 were removed from overall analyses as three had reported seeing the problem before, and 20 later reported that they had thought to use exponents, but believed it was against the rules (as it may include an additional carrot "^"

symbol). Data from 257 students were used in analyses of the lab phase of the study. A total of 86 students solved in the lab— a 33% solve rate compared to a solve rate of 22% in study 1 (see Figure 6). Students who *did not* solve in the lab were enrolled in the wild phase of the study (*n*= 171). Of these students, five students were removed from analyses on wild solve rates because they never took the wild survey (after repeated attempts to follow up), a further 16 participants were removed from analyses as they had reported googling the problem or solving with someone else, and three students were omitted as they failed the memory test. This resulted in a sample size of *n* = 147 students for wild solve analyses (control = 72 participants, LD = 75 participants). Across conditions, 70 students solved in the wild (48%), a comparable rate to study 1 (52%).

Effect of incubation break. To test my first research question about the effect of condition on solve rate, I used binary logistic regression to predict lab solve from condition. This regression tests for the relationship between condition and odds of solving the math puzzle. If the incubation break facilitated incubation and insight within the lab, the model would be statistically significant with a positive *b* coefficient (indicating that being in the LD group rather than the control group is positively related to solving in the wild). Replicating the results from study one, the model was *not* significant (LD vs. Control p = .62), indicating that there was no effect of incubation condition on solve rates *in the lab*. To test for differences in solve rates in the wild, I ran a separate binary logistic regression predicting wild solve rates from condition. If the results from study 1 replicated, this model would be significant with a negative *b* coefficient, indicating worse predicted solve rates for LD students compared to control students. However, this was not what I found, as the model was not significant (p = .28), suggesting that there was no statistically significant differences between the two conditions for wild solve rates in study 2. Table 3 shows the percent of the condition that solved in the lab and the wild for each study.

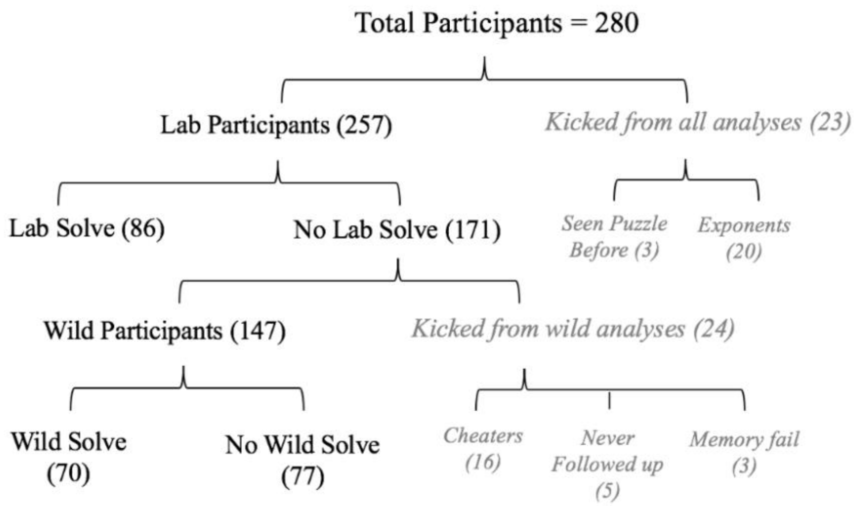

Figure 5: Division of total sample across study. Sections that appear in gray represent participants who were removed from analyses.

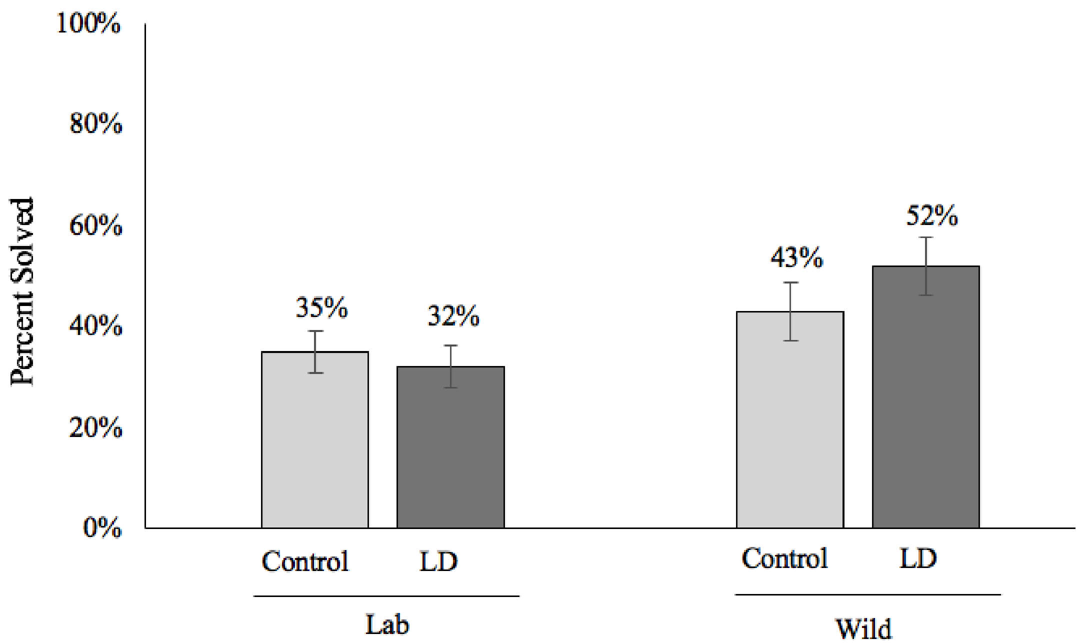

Figure 6: Solve rates in the lab and in the wild by condition for study 2.

Table 3: Percent of Control condition and LD condition who solved in lab and wild by study.

	Study 1		Study 2	
	Lab Solve	Wild Solve	Lab Solve	Wild Solve
Sample Size	231	166	257	147
% solve: Control	18%	64%	35%	43%
% solve: LD	22%	44%	32%	52%

Trait individual measure differences and lab solve rate. I also examined whether trait individual differences (e.g. math anxiety, PANAS, openness) could predict solve rates in the lab (see Table 4 for descriptive statistics, and Table 5 for logistic regression statistics). Similar to study one, the PANAS Negative scale, which captures a participant's trait tendency to experience negative emotions, was not significantly related to solve rates in the lab. On the other hand, the relation between math anxiety and solve rates in the lab was "marginally" significant ($p = .053$), with a similar strength of evidence as study 1 ($p = .048$). Both directionalities suggest that greater math anxiety is generally associated with lower expected odds of solving in the lab. Moreover, whereas study 1 found a negative relationship between PANAS Positive, and lab solve, study 2 did not find this relationship.

In terms of the new trait measures, I found that intellect was statistically significantly related to lab solve, such that for every one unit increase in intellect, the expected odds of solving in the lab increased by 63%. Alternatively, intellect's counterpart – openness – was unrelated to lab solve. The second new measure, curious-I, was also positively related to lab solve, such that for every one unit increase on curious-I, the expected odds of solving increased by 61%. Alternatively, curious-D was not related to solve rate in the lab . Creativity anxiety also did not relate to solve rate in the lab.

Trait individual differences related to wild solve rates. In this section I continue to examine trait individual difference measures, but examine whether these measures relate to solve rate in the wild. In similar fashion to study 1, none of the trait individual differences predicted wild solve rates, except one. In study 1 math anxiety did not predict solve rates in the wild, however in study 2, this relationship was marginally significant ($p = .048$). None of the new trait measures in study 2 predicted solve rate in the wild (i.e., intellect, openness, creativity anxiety, curious-I, curious-D). Collectively, these results mirror those in study 1, which generally found trait measures were unable to predict solve rate in the wild.

State measures related to wild solve rates. In this section I examine the relationship between the state-level measures and solve rates in the wild. As a reminder, only participants who did not solve the puzzle in the lab were asked to complete these measures (and thus, state measures were not used to predict lab solve). Consistent with findings in study 1, I found that there was no relationship between solve rate in the wild and puzzle stress, puzzle positive, puzzle difficulty, and the extent to which participants wanted to figure out the solution on their own. Study 1 found that feelings of relief that the problem-solving attempt had ended in lab was *negatively* associated with wild solve rates, and level of curiosity to know the solution was *positively* related to wild solve rates. In study 2, neither of these measures predicted solve rates in the wild.

Table 4

Descriptive statistics of trait individual difference measures used for lab and wild solve analyses in study 2, and state measures used in wild solve analyses.

	n	mean	*s*	median	min	max	range
Trait Individual Differences							
Math Anxiety	257	2.57	0.72	2.4	1.1	4.8	3.7
PANAS Positive	257	16.74	3.05	17	8	24	16
PANAS Negative	257	11.44	3.15	11	5	21	16
Openness	257	3.80	0.63	3.8	2	5	3
Intellect	257	3.29	0.65	3.3	1.5	4.8	3.3
Creativity Anxiety	257	2.70	0.92	2.6	1	5	4
Curious (Interest)	257	2.69	0.63	2.5	1.3	4	2.8
Curious (Deprivation)	257	2.33	0.66	2.3	1	4	3
State Measures							
Puzzle Stress	147	4.27	1.50	4	1	7	6
Puzzle Positive	147	5.10	1.56	5	1	7	6
Relieved	147	3.31	1.82	3	1	7	6
Puzzle Difficulty	147	4.14	0.54	4	3	5	2
Figure out on Own	146	3.56	1.29	4	1	5	4
Curiosity	147	6.41	1.06	7	1	7	6
Lab Impasse	147	0.78	0.41	1	0	1	1
Lab Impasse (Continuous)	147	3.09	0.69	3	1	4	3

Note: Math Anxiety, openness, intellect, creativity anxiety, curious (Interest) and curious (Deprivation) scores are averaged across a series of items, whereas PANAS items are summed across 5 items.

Table 5

Outcomes of logistic regressions predicting lab solve and wild solve from individual difference measures in study 2

	Lab Solve			Wild Solve		
	b estimate	SE	Sig.	*b* estimate	SE	Sig.
Trait Individual Differences						
Math Anxiety	-0.37	0.19	$p = .05$	-0.45	0.23	$p = .05$
PANAS Positive	0.02	0.04	$p = .62$	-0.02	0.05	$p = .76$
PANAS Negative	-0.05	0.04	$p = .26$	-0.01	0.05	$p = .92$
Openness	0.08	0.21	$p = .69$	-0.11	0.27	$p = .67$
Intellect	**0.60**	**0.21**	$\boldsymbol{p < .01}$	-0.07	0.26	$p = .79$
Creativity Anxiety	-0.07	0.15	$p = .64$	-0.05	0.18	$p = .76$
Curious (Interest)	**0.43**	**0.21**	$\boldsymbol{p = .04}$	-0.13	0.27	$p = .65$
Curious (Deprivation)	0.32	0.20	$p = .11$	-0.00	0.27	$p = .99$
State Measures						
Puzzle Stress	--	--	--	-0.91	0.51	$p = .09$
Puzzle Positive	--	--	--	-0.03	0.10	$p = .77$
Relieved	--	--	--	0.05	0.09	$p = .61$
Puzzle Difficulty	--	--	--	-0.39	0.31	$p = .22$
Figure Out on Own	--	--	--	-0.06	0.13	$p = .63$
Curious	--	--	--	-0.12	0.16	$p = .43$
Lab Impasse	--	--	--	0.19	0.40	$p = .62$
Lab Impasse (Continuous)	--	--	--	0.10	0.24	$p = .67$

Note: Output that appears in bold represent statistically significant findings.

Lab impasse. A new and unique research question addressed by study 2 was whether reaching an impasse in the lab boosted solve rates in the wild. As a reminder, students who did not solve the problem in the lab were asked whether or not they had reached an impasse, and rated how great of an impasse they had reached. I found that neither the binary measure of lab impasse (yes/no), nor the continuous measure, was predictive of solve rate in the wild ($p = .62$, $p = .66$ respectively). Surprisingly, I found that condition was significantly related to reaching an impasse in lab, but in the opposite direction predicted. Specifically, students assigned to the control condition were less likely to report an impasse in lab ($p < .001$), and LD students reported reaching a greater impasse ($p <.001$) compared to control students. The interaction between lab impasse and condition on solve rate in the wild was not significant ($p = .40$). Results related to bringing oneself to impasse in the wild are discussed in a subsequent section.

Math background on lab solve. For study 2, students overall had high levels of math background with 91% of the overall sample having taken and passed calculus. This is consistent with study 1 (where 88% had taken and passed calculus). Across both studies, specific math courses taken solve rate in the lab or wild (e.g. taking calculus did not relate to solving the math puzzle). The relationship between self-reported average math grade and solve rate in the lab was not significant in study 1 but was significant in study 2 ($p < .005$). For every one unit increase on reported math GPA (one unit being B to B+, B+ to A-, A- to A, etc.) the expected odds of solving increased by 59%. Furthermore, the relationship between self-reported GPA and solve rate in the wild was not significant in study 1 but was significant in study 2 ($p = .02$). For every one unit increase in self-reported GPA, the expected odds of solving outside the lab increased by 58%.

Part II

The first set of results examined trait state differences captured in the lab, and their relationship to solve rate in the lab and in the wild. In this section, I shift focus away from measures captured in lab and toward measures taken in the wild.

How often thought about problem. Replicating the findings from study 1, students who reported thinking about the problem to a greater degree in general (*"How often did you think about the math problem since you left the lab session?")* were more likely to solve in the wild (p <.001). In contrast, new measures of their total estimated thinking time about the problem and the number of reported independent thinking episodes *were not* related to solve rate in the wild *(p= .82 and p* =.14 respectively). These results suggest that students' general perceptions of how often they thought about the math puzzle, but not their quantitative estimates of time thinking about the puzzle, predicts solve rate in the wild.

Tried to solve in the wild. Participants were also asked explicitly if they *tried to solve* the problem after leaving the lab, as this may differ from reports of generally thinking about the puzzle. Across solvers and non-solvers in the wild, 74% of participants reported trying to solve the problem in the wild. Of the students who reported trying to solve in the wild, 37% reported that they even wrote down the symbols and digits of the puzzle in an attempt to solve. Students who reported trying to solve outside the lab showed greater solve rates compared to those who did not try to solve outside the lab (p <.001), such that there is an expected increase of wild solve rate of 83% for students who reported either trying to solve the problem in their head or on paper compared to those who did not report trying to solve it at all.

Wild impasse. Another measure included for analyses was whether students brought themselves to impasse after leaving the lab (regardless if they already reached an impasse).

Across all participants enrolled in the wild portion of the study, 70% of participants reported bringing themselves to impasse in the wild, but this was not predictive of whether students solved in the wild (p = .17). Interestingly, while there was a difference between the two conditions for lab impasse, there was no difference between the two conditions when it came to bringing oneself to impasse outside the lab (p = .13)

Context for solving in the wild.

Consciously thinking of the puzzle. Wild solvers were asked whether they were consciously thinking of the puzzle before solving. In study 2, 89% of wild solvers reported that they were thinking of the puzzle right before solving. This is comparable to study 1, which found that 82% of wild solvers reported consciously thinking of the puzzle before solving.

How solution came. In study 1, participants were presented with three statements designed to describe an insight moment experience, an analytical problem-solving experience, and a mix of both, and participants selected which one best described their experience. For study 2, these measures were included, but instead of asking students to choose the best description that matched their experience, they were allowed to check all that applied. Over half (56%) of students agreed that “I could feel myself slowly getting closer to the solution, until the solution came” (analytical problem-solving). Within the same sample, 69% reported that “The solution seemed to come out of nowhere” (insight), and 51% agreed that “I tried many different things until it seemed like there were no solutions. Then it came to me” (mix). This conceptually replicates the findings in study 1, where there was good representation among all three categories (which were roughly chosen equally among participants). Results further revealed that 33% of students freely chose only one of these options, while 46% selected two and 17% selected all three options. This suggests that some students felt a mix of these experiences, or phases of these experiences leading up to their solve experience or were simply unable to adequately describe their experience.

To better capture whether students subjectively experienced an *aha* moment, I asked them to agree or disagree that their solve experience felt like an *aha!* moment. The majority (89%) agreed that their solve moment felt like an *aha!* moment. However, when asked if "the solution came suddenly and unexpectedly," only 71% agreed, suggesting some students experienced an *aha!* moment without the sense of suddenness and unexpectedness.

Wild solve activities. In addition to reporting information about what best characterized their solution experience, wild solvers in study 2 were asked to check all activities they were engaged in before solving from a list generated from the open-ended responses in study 1. Previous research has found themes of the five b's of insight moments— bed, bath, buses, boring meetings, and bars. Overall, these insight activities tend to capture moments where one is at rest and not cognitively engaged with any particular task. Table 6 contains the full list and percent of participants who reported each activity.

Most notably, walking was the top activity for solving (46%), which may be explained by instances of students solving when they are walking away from the lab (see solve experiences section). The second and third most popular activity was waiting for something (30%) and taking a break or resting (24%). These accounts fit well with the findings of past research that *aha!* moments happening during brief moments of rest or breaks from engagements.

Table 6: Percent of all wild solvers who was doing each of the following activities before solving in study 2.

Activity	**% of Wild Solvers**	**Activity**	**% of Wild Solvers**
Walking	46%	Actively engaged with something (besides math problem)	10%
Waiting for something	30%	Drinking	9%
Taking a break / resting	24%	Listening to a lecture / sitting in class	6%
Talking to others	19%	Sitting on the bus / in the passenger seat of a car	4%
Sitting / Lying in bed	17%	Driving	1%
Eating	13%	Working on the problem with another person	1%
Doing homework / studying	11%	Showering	0%
Listening to a podcast / music	11%	Watching TV	0%
Checking email / social media	10%	Smoking	0%

Solve experiences (open-ended). Similar to study 1, wild solvers in study 2 reported a diverse array of experiences solving, with some consistent themes of how students arrived at the solution. For example, study 2 found that many students reported working on the problem outside of the lab. One student said:

> I was walking around for about 10 minutes until I sat down and started thinking about different methods to solve the problem and realized that I could do "to the power of" and took out different pieces of paper. I started focusing on combinations of to the power of and forgot about the addition symbol. As soon as I remembered about that symbol, I realized that 3 to the power of 2 = 4+5.

Other students echoed this conscious work theme by stating that they mentally thought of different ways to rearrange the numbers. A few students reported thinking about different kinds of math, which led them to think of PEMDAS (parentheses, exponents, multiplication, division, addition, and subtraction) which led them to think of exponents. Other students reached the solution by simply trying to mentally rearrange numbers. For instance, one participant said

> I was in the back of a car with friends who picked me up from campus after I participated in the study. I had continued thinking about the problem as I left the study and got into the car. I was mentally trying to rearrange the digits and symbols. I realized that exponents do not require additional symbols.

These examples illustrate how some students arrived at the solution by mentally grappling with the problem, in line with an analytical-problem solving approach. Additionally, other students reported having experiences more akin to spontaneous insights, where the solution felt like it came to them unexpectedly, without conscious effort. For example, one participant reported that their morning routine included dancing before the start class, which is when the insight moment came to them. The participant explained, “While I was dancing, I thought of the math problem and just brushed it off for a bit. 10 minutes later, the solution just came to me.” Another student reported a more traditional insight moment (e.g. not necessarily having the solution come without previous conscious thought), reporting

> “I was walking around outside back to my dorm room from dinner. I was think[ing] about my day. I thought about how I did not solve the problem and then it came to me in a couple of minutes. It just came to me.”

Study 2 also found evidence for some of the five B’s” of insight— bed, bathroom, bus, bars, and boring meetings (though evidence for the few students who reported drinking indicates that they were most likely drinking coffee). Although no participant reported solving in the bathtub/ shower, some participants reported solving in bed either before sleep or after awakening. For example, one participant recalled what they had done right before they solved, stating:

> “I was preparing for bed. I'd already gone through my night routine (meds, brushing teeth, shower, skincare). I climbed into bed to listen to My Brother, My Brother, and Me, the comedy podcast that I like to listen to in free time while walking to and from classes and before going to bed. I was drifting off to sleep while thinking about the problem.”

Another student reported “I had just woken up and remembered the problem. I decided to look at it one more time. All of a sudden I realized the solution to the problem. I was still very sleepy.”

When it came to solving on the bus, or in cars in general, several participants reported solving during this time. One of the most notable accounts comes from a participant who reported being stuck in traffic for an hour while trying to drive to visit their parents for the weekend. This participant explained that they had become bored, and began to think about the problem: “I visualized "23" in my head and thought about how you could use *any* mathematical method to solve it. Then, it hit me—I could use exponents.” Similarly, another participant reported that they had an hour-long bus ride to and from campus, and after daydreaming for a little while, they grabbed a notebook and started working on the math puzzle, finding the solution afterwards.

The theme of “boring meetings” was also common among wild solvers. For instance, one participant wrote:

> I was in class taking notes and listening to the professor talk. Some students were saying their opinions. Then we watched a video. At some point I got bored and I was thinking about the problem, then I started to try solving it. I started using all the numbers and combining them on a piece of paper until I thought I had an answer.

Another participant, in a moment of irony, reported solving during another campus experiment when they became bored. They stated, “while I was watching the video doing the experiment, I thought about what other types of mathematical rules don't need a sign, and I suddenly came up with the rule of power.”

Environmental description and hints. In addition to finding evidence of some of the five b's, participants in study 2 also reported hints in their environment had helped them to solve (see Figure 7 for an image). For instance, one participant reported:

> I went outside and sat [on] the benches that is intersected between Franz Hall and Psychology Building. I listened to music and stared at the UCLA tour group where my attention was caught when the tour guide pointed at the Kinsey Pavilion. I looked up and realized the equation written on the mural. Then I configured combinations of symbols and numbers to solve the equation.

Figure 7. The Kinsey Pavilion Mural with exponent formula

Another participant had reported doing a homework assignment that had an equation with an exponent, which helped them make the connection and solve. Other participants reported hints that were less direct. For example, one participant said:

> "I was watching Pewdiepie play a 12 hour minecraft stream. I went to heat up some ramen and as I was eating, my boyfriend called. I told him about the study… and I started staring at this corner on my desk. Then, I thought about the Pythagorean theorem and wrote down the 4 digits and 2 symbols to solve it."

In this case, the participant had seen the right angle of the desk corner, which had triggered the idea of Pythagorean's theorem, which uses exponents ($a^2 + b^2 = c^2$). The environmental hint of a desk corner is much weaker than seeing an exponent in the wild, but it still helped this student reach an insight moment. These qualitative accounts suggest that students do not necessarily need to see the solution to solve (often used in previous research), but can make remote connections given the *right* hint.

In addition to describing their solve experiences, students who solved in the wild were also asked to describe their environment and report if there was something in their environment that helped them solve. This was included as another measure to test the opportunistic assimilation hypothesis, which suggests that diverse environments are helpful for problem solving, as they provide useful environmental cues for individuals who have reached an impasse on a problem but are unable to solve (and thus, are more sensitized to information that could help them solve). Out of the 70 students who solved in the wild, only 15 reported that there was something in their environment that helped them solve. Two examples have already been provided above. Another example comes from a participant who said:

> My roommate next to me was doing his math homework on a late Sunday night, and I glanced at it a couple times out of curiosity. I can't say that this was the direct cause, but seeing exponents written down probably jogged my memory that they even existed.

To test one of the tenants of the opportunistic assimilation hypothesis – that students need to reach impasse before picking up on hints in the environment – I ran a binary logistic regression predicting whether a participant reported a hint in the wild from impasse. The resulting model was not significant, indicating that students did not need to reach impasse to pick up on hints.

Open-ended responses about the environment were then reviewed for any evidence of math-related stimuli in the environment, and were coded as to whether it may have possibly helped the student solve the problem. Only two participants who did not self-report a hint had listed information in their environment that could have been an environmental cue, resulting in 17 participants with a potential hint in the environment that aided them. This suggests than less than one out of four participants who solved in the wild may have benefitted from a hint in their environment.

Study 2 Discussion

The purpose of study 2 was to investigate whether the results of study 1 were replicable, and evaluate the predictive power of new trait individual difference measures that may better explain solve rates in the wild. Study 2 also included new measures of impasse and hints in the environment in order to evaluate aspects of the opportunistic assimilation hypothesis. With a close to identical procedure as study 1, the results of study 2 confirm some findings, while casting doubt on other important findings of study 1.

One of the most notable takeaways from study 2 was that I was unable to replicate the original finding that students who are not interrupted during their problem solving attempt in the lab (control condition) show higher solve rates in the wild compared to those whose thinking is broken up with an incubation break (LD condition). Results revealed that the condition students were assigned to did not explain solve rates in the lab or in the wild. One explanation for the discrepancy between studies for wild solve rates is that the results of study 1 were a type 1 error (false positive), or that the results of study 2 were a type 2 error (false negative). Without further follow-up, it is difficult to ascertain which of these possibilities is true, but it is worth remembering the real-life context in which this aspect of the study took place. Replicating research even in the strictest lab settings has proven to be a difficult task (e.g. Altmejd et al., 2019; Open Science Collaboration, 2015). Replication may be even more difficult for studies that take place outside the lab in the more chaotic context of everyday life.

Relatedly, study 2 found that both state and trait measures largely did not predict solve rates in the wild. The only significant predictor identified in study 2 was math anxiety, and the evidence was relatively weak (p = .048). As this is comparable to the strength of evidence of study 1 (p = .052), the combined results of the two studies suggest that math anxiety might play a small, but perhaps a significant role as to whether or not students solve in the wild. Further study is needed to explain why math anxiety seems to be teetering on the edge of significance. One possibility is that the math puzzle may not have been aversively stressful enough to trigger students' math anxiety more uniformly. While some math anxious students may show an acute negative reaction to even benign situations that require math (i.e., calculating a tip, see math symbols) others may not feel much anxiety towards situations like calculating a tip or solving the math puzzle. In fact, the majority of students in the sample felt quite positive about the math puzzle as they indicated that the puzzle was a positive challenge and showed a willingness to continue thinking about the puzzle, rather than the traditional avoidance response that has traditionally been found to characterize math anxious students.

Trait individual difference measures showed more promise at predicting solve rate within the lab rather than outside of the lab. Of the new trait individual measures included in study 2 – openness, intellect, curious-I, and curious-D, creativity anxiety – two were significantly related to solving in the lab. The first was intellect, a personality trait that measures a person's natural tendency to engage with abstract and semantic information through reasoning (Kaufman et al., 2016). It is not surprising that students higher in intellect would be show better odds at solving in the lab, as past research has been linked intellect to academic achievement, GPA, critical thinking, and effort regulation (see Bidjerano & Dai, 2007). In fact, the effect of intellect on lab solve was still significant, even when controlling for students' self-reported math GPA. In addition, students who reported greater scores on curious-I were more likely to solve in lab. This measure captures a person's curiosity to learn for the simple joy of learning something new. Research on curious-I has found that it is associated with a focus on mastery and learning, rather than performance (Litman, 2008), and is positively associated with the acceptance of uncertainty (Litman, 2010). Thus, the current finding aligns well with past research on curious-I, as the math puzzle is an ill-defined problem that students may have had an easier time solving if they enjoyed the challenge of solving something new and were unbothered by any ambiguity related to the problem.

Study 2 collected measures of impasse and information about the environment in order to evaluate whether students ability to reach an impasse might be driving the increased wild solve rates of students in the control condition. However, I did not find this to be the case. Reaching an impasse in the lab or the wild did not predict wild solve rates. There were differences between the two conditions in terms of lab impasse, such that students randomly assigned to the LD condition were more likely to report reaching an impasse while in the lab (the opposite direction predicted). Yet, there were no differences in impasse rates recorded in the wild between the two conditions. It is difficult to ascertain why students in the LD condition may be reaching a higher rate of impasse in the lab (or report one). Perhaps giving students a break from solving the problem provides students with the mental space to reflect on their lack of progress and lead to greater sense of impasse. If this is the case, then future research should consider using more objective measures of reaching an impasse that provide visible and measurable evidence of progress (i.e., pauses in writing, pauses in a think-out-loud paradigms).

Student descriptions of the context and method by which they solved the puzzle in their everyday lives revealed some interesting findings. For instance, some participants reported examples of environmental hints triggering insight moments. Yet, this was not the most predominant experience as only 17/70 students reported a hint (or provided information that signaled a hint was present at the time of solve). While not all students may have benefitted from hints, the opportunistic assimilation hypothesis may help to explain at least *some* students' solve experiences. In addition, students' description of the context for solving provided additional evidence for the observation that insights and solve moments occur during benign daily moments such as boring meetings, bus and car rides, and before and after sleep. No participants reported solving in the shower, which has been popularly referenced experience (see Kaufman & Gregoire, 2016; Ovington, Saliba, Moran, Goldring, & MacDonald, 2018) or while at a bar or when drinking alcohol (see Jarosz, Colflesh, & Wiley, 2012). These results reveal that problem solving happens quite regularly in the diverse context of informal settings, and is an area that would benefit from additional future research.

General Discussion

Math instruction is viewed by many students as a rigid subject in which success depends on conventional thinking and rote memorization (Rodriguez, 2019). However, mathematics draws heavily on creativity, insight, and long periods of thoughtful deliberation (Mann, 2006). Expert mathematicians who work at the forefront of mathematics can struggle with a problem for months or even years, later reporting sudden and seemingly unexplained insight experiences (Mackenzie, 2006). Students also report these moments in math, reporting *aha!* moments that reflect sudden solutions, or an understanding concepts (Liljedahl, 2005).

Beyond anecdotal evidence that creative insights occur in math classrooms, much of what we know about these moments is informed by research conducted in controlled laboratory settings that use abstract verbal tasks. Traditional laboratory experiments find that students can solve insight problems in a matter of minutes, but without further follow-up, our understanding is limited to those who can solve such problems in short periods of time, artificially approximating the creative experiences in the outside world. More naturalistic research methods suggest that authentic *aha!* moments are often the end result of complex cognitive processes that occur over the span of days or even weeks of struggle (Gable, Hopper, & Schooler, 2019). Because creative problem-solving can take naturally longer periods of time, what we learn from hour-long lab experiments we may not be able to generalize to the experiences that happen in everyday life.

In the current dissertation, I sought to paint a more holistic picture of the creative problem-solving process by expanding previous research through two studies. I shifted away from abstract lab tasks toward a mathematics puzzle task and built upon prior lab work by following students outside of the lab to capture information about their solve experiences in their everyday lives. In short, I attempted to diversify the lab research on creativity and insight moments by testing if previous findings could hold when students were faced with a creative math puzzle, and extend this body of work by providing rich information about how students come to experience *aha!* moments outside the lab.

Across both studies, I found that students randomly assigned to an incubation break did not show improved solve rates compared to students who were not given an incubation break in the lab. This finding contrasts previous studies that find that providing students a brief break from a problem partway through their solve attempt can facilitate insight within a controlled laboratory setting (see Sio & Ormerod, 2009). Instead, the results of study 1 revealed that students who are *not* provided this break during their problem solving attempt (control condition) showed a higher solve rate *in the wild*, relative to both experimental conditions. As seemingly anomalous results from control conditions have led to important discoveries and findings (Dunbar, 2001b; see also Buckner, Andrews-Hanna, & Schacter, 2008), I attempted to replicate this effect in study 2, but was unsuccessful. It is difficult to know whether the results of study 1 are the result of a type 1 error, or perhaps capture some real effect of an incubation condition vs. control condition on solving in the wild that study 2 was unable to capture. However, as the replication was unsuccessful, this finding is inconclusive and warrants further research.

Lab vs. Wild

Averaging across both studies, I found that 28% of students could solve the math puzzle in the lab. This is a promising sign that the math puzzle can be used as a suitable measure for studying math-related creativity as it produces good variability and requires aspects of creative thinking to solve (e.g., overcoming fixations, inhibit interference, and make remote connections to solve). In addition, it led students to report *aha!* moments when solving. Importantly, if the experiment had ended after the lab visit, it would seem that only about one in four students at a world-renown academic institution could reason creatively enough to solve a math problem that largely draws upon basic-level math. By following students in the wild, this inference changes. Whereas a lab-only inference would conclude than only 28% of participants could successfully solve the math puzzle, across both studies, I find that when we follow students in to the wild, another 32% of all students go on to solve, creating an overall solve rate of 60%.

The disconnect between lab experiments and everyday life in the wild is not a new revelation. Many researchers struggle to replicate classic lab findings in real-world contexts. For example, Fries, DeCaro, and Ramirez (2019) studied whether providing students with interesting but irrelevant “seductive details” during instruction helps with the learning of core material. In previous laboratory research, the use of seductive details has been found to distract attention away from core material and harm learning (see Sundararajan & Adesope, 2020). A similar negative impact of using seductive details was also found by Fries et al. (2019) when they attempted to replicate the finding in a low-stakes environment. However, when they attempted to approximate a more ecological setting using a high-stakes environment that mirrored the environmental pressures often found in classrooms, they found a very different pattern of results – the inclusion of seductive details largely *helped* with learning.

In a similar fashion, Hulleman and Cordray (2009) found that while an intervention to improve student motivation was successful in the lab (hedge’s g = .45), there was no effect when this intervention was applied within the classroom (hedge’s g = .05). While Hulleman and Cordray (2009) suggest that implementation and the diverse environment of the classroom may wash out effects from lab to the field, other researchers have argued that students might process information differently inside and outside the lab altogether. Dunbar (2001b), for instance, found that when doing analogies, students in the lab tend to focus on superficial details, but when they solve them outside the lab, they tend to focus more on underlying structure between them. Differences between the lab and the wild in the current dissertation (as well as past studies) may be due to differences in processing, differences in the environment, or even due to the fact that participants often have more time to think outside the lab. The larger picture is that the laboratory context is a critical and important step for investigating psychological phenomenon. Nevertheless, psychologists interested in understanding psychological processes related to education should also extend their work to the classroom or the wild as well, given the disparate results that can arise.

Trait and State Measures

A consistent pattern of results across the two studies was that trait individual difference measures could not predict whether a student would solve outside the lab. For example, study 1 found that only math anxiety and PANAS positive traits predicted solve rate within the lab, but neither could predict solve rate in the wild. In turn, study 2 found that measures of intellect and curious-I could predict solve rates in the lab, but similarly, failed to predict whether students would go on to solve in the wild.

As none of these measures predicted solve rate in the wild, it is likely that the wild environment in combination with the math puzzle task might be too varied for trait measures to explain. For example, reports from students who solved in the wild revealed that insight could potentially derive from different mechanisms, such as whether a student decides to work hard to solve the problem after leaving the lab, or if they happen to be aware enough of their environment to pick up on hints, or if they are processing the problem outside conscious awareness. Because of the various paths and processes that can lead to solving in the wild, identifying a trait or state measure that explains all potential mechanisms that lead to solving is very difficult.

A second point to consider is that many traditional individual difference measures are developed specifically for laboratory research, and are validated in the lab. As lab-based experiments fundamentally control variation (Falk & Heckman, 2009), trait measures may be able to explain more variation in the lab because there is simply less variation that needs explaining. However, the drawback to laboratory procedures is that such control over variation reduces the ecological validity when using these measures more broadly. While a number of psychological measures exist that were developed for the lab and predict field data as well (e.g., test anxiety, expectancy-value costs), future field research could benefit from utilizing psychological measures that were specifically validated using field data or informed by focus groups consisting of the populations of interest (Yeager et al., 2016).

Wild Solve Experiences

As previously mentioned, data collected from students who solved in the wild showed that students have a diverse array of solve experiences. Some students reported experiencing a spontaneous insight, where the solution suddenly came to them without consciously thinking about the problem beforehand. Others reported struggling with the problems outside of the lab, and solving either through conscious effort, or from hints in their environment (e.g. seeing an exponent sign). Some of the literature still views insight moments as synonymous with the sudden realization of a solution (Chein & Weisberg, 2014) or assumes that participants' *aha!* moments reflect specifically insight moment and not analytical problem-solving (e.g. Jung-Beeman et al., 2004; Kounios & Beeman, 2009). This dissertation shows that this assumption might not be entirely accurate, as 89% of students in study 2 reported having an *aha!* moment, but many of these students reported solving through an analytical problem approach (45% of those who reported an *aha!* moment), where they worked on the puzzle outside the lab. This corroborates past work that there is overlap between characteristics of insight experiences and analytical problem-solving experiences, suggesting there is more overlap between the two than previously thought (see Chein, Weisberg, Streeter, & Kwok, 2010). My findings reveal that, given an interesting challenge to overcome, students will take conscious action to solve the problem outside of the laboratory even when unprompted. This finding is especially interesting in the context of everyday life, as it suggests that sometimes students do not wait for insight to strike – but rather try to bring themselves to a solution through conscious work.

My findings align better with work that proposes that *aha!* moments can happen both when solutions suddenly come to mind, but also when someone discovers a new direction of thinking about the problem (Csikszentmihalyi & Sawyer, 2014; Fleck & Weisberg, 2013). For example, some students reported thinking of PEMDAS operations, which led them to think of exponents. Their *aha!* moment may have been to think outside of addition, rather than to suddenly come up with the idea of 3^2. This aligns with more recent work done in lab experiments using RAT items that questions the assumption that solutions always "pop" into students' head as an insight moment. Rather, insights could be the result of a series of different strategies that eventually lead to the correct solution (Davelaar, 2015). Collectively, this dissertation contributes evidence to the accumulating idea that insight moments can reflect an array of different experiences, and there is not a "one-insight-fits-all" characterization.

Opportunistic Assimilation

In study 1, participants assigned to the control condition showed stronger solve rates in the wild compared to those assigned to incubation conditions. The opportunistic assimilation hypothesis may explain this finding, as it posits that insight moments arise when individuals exhaust ideas to solve a problem and reach an impasse, which then sensitizes them to helpful hint in their environment that they later encounter (Seifert, et al., 1994). It could have been that control participants were more likely to reach impasse because they were not interrupted with an incubation break halfway through their lab solve attempt, and that this in turn, sensitized them to hints in the environment that helped them solve. We saw that environmental hints helped some participants solve the math puzzle in both studies. For example, in study 1, a participant reported attending a lecture about "Form and Power" which sparked an *aha!* moment for them that taking the power of a number was the solution to the problem.

Much of the literature that focuses on the role of hints on insight moments has found that they generally do not help. For example, Smith et al., (2012) provided students with hints to RAT Items during incubation breaks, and found that solve rates were not significantly different between the items that students did and did not receive hints for. This null finding is difficult to interpret though, as students were given multiple problems to solve in a single session, which may make sensitization to a hint difficult (as students have a limitation on the number of "open goals" they can process; Moss et al., 2007). In addition, students were only allowed 10 seconds to read each RAT item and attempt to solve. Though this has been commonly accepted as a standard timing for a RAT item, it does not allow students the opportunity to immerse in the problem and reach an impasse (Moss et al., 2011), which is a fundamental aspect of the opportunistic assimilation hypothesis.

With student attention divided between numerous problems, and without enough time to struggle with these problems, it is difficult to evaluate the effectiveness of hints and make conclusions about the opportunistic assimilation hypothesis from much of the extant literature. The current dissertation circumvents both of these issues and provides an earnest attempt to investigate the opportunistic assimilation hypothesis by providing students the opportunity to reach impasse, allowing them to go back into their lives and experience the diverse environment of everyday life, and identifying whether hints encountered by chance helped students solve.

In study 2, I found that impasses and condition did not predict whether a student reported a hint in the wild, and that only a small portion of students who solved in the wild reported a hint (or provided an environmental description the described a potential hint). Thus, there is not enough evidence to show that the opportunistic assimilation hypothesis is responsible for *all* solving events in the wild. Despite this, a small number of students continued to report that environmental hints had helped them solve. For example, in study 2, some participants reported seeing exponents in the wild, or even reported that seeing a shape reminded them of Pythagorean's theorem, which made them think of exponents.

Ultimately, the findings of this study align with previous lab research findings, such that hints are not significantly related to the odds that a student solves the math puzzle. But they do still seem to play a small role. Perhaps a more interesting line of questions about the opportunistic assimilation hypothesis is not whether hints can explain *all* insight moments or solve experiences, but how common are hint-related insight experiences? What percent of students are aware of such environmental hints? And how do we better measure hints in the environment without relying on self-reports?

Limitations

The work reported in this dissertation represents an important first step toward bridging the literature between creative problem-solving, the field, and mathematics research. Yet, there are some limitations that need to be highlighted. The first is that the two current studies used a math puzzle to approximate students' ability to reason creatively in mathematics. Although an authentic math problem (e.g. functions, proofs) would better speak to the authentic nature of creative problem-solving in math, the ability to solve such problems would be undoubtedly influenced by students' prior math knowledge. Using a more difficult math problem would have inadvertently rewarded students who had higher math knowledge, and the goal for this study was to study how the everyday students came to insights in math, not just students who have taken many math courses. In addition, a more advanced math problem could require a solution that was not within all students' knowledge base. Thus, to avoid problems with students lacking awareness or understanding of mathematical concepts needed for solving, the current math puzzle was used.

Another limitation of this study is that I relied on students' self-reported experiences to characterize solving in the wild, which is subject to bias and human error. For example, to evaluate the potential aid of environmental hints in the wild, students could only report the presence of a hint if they were aware of a hint, or aware enough to report them in the description of their environment. Likewise, their reports of solve experiences in the wild may not be entirely accurate, even though most participants took the wild follow-up survey immediately after finding the solution. Without invasive data collection methods (e.g. visual and audio recordings), it is difficult to estimate how accurate their accounts are.

A third limitation of these studies was that the two samples consisted entirely of UCLA undergraduate students. While the population of interest for this dissertation are students, undergraduate students at UCLA might not encompass the full range of trait and state measures compared to students at a community college, CSU, or even in K-12. For example, students who rate themselves as math anxious at UCLA compared to those who rate themselves as math anxious in community college might represent fundamentally different levels of underlying math anxiety. As UCLA students must meet strict criteria in order to earn admission, the results of this dissertation do not necessarily generalize to all students.

Future Directions

The findings of this study inform a host of future directions. For example, future research should consider administering math problems or concepts that align with students' current knowledge, requiring them to struggle in their zone of proximal development to capture insight moments. This work would align with the interest of educators who strive to identify how students finally come to understand and solve problems right at the edge of their understanding. Using more authentic problems could also provide more nuanced context as to when students have insights about concepts in different areas of mathematics, such as algebra, calculus, or statistics. This information could then be used to bolster mathematical pedagogy in these different domains of mathematics.

Relatedly, researchers can utilize technology in classrooms to capture insight experiences when they happen, such as by probing students using online textbooks or even using screening technology such as facial recognition software to identify the exact moment students have an insight moment. This would not only help us to study insight moments exactly when they occur, but knowing when these insights happen can help educators identify content in the class that helps students to make connections. Along these same lines, educators might be able to identify long stretches of time when students *are not* having these moments. When identified, educators can either revise materials to help students make more explicit connections to promote understanding, or even provide struggle opportunities and hints to help trigger insight moments for students.

Another interesting follow-up to this line of research would be to study how reaching insight moments in mathematics affects subsequent engagement and persistence of students. *Aha!* moments represent positive experiences, and are often the reward of hard work on the part of the student. Thus, they may likely help motivate students to stay engaged and persist even when an answer or solution does not feel close. Despite previous findings that *aha!* moments sometimes constitute the *only* positive experiences in math classrooms (Liljedahl, 2004), we know nothing about the critical role they may play in students' experiences in mathematics classrooms. Future work can remedy this by studying insight moments as a process of interest that goes on to affect subsequent outcomes educators care about, rather than an outcome in and of itself.

Conclusion

Understanding the methods by which we promote creative thinking in mathematics is an important mission for psychologists who wish to create a bridge between creativity and education. This dissertation highlights the complex nature of creative problem-solving in mathematics, and how different aspects of data collection (lab and wild) can contribute to a richer understanding of students' creative cognition. Embedding math creativity research within ecological context will allow us to make important contributions that will translate well in the various ways in which students reason about mathematics.

References

Altmejd, A., Dreber, A., Forsell, E., Huber, J., Imai, T., Johannesson, M., ... & Camerer, C. (2019). Predicting the replicability of social science lab experiments. *PloS one, 14*(12). https://doi.org/10.1371/journal.pone.0225826

Barnes, M. (2000). 'Magical' moments in mathematics: Insights into the process of coming to know. *For the Learning of Mathematics, 20*(1), 33-43.

Beaty, R. E., & Silvia, P. J. (2012). Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks. *Psychology of Aesthetics, Creativity, and the Arts, 6*(4), 309. doi: 10.1037/a0029171

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. doi: 10.1080/10400410802391314

Benedek, M., & Jauk, E. (2018). *22 Spontaneous and Controlled Processes.* The Oxford Handbook of Spontaneous Thought: Mind-Wandering, Creativity, and Dreaming, 285. doi: 10.1093/oxfordhb/9780190464745.013.22

Berlyne, D. E. (1954). A theory of human curiosity. *British Journal of Psychology. General Section*, *45*(3), 180-191. https://doi.org/10.1111/j.2044-8295.1954.tb01243.x

Bidjerano, T., & Dai, D. Y. (2007). The relationship between the big-five model of personality and self-regulated learning strategies. *Learning and Individual Differences, 17*(1), 69-81. https://doi.org/10.1016/j.lindif.2007.02.001

Bos, M. W., Dijksterhuis, A., & Van Baaren, R. B. (2008). On the goal-dependency of unconscious thought. *Journal of Experimental Social Psychology, 44*(4), 1114-1120. doi: 10.1016/j.jesp.2008.01.001

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network. *Annals of the New York Academy of Sciences*, *1124*(1), 1-38. doi: 10.1196/annals.1440.011

Burton, L. (1998). The practices of mathematicians: What do they tell us about coming to know mathematics?. *Educational Studies in Mathematics, 37*(2), 121. doi: 10.1023/A:1003697329618

Cacioppo, J. T., Petty, R. E., & Kao, C. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*(3), 306-307. doi: 10.1207/s15327752jpa4803_13

Cahan, D. (1995). *Hermann von Helmholtz and the Foundations of Nineteenth-Century Science.*

Chein, J. M., & Weisberg, R. W. (2014). Working memory and insight in verbal problems: Analysis of compound remote associates. *Memory & cognition*, *42*(1), 67-83. https://doi.org/10.3758/s13421-013-0343-4

Chein, J. M., Weisberg, R. W., Streeter, N. L., & Kwok, S. (2010). Working memory and insight in the nine-dot problem. *Memory & Cognition*, *38*(7), 883-892. https://doi.org/10.3758/MC.38.7.883

Craft, A., Gardner, H., & Claxton, G. (Eds.). (2007). *Creativity, wisdom, and trusteeship: Exploring the role of education*. Corwin Press. London, England: University of California Press. ISBN: 9780520083349

Csikszentmihalyi, M., & Sawyer, K. (2014). *Creative insight: The social dimension of a solitary moment.* In The systems model of creativity (pp. 73-98). Springer, Dordrecht. doi: 10.1007/978-94-017-9085-7_7

Daker, R. J., Cortes, R. A., Lyons, I. M., & Green, A. E. (2019). Creativity anxiety: Evidence for anxiety that is specific to creative thinking, from STEM to the arts. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000630

Davelaar, E. J. (2015). Semantic search in the remote associates test. *Topics in Cognitive Science, 7*(3), 494-512. https://doi.org/10.1111/tops.12146

DeCaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *Journal of Experimental Child Psychology, 113*(4), 552-568. doi: 10.1016/j.jecp.2012.06.009

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93*(5), 880. doi: 10.1037/0022-3514.93.5.880

Dijksterhuis, A., & Meurs, T. (2006). Where creativity resides: The generative power of unconscious thought. *Consciousness and Cognition, 15*(1), 135-146. doi: 10.1016/j.concog.2005.04.007

Dodds, R. A., Smith, S. M., & Ward, T. B. (2002). The use of environmental clues during incubation. *Creativity Research Journal, 14*(3-4), 287-304. doi: 10.1207/S15326934CRJ1434_1

Dorfman, J., Shames, V. A., & Kihlstrom, J. F. (1996). Intuition, incubation, and insight: Implicit cognition in problem solving. *Implicit Cognition*, 257-296. doi: 10.1093/acprof:oso/9780198523109.003.0007

Dunbar, K. (2001a). What scientific thinking reveals about the nature of cognition. *Designing for science: Implications from everyday, classroom, and professional settings*, 115-140.

Dunbar, K. (2001b). The analogical paradox: Why analogy is so easy in naturalistic settings yet so difficult in the psychological laboratory. *The analogical mind: Perspectives from cognitive science*, 313-334.

Elliot, A. J., & McGregor, H. A. (1999). Test anxiety and the hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and social Psychology*, *76*(4), 628. https://doi.org/10.1037/0022-3514.76.4.628

Ervynck, G. (2002). Mathematical creativity. In *Advanced mathematical thinking* (pp. 42-53). Springer, Dordrecht. https://doi.org/10.1007/0-306-47203-1_3

Falk, A., & Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science, 326*(5952), 535-538. https://doi.org/10.1126/science.1168244

Feldman, D. H., Csikszentmihalyi, M., & Gardner, H. (1994). *Changing the world: A framework for the study of creativity.* Westport, CT, US: Praeger Publishers/Greenwood Publishing Group.

Fleck, J. I., & Weisberg, R. W. (2013). Insight versus analysis: Evidence for diverse methods in problem solving. *Journal of Cognitive Psychology, 25*, 436–463. doi: 10.1080/20445911.2013.779248

Fries, L., DeCaro, M. S., & Ramirez, G. (2019). The lure of seductive details during lecture learning. *Journal of Educational Psychology*, *111*(4), 736. https://doi.org/10.1037/edu0000301

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. https://doi.org/10.1177/0956797618820626

Gick, M. L., & Lockhart, R. S. (1995). Cognitive and affective components of insight. In Sternberg, R.J., Davidson, J.E. (Eds.), *The Nature of insight.* (pp. 197-228). Cambridge, MA: MIT Press.

Gilhooly, K. J. (2016). Incubation and intuition in creative problem solving. *Frontiers in Psychology, 7*, 1076. doi: 10.3389/fpsyg.2016.01076

Gilhooly, K. J., Georgiou, G., & Devery, U. (2013). Incubation and creativity: Do something different. *Thinking & Reasoning, 19*(2), 137-149. doi: 10.1080/13546783.2012.749812

Hadamard, J. (1945). *The mathematician's mind*. Princeton: Princeton University Press.

Hardy, G. H. (1946). The Psychology of Invention in the Mathematical Field. doi: 10.2307/3608500

Hatano, G., & Inagaki, K. (1984). Two courses of expertise. *Research and Clinical Center for Child Development Annual Report, 6*, 27-36.

Haylock, D. W. (1987). A framework for assessing mathematical creativity in school children. *Educational Studies in Mathematics, 18*(1), 59-74. doi:10.1007/BF00367914

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. *Second handbook of research on mathematics teaching and learning,* 1, 371-404.

Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. K. (2003). The abbreviated math anxiety scale (AMAS) construction, validity, and reliability. *Assessment, 10*(2), 178-182. doi: 10.1177/1073191103010002008

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, *2*(1), 88-110. https://doi.org/10.1080/19345740802539325

Jarosz, A. F., Colflesh, G. J., & Wiley, J. (2012). Uncorking the muse: Alcohol intoxication facilitates creative problem solving. *Consciousness and Cognition, 21*(1), 487-493. https://doi.org/10.1016/j.concog.2012.01.002

Johnson, J. A. (1994). Clarification of factor five with the help of the AB5C model. *European Journal of Personality, 8*(4), 311-334. doi: 10.1002/per.2410080408

Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R. et al. (2004). Neural activity when people solve verbal problems with insight. Public Library of Science Biology, 4, 1 – 23. https://doi.org/10.1371/journal.pbio.0020097

Kapur, M. (2014). Productive failure in learning math. *Cognitive Science, 38*(5), 1008-1022. https://doi.org/10.1111/cogs.12107

Kaufman, S. B., & Gregoire, C. (2016). *Wired to create: Unraveling the mysteries of the creative mind.* Penguin.

Kaufman, S. B., Quilty, L. C., Grazioplene, R. G., Hirsh, J. B., Gray, J. R., Peterson, J. B., & DeYoung, C. G. (2016). Openness to experience and intellect differentially predict creative achievement in the arts and sciences. *Journal of personality, 84*(2), 248-258. doi: 10.1111/jopy.12156

Kounios, J., & Beeman, M. (2009). The Aha! moment: The cognitive neuroscience of insight. *Current Directions in Psychological Science, 18*(4), 210-216.

Kounios, J., & Beeman, M. (2015). *The eureka factor: Creative insights and the brain*. Random House.

Lawson, A. E. (2001). Promoting creative and critical thinking skills in college biology. *Bioscene, 27*(1), 13-24.

Liljedahl, P. (2005). Mathematical discovery and affect: The effect of AHA! experiences on undergraduate mathematics students. *International Journal of Mathematical Education in Science and Technology, 36*(2-3), 219-236. doi: 10.1080/00207390412331316997

Liljedahl, P. G. (2004). The AHA! experience: Mathematical contexts, pedagogical implications (Doctoral dissertation, Theses (Faculty of Education)/Simon Fraser University).

Liljedahl, P., & Sriraman, B. (2006). Musings on mathematical creativity. *For The Learning of Mathematics, 26*(1), 17-19.

Litman, J. A. (2008). Interest and deprivation factors of epistemic curiosity. *Personality and Individual Differences, 44*(7), 1585-1595. https://doi.org/10.1016/j.paid.2008.01.014

Litman, J. A. (2010). Relationships between measures of I- and D-type curiosity, ambiguity tolerance, and need for closure: an initial test of the wanting-liking model of information seeking. *Personality and Individual Differences, 48,* 397–410. doi:10.1016/j.paid.2009.11.005.

Mackenzie, D. (2006). The Poincaré conjecture-proved. *Science, 314*(5807), p. 1848-1849. doi: 10.1126/science.314.5807.1848. doi: 10.1126/science.314.5807.1848

Mann, E. L. (2006). Creativity: The essence of mathematics. *Journal for the Education of the Gifted, 30*(2), 236-260.10.4219/jeg-2006-264

McCrae, R. R., & Costa, P. T. (1997). Personality trait structure as a human universal. *American Psychologist, 52*(5), 509. doi: 10.1037/0003-066X.52.5.509

Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and noninsight problem solving. Memory & Cognition, 15, 238-246. doi: 10.3758/BF03197722

Miller, S. (n.d.). Sumthing's Up. Retrieved June 06, 2019, from https://mathriddles.williams.edu/?p=528#comments

Moss, J., Kotovsky, K., & Cagan, J. (2007). The influence of open goals on the acquisition of problem-relevant information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(5), 876. doi: 10.1037/0278-7393.33.5.876

Moss, J., Kotovsky, K., & Cagan, J. (2011). The effect of incidental hints when problems are suspended before, during, or after an impasse. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(1), 140. doi: 10.1037/a0021206

Mumford, M. D., Lonergan, D. C., & Scott, G. (2002). Evaluating creative ideas: Processes, standards, and context. Inquiry: *Critical Thinking Across the Disciplines, 22*(1), 21-30. doi: 10.5840/inquiryctnews20022213

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251). https://doi.org/10.1126/science.aac4716

Ovington, L. A., Saliba, A. J., Moran, C. C., Goldring, J., & MacDonald, J. B. (2018). Do people really have insights in the shower? The when, where and who of the Aha! Moment. *The Journal of Creative Behavior, 52*(1), 21-34. https://doi.org/10.1002/jocb.126

Poincaré, H. (1946). *The foundations of science.* Lancaster, PA: The Science Press.

Ramirez, G. (2017). Motivated forgetting in early mathematics: A proof-of-concept study. *Frontiers in Psychology, 8,* 2087. doi: 10.3389/fpsyg.2017.02087

Ramirez, G., McDonough, I. M., & Jin, L. (2017). Classroom stress promotes motivated forgetting of mathematics knowledge. *Journal of Educational Psychology, 109*(6), 812. doi: 10.1037/edu0000170

Ritter, S. M., & Dijksterhuis, A. (2014). Creativity-the unconscious foundations of the incubation period. *Frontiers in Human Neuroscience, 8*. doi: 10.3389/fnhum.2014.00215

Rodriguez, B. K. N. (2019). Math: The girl that all the nerds want (Order No. 13899222). Available from Dissertations & Theses @ University of California; ProQuest Dissertations & Theses A&I; ProQuest Dissertations & Theses Global. (2247139447). Retrieved from https://search.proquest.com/docview/2247139447?accountid=14512

Runco, M. A. (2014). *Creativity: Theories and themes: Research, development, and practice*. Elsevier.

Sadler-Smith, E. (2015). Wallas' four-stage model of the creative process: More than meets the eye?. *Creativity Research Journal, 27*(4), 342-352. doi: 10.1080/10400419.2015.1087277

Saucier, G. (1992). Openness versus intellect: Much ado about nothing?. *European Journal of Personality, 6*(5), 381-386. doi: 10.1002/per.2410060506

Savic, M. (2012). *Proof and proving: Logic, impasses, and the relationship to problem solving*. New Mexico State University.

Savic, M. (2015). The incubation effect: How mathematicians recover from proving impasses. *The Journal of Mathematical Behavior, 39*, 67-78. doi: 10.1016/j.jmathb.2015.06.001

Savic, M. (2016). Mathematical Problem-Solving via Wallas' Four Stages of Creativity: Implications for the Undergraduate Classroom. *The Mathematics Enthusiast, 13*(3), 255-278.

Schooler, J. W., & Melcher, J. (1995). The ineffability of insight. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 97-133). Cambridge, MA, US: The MIT Press.

Segal, E. (2004). Incubation in insight problem solving. *Creativity Research Journal, 16*(1), 141-148. doi: 10.1207/s15326934crj1601_13

Seifert, C. M., Meyer, D. E., Davidson, N., Patalano, A. L., & Yaniv, I. (1994). Demystification of cognitive insight: Opportunistic assimilation and the prepared-mind hypothesis.

Shen, W., Yuan, Y., Liu, C., & Luo, J. (2016). In search of the "Aha!" experience: Elucidating the emotionality of insight problem-solving. *British Journal of Psychology, 107*(2), 281-298. doi: 10.1111/bjop.12142

Silvia, P. J. (2008). Discernment and creativity: How well can people identify their most creative ideas?. *Psychology of Aesthetics, Creativity, and the Arts, 2*(3), 139. doi: 10.1037/1931-3896.2.3.139

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2019). P-curve won't do your laundry, but it will distinguish replicable from non-replicable findings in observational research: Comment on Bruns & Ioannidis (2016). *PloS One, 14*(3). https://doi.org/10.1371/journal.pone.0213454

Simonton, D. K. (2018). Creative ideas and the creative process: Good news and bad news for the neuroscience of creativity. *The Cambridge handbook of the neuroscience of creativity,* 9-18. doi: 10.1017/9781316556238.002

Sio, U. N., & Ormerod, T. C. (2009). Does incubation enhance problem solving? A meta-analytic review. *Psychological Bulletin, 135*(1), 94. doi: 10.1037/a0014212

Sio, U. N., & Rudowicz, E. (2007). The role of an incubation period in creative problem solving. *Creativity Research Journal, 19*(2-3), 307-318. doi: 10.1080/10400410701397453

Smith, S. M. (1995). Fixation, incubation, and insight in memory and creative thinking. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 135-146). Cambridge, MA: MIT Press.

Smith, S. M., & Blankenship, S. E. (1989). Incubation effects. *Bulletin of the Psychonomic Society, 27*(4), 311-314. doi: 10.3758/BF03334612

Smith, S. M., & Blankenship, S. E. (1991). Incubation and the persistence of fixation in problem solving. *The American Journal of Psychology*, 61-87. doi: 10.2307/1422851

Smith, S. M., Sifonis, C. M., & Angello, G. (2012). Clue insensitivity in remote associates test problem solving. *The Journal of Problem Solving, 4*(2), 3. doi: 10.7771/1932-6246.1124

Sundararajan, N., & Adesope, O. (2020). Keep it coherent: A meta-analysis of the seductive details effect. *Educational Psychology Review*. https://doi.org/10.1007/s10648-020-09522-4

Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of Cross-Cultural Psychology, 38(*2), 227-242. doi: 10.1177/0022022106297301

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making, 11*(1), 99. doi: 10.1037/t49856-000

Treffinger, D. J., Isaksen, S. G., & Stead-Dorval, K. B. (2005). *Creative problem solving: An introduction*. Prufrock Press Inc.

Wallas, G. (1926). *The art of thought*. Kent, England: SOLIS Press. ISBN: 978-1-91046-05-7

Ward, T. B., Smith, S. M., & Finke, R. A. (1999). Creative cognition. *Handbook of creativity,* 189, 212. doi: 10.1017/CBO9780511807916.012

Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., ... & Trott, J. (2016). Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of Educational Psychology, 108*(3), 374. https://doi.org/10.1037/edu0000098

Zeigarnik, B. (1938). On finished and unfinished tasks. A source book of Gestalt psychology, 1, 300-314. doi: 10.1037/11496-025

Zhong, C. B., Dijksterhuis, A., & Galinsky, A. D. (2008). The merits of unconscious thought in creativity. *Psychological Science, 19*(9), 912-918. doi: 10.1111/j.1467-9280.2008.02176.x