Inverkan av olika verbaliseringsprotokoll på insiktsproblemlösning inom spatiala och verbala domäner

The impact of different verbalization protocols on insight problem solving in spatial and verbal domains

Valery Nkenguruke

Handledare/supervisor
Andreas Lind

KOGM21

2025-06-05

Masteruppsats i kognitionsvetenskap
Avdelningen för kognitionsvetenskap
Filosofiska institutionen
Lunds universitet

Master's thesis in cognitive science
Cognitive Science department
Department of Philosophy
Lund University

The impact of different verbalization protocols on insight problem solving in spatial and verbal domains

Valery Nkenguruke
va2518nk-s@student.lu.se

This study investigated how three verbalization protocols, Silent, Type 1 (think-aloud), and Type 3 (explanatory verbalization), influence problem-solving performance on insight and non-insight problems within both spatial and verbal domains. Twenty-eight participants completed twelve problems (three per category) under each condition. Informed by theories of insight, verbal overshadowing, and cognitive load in working memory, the study assessed solution rates and log-transformed solution times. Verbalization did not significantly alter solution rates for any problem type. However, it imposed a clear time cost, significantly increasing solution times for spatial problems, an effect most pronounced under the cognitively demanding explanatory (Type 3) condition. Given the study's sample size and specific design constraints, these findings should be considered preliminary. This study highlights the need for further research to elucidate how various verbalization types interact with the cognitive and representational demands inherent in problem-solving.

1 Introduction

To be a cognitive agent is to be a problem solver, and perhaps the most characteristic feature of humans is the capacity to act as general problem solvers (Newell & Simon, 1972; Holyoak, 1995). Alongside this problem-solving capacity stands language as another defining feature of human cognition. This dual view is echoed in the seminal work of Newell and Simon (1972), who, within the information-processing framework, conceptualized problem-solving as a systematic navigation through a problem space, moving from an Initial State (IS), the starting point, to a Goal State (GS) via a series of operators (any cognitive or physical action that transforms the current state of a problem into a new state). Language functions as one such operator, enabling the transformation and structuring of internal representations during problem-solving (Newell & Simon, 1972; Vygotsky, 1962; Clark, 1998).

Within this problem space framework, problems can be either well-defined or ill-defined. In a well-defined problem, all elements of the problem space, the IS, GS, and (allowable) operators, are explicitly specified. The task is then a matter of applying the operators and traversing the problem space. In an ill-defined problem one or more of these elements is unclear or undefined. In this case the first cognitive challenge shifts from navigating a known path to first establishing an appropriate representation of the problem, finding the right frame or perspective of the problem itself (Kaplan & Simon, 1990). The moment this new perspective clicks into place, often unexpectedly, often after repeated mis-framings, and often accompanied by a sense of sudden clarity, one experiences what the Gestalt psychologists called an insight (e.g., Köhler, 1925; Wertheimer, 1945/1959; Koffka, 1935; for a comprehensive review, see Sternberg & Davidson, 1995).

This Gestalt notion of insight describes a process of sudden, non-sequential restructuring, one that operates differently from deliberate, step-by-step analysis (Ohlsson, 1992; Kounios & Beeman, 2014). Its abrupt and unforced nature stands in stark contrast to the deliberate, sequential, and rule-governed structure of language. This cognitive contrast between insight, a pivotal mechanism of problem-solving, and language, arguably the most influential cognitive operator, affords the present investigation its guiding question: What is the impact of actively engaging in linguistic processes, such as verbalizing one's thoughts, while solving (ill-defined) problems that require insight? If the cognitive mechanisms underlying insight are inherently non-verbal, it stands to reason that requiring verbalization could interfere with or disrupt these mechanisms.

Evidence for such disruption comes from the well-documented Verbal Overshadowing Effect, the finding that verbalizing certain kinds of experiences can impair subsequent cognitive performance. This effect was first demonstrated in the domain of face recognition: participants who verbally described a previously seen face were later less accurate in identifying it, compared to those who remained silent (Schooler & Engstler-Schooler, 1990). The phenomenon suggests that putting non-verbal (perceptual in this case) experiences into words can disrupt the underlying cognitive processes involved in encoding or retrieving them (Schooler et al., 1993; Chin & Schooler, 2008).

The possibility of verbal overshadowing becomes particularly relevant in the context of problem-solving, especially when viewed through the distinction between well-defined problems, often referred to as analytical problems, and ill-defined problems, which often take the form of so-called insight problems: problems specifically designed to evoke the kind of restructuring characteristic of insight (e.g., riddles). Analytical problems are structured so that their solutions can be reached through a sequence of logical, step-by-step operations. These steps are typically easily verbalizable and align well with the linear structure of language. Insight problems, by contrast, often depend on sudden representational shifts or non-reportable mental operations, suggesting a reliance on less verbal, more Gestalt-like processing (Ohlsson, 1992; Schooler et al., 1993).

Moreover, although solving insight problems, in general, involves overcoming an initial inappropriate framing through restructuring, the nature of this process and the cognitive systems involved may differ substantially across problem domains. Spatial insight problems, like the Nine-Dot puzzle (connect nine dots in a 3x3 grid with four straight lines without lifting the pencil), typically require overcoming perceptual biases or assumptions about spatial boundaries, likely engaging primarily visuospatial processing systems (Ash & Wiley, 2006; Chronicle et al., 2004). Verbal insight problems, such as riddles, conversely, often hinge on reinterpretations of meaning or the activation of less obvious semantic associations, presumably relying more heavily on linguistic and semantic networks (Bowden & Jung-Beeman, 2003; Kounios & Beeman, 2014). While both problem types share the need for restructuring away from an initial framing, these distinct representational formats (visuospatial vs. linguistic/semantic) and potentially different underlying cognitive systems raise the critical possibility that spatial and verbal insight might exhibit differential susceptibility to external factors that influence ongoing processing, such as the processing shifts induced by concurrently verbalizing thoughts during problem-solving attempts.

This raises key questions for this study: Does the act of verbalizing interfere more with insight problem-solving, which may depend on non-verbal restructuring, than with analytical problem-solving, which might align better with use of deliberate, rule application and step-by-step reasoning? And how does this further differ when considering different domains, i.e. spatial vs verbal? Additionally, not all thinking aloud is the same, and how verbalization instructions are framed likely matters. Foundational work by Ericsson and Simon (1993) highlighted that simply verbalizing thoughts already in verbal form might impact cognition differently than explaining one's reasoning or translating non-verbal thoughts into words. This distinction between think-aloud protocols and more demanding explanatory verbalization protocols, which differ in their potential to influence ongoing thought processes, is central to the current investigation.

Research into how verbalization affects problem-solving has yielded a complex and sometimes contradictory picture. Although concurrent thinking aloud is regarded as a valuable research tool (Ericsson & Simon, 1993), some key studies report that verbalizing can impair performance, particularly for insight problems, perhaps by disrupting non-verbal restructuring processes (e.g., Schooler et al., 1993). Yet, other findings suggest verbalization, especially more explanatory forms, can sometimes facilitate problem-solving, possibly by enhancing metacognitive monitoring or strategy use (e.g., Kiyokawa et al., 2023; Fleck & Weisberg, 2004). This underscores that the impact of thinking out loud on problem solving likely depends significantly on what kind of problem is being solved and how one is asked to talk about it.

Despite this existing research, a gap in the literature remains. While influential studies have examined verbalization effects (e.g., Schooler et al., 1993; Gilhooly et al., 2010), none have systematically compared the effects of different types of concurrent verbalization, specifically, minimally intrusive think-aloud (Type 1) versus more demanding explanatory protocols (Type 3), on performance across both insight and non-insight problems and across both spatial and verbal domains within a single experimental design. Gilhooly et al. (2010), for instance, critiqued Schooler et al. (1993) on two fronts: first, for failing to adequately distinguish between different types of verbalizations (e.g., type 1 vs. type 3) and second, for potentially conflating problem domain (verbal vs. spatial) with insight classification, since many of Schooler's "insight" problems were spatial and most "non-insight" problems were verbal. Although Gilhooly et al. attempted to address these confounds by systematically crossing insight/non-insight with verbal/spatial domains, their study applied only a single verbalization condition, namely Type 1 (direct concurrent think-aloud).

Although Gilhooly's critique of Schooler highlights important methodological limitations such as insufficient differentiation between variables, their own design fell short by not testing whether different types of verbalizations, particularly more metacognitive or explanatory forms (Type 3), might interact differently with problem type and domain. This leaves open key empirical questions: Does the cognitive impact of verbalization on problem solving depend on the kind of verbalization used, and how does this interact with both the nature (insight vs. non-insight) and the representational domain (spatial vs. verbal) of the problem? The present study aims to directly address this gap through a factorial experimental design that contrasts two distinct verbalization protocols (Type 1 vs. Type 3) against a silent control, across both insight and non-insight tasks, and within both spatial and verbal domains.

The structure of this thesis is as follows: Chapter 2 reviews the relevant background literature. Chapter 3 details the experimental methodology. Chapter 4 presents the statistical results. Finally, Chapter 5 discusses the findings in relation to the research questions and existing literature, considers broader implications, acknowledges limitations, and suggests directions for future research.

2 Theoretical Background and Literature Review

This chapter reviews the theoretical foundations and empirical literature informing the study's central question: how verbalization, the primary mechanisms for explicit, analytical cognition, interacts with insight, a process often marked by sudden, non-verbal restructuring. The review begins by defining insight and tracing its historical and cognitive foundations, before moving into theoretical accounts of problem-solving, the role of working memory, and the mechanisms of verbalization. Emphasis is placed on how different types of verbalizations may either support or impede insight, depending on the problem domain and the cognitive processes involved. The chapter concludes by formally stating the research questions and hypotheses that arise from this synthesis.

Defining Insight

Insight represents a distinct and intriguing mechanism of problem solving that has captivated researchers since the early days of psychology. While often contrasted with more incremental, analytical approaches, achieving a precise definition of insight has proven challenging, leading to ongoing debate (Weisberg, 1995). Nevertheless, a core set of characteristics generally associated with insight can be identified, drawing both from historical accounts and contemporary research.

The systematic investigation of insight originated largely within the Gestalt school of psychology. Researchers such as Köhler (1925), Wertheimer (1945/1982), and Duncker (1945) argued against purely associationist or behaviorist accounts of problem solving (e.g., Thorndike, 1911), which emphasized trial-and-error learning. Instead, they proposed that solutions to certain problems arise not through gradual accumulation of knowledge or stimulus-response pairings, but through a process of restructuring the perceptual or mental representation of the problem situation. This restructuring involves perceiving the problem elements and their interrelations in a fundamentally new way, overcoming an initial misleading perception or fixation that blocks progress (Ohlsson, 1984; Wertheimer, 1982). Such restructuring often occurs after an impasse is reached, a point where previous attempts based on the initial representation have repeatedly failed (Ohlsson, 1992; Smith, 1995). Köhler's (1925) seminal observations of chimpanzees exhibiting sudden, apparently intelligent solutions to obtain out-of-reach rewards (e.g., using sticks, stacking boxes) after periods of unsuccessful attempts, and Duncker's (1945) work demonstrating functional fixedness (e.g., the difficulty in perceiving a tack box as a potential shelf in the candle problem), provided early empirical grounding for these concepts.

The moment of restructuring is often accompanied by a subjective experience of suddenness and surprise, commonly termed the "Aha!" experience or illumination (the idea of a light bulb going on) (Kaplan & Simon, 1990; Kounios & Beeman, 2009; Sternberg & Davidson, 1995). This feeling contrasts sharply with the experience of solving analytical problems, where progress typically feels incremental and predictable (Metcalfe & Wiebe, 1987). Metcalfe and Wiebe provided empirical support for this distinction using "feeling-of-warmth" ratings, finding that participants solving non-insight problems showed gradually increasing confidence as they neared the solution, whereas those solving insight problems often reported feeling "cold" (far from solution) until immediately before solving, indicating a lack of conscious awareness of incremental progress. This sudden transition from incomprehension to comprehension is a hallmark characteristic frequently used to identify insight (Davidson, 1995; Sternberg & Davidson, 1995). However, the "Aha!" experience itself is not synonymous with the underlying cognitive process of restructuring; rather it is the process's subjective correlate, and its presence or intensity may vary (Weisberg, 1995; Webb, Little, & Cropper, 2016).

Insight seen as Restructuring and Re-framing in the Context of Ill-Defined Problems

The process of restructuring, identified by Gestalt psychologists as central to insight, takes on particular importance when individuals confront ill-defined problems. As mentioned in Chapter 1, such problems, where initial states, goals, or operators are ambiguous (Simon, 1973), often require more than a linear application of known procedures. Solvers rarely frame these problems optimally from the outset. Instead, they must engage in significant cognitive work to (re)formulate or (re)frame their understanding to create a tractable problem space (Kaplan & Simon, 1990). In this context, insight is the cognitive process central to this act of re-framing. It involves a fundamental reconfiguration of the problem's representation, a perspectival shift that brings different elements to the foreground and affords a new understanding (much like perceiving the alternative image in the Necker cube or Rubin vase).

Although reconfiguration can occur at any level of the problem space i.e. IS, GS, intermediary state and operators, the way a solver initially frames the IS is often the most consequential step, as it defines the structure of the problem space and dictates which solution paths are perceived as accessible. Mis-framing, particularly of the IS, typically leads to the misidentification of intermediate states and appropriate operators, often culminating in a frustrating state of impasse where progress halts (Ohlsson, 1992; Duncker, 1945). Consider. for example, the riddle: "A hole is 6 meters long, 2 meters wide, and 1 meter deep. How many cubic meters of soil are in it?" Many solvers initially approach this as a simple arithmetic problem thereby mis-framing it and thus failing to recognize it as a semantic puzzle: a hole, by definition, contains no soil. Similarly, the Nine-Dot puzzle often leads to impasse because solvers implicitly assume that lines must stay within the perceived boundary formed by the dots. Such mis-framing-induced impasses, where the initial representation of the problem state, goal, or operators is misleading, perhaps guided by habit or flawed heuristics, are precisely the situations where insight becomes most critical for achieving a solution.

Insight problems, commonly used in research, can thus be understood as a specific class of problems, often ill-defined or quasi-well-defined (Weisberg, 1995; Kaplan & Simon, 1990), in which the most readily accessible or heuristically suggested initial framing is not just incomplete but actively misleading. This misleading representation often creates a functional fixedness or mental set, blocking access to necessary operators or pathways and leading to the type of impasse previously discussed (Ohlsson, 1992; Weisberg, 1995). Solving an insight problem, therefore, typically requires abandoning this initial unproductive frame and fundamentally restructuring the problem representation to overcome the fixation (Smith, 1995). In this view, insight is about achieving an effective problem representation when the initial one fails (Davidson & Sternberg, 1995).

This process of representational shift critically distinguishes insight from more analytical problem-solving, which generally involves search within a stable and adequate initial problem representation (Newell & Simon, 1972). While the outcome of insight is often a clearer understanding and a seemingly rapid path to solution, the underlying process involves this pivotal, non-obvious, change in perspective (Sternberg & Davidson, 1995). Understanding the factors that trigger or impede this re-framing process is therefore central to a comprehensive understanding of insight.

How do people solve problems: The role of working memory and Interactive Processes

The problem space framework (Newell & Simon, 1972), previously introduced, provides a foundation for understanding how individuals solve problems by navigating from an initial state to a goal state using various operators or actions. This process often involves heuristic search strategies such as means-ends analysis (reducing differences between the current and goal state) or hill-climbing (choosing steps that lead progressively closer to the goal), to find a solution path (Newell & Simon, 1972). The effectiveness of these strategies heavily depends on how the solver initially represents or frames the problem. A well-formed representation allows for the efficient application of known operators, whereas a flawed or incomplete representation can, as previously discussed, lead to an impasse, a state where no further progress seems possible (Ohlsson, 1992; Weisberg, 1995).

Navigating this problem space centrally involves working memory (WM), defined as a limited-capacity system for temporarily holding and manipulating information relevant to the task at hand (Baddeley, 1986). WM plays an important role in maintaining goal information, tracking intermediate states, retrieving relevant knowledge from long-term memory, and executing cognitive operations (Smeekens, 2016; Fleck, 2008). Its efficient functioning is considered essential, not only for analytical thought but also for the complex cognitive achievements involved in insight, such as suppressing misleading information during fixation or manipulating problem elements during representational restructuring (Gilhooly & Fioratou, 2009).

However, emerging research suggests that insight and analytic problem-solving may recruit distinct WM subcomponents. A key study by Fleck (2008) revealed that analytic problem-solving is strongly supported by overall working memory capacity, which indicates central executive involvement in functions like active monitoring, updating, and strategy implementation. It is also supported by verbal working memory, reflecting the typically sequential and propositional nature of analytic tasks. Insight problem-solving, in contrast, appears more dependent on visuospatial sketchpad resources and processes such as representational change, especially when restructuring involves mentally manipulating spatial elements or relaxing inappropriate constraints, though it also relies on verbal Short-Term Memory (STM) capacity. Fleck (2008) interpreted this distinction as supporting the view that restructuring in many insight problems occurs more spontaneously; in this view, STM is primarily needed to maintain the solution once it enters awareness, rather than for an all-encompassing, WM-heavy active search for the new representation itself. Furthermore, the necessity of suppressing misleading cues or default interpretations, a process often central to achieving insight, also places selective demands on the inhibitory control components of working memory.

Extending this distinction, Korovkin et al. (2018) demonstrated that verbal reports of insight do not always align with behavioral performance, thereby highlighting the cognitive load imposed by verbalization itself. Their observations indicated that participants solving insight problems exhibited patterns suggestive of implicit, non-verbal processes driving performance. In contrast, analytic solvers showed greater alignment between verbalization and behavior, consistent with their increased reliance on consciously accessible WM resources. This mismatch between felt insight and the actual cognitive strategy employed suggests that concurrent verbalization tasks, which engage verbal WM and metacognitive control, might interfere more significantly with insight problem-solving, especially when solutions depend on restructuring processes that are not easily verbalized. Collectively, these findings underscore that WM's role in problem-solving is not unitary; task-specific demands and verbalization instructions can differentially tax WM subsystems, thereby shaping performance outcomes.

Fleck's (2008) finding that verbal STM predicted success on her overall set of insight problems was interpreted as supporting a more spontaneous restructuring process. However, she also acknowledged considerable heterogeneity among these problems. This acknowledgment highlights that different insight problems, particularly those varying by domain (spatial vs. verbal; a distinction elaborated later), might engage WM components differently. For instance, while Fleck’s overall result pointed to verbal STM, theoretical accounts and other research (e.g., Korovkin et al., 2018; Gilhooly & Murphy, 2005) emphasize the probable importance of distinct storage systems: visuospatial systems (like the visuospatial sketchpad) are considered crucial for processing and restructuring spatial insight problems, while phonological storage is deemed important for verbal insight problems. Korovkin et al. (2018) did indeed find that insight problem solving loaded more heavily on modal-specific storage systems than on the central executive. Thus, while the central executive component of WM might be less consistently implicated in the core restructuring phase of insight compared to analytic problem solving, the specific storage systems of WM (verbal or visuospatial) are likely crucial depending on the problem's domain and representational demands.

Beyond internal cognitive manipulations, human problem-solving is often an interactive process where individuals actively engage with their environment to facilitate cognition. Kirsh and Maglio (1994) differentiated between pragmatic actions (performed to bring the agent physically closer to a goal) and epistemic actions (performed primarily to uncover information, reduce cognitive load, or simplify the problem by changing its external representation). Examples of epistemic actions include physically rearranging puzzle pieces, making notes, or manipulating objects in a problem (like rotating Tetris pieces); such actions help solvers perceive new relationships, test hypotheses, or offload memory demands. While not directly achieving the goal, these actions modify the problem's external structure, making it more amenable to the solver's cognitive processes.

This interaction with the environment can be seen as a broader strategy where agents modify their surroundings to make the world fit their cognitive capacities and simplify information processing. By externalizing information or physically encoding problem states, individuals create "epistemic structures" in the environment that can reduce the need for complex internal mental simulations or extensive memory search. This interaction between internal cognitive processes and external actions underscores that problem-solving is not purely an "in-the-head" phenomenon but often involves a dynamic interplay between the solver and their environment. Such epistemic engagement can be particularly pivotal when an impasse is reached due to flawed initial representations. By enabling solvers to physically manipulate or externally re-represent problem elements, these interactive strategies can help reveal new affordances or facilitate the relaxation of self-imposed constraints. In doing so, epistemic actions can directly contribute to the fundamental shift in representation, the restructuring, that is characteristic of insight. However, the concurrent demand of verbalizing one's thoughts may interfere with this beneficial function of offloading or simplifying internal cognition, thereby potentially obstructing the representational shifts or restructuring epistemic actions aim to facilitate.

Theoretical Accounts of Insight Mechanisms

While Gestalt psychologists provided foundational concepts like restructuring and impasse, subsequent research, particularly within the information-processing tradition, has sought to specify the cognitive mechanisms underlying these phenomena more precisely. Understanding these mechanisms is often framed by a central, ongoing discussion in insight research: The "Special Process" vs. "Business-as-Usual" Debate (Davidson, 1995; Weisberg, 1995, 2015).

The Special Process View, which inherits the legacy of Gestalt psychology, maintains that insight involves cognitive processes qualitatively distinct from those underlying analytical thought (Davidson, 1995; Metcalfe & Wiebe, 1987). Proponents often emphasize the suddenness and subjective discontinuity of insight, suggesting it may involve non-conscious processing, implicit activation spreading through semantic networks, or holistic processing modes that are distinct from deliberate, step-by-step analysis (Bowden, Jung-Beeman, Fleck, & Kounios, 2005; Kounios & Beeman, 2014). Cognitive neuroscience has provided some supporting evidence, indicating distinct neural correlates for insight solutions; these may involve right-hemisphere processing or anterior cingulate cortex activity linked to conflict detection and attentional shifts (Jung-Beeman et al., 2004; Kounios et al., 2006, 2008; Subramaniam et al., 2009). The subjective feeling of suddenness and the lack of predictive "feeling-of-warmth" (Metcalfe & Wiebe, 1987) are also frequently cited as evidence for processes operating outside conscious analytical control.

The Business-as-Usual (or Nothing Special) View, in contrast, argues that insight does not rely on unique cognitive mechanisms, instead, it emerges from applying ordinary cognitive processes, such as memory retrieval, heuristic search, logical inference, and analogy, to overcome a representational impasse (Weisberg, 1995). Proponents argue that the apparent suddenness can result from reaching a knowledge threshold or successfully executing the final step in a complex reasoning chain, which may have involved considerable prior, albeit perhaps unsuccessful, analytical effort (Weisberg, 1992). Evidence supporting this view often comes from detailed analyses of problem-solving protocols showing incremental steps (Fleck & Weisberg, 2004), demonstrations that specific prior knowledge or hints can facilitate solutions previously attributed to insight (Weisberg & Alba, 1981), and findings linking insight performance to general cognitive abilities like working memory and fluid intelligence (DeYoung, Flanders, & Peterson, 2008; Gilhooly & Fioratou, 2009). This perspective suggests that insight problems are difficult not because they demand special processes, but because they often require accessing less obvious knowledge or overcoming strong but inappropriate heuristics using standard cognitive machinery.

Recognizing the limitations of both extreme positions, many researchers now advocate for integrated theories (Weisberg, 2015). These perspectives acknowledge that while insight solutions may rely on standard cognitive building blocks (e.g., memory, attention, executive functions), the process of overcoming impasse and achieving representational change may emerge from how these processes interact in particular, perhaps characteristic, ways. For example, defocused attention might enable solvers to notice previously overlooked aspects of a problem (Ansburg & Hill, 2003; Schooler & Melcher, 1995), or specific attentional control dynamics could facilitate escaping mental fixation (Chein & Weisberg, 2014). Another proposed mechanism is the interplay between implicit and explicit processes, where unconscious associations influence conscious reasoning. This interaction has been compared to a dialogue between two layers of the mind: while the conscious self struggles with a problem, intuitive, background processes may quietly suggest novel connections, eventually giving rise to the "Aha!" moment (Kounios & Beeman, 2014). These mechanisms, though not necessarily unique to insight, appear to be particularly characteristic of it. Thus, insight may involve both incremental search within an initially flawed representation and a distinct restructuring event triggered when specific cognitive conditions align (Weisberg, 2015; Bowden et al., 2005).

Within this broader context of debate and integration, specific theories have been proposed to detail the mechanisms of insight. One prominent example is Representational Change

Theory (RCT). Building upon the Gestalt notion of restructuring and Newell and Simon's (1972) problem space framework, Ohlsson (1992, 1984b) proposed RCT as a detailed information-processing account. RCT posits that an impasse occurs when the initial problem representation, which is constructed from prior knowledge activated by the problem statement, fails to contain the solution path. The solver becomes stuck because the operators available within this representation are insufficient to reach the goal state. Ohlsson argued that overcoming such an impasse necessitates changing the problem representation itself. He proposed two primary mechanisms for this: Constraint Relaxation and Chunk Decomposition.

Constraint Relaxation involves abandoning inappropriate, self-imposed constraints (often implicit and based on prior knowledge or misinterpretation) that unduly limit the perceived problem space. Relaxing these constraints, such as the assumption that lines must stay within the dot boundary in the Nine-Dot puzzle, can expand the problem space to include solution paths. Chunk Decomposition, on the other hand, refers to breaking down perceptual or conceptual chunks, tightly integrated units of information typically treated as single entities, into their constituent elements. This process allows elements to be reconfigured or used in novel ways, thereby overcoming functional fixedness (e.g., decomposing a “tack box” to perceive the “box” element as a potential shelf).

RCT suggests that these representational changes make new operators applicable or reveal new pathways, allowing the impasse to be broken, often suddenly. The theory provides a computationally plausible mechanism for restructuring, and empirical work manipulating constraints or chunking has supported these mechanisms.

Complementing these frameworks, other perspectives elaborate on specific mechanisms or triggers relevant to insight. Progress Monitoring Theory, for example (MacGregor, Ormerod, & Chronicle, 2001), suggests that a perceived lack of progress towards a goal can trigger the abandonment of current strategies and prompt a search for alternative approaches, potentially leading to restructuring. Similarly, and particularly relevant for verbal insight, Spreading Activation Theories (Bowden & Jung-Beeman, 2003; Kounios & Beeman, 2014) propose that unconscious activation diffusing through semantic networks can enable distant concepts to connect and reach awareness, causing solutions to seemingly “pop out”.

These discussed diverse theoretical accounts collectively underscore the multifaceted nature of insight. A significant challenge remains in dissociating the contributions of these varied cognitive processes and understanding their interplay. The present study, by examining how different types of cognitive load (induced by verbalization) interact with distinct types of insight problems (spatial vs. verbal), aims to provide empirical data that can help test assumptions and evaluate predictions derived from this broader range of theoretical viewpoints, contributing to a more nuanced understanding of the mechanisms underlying insight.

Domain Specificity in Insight: Spatial versus Verbal Problems

While the general mechanisms of restructuring and impasse are considered central to insight across various tasks, research suggests that the specific cognitive resources and representational formats involved may differ depending on the problem domain. A primary distinction often drawn in the literature is between spatial insight problems and verbal insight problems (Dow & Mayer, 2004; Gilhooly & Murphy, 2005; Salmon-Mordekovich & Leikin, 2023). Understanding the potential differences between these domains is crucial, as this understanding directly bears on questions about the generality of insight mechanisms and their potential differential susceptibility to factors like verbalization.

Spatial insight problems typically involve visual or geometric stimuli and require solvers to manipulate spatial relationships or overcome perceptual constraints. Classic examples include the Nine-Dot problem, matchstick arithmetic problems (which involve moving one matchstick to make an equation true), and geometric dissection puzzles (e.g., divide a shape into four identical smaller shapes, as cited in Weisberg, 1995). Solving such problems often necessitates overcoming perceptual grouping biases (like seeing the nine dots as a square), relaxing implicit assumptions about boundaries (e.g., lines must stay within the square), or mentally manipulating objects in non-standard ways (e.g., forming a 3D structure from matchsticks; Scheerer, 1963; Knoblich et al., 1999; Weisberg, 1995). Success is thought to rely heavily on visuospatial working memory, mental imagery, spatial attention, and the ability to inhibit prepotent perceptual interpretations (Ash & Wiley, 2006; Gilhooly & Fioratou, 2009; Chronicle al., 2004).

Verbal insight problems, in contrast, rely primarily on linguistic or semantic processing. Examples include riddles that exploit lexical ambiguity (e.g., the “marrying man” problem; Weisberg, 1995, Appendix B, problem 7), anagrams (though their classification as insight problems is debated; Weisberg, 1995), and the Compound Remote Associates Test (RAT; Bowden & Jung-Beeman, 2003; Mednick, 1962), which requires finding a single word associated with three seemingly unrelated cue words (e.g., finding “tree” for “apple/family/house”). Solving such problems often involves accessing less frequent meanings of words, inhibiting dominant but incorrect semantic associations, activating broad semantic fields to find distant connections (a process often conceptualized as spreading activation; Bowden et al., 2005; Kounios & Beeman, 2014), or reinterpreting the problem statement’s linguistic structure (Bowden et al., 2005; Kounios & Beeman, 2014; Ansburg & Dominowski, 2000). Performance on these tasks is generally linked to verbal abilities, semantic memory access, and potentially processes related to suppressing interference from strong but irrelevant verbal information (DeYoung et al., 2008; Mendelsohn, 1976).

Empirical evidence supports the idea that spatial and verbal insight problems engage partially distinct cognitive resources. Factor analyses and cluster analyses of performance across various insight tasks often reveal separate factors or clusters corresponding roughly to spatial and verbal domains (Dow & Mayer, 2004; Gilhooly & Murphy, 2005; Cunningham, Mac-

Gregor, Gibb, & Haar, 2009). Also, individual difference studies show differential correlational patterns; for instance, spatial insight performance may correlate more strongly with measures of spatial ability or visuospatial working memory, while verbal insight performance correlates more strongly with verbal fluency or vocabulary (Gilhooly & Murphy, 2005; Gilhooly & Fioratou, 2009). Training studies have also demonstrated domain specificity, where training on strategies for solving one type of insight problem (e.g., spatial) improves performance primarily on that type, with limited transfer to other domains (Dow & Mayer, 2004).

This distinction between spatial and verbal insight domains is theoretically significant. If these problem types rely on different representational systems (visuospatial vs. linguistic/semantic) and engage distinct cognitive resources, they might be differentially affected by interventions or concurrent tasks. Specifically relevant to the present study, the requirement to verbalize thought processes might interact differently with spatial versus verbal re-framing. Articulating spatial transformations or overcoming perceptual Gestalts might be particularly hindered by engaging the linguistic system, whereas verbalizing thoughts during a verbal insight task might be less disruptive or even potentially helpful by reinforcing semantic processing (Schooler, 2002; Gilhooly et al., 2010). Investigating this potential interaction forms a core part of the rationale for the current research.

Rationale for Problem Selection

For researchers studying insight, using reliable problems is essential to ensure they genuinely elicit insight. Weisberg (1995) offers a taxonomic framework for identifying such problems, beginning with the question of whether a solver experiences discontinuity, a shift from unsuccessful attempts to a successful approach. Importantly, not all discontinuities qualify. Insight requires that this shift be the result of a representational change, a transformation in how the problem's elements operators, or goals are mentally conceived, i.e. a reconstruction of the problem space.

According to Weisberg (1995), pure insight problems are those in which restructuring is the only viable path to a solution; in these cases, the initial representation inevitably leads to an impasse and must be fundamentally altered for progress to occur. Examples include the "Radiation Problem," which requires a shift from a single high-intensity beam to multiple low-intensity rays converging from different directions, or certain linguistic riddles like the "Marrying Man" problem, which require a reinterpretation of key terms. In contrast, hybrid insight problems, such as the Nine-Dot Problem, can potentially be solved either through restructuring (e.g., by relaxing the assumption that lines must stay within the square) or, alternatively, through non-insightful strategies like trial-and-error, where participants might stumble upon the correct solution. Weisberg places such problems in a distinct category, arguing that the possibility of a procedural solution without representational change disqualifies them as pure insight tasks.

However, this exclusion of problems that involve only discontinuity, i.e., a sudden shift in approach without a clearly identifiable representational restructuring, has been debated (Öllinger et al., 2008). These ambiguities raise questions about whether "pure" and "hybrid" insight tasks can always be meaningfully distinguished at the level of task structure, or whether insight is better understood as a process that can occur differently across individuals.

While acknowledging Weisberg's (1995) taxonomy, the problem selection process in the present study, consistent with common practice in the field, includes tasks considered "pure" insight alongside potentially "hybrid" ones that are well-established in prior literature. This approach was adopted partly because, as Weisberg (1995) himself noted, providing a comprehensive list of "pure" insight problems that also systematically vary across domains (e.g., spatial and verbal) is difficult. Therefore, rather than attempting to enforce strict theoretical purity, this study recognizes the ongoing debates surrounding insight problem taxonomy. It intends to use this inherent variability as a lens for interpreting both performance data and qualitative reports, an issue that will be revisited in the Discussion chapter.

The Interplay of Insight and Verbalization

As established in the previous section, insight problem-solving is characterized by sudden restructuring processes that often rely on non-verbal, holistic forms of cognition, particularly in spatial domains. In contrast, language, our primary cognitive tool for expressing and structuring explicit thought, is inherently sequential and analytic. This fundamental difference invites the question: what happens when language-based processes, such as verbalization, are introduced into tasks that may depend on non-verbal restructuring? Could talking aloud interfere with the very mechanisms insight requires?

This concern forms the basis of the Verbal Overshadowing Effect (Schooler & Engstler-Schooler, 1990; Schooler et al., 1993), a phenomenon where verbalizing certain experiences, especially those reliant on perceptual or holistic information, can disrupt cognitive performance. Although initially observed in face recognition, similar disruptive effects have been shown in insight problem-solving, where requiring verbalization have been found to impair performance on insight tasks. (Schooler et al., 1993; Gilhooly et al., 2010). Verbal overshadowing becomes particularly relevant when comparing problem types that differ in their representational formats. Standard analytical problems often benefit from verbal processing, whereas insight problems may suffer under its weight, particularly when restructuring is required in visuospatial rather than linguistic domains (Bowden & Jung-Beeman, 2003; Kounios & Beeman, 2014).

Verbalization, however, is not a monolithic process. Seminal work by Ericsson and Simon (1993) established important distinctions between different types of concurrent verbalization and their predicted cognitive consequences. Their model of verbal protocol analysis remains the most influential theoretical framework for understanding how verbalization interacts with cognition.

Ericsson and Simon's Model of Verbalization

The theoretical basis for the modern use of verbal protocols rests heavily on the model proposed by Ericsson and Simon (1993). Their model is situated within a general information-processing architecture comprising sensory stores, a limited-capacity short-term memory (STM) or working memory, and a vast long-term memory (LTM). Central to their model is the concept of heeded information: information that is currently held in focal attention within STM. This includes information recently acquired from sensory input, retrieved from LTM, or generated as the output of intermediate cognitive operations (Ericsson & Simon, 1993, p. 11).

Ericsson and Simon's (1993) core claim is that verbal reports are generated from the subset of information that is currently or was recently heeded (i.e., present in STM). Crucially, they argue that the act of verbalizing this already attended information need not fundamentally alter the sequence or content of the primary task-related thought processes, provided certain conditions are met. Their model distinguishes between different levels of verbalization based on the directness of the mapping between heeded information and the verbal output:

- Level 1 Verbalization (Direct Articulation): This involves the direct vocalization of information that is already present in STM in a verbal or propositional code (e.g., inner speech). Ericsson and Simon (1993) posit that when information is attended to in such a readily articulable format, vocalizing it overtly requires minimal additional processing time or capacity beyond the articulation itself. This forms the basis for "talk aloud" instructions, which aim to capture this direct stream of verbally encoded thought.
- Level 2 Verbalization (Recoding/Explication): This level applies when the heeded information is not in a verbal code (e.g., visual imagery, spatial representations, sensory experiences). Verbalizing such information requires the participant to engage in additional recoding processes to translate the non-verbal content into a verbal description or label (Ericsson & Simon, 1993). This recoding takes additional time and cognitive effort. "Think aloud" instructions often encompass both Level 1 and Level 2 processes, asking participants to verbalize whatever passes through their mind, which may require some description of non-verbal thoughts. While requiring extra processing for recoding, Ericsson and Simon argue that Level 2 verbalization, if focused solely on describing attended content without further elaboration, should not alter the sequence of the primary task-related thoughts, although it may slow down the overall process.
- Level 3 Verbalization: This level involves verbalizations that go beyond merely reporting the content currently in focal attention. It includes requests for participants to explain their thoughts, justify their actions, interpret their strategies, or report on motives or causes (Ericsson & Simon, 1993). Generating such reports requires accessing information beyond what is currently heeded, retrieving related knowledge from LTM, making inferences, and synthesizing explanations. Ericsson and Simon predict that these additional cognitive operations, demanded by Level 3 instructions, will significantly alter the sequence and content of the primary task-related thought processes, as attention must be redirected towards generating the explanation rather than solely executing the task.

Operationalizing Verbalization Types in the Present Study: Type 1 (Think-Aloud) vs. Type 3 (Explanatory)

The present study operationalized two distinct concurrent verbalization conditions based on Ericsson and Simon's framework:

- Type 1 (Think-Aloud): This corresponds closely to Ericsson and Simon's conceptualization of Level 1 and potentially Level 2 verbalization using instructions that minimize elaboration. Participants are typically instructed to continuously verbalize any thoughts that enter their mind while performing the task, without attempting to plan or explain them (Ericsson & Simon, 1993; Fox et al., 2011). Theoretically, this type of verbalization is predicted to have minimal impact on the accuracy and sequence of thought processes, although it might slightly increase overall task time due to articulation and potential minor recoding demands.
- Type 3 (Explanatory): This condition aligns with Ericsson and Simon's (1993) Level 3 verbalization. Instructions typically direct participants to explain their reasoning, strategies, or the rationale behind their actions, often as if elucidating the process to another person (Ericsson & Simon, 1993; Fox et al., 2011). As previously discussed, this type of verbalization necessitates additional processing that extends beyond merely reporting attended thoughts. It demands metacognitive reflection, retrieval of justifications, organization of explanations, and potentially the adoption of a more analytical or deliberative processing mode. Consequently, Type 3 verbalization is predicted to be reactive; that is, it is likely to alter the course, speed, and possibly the outcome of the primary cognitive task being performed.

A meta-analysis by Fox et al. (2011) offered strong empirical support for this theoretical distinction. Reviewing numerous studies, they found that Type 1 think-aloud procedures generally did not significantly affect task accuracy when compared to silent control conditions. Type 3 procedures, in contrast, were frequently associated with significant changes in performance accuracy, sometimes positive, sometimes negative, and often increased solution times, thereby confirming their reactive nature. This established theoretical and empirical distinction between Type 1 and Type 3 verbalization, particularly their differential potential for reactivity, provides a crucial rationale for their comparison in the current study. Such a comparison is especially relevant in the context of insight problem-solving, where cognitive processes might be particularly sensitive to concurrent demands.

The present study

The preceding review has synthesized literature on the nature of insight, its historical roots, and its defining characteristics, particularly in relation to ill-defined problems and representational re-framing. Key theoretical accounts of insight mechanisms were explored, including Representational Change Theory and the ongoing debate between "special process" and "business-as-usual" perspectives, alongside complementary theories like Progress Monitoring and Spreading Activation. Additionally, the distinction between spatial and verbal insight domains was established, and the theoretical underpinnings of verbalization as a research method were detailed. This comprehensive review highlights a specific need for research that systematically investigates how different verbalization demands interact with these varied facets of insight problem-solving.

Based on the theoretical background reviewed in this chapter, the present study aims to address the following research questions:

1. General Effect of Concurrent Verbalization:
   a. Compared to a silent control condition, does engaging in concurrent verbalization lead to significantly different objective performance (i.e., solution rate and solution time) on insight problems versus non-insight problems?
   b. Does the impact of this general concurrent verbalization (on solution rate and solution time) differ significantly when comparing performance on spatial problems versus verbal problems?
2. Effect of the different Verbalization Types:
   a. How do different verbalization types, Type 1 (think-aloud) versus Type 3 (explanatory), affect performance on insight problems compared to non-insight problems, and how do their effects compare to a silent baseline and to each other?
   b. How do these specific effects of Type 1 versus Type 3 verbalization (and their comparison to a silent baseline) manifest differently for spatial problems versus verbal problems?
3. Higher-Order Interactions: Are there significant interaction effects involving Verbalization Condition (Silent, Type 1, Type 3), Problem Type (Insight vs. Non-Insight), and Problem Domain (Spatial vs. Verbal)? (This includes an interest in other notable two-way interactions, such as Problem Type × Problem Domain, that might emerge independently or in conjunction with verbalization effects.)

Drawing from the theoretical background and the research questions outlined above, the following hypotheses were formulated:

H1. General Effect of Concurrent Verbalization (Related to RQ1):

- H1a - Solution Rate: It was hypothesized that concurrent verbalization (pooling Type 1 and Type 3) would significantly reduce solution rates for insight problems compared to the silent control condition. The impact on non-insight problems was expected to be less detrimental or potentially neutral.
- H1b - Solution Time: It was hypothesized that concurrent verbalization would significantly increase solution times across all problem types compared to the silent condition. This increase was anticipated to be more pronounced for spatial problems than for verbal problems, reflecting potentially greater interference with visuospatial processing.

H2. Effect of Verbalization Type (Related to RQ2)

- H2a - Solution Rate: It was hypothesized that Type 3 (explanatory) verbalization would lead to a greater decrease in solution rates for insight problems compared to both Type 1 (think-aloud) verbalization and the silent condition, owing to its higher cognitive demands and reactivity.
- H2b - Solution Time: It was hypothesized that Type 3 verbalization would produce significantly longer solution times than both Type 1 verbalization and the silent condition, with this effect being particularly prominent for spatial problems.

H3. Higher-Order Interaction Effects (Related to RQ3)

- H3a - Three-Way Interaction: A significant three-way interaction was hypothesized among Verbalization Condition, Problem Type, and Problem Domain. Specifically, the most substantial performance detriment (in terms of both solution rates and times) was predicted for spatial insight problems when performed under Type 3 (explanatory) verbalization.
- H3b - Problem Type × Domain Interaction: A significant two-way interaction was hypothesized between Problem Type and Problem Domain. It was expected that the performance difference between insight and non-insight problems would vary across domains, with this difference anticipated to be larger in the spatial domain, where spatial insight problems are presumed to be more vulnerable to cognitive interference than verbal insight problems.

3 Method

Participants

Twenty-eight undergraduate students (20 female, 8 male) from Lund University participated in this study. Their ages ranged from 23 to 42 years (mean age = 25.5). Recruitment utilized convenience sampling within the university population. As compensation for their time, participants received candy. Inclusion criteria required participants to be fluent in either Swedish or English, as task materials were available in both languages. Ultimately, two participants completed the experiment in English, while the remaining 26 conducted it in Swedish. All individuals provided written informed consent before participation.

Materials

Problem-Solving Tasks

Specific problems were selected based on their prevalent use in insight and problem-solving literature, and their established classification as requiring either insight (restructuring) or analytical processing. In total, 12 problem-solving tasks were chosen, with three problems selected for each of four categories.

The first category, Spatial Insight (SI) problems, included the Nine-Dot Problem, which requires connecting nine dots arranged in a 3x3 square using four straight lines without lifting the writing utensil (Weisberg, 1995); the Nine Dogs Pen Problem, involving arranging nine dogs in a square pen such that there is an odd number of dogs along each side (Weisberg, 1995); and the Triangle of Coins Problem, where the task is to invert the direction of a triangle composed of ten coins by moving only three coins (Metcalfe, 1986b; Weisberg, 1995).

The second category consisted of Spatial Non-Insight (SNI) Problems. These were the Trace Problem, where participants trace a complex geometric figure without lifting the pen or retracing any line (adapted from Webb, Little, Cropper, & Roze, 2017); a 5-disc version of the Tower of Hanoi puzzle, (adapted from Fleck, 2008); and the Count the Squares problem, which requires determining the total number of squares embedded within a complex grid figure (Gilhooly and Murphy, 2005).

The third category was Verbal Insight (VI) Problems. It included: the Hole Problem, which asks for the volume of soil in a hole of specified dimensions; the Policeman Problem, requiring an explanation for why a police officer observes multiple traffic violations by a woman but makes no arrest (adapted from Gardner, 1978; Weisberg, 1995); and the Chess Problem, which involves explaining how two men can play five chess games with each winning an equal number without any draws (adapted from Sternberg & Davidson, 1982; Weisberg, 1995).

Finally, the fourth category was Verbal Non-Insight (VNI) Problems. It included: the Bachelors' Dinner Problem, where participants deduce which bachelor hosted dinner on which night and served which meal based on logical constraints (Gilhooly et al., 2010); the Crime Suspect Problem, which involves determining who committed a crime and who is telling the truth from four statements where only one is true (Gilhooly and Murphy, 2005); and the Flowers Problem, requiring deduction of which partner gave which type of flowers to which woman based on logical constraints (Schooler et al., 1993; Gilhooly et al., 2010).

Practice Problems

Two problems were selected as practice tasks: the Mutilated Checkerboard Problem (Kaplan & Simon, 1990) and the Wolf, Sheep, and Cabbage Problem (Gilhooly et al., 2010). These were used to familiarize participants with the verbalization procedures before they encountered experimental blocks requiring verbalization. These tasks aimed also to help participants understand the general problem-solving interface and the specific requirements of either thinking aloud or explaining their thoughts.

Post-test Questionnaire

At the end of the session, a Post-Experiment Questionnaire was administered to gather feedback on participants' overall experience. This questionnaire collected general feedback on several aspects, including overall task engagement; the occurrence of any "Aha!" experiences; perceived difficulty trends; familiarity with any of the problems; comfort levels with verbal versus spatial problem types; clarity of instructions; subjective experiences under different verbalization conditions; perceived effects of verbalization; and any encountered difficulties or distractions.

Equipment

Tasks were built and presented on a computer screen using Gorilla (Anwyl-Irvine et al., 2020), an online experiment builder. Solution times were recorded with an integrated software timer. Audio recordings were made during the verbalization conditions using a phone's voice recorder (iPhone 13, iOS version 18.4.1).

Design

This study employed a within-subjects experimental design. The primary independent variables were Verbalization Condition and Problem Type. Verbalization Condition had three levels: Silent, Type 1 Think-Aloud, and Type 3 Explanatory. Problem Type encompassed four levels: Spatial Insight (SI), Spatial Non-Insight (SNI), Verbal Insight (VI), and Verbal Non-Insight (VNI). This latter variable could alternatively be conceptualized as two distinct factors: Domain (Spatial, Verbal) and Insight Requirement (Insight, Non-Insight).

A specific counterbalancing scheme was implemented to control for potential order and carry-over effects between the two active verbalization conditions. To keep the baseline measurement consistent across all participants and mitigate potential fatigue or practice effects, the Silent condition block was always presented in the fixed second position, between the two verbalization blocks. The Type 1 and Type 3 verbalization conditions were thus counterbalanced across participants, appearing as either the first or the third block.

Each participant completed three experimental blocks, one for each verbalization condition. Within every block, participants attempted four problems (one SI, one SNI, one VI, one

VNI), with the presentation order of these four problems randomized within the block. The assignment of specific problems to condition blocks was also randomized across participants, ensuring each problem appeared roughly equally frequently under each verbalization condition across the sample. The primary dependent variables were Solution Rate, defined as whether the participant reached the correct solution (coded dichotomously: solved/unsolved) and solution time defined as the time elapsed from problem presentation until the correct solution was articulated or indicated, recorded only for correctly solved problems.

Procedure

Participants were tested individually in a quiet laboratory setting at the Humanities lab of Lund University. The session proceeded as follows:

Initially, during the Introduction and Informed Consent phase, the researcher greeted each participant, explained the study's general purpose, and provided an informed consent form for review and signature. Before starting an experimental block requiring verbalization (i.e., either the Type 1 Think-Aloud or Type 3 Explanatory condition), participants received an introduction to the specific verbalization requirements for that block. Following these instructions, participants completed one of two designated practice problems, assigned via a counterbalancing scheme, using the relevant verbalization technique. For example, if a participant was assigned the Mutilated Checkerboard problem to practice with Type 1 Think-Aloud instructions before their first verbalization block, they would subsequently practice the Wolf, Sheep, and Cabbage problem with Type 3 Explanatory instructions before their other verbalization block (or vice-versa, according to the counterbalancing of both verbalization condition order and practice problem assignment). During this practice, the experimenter provided feedback and clarification to ensure the participant understood the verbalization instructions. No practice problems were given before the Silent condition block.

The main experiment consisted of three task blocks. Each block included four problems (one SI, one SNI, one VI, and one VNI), presented in a randomized order. Participants performed each block under one of the three verbalization conditions. The Silent condition was consistently administered as the second block. The order of the Type 1 Think-Aloud and Type 3 Explanatory verbalization conditions, which were administered as the first and third blocks, was counterbalanced across participants.

The Verbalization Instructions provided were as follows: For the Silent Condition, participants were instructed to solve the problems silently without speaking. For the Type 1 Think-Aloud Condition, participants were instructed to "think aloud constantly", verbalizing any thoughts that entered their minds as they worked on the problem, without trying to explain or justify them. If they fell silent for a prolonged period, they were reminded to "keep talking". For the Type 3 Explanatory Condition, participants were instructed to "explain their thoughts and reasoning aloud", as if teaching someone how they were attempting to solve the problem, thereby justifying their steps and strategies; they were similarly reminded to continue explaining if prolonged silences occurred. Verbatim instructions administered to participants are available in Appendix A.

For task presentation and timing, each problem was displayed on the computer screen, and participants had a maximum of 3 minutes to solve each one. This time limit drew upon prior studies, such as DeYoung et al. (2008) (2-minute limit) and Gilhooly et al. (2010) (4-minute limit). A 3-minute duration was chosen as a pragmatic midpoint, aiming to balance sufficient time for insight with the minimization of participant fatigue and the maintenance of procedural consistency across tasks. If a problem was not solved within this time limit, they were advanced to the next problem, and the trial was recorded as unsolved.

Data recording involved noting the Solution Rate (solved/unsolved) for each problem. Solution times were measured from problem onset until the participant stated or indicated the correct solution. Verbalizations during the Type 1 and Type 3 conditions were audio-recorded. After completing all three blocks, participants filled out the Post-Experiment Questionnaire, as detailed in the Materials section. The entire session lasted approximately 55-60 minutes for each participant.

4 Results

This section presents the findings related to the three research questions (RQ1–RQ3), which investigate how different types of verbalizations impact problem-solving performance across various problem types (Insight vs. Non-Insight) and domains (Spatial vs. Verbal).

The analysis utilized two primary dependent measures: (1) Solution Rate (the proportion of problems solved correctly, coded as Success), and (2) Log-Transformed Solution Time (Log_SolutionTime), recorded only for correct responses. Given the within-subjects design, all statistical models were chosen to properly account for repeated measures from the same participants. Specifically, solution rates (a binary outcome) were analyzed using Generalized Estimating Equations (GEE) with a binomial distribution and a logit link function, as this method is robust for analyzing correlated binary data from repeated-measures designs. Log-transformed solution times were analyzed using Linear Mixed Models (LMM), which are ideal for handling continuous data where observations are clustered within participants. In the LMMs, participant was treated as a random intercept to account for baseline individual differences in solving speed. In all figures presented in this chapter, error bars represent 95% Confidence Intervals.

An alpha level of .05 was adopted for determining statistical significance, and marginal trends are reported for p-values $< .10$. The analysis included data from 28 participants, totaling 336 trials (12 tasks × 28 participants). It is important to note that the findings, particularly non-significant results, should be interpreted in the context of the study's statistical power. A sensitivity power analysis was conducted to determine the magnitude of effects this study was equipped to detect, and it will be presented at the end of this chapter to provide this crucial context.

RQ1: General Effect of Concurrent Verbalization

Solution Rates

Fig.1 displays the mean solution rates for insight and non-insight problems under Silent versus general Verbalizing conditions. Visually, for insight problems, mean solution rates seem slightly lower under verbalization compared to silence; conversely, for non-insight problems, they appear marginally higher. However, in both instances, the error bars show considerable overlap.

A GEE model (Success ~ GeneralVerbalization * ProblemType) revealed no significant main effect of general verbalization (vs. Silent) for insight problems (the reference category for ProblemType) on the log-odds of success (Coefficient = –0.227, SE = 0.256, z = –0.887, p = .375). The GeneralVerbalization × ProblemType interaction was also not significant (Coefficient = +0.417, SE = 0.359, z = 1.162, p = .245), indicating that the effect of general verbalization on solution rates did not reliably differ between insight and non-insight problems. Likewise, there was no significant difference in solution rates between Non-Insight and Insight problems under silent conditions (Coefficient = –0.426, SE = 0.320, z = –1.334, p = .182).

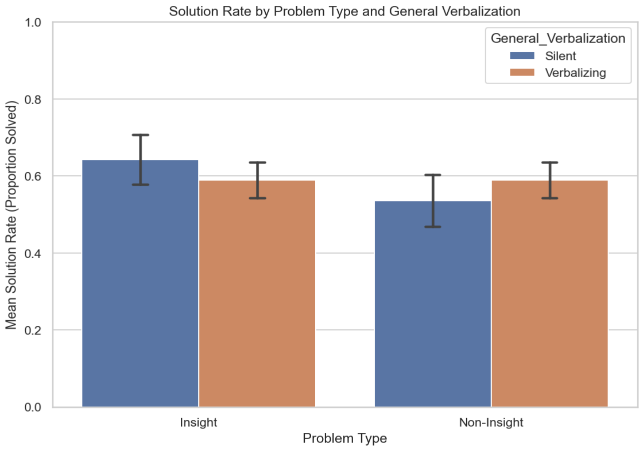

Figure 1. Mean solution rates for insight and non-insight problems under silent conditions versus general verbalization.

Fig. 2 shows the mean solution rates for spatial and verbal problems under Silent versus general Verbalizing conditions. Visually, for spatial problems, mean solution rates seem slightly lower with verbalization compared to silence. For verbal problems, they appear marginally higher, although the error bars again indicate considerable overlap.

A GEE model (Success ~ GeneralVerbalization * ProblemDomain) indicated no significant main effect of general verbalization (vs. Silent) for spatial problems (the reference category for ProblemDomain) on the log-odds of success (Coefficient = –0.177, SE = 0.295, z = –0.600, p = .548). The GeneralVerbalization × ProblemDomain interaction was also not significant (Coefficient = +0.329, SE = 0.397, z = 0.830, p = .406), suggesting that the effect of general verbalization on solution rates did not reliably differ between spatial and verbal problems. Finally, there was no significant difference in solution rates between Verbal and Spatial problems under silent conditions (Coefficient = +0.126, SE = 0.308, z = 0.409, p = .682).

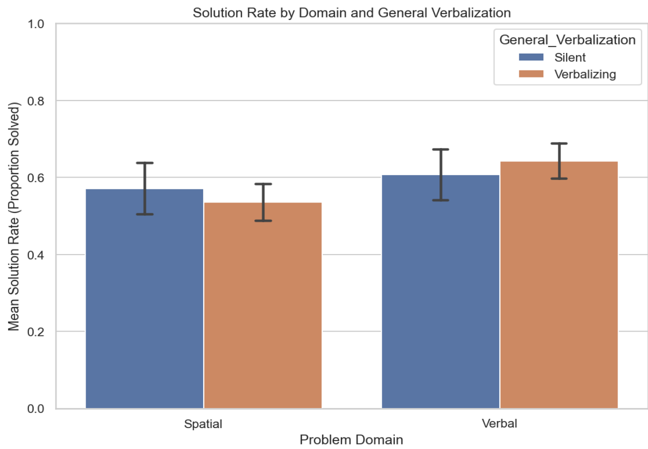

Figure 2. Mean solution rates for spatial and verbal problems under silent conditions versus general verbalization.

In summary, general concurrent verbalization did not significantly affect solution rates when compared to silence, nor did its impact significantly vary between insight versus non-insight problems or between spatial versus verbal domains.

Solution Times

Fig.3 displays the distributions of log-transformed solution times for insight and non-insight problems under Silent versus general Verbalizing conditions. For both insight and non-insight problems, median log solution times are visibly higher with verbalization than under silence. Variability appears smaller in the verbalizing condition for both problem types.

A Linear Mixed Model (LogSolutionTime ~ GeneralVerbalization * ProblemType), with Silent and Insight as reference categories and using only correct trials, revealed a marginal trend suggesting that general verbalization (vs. Silent) increased log-transformed solution times for insight problems (Coefficient = +0.349, SE = 0.185, t(261) = 1.889, p = .060). There was no significant interaction between GeneralVerbalization and ProblemType (Coefficient = –0.050, SE = 0.267, t(261) = –0.188, p = .851), indicating that this verbalization-related slowdown did not reliably differ between insight and non-insight problems. Additionally, under silent conditions, non-insight problems took significantly longer to solve than insight problems (Coefficient = +0.751, SE = 0.221, t(261) = 3.398, p < .001).

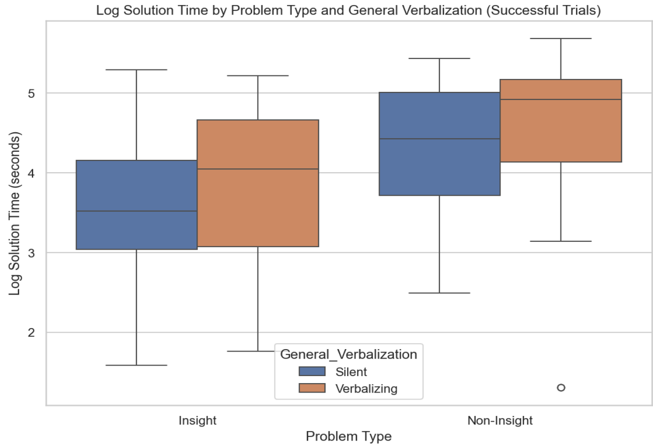

Figure 3. Log solution time for correct trials for insight and non-insight problems under silent conditions versus general verbalization.

Fig. 4 shows the distributions of log-transformed solution times for spatial and verbal problems under Silent versus general Verbalizing conditions. For both spatial and verbal problems, median log solution times are visibly higher with verbalization compared to silence, showing slightly greater variability when verbalizing. The impact of verbalization appears more pronounced for spatial problems. Overall, spatial problems tend to exhibit lower median log solution times than verbal problems.

A Linear Mixed Model (LogSolutionTime ~ General Verbalization * Problem Domain), with Silent and Spatial as reference categories and using only correct trials, revealed that general verbalization significantly increased log-transformed solution times for spatial problems (Coefficient = +0.415, SE = 0.201, t (261) = 2.069, p = .039). The General Verbalization × Problem Domain interaction was not significant (Coefficient = –0.141, SE = 0.276, t(261) = –0.512, p = .609), suggesting that the verbalization-related slowdown was statistically comparable for verbal and spatial problems. Under silent conditions, verbal problems took significantly longer to solve than spatial problems (Coefficient = +0.678, SE = 0.225, t(261) = 3.012, p = .003).

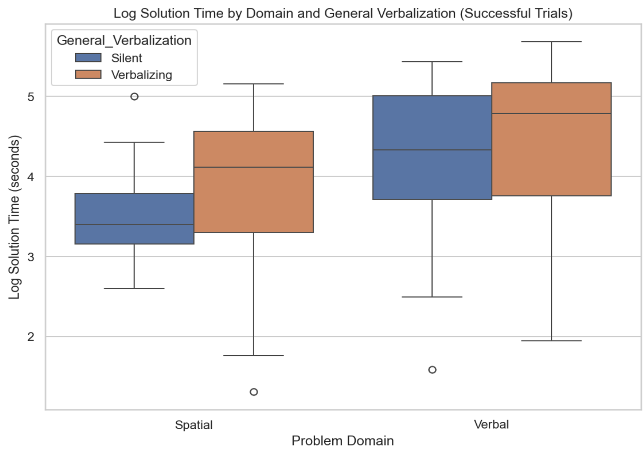

Figure 4. Log solution time for correct trials for spatial and verbal problems under silent conditions versus general verbalization.

In summary, general concurrent verbalization tended to increase log-transformed solution times compared to silence. This increase was statistically significant for spatial problems (and, due to the non-significant interaction, analogous for verbal problems). Under silent conditions, non-insight problems were solved significantly more slowly than insight problems, and verbal problems were solved significantly more slowly than spatial problems.

RQ2: Effect of the different types of Verbalizations

Solution Rates

Fig. 5 visualizes how different verbalization types (Silent, Type 1, Type 3) appear to affect solution rates for insight versus non-insight problems. For insight problems, visual inspection suggests that Type 3 verbalization yielded slightly higher solution rates than Type 1, though both were below the silent condition. For non-insight problems, both verbalization types seem to have slightly outperformed the silent condition, with Type 1 appearing marginally higher than Type 3.

A GEE model (Success ~ VerbalizationCondition * Problem Type) revealed no significant effect of Type 1 (think-aloud) verbalization (vs. Silent) on the log-odds of success for insight problems (Coefficient = –0.300, SE = 0.295, z = –1.016, p = .310). Type 3 (explanatory) verbalization also did not significantly alter success on insight problems (Coefficient = –0.153, SE = 0.322, z = –0.474, p = .636). The main effect of Problem Type was not significant (Coefficient = –0.426, SE = 0.320, z = –1.335, p = .182), indicating no difference between non-insight and insight problems under Silent. Finally, neither the Type 1 × Non-Insight interaction (p = .247) nor the Type 3 × Non-Insight interaction (p = .453) was significant, suggesting that verbalization type did not differentially affect solution rates for insight versus non-insight problems.

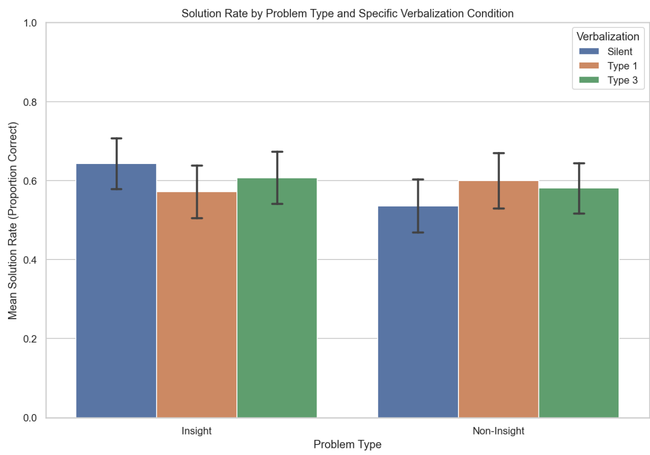

Figure 5. Mean solution rates for insight versus non-insight problems broken down by verbalization type.

Fig. 6 visualizes how different verbalization types appear to influence solution rates for spatial versus verbal problems. For spatial problems, a visual trend suggests that Type 3 verbalization slightly outperformed Type 1, although both conditions resulted in lower solution rates than the silent condition.

For verbal problems, the visual pattern indicates that Type 1 and Type 3 verbalizations yielded similar performance, with both appearing slightly above the silent baseline.

A GEE model (Success ~ Verbalization Condition * Problem Domain) revealed no significant effect of Type 1 verbalization (vs. Silent) on success for spatial problems (Coefficient = –0.188, SE = 0.369, z = –0.508, p = .611) or of Type 3 verbalization (Coefficient = –0.168, SE = 0.339, z = –0.497, p = .619). The main effect of ProblemDomain was not significant (Coefficient = +0.126, SE = 0.308, z = 0.409, p = .682), indicating no Silent-condition difference between Verbal and Spatial. Interaction terms Type 1 × Verbal (p = .554) and Type 3 × Verbal (p = .389) were also non-significant, showing that verbalization type did not significantly affect success in either domain.

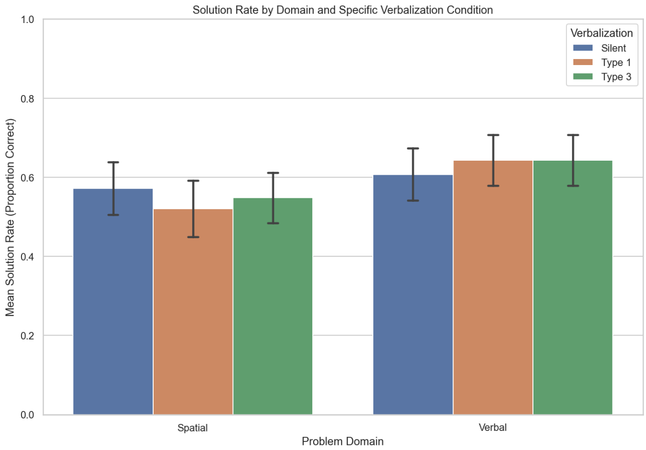

Figure 6. Mean solution rates for spatial versus verbal problems down by different verbalization type.

In summary, neither Type 1 nor Type 3 verbalization significantly affected solution rates compared to Silent, regardless of problem type (insight vs. non-insight) or domain (spatial vs. verbal). There were no significant Verbalization Condition × Problem Type or Verbalization Condition × Problem Domain interaction effects on success.

Solution Times

Fig. 7 displays the log-transformed solution times for correct trials, broken down by problem type (insight vs. non-insight) and verbalization condition (Silent, Type 1, Type 3). For insight problems, both Type 1 and Type 3 verbalizations exhibit longer median solution times compared to the silent condition. Visually, Type 3 shows a slightly higher median and a narrower interquartile range (spread) than Type 1 for these problems. A similar pattern is apparent for non-insight problems, although the visual differences between the verbalization types seem less pronounced.

A Linear Mixed Model (LogSolutionTime ~ VerbalizationCondition * ProblemType), using all trials, revealed that Type 1 verbalization did not significantly change log-transformed solution times on insight problems compared to Silent (Coefficient = +0.099, SE = 0.183, t(330) ≈ 0.54, p = .589). In contrast, Type 3 verbalization produced a significant increase in log-transformed solution times on insight problems (Coefficient = +0.377, SE = 0.183, t(330) ≈ 2.07, p = .039). Under silent conditions, non-insight problems were solved significantly more slowly than insight problems (Coefficient = +1.028, SE = 0.183, t(330) ≈ 5.62, p < .001). Neither the Type 1 × Non-Insight interaction (Coefficient = +0.016, p = .952) nor the Type 3 × Non-Insight interaction (Coefficient = –0.197, p = .441) was significant, indicating that the impact of verbalization type on solution time did not reliably differ between insight and non-insight problems.

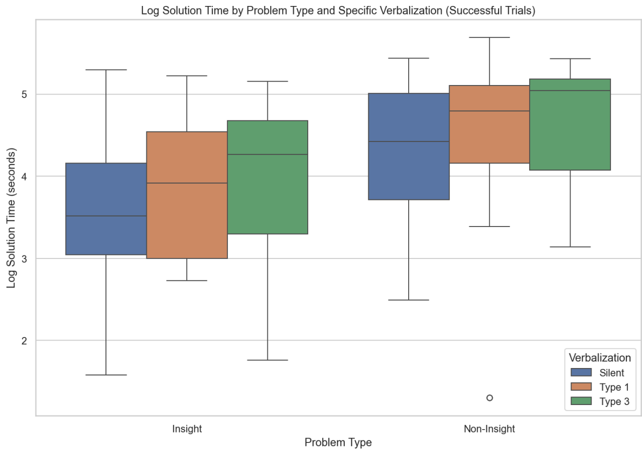

Figure 7. Log solution times for correct trials for insight versus non-insight problems broken down by verbalization type.

Fig. 8 displays the log-transformed solution times for correct trials across problem domains (spatial vs. verbal) and verbalization types (Silent, Type 1, Type 3). For spatial problems, visual inspection indicates that Type 3 verbalization resulted in the longest median solution times, while Type 1 verbalization led to slightly shorter, though still elevated, times compared to silence. In contrast, for verbal problems, median solution times appeared similarly high across all three verbalization conditions.

A Linear Mixed Model (LogSolutionTime ~ VerbalizationCondition * ProblemDomain), using all trials, revealed that Type 1 verbalization led to a non-significant increase in log-transformed solution time for spatial problems compared to Silent (Coefficient = +0.279, SE = 0.199, t(330) ≈ 1.41, p = .160). Type 3 verbalization led to a significant increase in solution time on spatial problems (Coefficient = +0.599, SE = 0.188, t(330) ≈ 3.19, p = .001). There was a significant main effect of ProblemDomain under Silent: verbal problems took significantly longer than spatial problems (Coefficient = +1.038, SE = 0.192, t (330) ≈ 5.41, p < .001). The Type 1 × Verbal interaction was not significant (Coefficient = –0.440, SE = 0.276, t (330) ≈ –1.59, p = .111), whereas the Type 3 × Verbal interaction was significant (Coefficient = –0.572, SE = 0.268, t (330) ≈ –2.13, p = .033). Thus, while Type 3 verbalization significantly slowed spatial problem solving, its relative impact was reduced for verbal problems.

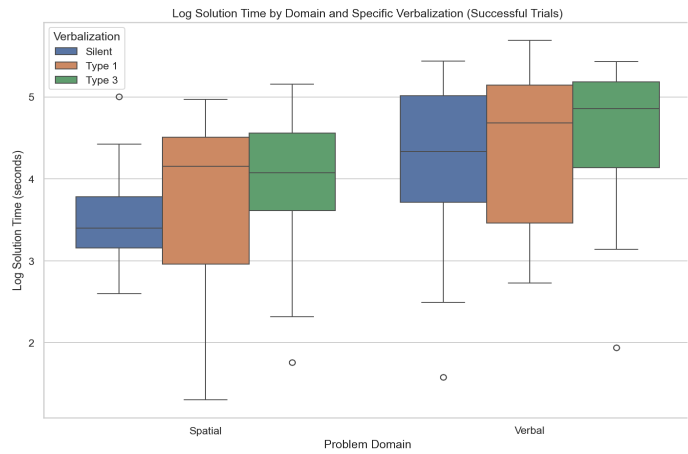

Figure 8. Log solution times for correct trials for spatial versus verbal problems broken down by verbalization type.

In summary, both Type 1 and Type 3 verbalization showed trends toward longer solution times but only Type 3 verbalization produced a statistically significant slowdown, and this was observed specifically for insight and spatial problems. A significant Type 3 × ProblemDomain interaction confirmed that its disruptive effect was stronger in the spatial domain than in the verbal domain.

RQ3: Interaction Effects: Problem Type × Domain × Verbalization

Solution Rates

Figure 9 visually suggests domain-specific patterns for solution rates; for instance, verbalization appears more detrimental to performance on spatial insight problems but somewhat beneficial for verbal non-insight tasks.

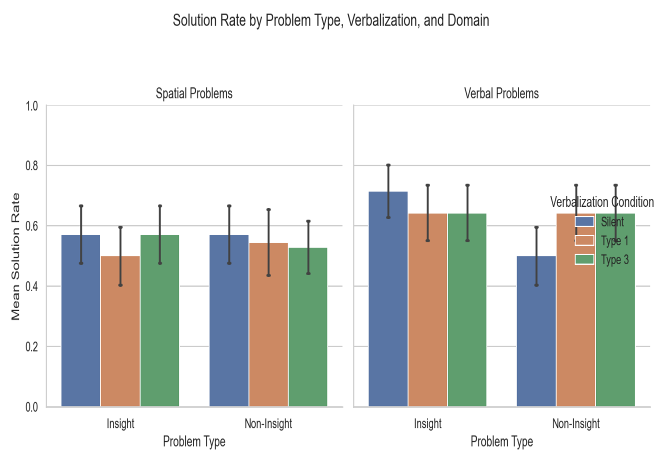

Figure 9. Mean solution rates of different verbalization types across problem type and domain

A GEE model (Success ~ VerbalizationCondition * ProblemType * ProblemDomain), with Silent, Insight, and Spatial as reference categories, revealed that the three-way interaction involving Type 1 verbalization (Verbalization[T.Type 1] × ProblemType[T.Non-Insight] × ProblemDomain[T.Verbal]) was not significant (Coefficient = +0.675, SE = 0.540, z = 1.250, p = .211). Similarly, the three-way interaction involving Type 3 verbalization (Verbalization[T.Type 3] × ProblemType[T.Non-Insight] × ProblemDomain[T.Verbal]) was not significant (Coefficient = +1.226, SE = 0.915, z = 1.340, p = .180). However, the model did reveal a significant two-way interaction between ProblemType and ProblemDomain (ProblemType[T.Non-Insight] × ProblemDomain[T.Verbal]: Coefficient = –0.958, SE = 0.445, z = –2.154, p = .031), indicating that the difference in success between insight and non-insight problems varied significantly between spatial and verbal domains. Specifically, the negative coefficient indicates lower success for non-insight problems in the verbal domain relative to the spatial domain. All other main effects and lower-order interactions, including those involving Type 1 and Type 3 verbalization individually, were non-significant.

Solution Times

Fig. 10 displays the log-transformed solution times across all conditions. For spatial insight problems, median solution times for Type 1 and Type 3 verbalization are comparable, though Type 3 shows substantially greater variability. This pattern is clearer for spatial non-insight problems, where Type 3 verbalization appears slower and more variable than Type 1. Conversely, in the verbal domain, Type 3 seems to produce a more pronounced slowdown for insight problems, while all conditions appear similar for non-insight problems.

A three-way Linear Mixed Model (LogSolutionTime ~ VerbalizationCondition * ProblemType * ProblemDomain) revealed a significant three-way interaction for the Type 3 contrast (Type 3 × Non-Insight × Verbal: Coefficient = –1.007, SE = 0.476, t(324) = –2.12, p = .034), whereas the corresponding interaction for Type 1 was not significant (p = .354). To decompose this interaction, a follow-up analysis on the spatial-insight subset showed that Type 3 verbalization led to a marginal increase in solution time compared to Silent (Coefficient = +0.387, SE = 0.218, t(81) = 1.77, p = .076), while Type 1 verbalization was non-significant (Coefficient = +0.236, SE = 0.218, t(81) = 1.08, p = .279). Additionally, the main three-way model revealed a marginal two-way interaction between ProblemType and ProblemDomain (Coefficient = +0.558, SE = 0.340, t(324) = 1.64, p = .100), indicating that the difference in solution time between non-insight and insight problems was more pronounced in the verbal domain than in the spatial domain. All other main effects and two-way interactions were non-significant.

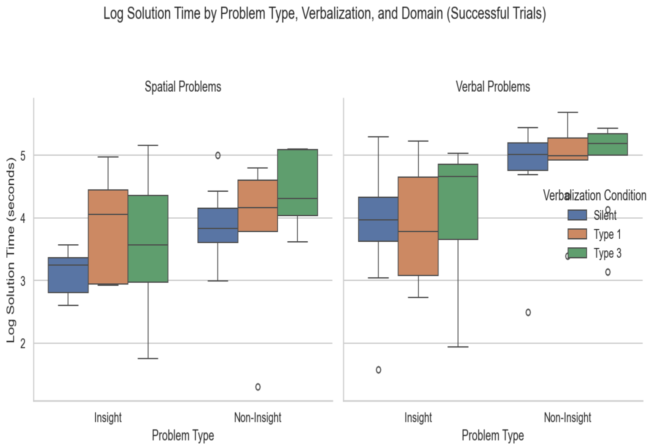

Figure 10. Log solution times for correct trials across verbalization types broken down by domain and problem type.

In summary, while the three-way interaction for Type 1 was non-significant, Type 3 verbalization exhibited a significant three-way interaction, consistent with H3a. The marginal two-way ProblemType × ProblemDomain interaction suggests a larger insight vs. non-insight difference in solution time for verbal than for spatial problems.

Power analysis

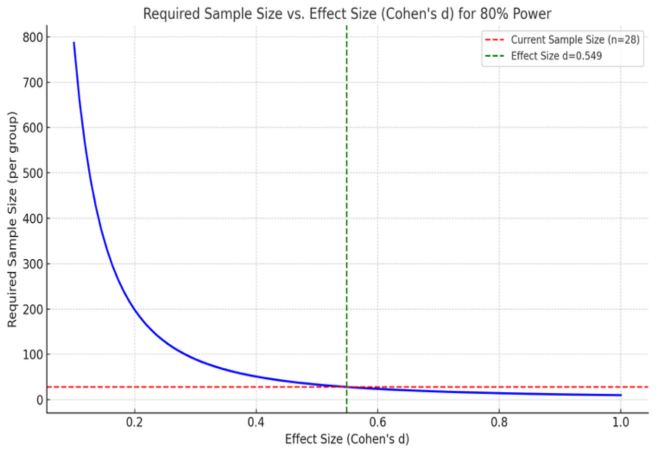

Figure 11. Power analysis plot showing the required sample size for various effect sizes (Cohen's *d*) to achieve 80% power at α = 0.05.

Discussion

This study investigated the impact of three concurrent verbalization conditions, Silent, Type 1 (think-aloud), and Type 3 (explanatory), on problem-solving performance across four distinct categories. The findings present a multifaceted picture that informs theories of insight, verbal overshadowing, and cognitive load in working memory. The following section interprets the results for each research question by systematically discussing the outcomes of the study's formal hypotheses.

Summary of Main Findings

The first research question addressed the general effect of verbalization. The findings showed no significant impact on solution rates for any problem type or domain. This fails to support H1a, which predicted that verbalization would decrease solution rates for insight problems. This null result for accuracy challenges strong versions of the Verbal Overshadowing Effect, suggesting that the core mechanisms for reaching a solution remain robust even under the load of verbalization. However, engaging in verbalization generally increased solution times. This slowing effect was statistically significant for spatial problems ($p = .039$) and showed a marginal trend for insight problems ($p = .059$). This provides partial support for H1b, confirming that verbalization imposes a time cost consistent with a cognitive load model. The lack of a significant interaction means we cannot conclude this time cost was significantly more pronounced for spatial problems over verbal ones.

Secondly, the study addressed the specific effects of different verbalization types (RQ2). In line with the general findings, solution rates were not significantly altered by either Type 1 or Type 3 verbalization, failing to support H2a. For solution times, however, a clearer picture of interference emerged. Explanatory (Type 3) verbalization was found to be uniquely disruptive, significantly increasing solution times for both insight problems ($p = .039$) and, even more strongly, for spatial problems ($p = .001$) when compared to silence. Crucially, a significant Type 3 × Domain interaction ($p = .033$) was also found. This interaction confirms that the disruptive time cost of explaining was significantly stronger for spatial problems than for verbal problems, providing robust support for H2b and highlighting the particular vulnerability of non-verbal processing to this type of verbal interference.

Thirdly, the study tested for higher-order interactions. The primary hypothesis of a three-way interaction between verbalization, problem type, and domain (H3a) was not supported for either solution rates or solution times. The predicted "worst-case scenario", a substantial performance deficit for spatial insight problems under Type 3 verbalization, did not occur. This suggests the interplay between these factors is more subtle than hypothesized, or that the study lacked the statistical power to detect such a complex effect.

However, the analysis did reveal a significant two-way Problem Type × Domain interaction for solution rates ($p = .031$). This finding partially supports H3b. While the hypothesis correctly predicted an interaction between these factors, the nature of the interaction was opposite to what was expected. The hypothesis predicted the largest performance difference in the spatial domain; instead, the results indicate that non-insight problems were solved significantly less often than insight problems primarily in the verbal domain. This unexpected result is valuable, as it suggests that the cognitive distinction between insight and non-insight processing is itself domain dependent.

Interpretation of Findings in Relation to Research Questions and Literature

The findings of this study, which explored the impact of different verbalization types on insight and non-insight problem-solving across spatial and verbal domains, present a complex picture demanding careful interpretation. This interpretation must consider not only existing theories but also maintain a critical awareness of specific characteristics of the current experimental design that likely influenced the observed outcomes and the statistical power to detect certain effects.

Notably, this study was conducted with a sample of 28 participants. An illustrative sensitivity power analysis (for a paired t-test scenario, with $\alpha = .05$ and power = .80) indicated that this sample size provided adequate power to detect effect sizes (Cohen's d) of approximately 0.549 or larger (see fig. 11). Such an effect size is generally considered medium to large (Cohen, 1988). Consequently, the study was likely underpowered to reliably detect true effects of a smaller magnitude. This limitation is an important consideration when interpreting non-significant findings for main effects or interactions, particularly where visual trends were apparent or where existing theory might predict more subtle effects.

Additionally, the problem set consisted of 12 unique tasks, with each specific problem sub-category (e.g., Spatial Insight) represented by only three distinct problems. This scale is modest when compared to some larger investigations, for instance, Gilhooly, Fioratou, & Henretty (2010) reportedly utilized around 32 problems (8 in each sub-category) with approximately 80 participants, and inherently affects statistical power. Thus, non-significant results, especially for complex interactions or subtle effects, should be interpreted cautiously, as they may reflect insufficient statistical power rather than a definitive absence of an effect.

Further complicating the interpretation is the participants' prior familiarity with some of the problems used. As Appendix C reveals, several participants reported previous encounters with tasks such as the Tower of Hanoi, the Nine-Dot puzzle, or some of the verbal non-insight problems (e.g., participant 13295350 stated, "I have seen some before like Hanoi, nine-dot problems...I don't remember the solution...But I have done these matching questions before...they felt familiar anyway"). Although participants often stated they did not recall the solutions, such prior exposure could still have influenced their initial approach, reduced problem novelty, subtly guided their search (potentially making the problems less ill-defined), and interacted with verbalization effects in ways not easily controlled. Ideally, a larger number of task items per sub-category or a larger sample size would have allowed for the exclusion of trials involving significant problem familiarity, thereby offering better control for this variable. This pre-existing knowledge, even if partial, adds another layer of individual variability that may affect aggregate performance measures and the clarity of experimental effects.

Furthermore, as detailed in the literature review, the nature and selection of problems are central considerations. This study aimed to employ problems reflecting distinct "insight" versus "non-insight" categories. However, the purity of insight problems is a complex issue, as many tasks may be hybrids solvable through non-restructuring pathways like trial-and-error, rather than solely through insight (Weisberg, 1995). Qualitative data from the post-experiment questionnaire (see Appendix C) indicated instances of such mixed problem-solving approaches, even for tasks designated as insight problems. For example, some participants described solving the Triangle problem via attempts that felt like trial-and-error but still culminated in a subjective "Aha!" moment. This observation aligns with Weisberg's (1995) concept of discontinuity, which can be experienced without necessarily involving deep conceptual reframing in all instances. The potential for "erroneous insights", an "Aha!" experience not corresponding to a correct solution, as reported by one participant (13295350: "When I had strong aha experiences, the answer was always wrong …" Appendix C), further complicates the clear separation of these cognitive processes.

As important, the temporal dynamics of the specific insight and non-insight problems used in this study also differ, affecting how solution time data should be interpreted. Non-insight problems, including verbal puzzles or spatial non-insight tasks like "Count the Squares" and the "Tower of Hanoi", typically require sequential, cumulative steps that naturally result in longer solution durations, even when the tasks are straightforward. Insight problems, conversely, often involve an impasse followed by a sudden reframing, with the final solution step occurring rapidly (e.g., in the "Hole Problem", where grasping the semantic twist directly yields the answer). Therefore, comparing solution times across these distinct categories demands caution, as the underlying cognitive processes being timed differ substantially and could therefore skew solution time comparisons independently of any verbalization effects.

The complexity described above is further compounded by the lack of standardized difficulty ratings across tasks. Although problem selection was guided by prior literature and digital feasibility, task difficulty was not formally scaled or balanced across conditions. Some insight problems, notably the "Nine-Dot" or "Nine Dogs Pen" puzzles, proved particularly challenging and were potentially too difficult to reliably elicit successful restructuring within the allotted time constraints.

These methodological limitations, including the hybrid nature of some insight problems, the potential for insight-like experiences in non-insight tasks, differing solution time profiles, and variable task difficulty, necessitate caution against overinterpreting quantitative outcomes without considering the nuanced cognitive dynamics at play during problem-solving. With these crucial methodological considerations in mind, the specific findings related to the research questions can be explored.

The initial observation that general concurrent verbalization tended to increase solution times without significantly impacting solution rates (RQ1) aligns with a common pattern in verbalization research. This time cost, particularly evident for spatial problems where verbalization significantly increased solution times ($p = .039$), is consistent with established theoretical perspectives. As detailed in the theoretical background, Ericsson and Simon's (1993) model predicts that even low-level verbalization introduces processing demands by engaging working memory, especially when recoding non-verbal content is required. Participant feedback in this study echoed these demands; for instance, one participant under the explanatory (Type 3) verbalization condition remarked, “First I have to explain why I’m doing this... so I can move forward... but maybe that’s not going to work,” suggesting that the act of explanation temporarily interrupted their problem engagement.

The general increase in solution times due to verbalization, as predicted by models like Ericsson and Simon's (1993) due to added processing demands, manifested significantly for spatial problems, not only under general verbalization (RQ1) but more specifically under Type 3 verbalization when compared to silence (RQ2, $p = .022$). Given that Type 3 verbalization significantly increased solution times for spatial problems when compared to silence ($p = .022$), while Type 1 only showed a non-significant trend in the same direction for spatial problems ($p = .201$), it is plausible that the significant time cost observed for spatial problems under general verbalization was predominantly driven by the more demanding nature of the Type 3 explanatory protocol. The higher cognitive load and inherent need to structure a narrative associated with Type 3 verbalization likely contributed more substantially to this pooled effect than the less intrusive Type 1 think-aloud.

This susceptibility of spatial tasks to verbalization-induced slowing aligns with working memory literature (e.g., Fleck, 2008; Korovkin et al., 2018), which suggests that spatial processing, often reliant on visuospatial sketchpad resources and less directly verbal processes, can be particularly disrupted or overloaded by concurrent verbal demands, especially the more reactive Level 3 explanations.

However, an important nuance for interpreting H1b (related to general verbalization) and H2b (related to Type 3 verbalization), which hypothesized that these time costs would be more pronounced or particularly strong for spatial problems, comes from the interaction analyses. Neither the General Verbalization × Domain interaction for solution times ($p = .610$) nor the specific Type 3 × Domain_Verbal interaction for solution times ($p = .599$) were statistically significant. This indicates that while verbalizing (both generally and specifically Type 3) demonstrably slowed performance on spatial tasks compared to silence, the current data do not statistically confirm that this slowing effect was significantly greater for spatial problems than for verbal problems. The observed time increases for spatial tasks were significant in their own right, but their difference in magnitude compared to the (non-significant) increases for verbal tasks did not reach statistical significance within this study's power constraints. Thus, while the cognitive load is a substantial factor for spatial tasks under verbalization, the hypothesis that this represents a uniquely magnified interference for spatial processing compared to verbal processing is not statistically supported by these interaction terms.

Regarding insight problems specifically, while neither Type 1 nor Type 3 verbalization led to statistically significant increases in solution time compared to silence, there were notable trends: Type 3 showed a marginal trend towards slower times ($p = .082$) and Type 1 a non-significant trend in the same direction ($p = .124$). These trends, though not conclusive, suggest a potential processing cost for insight problems too, particularly when explanatory demands are high, even if this cost did not manifest as uniformly or significantly as it did for spatial problems under Type 3 verbalization.

Transcription analysis further illuminated these effects, revealing participants under Type 3 instructions attempting to structure their explanations, a process potentially more laborious than the direct stream-of-consciousness output typical of Type 1 verbalization. For instance, one participant explicitly described their strategy of writing information down to manage variables, demonstrating a clear metacognitive and explanatory effort. Despite these added demands from verbalization, solution accuracy did not significantly decline across conditions. This suggests that participants may have adapted by segmenting their thought processes or by generally slowing their pace, rather than failing to solve the problems outright. Such adaptation could explain why verbalization incurred a measurable cost in solution time but not in solution accuracy, a pattern consistent with theories proposing that verbalization taxes overlapping working memory systems without necessarily blocking access to solutions.

The pattern of findings offers a complex lens through which to view Representational Change Theory (RCT) (Ohlsson, 1992) and the “Special Process” versus “Business-as-Usual” debate. The significant slowing of spatial problem-solving by Type 3 (explanatory) verbalization, and the general trend of verbalization increasing solution times for spatial problems, could suggest interference with the subtle, visuospatially dominant restructuring processes proposed by RCT and often associated with “special process” views. If insight, particularly spatial insight, relies on non-verbal representational shifts, then demanding concurrent verbal explanation (Type 3) might disrupt these delicate processes. This interpretation is anecdotally supported by participant feedback, such as one individual noting in the post-test summary, “Verbalization may disrupt performance more on spatial tasks than verbal ones”, aligning with a “special process” perspective where these non-verbal mechanisms are vulnerable.

However, the current study’s findings do not uniformly support a strong “special process” account to the exclusion of “business-as-usual” mechanisms. Importantly, verbalization (neither general nor specific types) did not significantly reduce solution rates for insight problems. This outcome contrasts somewhat with early verbal overshadowing literature that reported accuracy detriments for insight tasks (e.g., Schooler et al., 1993). The present findings for solution rates are perhaps more aligned with Gilhooly et al. (2010), who, using only Type 1 verbalization, also found limited detrimental effects on insight problem accuracy, particularly for verbal insight problems, though they did note some evidence of Type 1 hindering spatial insight. The current study extends Gilhooly et al. by including Type 3 verbalization and still finding no significant accuracy detriment for insight across domains. This resilience in accuracy, despite increased solution times, could be interpreted from a “business-as-usual” perspective (Weisberg, 1995): ordinary cognitive processes, including those involved in insight, might simply take longer under the dual-task load of verbalizing, but the fundamental problem-solving mechanisms eventually succeed. The marginal trend for Type 3 verbalization to slow insight solving (p=.082) without affecting accuracy further points to increased effort or time on task rather than a fundamental disruption of the insight mechanism itself. Thus, insight processes might be "special" in their subjective quality or reliance on particular representational shifts, but they may still operate via cognitive components that, while slowed, are not necessarily derailed by concurrent verbalization, especially less demanding Type 1 protocols. This is also supported by transcriptions capturing participants overcoming impasses on insight problems even while verbalizing.

Additionally, the nature of the insight problems employed in this study, some being potentially hybrid (Weisberg, 1995), implies that solutions might occasionally be found through persistent effort culminating in an "Aha!" feeling, rather than solely through profound, instantaneous restructuring. One could argue that insight might occur at any level of the problem space, not just for the initial or goal state, but also for intermediate steps or operator selection, even in problems not typically classified as "insight problems". This aligns with Newell & Simon's (1972) view of navigating a problem space where any component can initially be obscure. Such a perspective would further suggest that a strict dichotomy between problem types based on a single "insight" event may be overly simplistic and that a truly "pure" non-insight problem might be one whose problem space is completely well-defined.

Further complicating a simple view of insight versus non-insight processing is the significant two-way interaction found between Problem Type and Domain for solution rates (H3b only partially supported; p = .031). This interaction, independent of verbalization condition suggested that non-insight problems had lower solution rates relative to insight problems in the verbal domain, compared to the spatial domain where the difference in solution rates between insight and non-insight problems was minimal (and the main effect of problem type non-significant). This finding suggests that the "advantage" or nature of insight processing is not uniform. It might be that verbal insight problems, often hinging on specific semantic re-interpretations, are more distinctly different in their solvability, leading to higher success if the "trick" is found, compared to verbal non-insight problems which may involve more laborious logical deduction. In contrast, the spatial problems selected, both insight and non-insight, might have shared more overlapping processing demands or had a different distribution of difficulty, leading to less differentiation in their solution rates between problem types within that domain. This interaction underscores the importance of considering both problem type and domain concertedly when theorizing about problem-solving mechanisms and the nature of insight.

Beyond the quantitative performance metrics, the qualitative data from participant transcriptions provide noteworthy observations into the subjective and cognitive differences between Type 1 and Type 3 verbalization. These observations raise questions about the extent to which Type 3 can be considered truly "concurrent" in the same way as Type 1, and how each verbalization type interacts with the dynamic process of problem-solving. Analysis of transcriptions revealed distinct cognitive dynamics operating under the two active verbalization conditions. For example, as illustrated by transcriptions for tasks like the Tower of Hanoi (e.g., "Okay, then I'll do this. No. Yes. And then I'll do that."), participants in the Type 1 condition frequently produced more fragmented, reactive speech. This often took the form of a direct running commentary on their immediate actions or fleeting self-evaluations, such as "what am I doing?" or "no, that doesn't work," particularly during exploratory or trial-and-error phases. Such verbalizations closely align with Ericsson and Simon's (1993) conceptualization of Level 1 and Level 2 verbalizations, where the primary focus is on articulating currently heeded thoughts or describing ongoing non-verbal processing with minimal elaboration.

In contrast, during Type 3 verbalizations, participants frequently produced more structured narratives. Their verbalizations tended to include reflections on strategy, justifications for actions, or attempts to construct a coherent account of their problem-solving process; at times, they appeared to articulate thoughts after a sequence of actions or a period of internal deliberation. For instance, a participant working on a logic puzzle under Type 3 conditions might state a planned strategy by saying, "by writing down the order of them we will be able to take and just calculate the order here". This organizational effort, however, sometimes seemed to incur a cost. Some participants apparently paused their active problem-solving to "catch up" with the demand for explanation, suggesting a potential shift in attentional focus from direct task engagement towards the metacognitive act of explaining. This observation corresponds with the significantly longer solution times recorded for Type 3 verbalization compared to Type 1 for non-insight problems, where constructing an explanation for a multi-step analytical process is inherently more demanding.

These differences suggest that Type 1 and Type 3 verbalizations differ not merely in length but reflect fundamentally different ways of thinking. Type 1 verbalization appears more conducive to tracking concurrent, sometimes "messy" or iterative, cognitive processes. In contrast, Type 3, by its nature, imposes a more reflective, analytical framework. Although this reflective stance could be advantageous in certain contexts, for instance, by compelling a more systematic strategy review, potentially aiding in overcoming minor impasses or consolidating learning, it might be less compatible with, or even disruptive to, other cognitive activities. Such activities include the rapid, exploratory epistemic actions that characterize much of practical problem-solving.

Several transcription segments illustrated participants' reliance on such epistemic actions (external manipulations intended to reveal problem structure, test hypotheses, or reduce cognitive load). Particularly in spatial problems (e.g., Tower of Hanoi, Nine Dots, or tracing puzzles), participants often "tested their way forward" through sequences of physical or simulated actions, with verbalizations like "if I do this...". The strong desire for external aids like pen and paper, frequently voiced in post-test reflections, further highlights that problem-solving is rarely a purely internal mental activity but is often a dynamic interaction between internal representations and external scaffolds.

This interactive view of problem-solving raises critical questions about how different verbalization protocols engage with or interfere with these epistemic actions. Explanatory verbalization (Type 3), by requiring participants to step back and articulate reasons or structure a narrative, may compete for cognitive resources with, or interrupt the flow of, these exploratory epistemic actions. Participants sometimes commented on how the need to articulate their reasoning (Type 3) disrupted the flow of "trying things out." This indicates that Type 3 verbalization might not just slow down problem-solving due to added articulatory or metacognitive load, but could, in some instances, fundamentally shift the nature of the problem-solving process itself from one of active, embodied exploration to one of more detached, retrospective narration or justification. This potential shift could have implications for tasks that particularly benefit from rapid, iterative interaction with the problem environment, including some forms of insight problem-solving where trying out unusual actions can lead to representational change.

A final methodological point concerns whether participants consistently adhered to the intended verbalization styles for Type 1 (think-aloud) and Type 3 (explanatory), or whether these styles may have overlapped. This study did not include a formal coding analysis of the transcripts and given the inherently private and fluid nature of inner thought, it's impossible to confirm perfect adherence. Still, a qualitative reading of the transcriptions (see Appendix B) suggests that, overall, participants engaged in distinctly different verbal behaviors across the two conditions. Type 1 verbalizations tended to consist of short, fragmented remarks about immediate actions, while Type 3 often involved more structured explanations or explicit reasoning about strategy. The positioning of the silent condition between the two active verbalization blocks was intended to reduce potential carryover effects. Even so, some subtle cross-contamination is possible and should be acknowledged as a limitation when interpreting the findings.

Conclusion

This thesis set out to investigate the influence of different concurrent verbalization protocols, silent, minimal think-aloud (Type 1), and explanatory (Type 3), on insight and non-insight problem-solving across spatial and verbal domains. Drawing from foundational theories of problem-solving, the nature of insight, verbal overshadowing, and cognitive load associated with verbal reporting, the study aimed to contribute to understanding whether, and how, these distinct verbalization methods facilitate or hinder problem-solving, particularly when considering the interplay with problem type and domain.

The findings painted a complex and illuminating picture. Three key results emerged. First, verbalization consistently imposed a time cost on problem-solving but did not significantly impair solution accuracy, suggesting that the cognitive system can accommodate the load, albeit at a slower pace. Second, this cost was most pronounced under the most demanding condition: explanatory (Type 3) verbalization significantly increased solution times for spatial problems compared to silence, providing strong evidence for its reactive nature and its interference with visuospatial processing. Third, independent of verbalization, a significant interaction between problem type and domain revealed that the relative difficulty of insight versus non-insight problems is not uniform, but changes depending on whether the task is spatial or verbal.

These findings should be interpreted within the study's specific methodological context. With a sample size of 28 participants, the study was powered to detect medium-to-large effects, but more subtle interactions may have gone undetected. Additional challenges included the problem of precisely matching insight and non-insight problems for baseline difficulty, ensuring the "purity" of insight tasks against "hybrid" solution strategies (Weisberg, 1995), managing participants' prior familiarity with some of the problems used' and accounting for the differing temporal dynamics of how solutions unfold for these categories. As such, this study should be viewed as preliminary and exploratory, thus highlighting the need for continued methodological refinement in the field. Future research would benefit from larger and more diverse participant samples, a more extensive and pre-scaled battery of well-matched problems and a formal analysis of verbal protocols grounded in Ericsson and Simon's (1993) model.

Ultimately, this investigation was motivated by the cognitive tension between the sudden and non-verbal nature of insight and the sequential, analytical and linguistic nature of verbalization. The study' findings suggest the relationship is not one of simple opposition. Language does not entirely appear to eliminate the mechanisms of insight, but it does seem to burden it, particularly when explanatory demands are high, and the task is spatial in nature. Understanding this interaction, between this prototypical tool for thought, reflection and communication and non-verbal cognitive functions involving restructuring, visuospatial reasoning, remains a vital research pursuit. While the current study offers preliminary steps and underscores important methodological considerations, the journey towards a comprehensive model of these interactions is by no means complete. Especially now with the prevalent growth of language-based AI models, the question of how language scaffolds or disrupts problem solving, and, by extension, cognition itself, has only grown more urgent. This study offers a modest but meaningful contribution to that exploration

References

Ansburg, P. I. and Hill, K. (2003). Creative and analytic thinkers differ in their use of attentional resources. Personality and Individual Differences, 34(7), 1141-1152.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., and Evershed, J. K. (2020). Gorilla Experiment Builder (www.gorilla.sc): A fast, flexible tool for online behavioural research. Behavior Research Methods, 52(1), 315-325.

Ash, I. K. and Wiley, J. (2006). The nature of restructuring in insight: An individual-differences approach. Psychonomic Bulletin & Review, 13(1), 66-73.

Baddeley, A. (1986): Working Memory, Clarendon Press, Oxford, UK.

Bowden, E. M., Jung-Beeman, M., Fleck, J. and Kounios, J. (2005). New approaches to demystifying insight. *Trends in Cognitive Sciences*, 9(7), 322-328.

Cunningham, J. B., MacGregor, J. N., Gibb, J., and Haar, J. M. (2009). The roles of insight and analysis in the creative process and the development of an insight problem-solving test. Creativity Research Journal, 21(2-3), 260-272.

Chein, J. M. and Weisberg, R. W. (2014). Working memory and insight in verbal problem solving. *Psychonomic Bulletin & Review*, *21*(3), 740-747.

Chin, J. and Schooler, J. W. (2008). Verbal overshadowing. In Durso, F. N., Nickerson, R. S., Schvaneveldt, R. W., Condor, T. L., Kellogg, S. T. and McNeese, M. (Eds.), *Handbook of Applied Cognition* (2nd ed., pp. 593-610). Wiley, Chichester, UK.

Chronicle, E. P., MacGregor, J. N. and Ormerod, T. C. (2004). What makes an insight problem? The roles of heuristics, goal conception, and solution recoding in knowledge-lean problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 14-27.

Clark, A. (1998). Being There: Putting Brain, Body, and World Together Again. MIT Press.

Davidson, J. E. (1995). The suddenness of insight. In R. J. Sternberg and J. E. Davidson (Eds.), *The Nature of Insight* (pp. 125-155). MIT Press.

Duncker, K. (1945). On problem solving. *Psychological Monographs*, 58(5, Whole No. 270).

Ericsson, K. A. and Simon, H. A. (1993): *Protocol Analysis: Verbal Reports as Data (Revised Edition)*, MIT Press, Cambridge, MA.

Fleck, J. I. and Weisberg, R. (2004). The use of verbal protocols as data: an analysis of insight in the candle problem. *Memory & Cognition*, 32, 990-1006.

Fleck, J. I. (2008). Working memory demands in insight versus analytic problem solving. *European Journal of Cognitive Psychology*, 20(1), 139-176.

Fox, M. C., Ericsson, K. A. and Best, R. (2011). Verbal protocols are reactive only when they alter the normal course of cognitive processing: A meta-analysis. *Psychological Bulletin*, 137, 144-167.

Gardner, M. (1978). Aha! Insight. Scientific American / W. H. Freeman.

Gilhooly, K. J. and Fioratou, E. (2009). Verbalization and problem solving: Insight and spatial factors. In Taatgen, N. A. and van Rijn, H. (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1628-1633). Cognitive Science Society, Austin, TX.

Gilhooly, K. J., Fioratou, E., Henretty, N. and Kouider, S. (2010). Verbalization and problem solving: insight and spatial factors. *British Journal of Psychology*, 101, 473-485.

Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R., Reber, P. J. and Kounios, J. (2004). Neural activity when people solve verbal problems with insight. *PLOS Biology*, 2(4), E97.

Kaplan, C. A. and Simon, H. A. (1990). In search of insight. *Cognitive Psychology*, 22, 374-419.

Kirsh, D. and Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, 18(4), 513-549.

Kiyokawa, S., Kashiwagi, M., and Onodera, Y. (2023). Explanatory verbalization promotes problem-solving performance. Thinking & Reasoning, 29(2), 241-262.

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

Kounios, J., Frymiare, J. L., Bowden, E. M., Fleck, J. I., Subramaniam, K., Parrish, T. B., and Jung-Beeman, M. (2006). The prepared mind: Neural activity prior to problem presentation predicts subsequent solution by insight. Psychological Science, 17(10), 882-890.

Kounios, J., Fleck, J. I., Green, D. L., Payne, L., Stevenson, J. L., Bowden, E. M., and Jung-Beeman, M. (2008). The origins of insight in resting-state brain activity. Neuropsychologia, 46(1), 282-291.

Koffka, K. (1935): *Principles of Gestalt Psychology*, Harcourt, Brace, New York, NY.

Korovkin, S., Vladimirov, I., Chistopolskaya, A. and Savinova, A. (2018). How working memory provides representational change during insight problem solving. *Frontiers in Psychology*, 9, Article 1864.

Kounios, J., & Beeman, M. (2014). *The cognitive neuroscience of insight.* Annual Review of Psychology, 65, 71–93.

Knoblich, G., Ohlsson, S., Haider, H. and Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(6), 1534-1555.

Köhler, W. (1925): *The Mentality of Apes* (E. Winter, Trans.), Kegan Paul, Trench, Trubner & Co., London.

MacGregor, J. N., Ormerod, T. C. and Chronicle, E. P. (2001). Information processing and insight: a process model of performance on the nine-dot and related problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(1), 176-201.

Mednick, S. A. (1962). The associative basis of the creative process. Psychological Review, 69(3), 220-232.

Mendelsohn, G. A. (1976). Associative and attentional processes in creativity. Journal of Personality, 44(2), 341-369.

Metcalfe, J. (1986b). Premonitions of insight predict impending error. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12(4), 623-634.

Metcalfe, J. and Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Memory & Cognition*, 15, 238-246.

Newell, A. and Simon, H. A. (1972): *Human Problem Solving*, Prentice Hall, New York, NY.

Thorndike, E. L. (1911): *Animal Intelligence; Experimental Studies*, The Macmillan company, New York.

Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. In Keane, M. and Gilhooly, K. (Eds.), *Advances in the Psychology of Thinking*, Vol. 1 (pp. 1-44). Harvester Wheatsheaf, London, UK.

Ohlsson, S. (1984). Restructuring revisited: II. An information processing theory of restructuring and insight. *Scandinavian Journal of Psychology*, 25(2), 117-129.

Salmon-Mordekovich, K., and Leikin, R. (2023). Spatial insight problem solving in geometry: The case of multiple-solution tasks. The Journal of Mathematical Behavior, 70, 101064.

Scheerer, M. (1963). Problem-solving. Scientific American, 208(4), 118-128.

Schooler, J. W. (2002). The fluidity of consciousness: From opaque thoughts to transparent words. *Trends in Cognitive Sciences*, 6, 339-344.

Schooler, J. W. and Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, 17, 36-71.

Schooler, J. W. and Melcher, J. (1995). The ineffability of insight. In Smith, S. M., Ward, T. B. and Finke, R. A. (Eds.), *The Creative Cognition Approach* (pp. 97-133). MIT Press, Cambridge, MA.

Schooler, J. W., Ohlsson, S. and Brooks, K. (1993). Thoughts beyond words: when language overshadows insight. *Journal of Experimental Psychology: General*, 122, 166-183.

Simon, H. A. (1973). The structure of ill structured problems. *Artificial Intelligence*, 4(3), 181-201.

Smeekens, B. A. and Kane, M. J. (2016). Working memory capacity, mind wandering, and creative cognition: An individual-differences investigation into the benefits of controlled versus spontaneous thought. *Psychology of Aesthetics, Creativity, and the Arts*, 10(4), 389-415.

Smith, S. M. (1995). Getting into and out of mental ruts: A theory of fixation, incubation, and insight. In Sternberg, R. J. and Davidson, J. E. (Eds.), *The Nature of Insight* (pp. 229-251). MIT Press, Cambridge, MA.

Sternberg, R. J., and Davidson, J. E. (1982). The mind of the puzzler. Psychology Today, 16(6), 37-44.

Sternberg, R. J., and Davidson, J. E. (Eds.). (1995). The Nature of Insight. MIT Press.

Subramaniam, K., Kounios, J., Parrish, T. B., and Jung-Beeman, M. (2009). A brain mechanism for facilitation of insight by positive affect. Journal of Cognitive Neuroscience, 21(3), 415-432.

Vygotsky, L. S. (1962). Thought and Language. MIT Press.

Webb, M. E., Little, D. R. and Cropper, S. J. (2016). Insight is not in the problem: Investigating insight in problem solving across task types. *Frontiers in Psychology*, 7, Article 1424.

Webb, M. E., Little, D. R., Cropper, S. J. and Roze, K. (2017). The contributions of convergent thinking, divergent thinking, and schizotypy to solving insight and non-insight problems. *Thinking & Reasoning*, 23(3), 235-258.

Weisberg, R. W. and Alba, J. W. (1981). Gestalt theory, insight, and past experience: Reply to Dominowski. *Journal of Experimental Psychology: General*, 110(2), 199-203.

Weisberg, R. W. (2015). On the "demystification" of insight: A critique of neuroimaging studies of insight. In Weisberg, R. W. (Ed.), *Creative Thinking and Problem Solving* (pp. 285-318). Springer, Cham.

Wertheimer, M. (1945): *Productive Thinking*, Harper, New York, NY.

Öllinger, M., Jones, G. and Knoblich, G. (2008). Investigating the effect of mental set on insight problem solving. *Experimental Psychology*, 55(4), 269-282.

Appendix A. The Verbalization Instructions

Verbalization Type	**Language**	**Instruction**
Silent	Swedish	I det här blocket vill jag att du arbetar med varje uppgift tyst, utan att yttra dina tankar högt. Fortsätt med problemet tills du har en lösning eller tills tiden tar slut. Förstår du vad du ska göra? När du är redo att börja, tryck på starta knappen!
Silent	English	In this block, I want you to work on each problem silently, without verbalizing your thoughts aloud. Work on the problem until you have a solution or until the time is up. Do you understand what I want you to do? When you are ready to begin, press start.
Type 1 (Think-Aloud)	Swedish	I det här blocket vill jag att du tänker högt medan du löser varje problem. Även om det kan kännas ovant, är det viktigt att du säger allt du tänker på, från det att du först ser uppgiften tills du ger ditt slutliga svar. Dina tankar kan vara halvfärdiga, mumlande eller osammanhängande – allt är okej. Du ska inte försöka planera eller ordna det du säger, och du behöver inte förklara dina tankar för mig. Tänk bara högt på ett naturligt sätt, som om du vore ensam och pratade för dig själv. Försök att prata hela tiden, men tvinga inte fram tankar om inget dyker upp just då – låt det komma av sig självt. Om du är tyst länge kommer jag påminna dig att fortsätta. Allt du säger spelas in. Fortsätt tills du löst uppgiften eller tills tiden är slut. Förstår du vad jag vill att du ska göra? När du är redo, tryck på starta för att börja med en övningsuppgift.
Type 1 (Think-Aloud)	English	In this block, I want you to think aloud while you solve each problem. Even if it feels strange, it's important that you say everything you're thinking from the moment you first see the problem until you give your final answer. Your thoughts might be half-finished, mumbled, or disjointed, any of that is fine. You don't need to plan or structure what you say, and you don't have to explain your thoughts to me. Just think aloud in a natural way, as if you were alone and talking to yourself. Try to keep talking, but don't force words if nothing comes to mind, let it flow by itself. If you're silent for too long, I will remind you to continue. Everything you say will be recorded. Keep going until you've solved the task or until time runs out. Do you understand what I'd like you to do? When you are ready to begin, press start.
Type 3 (Explanatory)	Swedish	I det här blocket vill jag att du noggrant rationaliserar varenda tanke och val du gör, ungefär som om du skulle undervisa någon i hur du löser uppgiften. Jag förstår att det kan kännas ovant att prata på det här sättet, men försök ändå beskriva varför du tänker som du gör och varför du väljer en viss väg eller lösning. Om du fastnar, berätta vad som känns svårt och vilka metoder du redan provat, samt varför de inte fungerade. Försök hela tiden resonera högt och motivera dina beslut så noggrant du kan. Om du blir tyst för länge kommer jag påminna dig att fortsätta. Allt du säger spelas in. Fortsätt så tills du hittat en lösning eller tills tiden tar slut. När du är redo, tryck på starta för att börja med en övningsuppgift.
Type 3 (Explanatory)	English	In this block, I want you to carefully rationalize every thought and choice you make, much like you would if you were teaching someone how to solve the task. I understand that it may feel strange to talk this way, but try to describe why you think the way you do and why you choose a certain path or solution. If you get stuck, tell me what feels difficult and what methods you have already tried, as well as why they did not work. Try to reason out loud all the time and justify your decisions as carefully as you can. If you remain silent for too long, I will remind you to continue. Everything you say is recorded. Continue like this until you find a solution or until time runs out. When you are ready to begin, press start.

Appendix B. (Transcriptions Excerpts

Partici-pant (Ex-ample ID)	Condition	Problem Type/Name	Excerpt (Swedish)	Excerpt (English Translation)
P1	Type 1 (Practice)	Tower of Ha-noi	”Den orangea lägger jag ner i mitten. Lägger den blåa eh i mitten också. Ja, då sitter den fast där. Ja, det var lite tricky”.	“I'll put the orange one in the middle. Put the blue one, uh, in the middle too. Yes, then it's stuck there. Yeah, that was a bit tricky.”
P1	Type 1 (Practice)	Hole Problem	”Hur många kubikmeter jord finns i ett 6 meter långt hål? Nej, men alltså jag vet inte. Jag kan inte sånt här... Jag vet inte hur man räknar ut kubik.”	“How many cubic meters of earth are in a 6-meter long hole? No, I mean, I don't know. I can't do things like this... I don't know how to calculate cubic [meters].”
P2	Type 3 (Practice)	River Cross-ing (Wolf, Sheep, Cab-bage)	”Om bonden lämnar vargen och kå-len kvar åker över med fåret först så kommer inte vargen äta kolet... Och nu åker jag tillbaka ensam... men gud alltså, penna och papper hade varit jäkligt trevligt här nu”.	“If the farmer leaves the wolf and the cabbage behind and goes over with the sheep first, the wolf won't eat the cabbage... And now I go back alone... but goodness, pen and paper would have been really nice here now.”
P3	Type 3 (Practice)	River Cross-ing (Wolf, Sheep, Cab-bage)	”...först måste jag ju då åka över med fåret... sen så åker jag tillbaks... då kommer jag ta dit kolen, och sen så tar jag med fåret tillbaks... Då tar jag varg... och lämnar kvar fåret... då åker jag tillbaks och hämtar fåret”.	“...first I have to go over with the sheep... then I go back... then I'll take the cabbage there, and then I'll take the sheep back... Then I take the wolf... and leave the sheep behind... then I go back and get the sheep.”
P3	Type 3	Counting Squares	”Så här har vi rutnät framför oss med massvis med kvadrater och kvadra-terna är överlappande... i det grövsta rutnätet är 4 * 4. Eh så där har vi då såklart 16 rutor. Men i de rutorna så kan vi också se att det finns eh vi kan slå ihop två av dem”.	“So here we have a grid in front of us with lots of squares and the squares are overlapping... in the coarsest grid it's 4x4. Uh so there we have, of course, 16 squares. But in those squares, we can also see that there are, uh, we can combine two of them.”
P4	Type 1	Trace Figure	”Om jag går varvet runt först och sparar den sista ner till vänster till ex-empel. Det går inte... Jag testar mig fram här. Jag kan inte gå upp för då låser jag av den översta. Så jag forts-ätter ner”.	“If I go all the way around first and save the last one down to the left, for example. That doesn't work... I'm try-ing things out here. I can't go up be-cause then I'll block off the top one. So I'll continue down.”
P5	Type 1 (Think-Aloud)	Mutilated Checkerboard (Practice)	”Tekniskt sett så borde ju gå eftersom att alltså om den täcker två... 62 / 2 är ju 31... Men någonstans så kommer det ju bli det. Hade de varit kvar hade det absolut inte gått... Nej, för den är inte geometriskt anpassad”.	“Technically it should work because if it covers two... 62 / 2 is 31... But somewhere it will be like that. If they [corners] had remained, it absolutely wouldn't have worked... No, because it's not geometrically adapted.”
P6	Type 3 (Explana-tory)	Nine Dogs Pen	”Nu ska jag då rita två ytterligare två kvadratiska stängsel så att varje hund hamnar i... varje hund ska hamna ett eget avskilt område... Den I'm lost for words. Last for thoughts känner jag.”	“Now I'm supposed to draw two addi-tional square fences so that each dog ends up in... each dog should end up in its own separate area... I'm lost for words. Lost for thoughts, I feel.”
P7	Type 1 (Think-Aloud)	Nine Dots	”Jag känner igen den här uppgifterna lite och är så här, åh, man ska rita. Alltså det är inte bara att hålla sig inom prickarna utan man ska typ dra det långt någon gång... Knepet är att man ska gå utanför...”	“I kind of recognize this task and it's like, oh, you have to draw. So it's not just about staying within the dots but you kind of have to draw it far out sometime... The trick is that you have to go outside...”

Appendix C. Translated Excerpts from the Post-questionnaire form

Partici-pant Private ID	Aha Experience	Describe the "Aha" moment	Easiest Condition	Hardest Condition	Problems that you were familiar with before the experiment?
13158317	Yes	the shape you were supposed to draw, or when you got the answer, then one of the riddles	think out loud felt best, I have a little bit of a hard time being quiet because sometimes it helps to say something out loud then you get more understanding I think	explaining wasn't the hardest but more challenging because then they didn't just have one thought, they had to develop it, getting in the way when you were close to a solution. then being quiet was the hardest for the reason I mentioned before	the square with dots, the sheep, etc. but not the solution
13188017	Yes	The triangle when I saw the symmetry, and when I understood the question of the hole problem	think-aloud, because it helped me "document" my thoughts in some way.	Explanatory, because then I needed to think about whether my statements were logical or "sufficiently" explanatory. Felt a bit more like an achievement.	The boat problem, the tower of hanoi
13195927	Yes	earth in hole - have no earth	explaining, because then I felt more organized, and good at	think-aloud, because then I had to both think and verbalize simultaneously, two tasks	towers draw the lines - 9 dot connect person with things
13206445	Yes	for example the chess problem. I had an aha experience after I finished reading. Same with the woman who walked and didn't drive. Pretty quickly after I read it I figured out that she walked and didn't drive. Both answers just popped into my head.	Silence was easiest because then you could concentrate completely on the task without having to explain what you were doing or thinking.	Explaining was the hardest, as it was difficult to explain how you thought while trying to think of new approaches. Often you didn't have a clear plan for why you were doing certain things.	
13295350	Yes	No, I was just testing myself. When I had strong aha experiences, the answer was always wrong, when I understood the answer to some (e.g. chess) there was no strong aha experience, just "then they must be like this" which was a neutral feeling.	Silently (but being able to write as a tool) is easiest for me, probably because I'm more of an introvert as a person. It was hard at first to have to verbalize when you're not used to it.	Explaining, because it put more pressure on you and forced you to evaluate your thoughts. It was especially difficult because I didn't know the tasks, and then it's hard to think out loud when you don't have any thoughts (which I experienced when it was difficult: brain freeze).	I have seen some before like hanoi, nine-dot problems but it was a long time ago and I don't remember the solution, but I remembered when the researcher said the solution that they were the way to do it. But I have done these matching questions before in work tests I think, they felt familiar anyway.

Appendix D. The use of generative AI

Generative AI, specifically OpenAI's ChatGPT, was used in this project primarily to assist with scripting and automation of data analysis workflows, including preprocessing, merging participant data, and calculating key outcome variables. Generative AI produced the initial scripts for data preprocessing, statistical modeling, and result summaries. Every AI-generated script was tested on small example datasets to confirm correct trial extraction, time conversion, and success coding. Model formulas for GEE and LMM were checked to ensure factors and reference levels matched the study design. Each model's output was compared to hand-calculated success rates and average solution times. Key results were also rechecked by fitting the same models in R to ensure consistent parameter values. Narrative descriptions of model findings were drafted by AI but then edited so that all coefficient values, standard errors, test statistics, and p-values matched exactly what the software produced. Any mismatch was corrected manually. All code and results were saved in version-controlled notebooks for full transparency. This process made it possible to use AI for efficiency while ensuring every number and interpretation was accurate.