A cognitive neuroscience perspective on insight as a memory process: Searching for the solution

Chapter · January 2022
DOI: 10.31234/osf.io/zv4dk

3 authors, including:

Maxi Becker
Humboldt-Universität zu Berlin
34 PUBLICATIONS  366 CITATIONS

Jasmin M. Kizilirmak
Deutsches Zentrum für Neurodegenerative Erkrankungen
44 PUBLICATIONS  310 CITATIONS

Some of the authors of this publication are also working on these related projects:

Learning from insight: Cognitive and neural processes View project

Midline structures in human memory View project

All content following this page was uploaded by Maxi Becker on 05 January 2022.

The user has requested enhancement of the downloaded file.
A cognitive neuroscience perspective on insight as a memory process: Searching for the solution

Maxi Becker¹, Roberto Cabeza¹,² and Jasmin M. Kizilirmak³,⁴

1) Humboldt University, Institute for Psychology, Berlin, Germany
2) Duke University, Durham, USA
3) German Center for Neurodegenerative Diseases, Göttingen, Germany
4) University of Hildesheim, Institute for Psychology, Germany

Abstract

What are the cognitive and brain processes that lead to an insight? In this chapter, we will describe the insight solution process from a neurocognitive perspective. Inspired by cognitive theories, we translate some of insight's main cognitive subprocesses (problem representation, search, representational change, solution) into related neurocognitive ones and summarize them in a descriptive framework. Those described processes focus primarily on verbal insight and are explained using the remote associates task. In this task, the solver is provided with several problem elements (e.g., drop, coat, summer) and needs to find the (remotely related) target that matches those cues (e.g., rain). In a nutshell, insight is the consequence of a problem-solving process where the target is encoded in long-term memory but cannot be retrieved at first because the relationship between the problem elements and the target is unknown, precluding a simple memory search. Upon problem display, the problem elements and a whole network of associated concepts are automatically activated in long-term memory in distinct areas of the brain representing those concepts (=problem representation). Insight is assumed to occur when automatic processes suddenly activate the target after control processes associated with inferior frontal gyrus and anterior cingulate cortex activation manage to overcome prior knowledge and/or perceptual constraints by revising the current activation pattern (=representational change).

Keywords: representational change, fMRI, problem solving, insight, search, aha experience, cognitive control

Content

1. Introduction ...........................................................................................................................................3
2. Insight as a memory process – a neurocognitive perspective ..............................................................6
   2.1 Problem representation – activation in semantic memory network .....................................................8
   2.2 Search – retrieval of solution candidates ............................................................................................9
   2.3 Representational change – cognitive control .....................................................................................11
   2.4 Solution – generation and evaluation of solution candidate ..............................................................13
3. Summary ..................................................................................................................................................14
4. Open questions & future directions ........................................................................................................15
5. References ...............................................................................................................................................16
1 Introduction

Why is the organic chemist August Kekulé (1829 – 1896) still known today far beyond his field of expertise? One of the reasons is because, he made a scientific discovery via insight, that is to say he suddenly became aware of a novel solution to a previously incomprehensible problem. Kekulé had long been searching for the structure of benzene, a liquid organic compound existing out of six carbon and six hydrogen atoms, used to produce industry chemicals. The sum formula of benzene C₆H₆ was already known. However, it was yet unclear how the atoms are spatially connected (Read, 1995). According to his own account, Kekulé suddenly discovered the structure of benzene in a day dream when he envisioned a snake biting its own tail (Sternberg & Davidson, 1983; Read, 1995). He suddenly realized that the six carbon atoms formed a closed circle¹. With this structural formula, he significantly contributed to the further development of organic chemistry (Read, 1995; Kekulé, 1866).

There have been different definitions of insight (see Kounios & Beeman, 2014). A commonly agreed upon definition of insight that will be used in this chapter is the sudden comprehension or solution of a problem that involves an AHA! experience (Danek et al., 2020; also see Bowden & Jung-Beeman, 2003). The AHA! experience is defined as the solver’s perception that the solution emerged suddenly, a strong belief that it is correct, and this AHA! experience is often accompanied by a feeling of pleasure (Topolinski & Reber, 2010; Danek & Wiley, 2017). The concept of insight is closely related to creativity which, according to Mednick, is “the forming of associative elements into new combinations, which either meet specified requirements or are in some way useful” (Mednick, 1962, p. 221). According to Dietrich and Kanso (2010), insight problem solving is a special phenomenon of creativity, where a novel solution is discovered by thinking processes that diverge from the routine.

How did Kekulé come up with such an insightful and creative solution and what were the processes that led to an insight? Insight solutions are considered to be largely (but not exclusively) a product of unconscious processing, because when they emerge, they seem to be disconnected from the ongoing stream of conscious thought (Kounios & Beeman, 2014; van Steenburgh et al. 2012). Research on the feeling-of-warmth, that is, the subjective closeness to the solution, suggests that it is precisely this disconnection from conscious thought that subjectively distinguishes an insight from a noninsight solution (Hedne et al., 2016; Kizilirmak et al., 2018; but see Reber et al., 2007 for a theoretical approach). In the latter, the solver has the impression that the solution was reached via incremental awareness like solving an arithmetic equation (Metcalfe & Wiebe, 1987; Kounios & Beeman, 2014).

For this reason, insight has fascinated insight researchers for more than a century now and explaining the underlying processes became the subject of various theories.

---

¹ It is important to note that Kekulé’s anecdote regarding the discovery process of benzene was part of an address he gave at a celebratory dinner in honor of this discovery multiple years later. It is still debated whether this anecdote is true or made up for presentation at that celebratory dinner (see Weisberg, 2006).
that has been used to study insight in the laboratory is the nine-dot problem (Kounios & Beeman, 2014; Danek, Wiley & Öllinger 2016; Öllinger et al., 2014). Note that throughout the chapter we will call tasks used to study insight insight tasks even though these tasks do not necessarily always elicit an insight in the sense of an AHA! experience. The challenge of the nine-dot problem, which is depicted in Figure 1, is to connect nine dots with four straight lines without lifting the pen once. The incorrect assumption that the line cannot be drawn outside the perceived Gestalt of a square is the main source of difficulty in solving the nine-dot problem.

Figure 1. Nine-dot problem: Problem and one possible solution to the nine-dot problem

There are several sources of problem difficulty in insight problem solving that prevent the solver from finding the solution immediately (Kershaw & Ohlsson, 2004). Among other things, these difficulties ought to explain the reason why solvers cannot estimate well how close they are to the solution – one distinguishing feature of insight compared to noninsight problem solving (Metcalfe & Wiebe, 1987, Kounios & Beeman, 2014). It was proposed that an inappropriate search heuristic adopted by the solver or an overly large problem space preventing the solver from searching for the solution exhaustively may be a source of problem difficulty (MacGregor, Ormerod & Chronicle, 2001; Kaplan & Simon, 1990; Weisberg & Alba, 1981). The problem space refers to the number of possible solutions to search for. According to the representational change theory (RCT), the main source of difficulty lies in an overly constrained problem representation (Ohlsson, 1984a,b; Ohlsson, 1992; Knoblich et al., 1999). Here, the solver technically has the required knowledge to solve the problem but adopts an initial mental representation of the problem that is biased due to false or solution-irrelevant assumptions about the task. In case of the nine-dot problem this refers to the incorrect assumption that the lines cannot be drawn outside of the imaginary box of the nine dots. Consequently, the problem space is too constrained and the solution is therefore not obvious (Ohlsson, 1992). An insight occurs when the solver manages to change the initial problem representation (Ohlsson, 1984a,b; Knoblich et al., 1999; Öllinger, Jones & Knoblich, 2014). In the case of the nine-dot task, representational change reflects the realization that the lines can also be drawn outside of the imaginary square of the nine dots.

According to the RCT (Öllinger et al., 2014), the main processes taking place during insight are the following: Upon problem display, the solver internally represents the problem by constructing a problem space. How solvers represent the problem, depends on how they perceive the problem and what they already know about it and similar problems. Once the problem is represented, the solver will search the problem space. Because the current problem space is overly constrained, the solver does not find the solution and runs into an impasse. The latter refers to the inability to generate more solution candidates given the current problem space. Finally, the solver either gives up or manages to generate the solution after changing the problem representation.
Why a neurocognitive perspective on insight?

When neuroscientific methods such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) found their way into psychology, more and more psychological phenomena began to be investigated at the neural level. The goal of this approach is to gain knowledge about brain processes that generate these psychological phenomena. Additionally, this knowledge is assumed to inform psychological theories leading to a more comprehensive and general explanation of these phenomena (Bechtel & Richardson, 2010).

Insight has been studied in great detail for almost a hundred years, with behavioral methods generating a variety of useful cognitive theories such as the RCT (Köhler, 1925; Duncker, 1945; Wertheimer, 1959; Sternberg & Davidson, 1995; Ohlsson, 1992; Weisberg, 2015). The basic goal of a neurocognitive perspective on insight is to also inform and complement such theories, ultimately leading to a better understanding of this phenomenon.

There has been an increasing interest in the neuroscience of creativity and insight in recent years, expanding our knowledge about the brain processes that generate insight in problem solving (Luo et al., 2007; Danek et al., 2016). However, as Dietrich and Kanso (2010) point out, neurocognitive studies investigating insight still provide diverging findings, making it difficult to draw coherent conclusions about the mechanisms of insight (see also Kounios & Beeman, 2014, Sprugnoli et al., 2017). One reason is that insight is most likely composed of an interplay between different cognitive processes. Therefore, insight can only be made tractable in the brain if it is further subdivided into different components that can be meaningfully associated with specific neurocognitive processes (Dietrich & Kanso, 2010). The other reason is that researchers encounter methodological and conceptual obstacles when trying to identify the neural basis of insight (Luo et al, 2007; Danek et al., 2016). It has been difficult to find adequate tasks for the systematic study of insight, because neuroimaging studies greatly depend on precise timing and repeatable behavior in well-controlled tasks across many trials. Classic insight problems like the nine-dot problem are unsuited, because they vary greatly in sources of task difficulty, take several minutes to solve, and consist of only one trial (see Luo et al., 2007).

In recent years, however, various tasks have emerged that have proven to be useful for examining insight neurocognitively (for a review see Sprugnoli et al., 2017), such as Chinese logographs, anagrams, matchstick problems, rebus puzzles, or the (compound) remote associate task (C/RAT).

The RAT and its different versions is the most popular verbal insight task in this context (Bowden & Jung-Beeman, 2003; Becker, Wiedeman, & Kühn, 2020). Note that in this chapter we will not differentiate between the different RAT versions. To solve a RAT problem, participants need to form word associations to correctly retrieve specific semantic knowledge from memory. Three cues are presented (e.g. *pine, pie, sauce*) and the task is to find a word that is associated to all three cues (*apple because pineapple, apple pie and applesauce*). This task elicits an AHA! experience (in about half of the cases) because the solver needs to think of distantly related lexical information to relate all three cues (Bowden & Jung-Beeman, 2003; Becker et al., 2020b). Hence, what is novel to the solver in RATs is the newly discovered association between the cue words via the solution word.

Scope of the chapter

In the current and the next chapter, we will propose a neurocognitive translation of insight’s main cognitive subprocesses from a long-term memory perspective, and summarize them in a descriptive framework (see Figure 2). In order to ensure a precise description of these processes and
meaningfully interpret findings from cognitive neuroscience, we decided to mostly restrict our scope
of explanation to verbal insight. The current chapter will concentrate on the cognitive and related
neuronal processes from the presentation of the problem until the solution is found. Inspired by RCT,
those subprocesses relate to: (1) problem representation; (2) search; (3) representational change, and
(4) solution. To illustrate those processes, we will mostly use the RAT as an example for describing
how the solution concept is retrieved from memory. The next chapter “A cognitive neuroscience
perspective on insight as a memory process: Encoding the solution” will concentrate on affective and
related neuronal processes of insight after the solution has been found.

Because the neurocognitive framework presented here has a specific focus on verbal insight, mostly
using the example of the RAT, this chapter is not a review in the classical sense and will not cover
the entire neurocognitive literature on insight (for more general neurocognitive reviews on insight,
see Kounios & Beeman, 2014; Shen et al., 2017; Sprurgoni et al. 2017).

**Figure 2.** Neurocognitive framework for verbal insight.

*Note. This framework roughly summarizes the neurocognitive processes of verbal insight described in this and the next
chapter: Once the problem is displayed, problem elements (concepts) and a whole network of associated concepts are
automatically activated in semantic memory (=problem representation). What gets activated depends on prior knowledge
and perception next to other factors. Automatic spreading activation processes continuously activate concepts in this
network. Additionally, the solver starts actively searching for the solution (=controlled search) retrieving possible
solution candidates. In most cases, too many solution-irrelevant and too few solution relevant concepts are activated in
memory (constrained problem representation) causing neither automatic activation nor controlled search processes to
successfully activate the solution. For this reason, enhanced cognitive/semantic control is necessary to change the
activation pattern in semantic memory (=representational change). Subsequently, insight is assumed to happen, when
automatic processes suddenly activate a solution candidate that is hereupon evaluated positively (i.e., it suffices all task
constraints). The positive evaluation triggers a whole cascade of additional processes (e.g., novelty detection, intrinsic
reward, positive affect) supporting encoding of the solution in long term memory. Those processes are further detailed in
the next chapter. Note, the timing of the proposed processes is not proportional to the length of the timeline; some
processes may happen simultaneously others extend over a longer period of time or may even be repeated.

2 Insight as a memory process – a neurocognitive perspective

Before describing the different insight solution steps from a neurocognitive perspective, it is
necessary to briefly characterize the semantic memory system and how concepts or ideas are
assumed to be represented in the brain because all further descriptions are based on it.
Semantic memory is a form of long-term memory including all acquired knowledge about the world, such as concepts or ideas, and it is believed to be stored in a complex network across the cortex (Tulving, 1983; Binder et al., 2011; Quillian, 1966; Dancker & Anderson, 2010). Concepts are the essential unit of word meaning and refer not only to the referents of words but also to their constituting meaningful features and to concepts that can only be described using several words. Priming experiments indicate that the semantic memory system is organized in a principled way, facilitating access to words during language production and comprehension (Collins & Loftus, 1975). One principle is that words used more frequently in language are accessed more quickly compared to words with lower frequency. Another principle is that words with similar meanings are organized together in the brain to facilitate access to related items (Huth et al., 2012). In fact, not only similar words but all kinds of knowledge, including conceptual information or life events are assimilated and aggregated in such a way to be quickly and easily accessed (Marshall, 1995). This principled structure can, in part, be explained by how concepts and ideas form and are represented in the brain.

On a cellular level, concepts or ideas are represented in the brain by simultaneous activation of a network of neurons – cell assemblies – which are interconnected by many or strong mutual excitatory synapses (Hebb, 1949; Kiefer & Pulvermüller, 2012). What determines a cell assembly is the connectivity structure between cells, i.e. how likely cells become co-activated in response to a concept or an idea (Wennekers & Palm, 2000). The likelihood for two cells becoming co-activated increases when the presynaptic cell continuously and persistently activates the postsynaptic cell (Hebb, 1949). This associative learning principle may not only explain how concepts in terms of new cell assemblies become encoded in long-term memory but also how different concepts as well as schemas become associated to each other. Schemas refer to meaningfully interrelated semantic information that influences the way people acquire and store information in memory (Marshall, 1995). Similar to concepts, schemas are assumed to be represented in the brain by simultaneous activation of a network of cell assemblies (Ghosh & Gilboa, 2014; Rumelhart et al., 1986, p.21). When two concepts or other kinds of knowledge elements are often co-activated within the same context (e.g., doctor, hospital), the cell assemblies representing each concept or knowledge element develop a stronger mutual excitatory connection. This in turn increases the likelihood that the activation of one cell assembly (e.g., representing the concept doctor) automatically activates an associated cell assembly (e.g., hospital).

However, the exact nature of how concepts, ideas or schemas are organized in the brain is still not known. There has been a longstanding debate in cognitive science about whether single neurons belong to one or more cell assemblies representing a concept. However, it is generally agreed that concepts are represented as networks of cell assemblies including different features which are coded in distinct areas of the brain (Kiefer & Pulvermüller, 2012; Davis et al., 2020; vanKesteren et al., 2012). Importantly, concepts as well as schemas are flexible mental entities whose features or different pieces of information are activated depending on prior knowledge, perception and situational constraints (see Section 2.1; Kiefer & Pulvermüller, 2012).

The ability to flexibly retrieve specific conceptual knowledge from semantic memory is important not only for language in general but also to think creatively and solve problems (Binder et al., 2009). When trying to solve a RAT problem, the solution word is technically known to the solver but she/he...
has difficulties retrieving it from semantic memory\(^2\). What determines difficulty in those seemingly simple verbal insight problems and how are these kinds of problems solved?

### 2.1 Problem representation – activation in semantic memory network

Based on neuroscientific knowledge of how concepts are stored and interconnected in the brain, the cognitive term \textit{problem representation} is translated into a neurocognitive one as \textit{activation of nodes in a semantic memory network}. When a verbal insight problem is presented to the solver, it is assumed that the problem elements and a complex network of associated concepts and schemas are automatically activated from semantic memory via spreading activation (see Figure 2; Bowden, 1997; Bowden & Beeman, 1998; Becker et al., 2020b; Anderson, 1983; Bowers et al., 1990; Cai et al., 2009). The \textit{problem elements} refer to all perceivable aspects of that problem (concepts, ideas or objects) that are presented to the solver during problem display. Importantly, not every associated concept gets activated equally but the activation strength highly depends on the amount of excitatory connections that the problem elements share with their associated concepts. In congruence with Hebb, the more frequent associations are assumed to be activated first (Simpon & Burges, 1985; Zempleni et al., 2007). In the case of the RAT, this means that not only the problem elements (\textit{pine, pie, sauce}) but also their frequent and semantic associations (e.g., \textit{pine} \rightarrow \textit{tree, wood}) are automatically co-activated. The concepts that are activated most strongly make up the current problem space where the solver will search for the solution (Becker et al., 2020b; Öllinger et al., 2014). However, the solution is often not a frequent but a more remotely related associate of all three cues. For this reason, it is less likely to be multiply co-activated by the cues and therefore less likely to be retrieved compared to the competing frequent associates (e.g., \textit{tree}; Bowers et al., 1990). Hence, the current problem space is too constrained in these kinds of problems because too many solution-irrelevant and/or too few solution-relevant associates of the cues are activated first. As a consequence, the solution seems non-obvious to the solver (Bowden et al., 2005; Reiter-Palmon et al., 1997).

In sum, what determines difficulty in solving RATs is the ability to detect the weak associative relationship between the cues and the solution. This ability and how the problem is internally represented by the solver are determined by several factors (Kershaw & Ohlsson, 2004). The most prominent factors are prior knowledge and perception (see Figure 2). However, also inter-individual differences regarding how those associative relationships are represented in the individual brain appear to influence problem representation.

\textit{Prior knowledge.} Prior knowledge relates to concepts and schemas that are associated with the problem elements modulating what will get activated upon the presentation of the problem due to their associative strength. For example, presenting a prime (e.g., \textit{air vs. tear}) before presenting an associated and ambiguous cue (e.g., \textit{drop}) in a RAT will bias an individual to more strongly represent one specific meaning over another (\textit{drop} as in \textit{teardrop} instead of \textit{drop} as in \textit{airdrop}) (Becker, Wiedemann & Kühn, 2020). Hence, a set of features representing the concept \textit{drop} will be preferably activated (liquid quantity) in the brain as a function of the prime. Becker and colleagues used this priming effect to specifically bias participants’ problem representation when solving a modified RAT problem (Becker et al., 2020). Together with the three cue words (\textit{drop, coat, proof}), the participants received a prime (\textit{air vs. tear}) that was related to the first cue word. The function of the prime was to

\(^2\) The expression “retrieving the solution from memory” translates to “finding the solution within the problem space” in cognitive insight theories.
either increase the amount of activated solution relevant (teardrop→raindrop) or irrelevant (airdrop→raindrop) associations. They found that increasing the likelihood of activating the solution-irrelevant compared to relevant associations of the primed cue significantly decreased the likelihood to retrieve the weakly associated solution word.

**Perception.** Our senses translate information about the environment into neuronal signals from which the brain generates perceptions (i.e., mental representations of the original stimulus) which enable us to build a representation of the world (Gazzaniga et al., 2019). Different visual features of an object such as its color or shape are processed independently of each other in early to late visual areas of the brain but they are integrated into a coherent representation (Gazzaniga et al., 2019). However, how we perceive objects, that is, which visual features of objects get chunked together into a coherent representation, is tightly coupled with experience and ultimately memory. *Perceptual chunking* refers to the process that binds certain visual features of objects through association (Miller, 1956; Knoblich et al., 1999). For example, through a lot of reading practice, we learn to chunk individual shapes to single letters and further chunk those letters to a word. This word is ultimately perceived as a single entity carrying meaning. The activation of perceptual chunks upon stimulus presentation is an automatic process and increases efficiency of information processing; for example, due to chunking reading is very fast (Miller, 1956). However, when faced with a new problem, perceptual chunking can also increase problem difficulty in perceptual insight tasks such as matchsticks problems. Note that perceptual chunking is not a source of difficulty in verbal insight such as RATs.

Another factor determining solution difficulty is how those associative relationships are represented in the individual brain. According to Mednick (1962), individuals with “flat” (broader) associative hierarchies are more likely to think of infrequent word associations compared to individuals with “steep” (few, common) associative hierarchies. In fact, there is evidence that the properties of a person’s simulated semantic network correlate with their RAT performance (Kenett et al., 2014). That is to say, the likelihood to solve a RAT item increases, the more directly two nodes are interconnected in the network, that is, when every node in the network takes just a few associations to reach any other node in the network (Kenett et al., 2014).

### 2.2 Search–retrieval of solution candidates

Once the problem is displayed and internally represented, the solver will dynamically search the problem space (Öllinger et al., 2014). An interplay between two kinds of processes is assumed to be involved in the search phase: automatic and controlled processes (see Figure 2; Ben-Zur, 1989). Some theorists have focused on only one of both types of processes (see Barr, 2018). However, there is a growing consensus that both of them play an important role in insight and creativity in general (Beaty et al., 2015; Wiley & Jarosz, 2012). Several dual-process models have been proposed, although the precise interplay between automatic and controlled processes has still not been clarified (Sowden, Pringle, & Gabora, 2015). A promising approach to fill this gap comes from computational cognitive neuroscience, formalizing the (verbal) insight process as the result of prior automatic and control processes in a semantic memory network, whose assumptions can be directly tested empirically (see Helie, 2010; Kajic et al., 2017; Olteanu et al., 2015).

**Automatic process.** This process refers to the spreading of activation across the semantic memory network and is assumed to take place subliminally, that is, below the threshold of the solver’s consciousness (see Figure 2). As mentioned in Section 2.1, when an insight problem is presented to the solver, not only are cell assemblies representing the problem elements automatically activated in
semantic memory but also their associated concepts. The spreading activation account of semantic memory proposes that the activation continues from the starting node (i.e., the cues in the RAT) over close associates to associations more and more remote until activation fades out (Anderson, 1983; Collins & Loftus, 1975). Solution candidates that are retrieved while searching for the solution are also believed to activate their close associates, further spreading activation through the semantic memory network.

Usually, the problem space is too large or biased in an inadequate direction, that is, too many solution-irrelevant or too few solution-relevant associates of the problem elements are activated, preventing this automatic process to activate the solution. However, it is assumed that there are cases specifically in RATs, where automatic processes alone can activate the solution without prior involvement of controlled processes because the problem space is not large enough or sufficiently biased (Becker, Kühn, Sommer, 2020). In this case, spreading activation from the cues or their associated concepts is strong enough to activate the solution (Bowers et al., 1990; Becker, Kühn, Sommer, 2020). If the activation exceeds a certain threshold, such that the cell assemblies representing the solution concept start firing, the solution may “pop” into the solver’s consciousness. RAT problems that are assumed to be solved mainly via automatic processes, giving rise to so-called pop-out solutions, are usually solved very fast, within the first 2-3s after problem representation (Becker, Kühn, Sommer, 2020; Nowick & Sherman, 2003). However, even though these pop-out solutions involve an AHA! experience – one indicator for insight – it is debated to what extent they represent an actual insight (Becker, Kühn, Sommer, 2020; Spiridonov, Loginov & Ardislamov, 2021). One of the reasons is because pop-out solutions do not involve representational change – one assumed cognitive key feature of insight – and have a different neurocognitive signature, including less memory search compared to solutions that do involve representational change (Becker, Kühn, Sommer, 2020; see also Section 2.3).

Controlled process. The controlled process in this context relates to the conscious semantic memory search for the solution (see Figure 2). Importantly, memory search in insight tasks is different compared to typical memory retrieval (Davelaar, 2015; Raaijmakers and Shiffrin, 1981; Hills et al., 2012). For example, regarding the latter, if one is asked which sea animal is a mammal, one could generate a few animal candidates (e.g., tuna, shark, whale) and then narrow the search by excluding false answers (e.g., the tuna lays eggs and has gills) until the solution is found. In contrast, in case of the RATs, it is difficult to organize a memory search because the topic of the answer is unknown. Whereas in the sea animal example one knows the answer is a mammal’s name, in the RAT problem pine/pie/sauce, it is not obvious how the words are related (e.g., by means of an overarching category, a common feature, or a common compound word). So how does a solver search in semantic memory in this context? Subjects produce solution candidates primarily on the basis of just one of the three cues at a time with occasional switching between the cues (Davelaar, 2015; Smith et al., 2013). Additionally, subjects adopt a local search strategy, that is to say, they generate new solution candidates based on their previous responses (Smith, Huber & Vul, 2013; Smith & Vul, 2015). Davelaar (2015) provides evidence that in addition to searching for solution candidates among associates that are closely related to each individual cue in RATs, participants also search for items that are associated with all three cues concurrently. The author proposes that the presentation of all three cues activates an overlapping memory pattern automatically that indirectly guides participants’ search behavior (Davelaar, 2015). This overlapping memory pattern describes those cell assemblies in the network that are most strongly activated as a function of all three cues. Hence, in line with a dual-process model, Davelaar implies that automatic processes (i.e., cell-assemblies automatically activated by the cues) guide active search for solution candidates.
Most studies investigate neural correlates of insight during solution retrieval. Little research has been conducted on the neural correlates of the search processes. In one study, the left inferior frontal gyrus (IFG) and left middle temporal gyrus (MTG) were identified as neural correlates for controlled semantic memory search while solving modified RATs (see Figure 3; Becker et al., 2020a). This finding is in line with the assumed role of the left IFG to guide controlled retrieval of semantic information (Wagner et al., 2001; Gold & Buckner, 2002). Language related studies repeatedly found the IFG to be related to semantic control in lexical decision tasks (Badre et al., 2005, 2007; Whitney et al., 2010a; Hampshire et al., 2010). Classically, the left MTG has been related to storing conceptual representations (Binder & Desai, 2011). However, it is often co-activated with the left IFG during retrieval of semantic information and may therefore also be associated with semantic control (Davey et al., 2015). Hence, either the left IFG controls retrieval of solution candidates stored in left MTG or both brain regions control retrieval of solution candidates represented across the cortex.

Figure 3. Neural correlates during search and solution while solving modified remote associates (adapted from Becker et al., 2020a).

Note. Left panel: Left IFG and MTG were more strongly activated while searching for the solution in modified RATs (generating potential but wrong solution candidates) compared to having found the final solution (generating the final solution). Right panel: Relative brain activity at the time of solution parametrically correlated with a variable quantifying representational change (this variable quantified how much an ambiguous cue word that had been primed towards a solution-irrelevant meaning had to be reinterpreted to build a meaningful compound word with the solution).

2.3 Representational change – cognitive control

While subjects are searching for the solution, they will often continuously fail and eventually experience an impasse, because the solution is not part of the problem space (Öllinger et al., 2014) or the latter is too big to find the solution.\(^3\) In the case of RATs, this translates to retrieving frequent but solution-irrelevant associations of the cues from memory (pine forest, pine tree, pine oil etc.) when less frequent associations (apple as in pineapple) are solution-relevant (Bowden et al., 2005). According to RCT, participants subsequently either give up or manage to find a solution by changing

\(^3\) Note that there are cases where neither impasse nor representational change are necessary to lead to an insight (see Fleck & Weisberg, 2004; Shen et al., 2019).
their problem representation\(^4\) (Öllinger et al., 2014), that is, changing the pattern of currently activated cell assemblies in the network (see Figure 2). One way to change the problem representation is by overcoming constraints caused by perceptual chunks\(^5\) or prior knowledge (Knoblich et al., 1999; Ohlsson, 1992; Öllinger et al., 2014).

**Overcoming constraints caused by prior knowledge.** For verbal insight, this process is assumed to involve enhanced semantic control in the verbal domain (Becker et al., 2020a). Semantic control refers to cognitive control directing the activation of semantic knowledge in a task-appropriate manner (Harvey et al., 2013). Cognitive control generally refers to the ability to regulate and coordinate thought and actions in accordance with internally maintained behavioral goals, and it is assumed to primarily depend on prefrontal cortex function (Miller, 2000). Semantic control is relevant in situations when automatic semantic retrieval is not enough, for example, because task-relevant information is only weakly activated (Badre et al., 2005) or only a subset of task-relevant information must be selected over a competing subset of irrelevant information (Noonan et al., 2010, 2013; Thompson-Schill et al., 1997; Whitney et al., 2010b). Retrieval and selection of solution-relevant semantic information in RATs is also difficult because initially, the solution is only weakly activated, while many solution-irrelevant associates are activated first due to their stronger associative strength with the cue words. Therefore, overcoming constraints caused by prior knowledge refers to: (a) inhibition of frequent but solution-irrelevant associations (e.g., tree as in pine tree); and (b) switching the attentional focus to infrequent but solution-relevant associations (e.g., apple as in pine apple; Becker et al. 2018; 2020a,b; Bowden et al., 2005; Dow and Mayer, 2004). Consistent with this idea, performance on RATs increases the better the solver is able to avoid retrieving solution candidates that are highly frequent associations of the cues (Gupta et al., 2012). Note, translating abstract cognitive processes such as representational change into more neurocognitive ones such as cognitive control reflecting inhibition and attentional shifts allows us to further decompose insight into processes that we already understand to a good extent (see Becker et al. 2020a,c). For example, the neuronal processes underlying cognitive control are well investigated using other cognitive tasks and better understood compared to representational change (Botvinick et al., 2001; Badre, 2008).

Studies investigating verbal insight reported that, as part of a wider distributed network, prefrontal areas such as the IFG are activated during the solution period (Becker et al., 2020a; Jung-Beeman et al., 2003; Luo, Niki and Phillips, 2004a, 2004b; Sandkühler & Bhattacharya, 2008; Tik et al., 2017; Zhao et al., 2013). However, very few studies have attempted to directly measure representational change for verbal insight. Becker and colleagues (2020a) manipulated the amount of representational change necessary to solve a modified RAT and identified bilateral IFG and the ACC together with other cortical und subcortical brain areas (see Figure 3). In this context, the authors attribute the function of the IFG to inhibiting solution irrelevant associations and retrieving the weakly activated solution (Becker et al., 2020a). These results are in line with the assumed general function of the left IFG, which is not only to retrieve semantic information, but also to select among competing semantic

\(^4\) Note that changing the problem representation is not the only strategy that can lead to insightful problem solving (for a more thorough discussion on this matter, see Ash et al., 2009).

\(^5\) Overcoming constraints caused by perceptual chunks refers to breaking up perceptual chunks into their elements (e.g., V \(\rightarrow\) V/) to make them available for novel recombination (Tang et al., 2015). In the context of problem solving, the shift in meaning introduced by the reorganization of perceptual elements can open up new possibilities to solve problems that seemed impossible to solve (see matchstick problems or the chunk decomposition task using Chinese characters, Knoblich et al., 1999; Luo et al., 2006; Wu et al., 2009, 2010, 2013). However, this cognitive control process refers to perceptual (not verbal) insight and will be therefore not explained in detail here.
alternatives held in working memory and resolve interference (Badre et al., 2005; Thomson-Schill et al., 1997, 1999). Note that IFG is not the only region involved in inhibition when overcoming constraints due to prior knowledge in verbal insight. For example, Di Bernardi Luft and colleagues (2018) provide robust evidence that alpha oscillations measured over the right temporal gyrus – a modality-specific brain region not classically related to cognitive control – also play a critical role in inhibiting obvious semantic word associations in RATs (see Section 2.2).

The anterior cingulate cortex (ACC) is often found in insight studies during solution (for a meta-analysis see Dietrich & Kanso, 2010). Its general function is attributed to conflict monitoring (Botvinick et al., 2004; Kerns et al., 2004). In the context of verbal insight, ACC activation during solution is interpreted as shifting the attentional focus to infrequent but solution-relevant associations (Becker et al., 2020a; Luo et al., 2004b; Zhao et al., 2011). Hence, both cognitive control regions, ACC and IFG, might be differentially involved in overcoming constraints caused by prior knowledge in insight. While the IFG may refer to inhibition of frequent but solution-irrelevant associations, the ACC may refer to switching the attentional focus to infrequent but solution-relevant associations. This assumption is in line with studies demonstrating that the IFG seems to respond specifically to item difficulty in verbal insight tasks with higher levels of activation in difficult compared to simple insight trials (Becker et al., 2020c; Luo et al., 2004a). The necessity to inhibit solution-irrelevant associations and select among competing alternatives should increase with increasing item difficulty. In contrast, when controlling for difficulty, the ACC is selectively activated in all insight trials (that involved representational change) compared to noninsight trials (Becker et al., 2020c; Luo et al., 2004a). Because the solution is usually an infrequent association of the cues, shifting the attentional focus to infrequent associations is always necessary for solving RATs and should therefore be independent of item difficulty.

2.4 Solution – generation and evaluation of solution candidate

Successfully inhibiting solution-irrelevant associations and switching the attentional focus to solution-relevant ones increases the likelihood for automatic processes to activate the weakly associated item that is the solution of the problem (Becker et al., 2020c; Bowers et al., 1990). Once this activation exceeds a certain threshold, the neurons representing the solution concept start firing and the solution is assumed to reach the solver’s awareness. Although the neural correlates of consciousness are a topic of ongoing debate (Koch et al., 2016), it is reasonable to assume that the firing of cell assemblies is causally and temporally associated with participants becoming aware of the solution. At this point, increased BOLD activity in the (right) lateral temporal cortex and a burst of gamma-band EEG activity over the same region have been found and have been interpreted as neural correlates for the insight experience, that is, when the weakly associated solution suddenly emerges into conscious awareness (Jung-Beeman et al., 2003; Kounios & Beeman, 2014). As already mentioned, conceptual representations are encoded in different brain regions across the cortex (Kiefer & Pulvermüller, 2012). In verbal insight studies, the activation of the weakly associated solution has

---

6 Note that there is evidence that the attentional focus as represented by the salience network including the ACC can even influence problem processing and the likelihood to solve a problem via insight in RATs before or during problem representation (Becker et al., 2020b; Kounios et al., 2006; Mennon & Uddin, 2010).

7 An exception are problems with pop-out solutions (see Section 2.2).

8 Gamma-band activity is assumed to be mechanism for binding information as it emerges into consciousness (Engel & Singer, 2001).
been linked to the lateral temporal cortex (Jung-Beeman et al., 2003; Shen et al., 2017; Tik et al., 2017). In particular, evidence suggests that the middle\(^9\) and superior temporal gyri represent, access and process abstract conceptual representations during insight (Binder & Desai, 2011; Shen et al., 2017).

Importantly, a potential reason for why the insight solution is experienced as sudden and seemingly disconnected from the ongoing stream of conscious thought is that the solution is activated by automatic processes\(^10\) (see Becker et al., 2020b,c; Kounios & Beeman, 2014). Solution retrieval via automatic processes distinguishes insight from noninsight solutions. The latter is not experienced as sudden and is assumed to be retrieved via conscious semantic memory search in a more trial and error manner (Becker et al., 2020c; Kounios & Beeman, 2014).

What further differentiates the solution from the solution candidate is how the retrieved information from memory is evaluated. The evaluation is a metacognitive process, where the solver checks whether the solution candidate meets all solution criteria. In the case of the RAT, the word apple counts as a solution because it is conceptually related to all three target words (pine, pie, sauce). If the solution candidate does not meet all task constraints, then the candidate is discarded and the search process starts again (see Figure 2). However, there are currently no studies specifically investigating the neural correlates of solution evaluation in insight.

Finally, retrieving the correct solution triggers a whole cascade of additional processes related to insight and further processing of the solution. One defining criterion for insight is the positive emotional response for suddenly having found the solution in the form of an AHA! experience. This positive emotional response during solution has been associated with dopaminergic midbrain structures such as the ventral tegmental area and substantia nigra as well as their direct projections such as the reward-related ventral striatum and the nucleus accumbens (Tik et al., 2017). Additionally, the hippocampus has been reported to be more strongly activated during insight relative to noninsight solutions (see Shen et al., 2017). Its role in the insight process is still a matter of debate but increasing evidence suggests that the hippocampus signals novelty in response to the newly discovered solution (Cabeza et al., 2020; Kizilirmak et al., 2016; Luo & Niki, 2003). The next chapter will further elaborate the roles of the emotional response and novelty detection concurrent with the insight solution, and their neural correlates, for learning by insight.

3 Summary

In the present chapter, we described the main cognitive processes as identified by RCT leading to the retrieval of the solution from a neurocognitive perspective, with a main focus on the RAT. That is to say, we drew the connection between the cognitive subprocesses of insight (problem representation, search, representational change and solution) and neurocognitive processes. On this basis, we devised a descriptive framework of insight from a memory perspective as illustrated in Figure 2. From this point of view, insight is the consequence of a problem-solving process where the solution is encoded

\(^9\) Note that the left MTG is often co-activated with the left IFG during retrieval of semantic information and may therefore not only represent storage of conceptual information but also play a role in cognitive control (Davey et al., 2015).

\(^10\) However, the weakly associated solution could not have been retrieved via automatic processes without prior enhanced cognitive control to overcome constraints caused by prior knowledge or perceptual chunks (Becker et al., 2020). Exceptions are pop-out solutions (see Section 2.2, automatic processes).
in long-term memory but cannot be retrieved at first. The retrieval process is difficult because the
topic of the answer is unknown, precluding a simple, controlled memory search. Retrieval is also
difficult due to constraints caused by prior knowledge or perception during problem representation
(next to other factors) hindering automatic processes to activate the solution. Insight is assumed to
occur when automatic processes activate the solution after the problem representation is changed,
that is, after control processes manage to overcome prior knowledge and/or perceptual constraints.
Because automatic processes operate below the level of conscious awareness, the solution seems to
emerge suddenly. Importantly, however, substantial (automatic and controlled) processing precedes
solution retrieval. Insight-related changes in BOLD activity before and at the beginning of the
solution process (after problem representation) provides further empirical evidence that insight is not
just a sudden solution at the end but precedes substantial prior processing (Becker et al., 2020a,b,
Kounios et al., 2006). In the following chapter (Kizilirmak & Becker, this volume), affective and
cognitive processes succeeding solution retrieval will be described in more detail, using the
neurocognitive framework as presented here. Additionally, the chapter will further focus on the role
of insight problem solving for long-term memory formation.

It is important to note that it is usually assumed that an insightful solution or a creative outcome is
novel to the respective individual, such as Kekulé’s discovery of the benzene ring was new to him.
This seems to contradict our claim that the solution itself is already encoded in long-term memory in
the case of RATs. Importantly however, what is novel to the solver in RATs is not the solution word
itself but the newly discovered association between the cue words via the solution word.

4 Open questions and future directions

In sum, a neurocognitive memory perspective provides a promising approach for exploring how a
solution suddenly emerges in the solver’s mind. However, despite the comprehensive description of
verbal insight given in this chapter, several questions still remain open. For reasons of brevity, we
will only name a few of them.

In contrast to Kekulé’s insight being purely visual, the description of insight presented here as a
spreading activation account in a semantic memory network is inherently language-oriented. Hence,
remains to be shown to what degree these neurocognitive processes (e.g., automatic and controlled
processes in long-term memory storing perceptual representations) can be generalized to Kekulé’s
perceptual insight (see Mayer, 1995 for a description of different types of insight). Another limitation
is that neuronal correlates of control processes, such as representational change and search, have only
been investigated using verbal tasks such as the RAT. Thus, even though cognitive control processes
and their associated neuronal correlates are assumed to be modality-independent, their involvement
in perceptual insight remains to be demonstrated. One promising approach would be to directly
compare the neural correlates during problem solving in RATs with their matched visual counterparts
(Becker & Cabeza, 2021; Olteteanu et al., 2020).

As we have shown using the example of RATs, insight is a complex process with different
interleaved automatic and control processes happening almost simultaneously. Specifically, during
solution, it is currently difficult to separate control processes related to representational change,
retrieving and evaluating the solution. This separation is made more difficult by the fact that the
hemodynamic changes measured in the MRI are slow and arise within the seconds range. Hence,
identified brain processes related to representational change in the RAT may be confounded with
evaluation or retrieval processes. More sophisticated experimental manipulations including methods with higher temporal resolution are required to tease these processes apart.

We described the general processes of representational change in terms of cognitive (semantic) control including an attentional shift from solution-irrelevant to relevant problem elements. However, it is currently unknown what exactly causes this attentional shift and how the brain “knows” which solution-irrelevant problem elements to inhibit. In this context, the role of neurotransmitters such as dopamine and noradrenaline has been discussed. First evidence suggests that moderate levels of dopamine in the striatum and prefrontal cortex facilitate insight and other types of creative processes (for a review see Boot et al., 2017). Furthermore, increased levels of noradrenaline in the cortex have been associated with cognitive flexibility in problem solving and creativity (Beversdorf et al., 1999). Cognitive flexibility is closely related to what we refer to as representational change in insight and it represents the ability to shift attention in order to process or respond to situations in different ways (Eslinger & Grattan, 1993). Specifically, elevated levels of noradrenaline in the prefrontal cortex are associated with a state of reduced attention on well-established cues or strategies and increased attention to novel or previously unattended cues (Aston-Jones, Rajkowski, & Cohen, 1999; Lin & Vartanian 2018; Sadacca et al., 2017). However, further evidence is necessary to substantiate those claims and link different neuromodulatory systems to subprocesses of insight.
References


Cognitive Neuroscience of Learning by Insight


Miller, E. (2000). The prefrontal cortex and cognitive control. *Nature reviews neuroscience*, 1, 59-65. [https://doi.org/10.1038/35036228](https://doi.org/10.1038/35036228)


