Chapter 3

Methods for Studying Human Thought

ARNDT BRÖDER

University of Mannheim

3.1 Introduction

As the other chapters of this book will reveal, the psychology of thinking is a fascinating research field which has discovered a lot of surprising insights into this faculty of the human mind. Overcoming the problems associated with investigating something "invisible" such as thoughts is an interesting philosophical problem and a research topic in itself. This chapter will start with the methodological foundation of cognitive psychology and the question as to why scientists do not just rely on people's reports about their thoughts as data. Then, I will provide an overview of the toolbox of methods that cognitive psychologists have developed for discovering insights into thinking. Most methods will be illustrated by one or two selected examples, but it should be kept in mind that the range of possible applications is much broader. There is no recipe as to how to do research on thinking, so psychologists can still be creative in developing new methods and in freshly combining old ones. This methodological challenge is one further aspect which makes research in cognitive science so intriguing.

Readers who want to recapitulate a few basics on the methods of psychology may want to consult Textbox 3.1 first.

Textbox 3.1: A brief primer of basic methods in empirical psychology

Psychological laws or hypotheses typically claim that one **independent variable (IV)** has some influence on another variable called the **dependent variable (DV)**. For example, it may be claimed that the more "deeply" information is processed, the better it will be remembered later (Craik & Lockhart, 1972). Here, the depth of processing is the IV, whereas memory performance is the DV. Theoretical psychological variables are themselves unobservable, but they may be **operationalized** by translating them into observable variables which are thought to represent the theoretical ones. For example, a shallow processing of information could entail counting the letters of written words, whereas deep processing is based on analyzing the meaning of the words. Likewise, memory performance may be measured by tallying the words someone can recall in a later test. If the hypothesis (or law) is true and the operationalization is adequate, both variables must show a covariation. Empirical tests of psychological hypotheses therefore assess whether such a predicted correlation exists. In a **correlation study**, researchers measure or observe both variables of interest and assess their covariation. However, the correlation in such a study does not allow the conclusion that the IV variation *caused* the DV change since they might both be influenced by a third variable.

https://doi.org/10.17885/heiup.470.c6666

For example, the motivation of a participant might influence both the learning strategy and the memory performance without a direct causal link between these variables. To test *causal* hypotheses, scientists try to run **experiments** whenever possible. Here, they can actively *manipulate* the IV (for example by instructing participants either to count letters or to find a meaningful associate to words). If participants are randomly assigned to the different experimental conditions (so that there are no systematic differences between them), an observed change in the DV has probably been *caused* by the variation in the IV. Experiments are therefore stricter tests of causal hypotheses than correlation studies.

3.2 A Natural Science of the Mind?

How can thoughts be studied scientifically? When reflecting on the natural sciences, we imagine researchers investigating *things* that can be *observed* or even *measured* in objective and precise ways. Thoughts, however, come as beliefs, imaginations, intentions, logical inferences, fantasies, insights, daydreaming, or plans, to name only a few of the many concepts associated with thinking. These immaterial "things" do not have a weight or size or electric charge that can be measured with physical instruments[1]. Furthermore, these thoughts are unobservable for outsiders and hence, they seem to evade an objective description.

Since they considered verbal reports based on so-called **introspection** (self-observation) as unreliable sources of data, philosophers and even the founder of Experimental Psychology, Wilhelm Wundt (1832-1920, see Figure 3.1), were convinced that higher cognitive processes like memory and thinking could not be studied with the methods of the natural sciences. Beginning with John B. Watson's (1913) "behaviorist manifesto", all internal psychological processes including thoughts were abandoned from scientific psychology for a few decades because verbal data were considered as subjective and thus not suited for scientific research (see Chapter 2, "History of the Field of the Psychology of Human Thought").

This state of affairs was unfortunate because in his groundbreaking experimental investigations of human memory, the German psychologist Hermann Ebbinghaus (1850–1909) had already shown how higher cognitive processes can be studied objectively without using subjective verbal reports as data. In principle, the methodological idea behind modern cognitive psychology foreshadowed by Ebbinghaus (1885) is simple: although cognitive processes like thoughts or memory traces are by themselves unobservable, they may lead to *observable consequences* in behavior which can be objectively noticed and described by different independent observers. Hence, hypotheses about these hidden or latent processes can be tested by setting up experiments and observations that target these predicted consequences of behavior as objective data. To use an example from memory research as founded by Ebbinghaus (1885), we may postulate that during the learning of new materials, these leave a hypothetical "trace" in memory which may vary in strength. This trace itself is unobservable, but one can show that it is "there", for example, when people are able to reproduce the material in a later memory test or even show faster responses to these stimuli in comparison to control stimuli they had not learned before. The test results (amount of recall or speed of reaction) are **indicators** of the memory strength, and they can serve as objective data for testing hypotheses about it. In the study of thinking, for example, the number of solved

[1] Most psychologists including myself believe for good reasons that all thoughts have a *material basis* since they strictly depend on processes in the brain. However, a belief or an insight, for example, have a psychological surplus dimension (a *meaning*) that cannot hitherto be reduced to electrical and chemical processes in the brain (some say it never will). The psychology of thinking benefits a lot from knowledge about the brain (see section 4.2.6), but it deals with the *semantics* (meaning) of thoughts in human behavior which is exactly this surplus dimension on top of the physical processes.

test items may be an indicator of a certain facet of intelligence (see Chapter 14, "Intelligence"), or the response to a logical puzzle may indicate whether someone followed the laws of logic or rather an intuitive sense of credibility of the conclusion's content (see Chapter 7, "Deductive Reasoning", belief bias).[2]

Hence, as in other natural sciences, psychologists can test hypotheses about unobservable variables by objectively observing or measuring their behavioral consequences. As the American psychologist Edward C. Tolman (1886-1959) argued, this kind of research strategy (later called **methodological behaviorism**) allows both (1) to use unobservable theoretical concepts in a scientific manner and (2) to do so without recourse to questionable introspective data. Basically, this view is still the methodological basis of modern cognitive psychology.

3.3 Why not just Ask People about their Thoughts?

Reading this introduction, you may wonder why psychologists do things in such a complicated way. Why don't we just ask the people about their thoughts to investigate thinking? They know best, don't they?

In fact, one of the first heated methodological debates in the then young science of Experimental Psychology was between Wilhelm Wundt (1907; 1908) and Karl Bühler (1908) about the value of *introspection* as a means of investigating thinking. Introspection literally means "viewing inside" and was used, for example, by psychologists of the Würzburg School of Psychology to gain insights into thought processes. Confronted with a thinking problem, the test person was asked to observe her own thinking processes and later report them to the researcher. In rare agreement, both Wundt (1907; 1908) and the founder of behaviorism, John B. Watson (1913), criticized the "interrogation method" as unscientific for the following reasons, still accepted by most psychologists today (see Massen & Bredenkamp, 2005; Russo, E. J. Johnson & Stephens, 1989):

1. Instrospection is prone to *memory errors*,
2. many thoughts cannot easily be *verbalized* (since they are based on images, for example),
3. some thoughts may even be *unconscious* (and hence, not detectable by introspection),
4. the observation of thoughts may lead to ***reactivity***, meaning that the act of observing changes the thinking process itself, and finally,
5. the verbalized observations are *subjective*, meaning that they cannot be scrutinized by independent observers (as is the case in other natural sciences).

With respect to reactivity, Wundt (1908) even doubted that it is logically possible to split one's consciousness into two independent parts, the thinker and the observer. And with respect to subjectivity, Watson (1913) bemoaned that, "There is no longer any guarantee that we all mean the same thing when we use the terms now current in psychology" (p. 163 f.).

In an attempt to vindicate verbal reports, a method less prone to memory error and reactivity called the **thinking-aloud method** was later championed by Ericsson and Simon (1993). Here, test persons are encouraged to verbalize everything that comes to mind in the thinking process without the instruction to explicitly "observe" their thoughts. These verbal protocols are later analyzed qualitatively, and Ericsson and Moxley (2019) provide extensive practical information on how to set up studies and how to analyze protocol data. However, this method does not solve problems 2, 3, and 5 of the above list, and even reactivity has been demonstrated in some studies (Russo et al., 1989; Schooler, Ohlsson & Brooks, 1993).

[2] This "indirect" measurement of theoretical variables is not unique to psychology, but also commonly used in other natural sciences, for example physics, where the mass of a particle may be inferred from its movement in a magnetic field, or the speed of distant stars by a shift of their spectral lines.

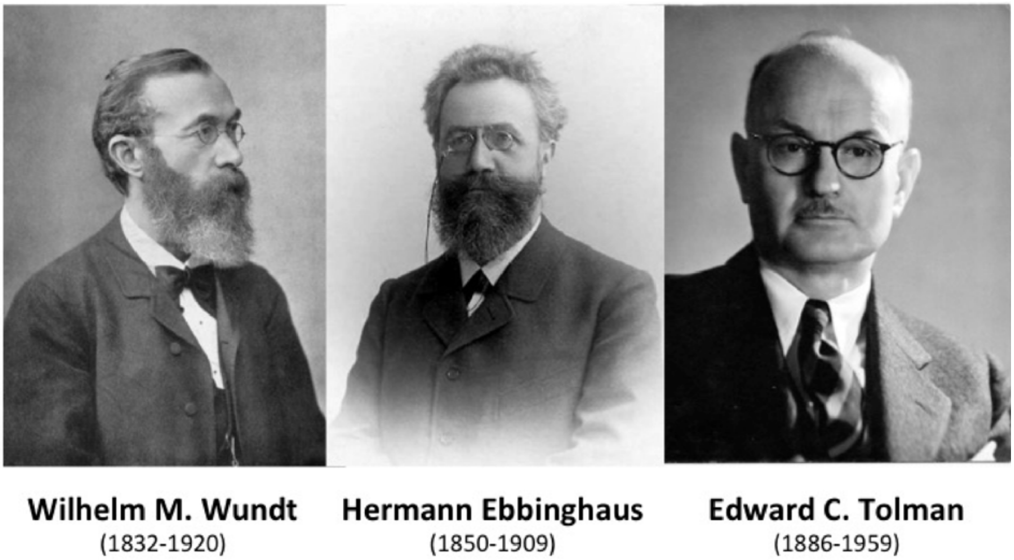

Figure 3.1: Three important methodological forethinkers of experimental cognitive psychology.

In light of the arguments above, are verbal data therefore worthless for investigating thought processes? This conclusion would be too harsh, especially with respect to thinking-aloud data. These and also classical introspective reports may be worthwhile in helping researchers to *generate* hypotheses about cognitive processes. In order to *test* these hypotheses empirically, however, one has to rely on objective data.

3.4 Objective Methods for Investigating Thought Processes

Psychologists have been quite creative in developing empirical methods for testing hypotheses about thought processes. The following section describes various methods. As we will see, although the methods can sometimes be subsumed under joint categories like, for example "response time analysis" (Section 4.2.1), the applications vary considerably depending on the specific task, theory, or hypothesis under scrutiny.

We will start with the simple idea that we can test hypotheses about thoughts by simply looking at the *outcomes* of the process, such as the quality or duration of a problem solution. The second and longest section will illustrate several methods that claim to more closely mirror the *processes* taking place during thinking. Finally, we will add very brief sections about *computer* simulations and *neuroscientific* methods in thinking research.

3.4.1 Outcome-based Methods

Observable behaviors like finding a problem solution, choosing an option or accepting a logical conclusion are the *results* of thought processes, but can they reveal information about the unobservable processes themselves? For example, large parts of research on creative problem solving (see Chapter 9, "Problem Solving") are based on a simple dependent variable, namely the percentage of participants who solved a problem, typically a hard-to-solve riddle. Whether this reveals insights into the processes involved depends on how you set up your study to test hypotheses. If you vary an *independent variable* which is believed to change certain thinking processes that either facilitate or impede successful problem solving, differences in solving rates between conditions in your experiment speak directly to your hypothesis at test. Next to simple solution

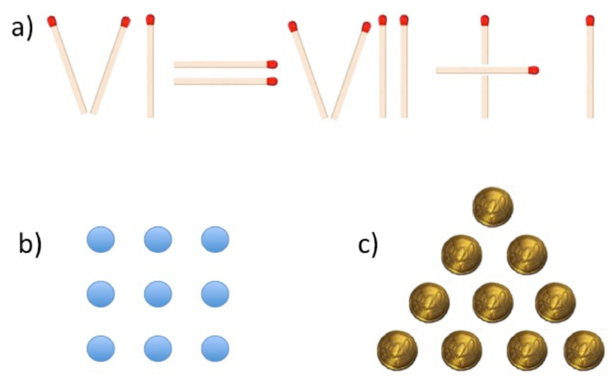

Figure 3.2: (a) Example of a matchstick puzzle - you are allowed to move only one matchstick to achieve a valid equation with Roman numerals, (b) The nine-dot problem: connect all dots with four straight lines without lifting the pen, (c) The ten-coins problem: turn the triangle upside down by moving only 3 coins.

rates and choices, more sophisticated methods utilizing behavioral outcomes allow conclusions about underlying processes by designing *diagnostic tasks* or even by the *model-based* disentangling of the processes involved. We will illustrate the three methods in turn with selected examples.

Simple Solution Rates. This issue has been controversial since Maier's (1931) anecdotal observation that unconscious "hints" can foster a problem solution. In more recent studies using matchstick puzzles (Knoblich & Wartenberg, 1998) or the notorious "nine-dots" and "ten-coins" problems (Hattori, Sloman & Orita, 2013; see Figure 3.2), researchers presented hints to the solution so briefly that they were not consciously registered by the participants. Still, in Hattori et al.'s study, solution rates for the nine-dots and ten-coins problems were tripled and increased fivefold, respectively, as compared to a control condition without these brief hints. On the premise that the hints were truly unconscious,[3] the outcome data therefore reveal a lot about the nature of problem solving processes. By simply registering success rates as the main dependent variable, numerous facilitating and impeding factors for creative problem solving have been identified (e.g. Bassok & Novick, 2012; Funke, 2003; see Chapter 9, "Problem Solving"). In a similar vein, large parts of reasoning research have used solution rates of logical arguments to investigate the factors which make logical problems easy or difficult (e.g. Johnson-Laird & Byrne, 1991) or to compare the cognitive abilities of different people.

Diagnostic task selection. Another example of how pure outcome measures may reveal information about latent processes uses the logic of *diagnostic tasks*, meaning that you choose tasks in a way that different processes or strategies predict different solutions or choices for a set of problems. You can then compare a subject's pattern of actual choices across these tasks with the predictions of the hypothetical strategies you are interested in. The strategy with predictions most "similar" to your actual data is presumably the one the participant used. There are different formal ways of assessing this similarity

[3] Whether this is the case with "subliminal" priming is still a matter of debate. I assume it to be true for the illustrative purpose of the example.

between predictions and data, and conclusions are subject to statistical error, but we will not deal with these complications here. As a general conclusion, it can be stated that pure outcome data may well provide information on detailed process hypotheses, given that these hypotheses make sufficiently different predictions for a set of tasks. An example of this research strategy is given in Textbox 3.2.[4]

Textbox 3.2: Which strategies do people use in memory-based inferences?

Bröder and Schiffer (2003) were interested in which strategies people use when they have to make decisions from memory. In their task, particpants had to compare different suspects in a hypothetical murder case and choose the one most likely to be the perpetrator. At the beginning of the experiment, participants had learned facts about the 10 suspects by heart (e.g., their blood type, their preferred cigarette and perfume brands, their vehicle). Later, they had received information about the evidence found at the crime scene. Based on the literature on decision strategies, the authors had identified 4 plausible strategies: the heuristic named *Take-the-best* (TTB; Gigerenzer & Goldstein, 1996) will look up the most important piece of evidence and base its decision on this evidence if it discriminates, otherwise, it will use the next most important evidence and so on. A *weighted additive rule* (WADD), in contrast, will look up all information and weigh it according to its importance. A *tallying rule* (TALLY) will compare the suspects simply on the number of matching pieces of evidence. Finally, participants might simply *guess*. The table shows three different task types, with the importance of the evidence decreasing from top to bottom:

	Item type 1		Item type 2		Item type 3	
Critical evidence	Suspect 1	Suspect 2	Suspect 3	Suspect 4	Suspect 5	Suspect 6
Blood type	match	-	match	match	match	-
Cigarette	-	match	match	-	match	match
Perfume	-	match	match	-	-	match
Vehicle	-	match	-	match	-	-

Across the three item types, TTB would predict the choices of Suspect 1, Suspect 3, and Suspect 5, whereas a partcipant using WADD would choose Suspects 2, 3, and 5. Someone relying on a pure tallying strategy would select Suspects 2 and 3, but be indifferent (guess with equal probability) between Suspect 5 and 6. Finally, pure guessers would select all suspects in equal proportions. Based on a few assumptions (see Bröder, 2010, for details), the probability of an empirical data pattern can be assessed for each hypothetical strategy, and the strategy with the highest probability of the observed data is diagnosed as the participant's strategy. Bröder and Schiffer (2003, Experiment 1) found a surprisingly high percentage (64%) of participants presumably using a simple TTB heuristic, and a later analysis of response times by Bröder and Gaissmaier (2007) fitted well with this interpretation (see Textbox 3.3).

[4] The more the predictions of various strategies differ, the firmer your conclusion about underlying strategies. A method for maximizing the diagnosticity of tasks is described in Jekel, Fiedler, and Glöckner (2011).

Model-based measurement of processes. Finally, detailed information about cognitive processes can be achieved by *measurement models* that formalize assumptions as to how latent processes interact to produce the behavioral outcomes. The processes are represented as *parameters* in a set of equations, and the values of these parameters are estimated from the observed data. This sounds quite abstract, so we provide an example depicted in Figure 3.3. This model formulated by Klauer, Musch, and Naumer (2000) was developed to investigate *belief bias* in syllogistic reasoning (see Chapter 7, "Deductive Reasoning"). Belief bias describes the phenomenon that people tend to accept plausible conclusions more readily than implausible ones, irrespective of the logical validity of the argument. For example, the syllogism "All vegetarians are peaceable. X is a vegetarian. Therefeore, X is peaceable" is a logically valid argument since the conclusion follows from the two premises. However, if "X" is replaced by "Mahatma Gandhi", people are more ready to accept the argument as valid than if X is replaced with "Adolf Hitler".[5] Klauer et al. (2000) formulated a processing tree model depicted in Figure 3.3 which decomposes participants judgments ("valid" vs "invalid") of four different types of syllogisms (valid and invalid arguments with plausible vs. implausible conclusion statements) into logical processes and biased guessing. Logical processes are represented by the r parameters, and guessing based on plausibility by the a parameters. Given certain assumptions and experimental procedures, the parameters can be estimated from the data, and they allow for diagnosing whether experimentally manipulated variables like time pressure, working memory load, the percentage of valid syllogisms in the task etc. affect logical abilities (reflected in r) or rather the readiness to accept conclusions irrespective of the logical validity (reflected in a).

Such measurement models have been developed for various tasks in cognitive psychology, including memory, perception, decision making, and logical thinking (see Batchelder & Riefer, 1999, and Erdfelder et al., 2009; for comprehensive overviews). If a measurement model has been validated in thorough experimental tests, it allows the drawing of very detailed conclusions about the underlying processes of observed behavior.

Evaluation: As we have seen, focusing on the outcomes of thought processes as objective data may yield much more evidence about the underlying processes than is evident at first glance. In the case of simple success rates as a dependent variable, an obvious advantage is that these are objectively measurable and do not require complex assumptions about their validity as measures. Diagnostic task selection and model-based disentanglement of processes need more assumptions (which should ideally be validated in systematic studies), but this comes with the payoff of sometimes quite detailed information about the underlying processes. As we will see in the next section, additional process measures can often enrich the data by adding valuable information.

3.4.2 Process-oriented Methods

As Schulte-Mecklenbeck et al. (2017, p. 446) have argued, "process models deserve process data". Since cognitive theories try to describe the processes that go on in our minds while thinking, it would be worthwhile eliciting data which more directly reflect these processes instead of just focusing on their results. Also, pure outcome data are often not diagnostic enough to differentiate between different theoretical models which may make the same predictions for many tasks (see, for example, item type 2 in Textbox 3.2, for which both TTB and WADD predict the same choice).[6] Although there is no con-

[5] Both historical persons were vegetarians. Hence, there is obviously something wrong with the first premise, but the *conclusion* has to follow from the premises *if* they were true.

[6] Some authors enthusiastic about process data evoke the impression that process data would be *necessary* to test process models in a sensible manner. As the preceding section 4.1 has shown, this is not the case, and I have argued elsewhere that outcome data are sufficient if they are diagnostic and formally linked to the process models under scrutiny (Bröder, 2000). I admit, however, that process data often increase the diagnosticity of the data and are therefore quite useful for research.

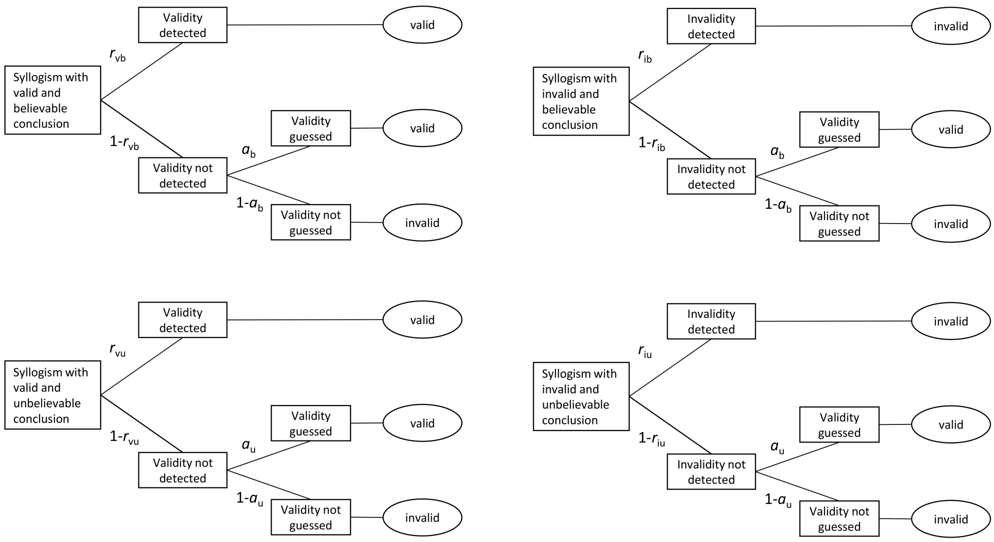

Figure 3.3: Multinomial processing tree model by Klauer et al. (2000) to assess logical reasoning and biased guessing in syllogisms. Each tree depicts processes for all the combinations of invalid vs. valid syllogisms with believable vs. unbelievable conclusions. Parameters *r* reflect reasoning, parameters *a* reflect biased guessing. ©American Psychological Association. Reprinted with permission.

sensus, yet, as to what a cognitive process actually *is* (see Newen, 2017), a defining feature of any kind of process is that it evolves over time. Hence, we will start with this most general property of cognitive processes, reflected in response time data.

3.4.2.1 Response Time Analysis

Response times are a major workhorse of cognitive psychology. They are useful for estimating the duration of component processes, or they can be analyzed as data to estimate cognitive parameters in decision models. Finally, they can be used to test cognitive theories.

Measuring the duration of cognitive processes. The first scientist to measure the duration of a simple cognitive process was presumably Frans C. Donders (1868) at the University of Utrecht in the Netherlands. We may smile today at his experimental setup, but in fact, this was a scientifc revolution because it pulled the actions of the mind into the realm of measurable natural science. He invented what later became known as the *subtraction method*: for example, using the regular oscillations of a tuning fork, he measured the simple reaction time of his colleague repeating a syllable like “ki” when the hearer knew in advance which syllable he would hear. In a second set of trials, the test person did *not* know in advance whether he had to repeat “ki”, “ku”, “ke”, or “ko”. Repeating the stimulus without knowledge took on average 46 ms (milliseconds) longer. Donders concluded that the difference was just the time needed to *choose* between the potential responses which was the only additional cognitive process needed in the second task. Shortly after this revolutionary invention, reaction time measurement for the analysis of simple processes became a fashionable method in the newly established psychological laboratories which also triggered technical developments for precise time measurement like Hipp’s chronoscope (see Figure 3.4). Although the subtraction method is preferably applied to percep-

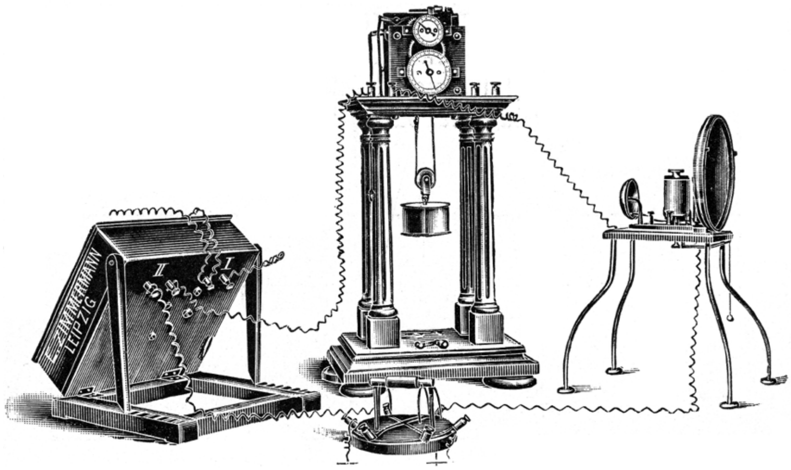

Fig. 249. Messung von Reproduktionszeiten
mit Ranschburgs Gedächtnisapparat, Chronoskop und Schallschlüssel.

Figure 3.4: An early experimental setup (c. 1900) for the precise measurement of verbal reaction times. The memory apparatus on the left displays a stimulus and starts the chronoscope (middle), the verbal reaction is recorded by the voicekey on the right which closes a circuit and stops the chronoscope (taken from Schulze, 1909).

tual tasks, there have been fruitful applications to processes of language understanding as well (Clark & Chase, 1972; 1974), showing that processes of sentence transformation and encoding a negation need certain amounts of time. Hence, the logic of the subtraction method in general is to contrast variants of speeded tasks that include or exclude specific component processes (such as negating a statement) and to generate a set of additive equations in order to estimate the durations of the component processes by simple difference calculations.

A severe limitation of the method is obviously to find tasks which can be designed to differ in only one process. To relax this requirement, S. Sternberg (1969) proposed the widely used *additive factors method* which can do without this specific task construction and merely requires a decomposition of a task into processing stages that can be selectively influenced by experimental factors.

Estimating parameters in cognitive models with reaction times. Sometimes, the researcher is not interested in the duration of processes per se, but reaction times are used as indicators for other aspects of cognition, such as ability or motivation. Particularly in research on decision making, various models have been developed that assume a process of *evidence accumulation* before a decision is made. For example, if I want to decide which of two bicycles to buy, I might sample evidence in favor or disfavor of each alternative (such as price, color, number of gears, weight etc.) until a certain subjective threshold of confidence favoring one option over the other is reached. Decision situations like these might be explained by accumulation models, like the *drift diffusion model* (DDF, Ratcliff, 1978) for simple perceptual and recognition decisions or the *decision field theory* DFT, (Busemeyer & Townsend, 1993) for more complex decisions (which would apply to the bicycle example). Figure 3.5 depicts the DDF, but the general idea is similar in other models as well. Donkin and Brown (2018) discuss variants of accumulation models, their similarities, and their differences.

These models were initially developed to explain the *speed-accuracy tradeoff*: in many tasks, people can sacrifice accuracy for higher speed, or they are slower and more accurate which depends both on their ability and their motivation to be accurate.

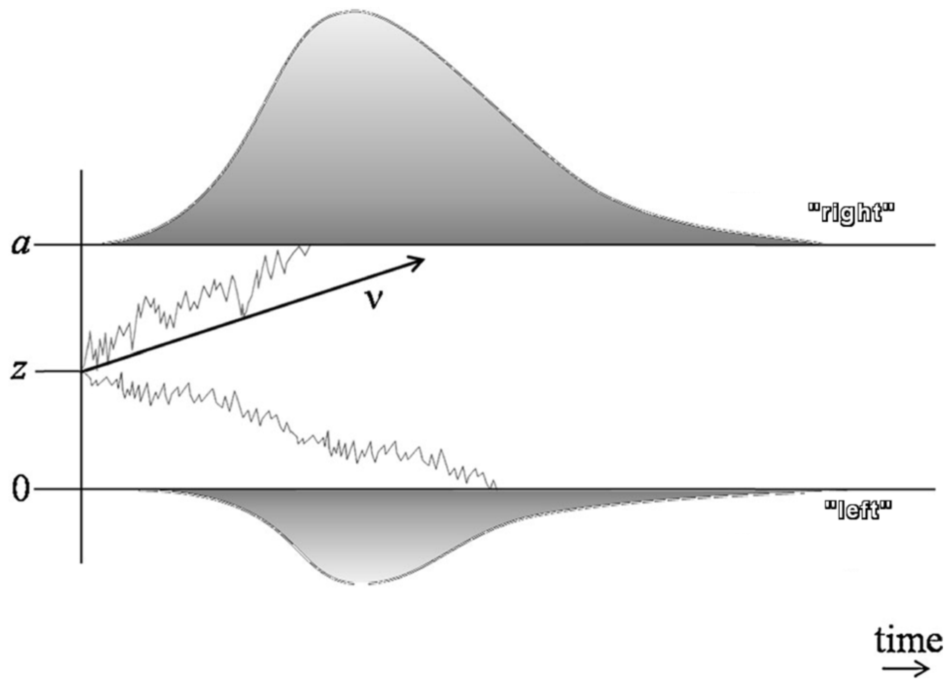

Figure 3.5: The drift diffusion model. When a stimulus with moving dots is presented, the person starts to sample perceptual evidence for the options "right" vs. "left" until a subjective evidence threshold is met. The drift rate *v*, and the distance of thresholds *a*, and the starting point *z* both determine the accuracy and the duration of the process.

Hence, only looking at error (or solution) rates or response times tells only half of the story. Suppose you have to decide in a perceptual task whether the majority of dots in a display with many randomly moving dots is moving to the right or to the left. According to the DDF, you start sampling perceptual evidence which, from time to time, may speak for one or the other direction, but on average, it will favor one of the decision options and approach the respective subjective threshold. The average speed of this accumulation process approaching one side is called the *drift rate v*, and it reflects the ease of the task (if you compare tasks) or the ability of the decision maker (if you compare people). The accuracy and the overall duration of the sampling process both depend on the *distance a* between the two subjective thresholds which is under the control of the participant who establishes a compromise between desired accuracy and speed. Furthermore, there may be a *bias z* favoring one of the answers (e.g. a tendency to respond "right" in the moving dots task), reflected in the starting point of the sampling process (an unbiased starting point is $z = a/2$, halfway between the boundaries). Although the mathematical concepts are quite complicated, various computer programs exist to estimate the parameters *v*, *a*, and *z* from empirical response time distributions associated with correct answers and errors. It has been shown in validation studies for various tasks that the parameters *v*, *z*, and *a* indeed primarily reflect task ease (or ability), bias, and motivation to be accurate, respectively (Arnold, Bröder, & Bayen, 2015; Voss, Rothermund, & Voss, 2004). The model has been successfully applied to various domains of cognitive research (Ratcliff & Smith, 2015).

Testing and validating cognitive models which make response time predictions. Finally, response time data are critical whenever a cognitive model explicitly or implicitly predicts certain response time patterns. The feature comparison model of categorization by Smith, Shoben, and Rips (1974) is a prominent example (see Figure 3.6). The model

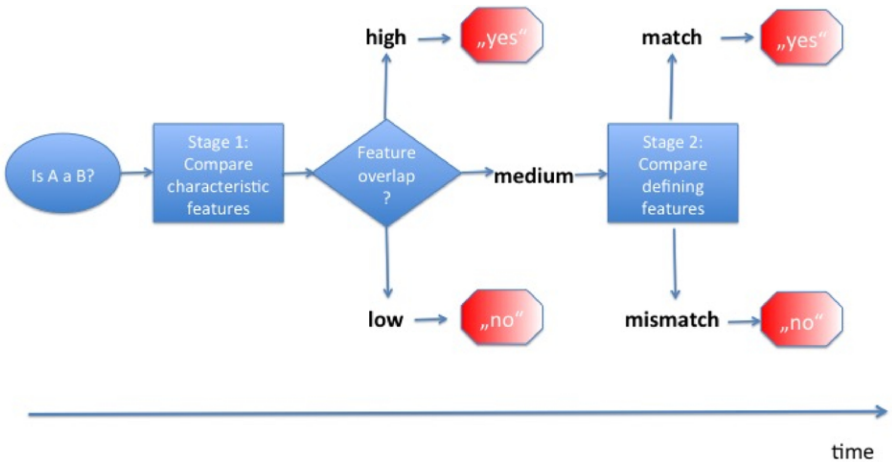

Figure 3.6: A simplified representation of the feature comparison model of categorization by Smith et al. (1974). If the object is sufficiently similar or dissimilar to the category, Stage 1 suffices for a decision. Medium similarity, however, invokes Stage 2 and hence, requires more time.

assumes that in order to categorize a stimulus, its various features are compared with the typical or *characteristic* features of the category. Hence, in deciding whether a robin is a bird, you may quickly find the answer because the characteristic features of birds in general and a robin in particular show a large overlap (can fly, has feathers and a beak, lays eggs, builds nests).

However, when asked whether a penguin is a bird, the feature overlap is smaller (since penguins do not fly and do not necessarily build nests), and the model predicts that you focus on the *defining* features in a second step (e.g. has feathers and a beak, lays eggs), excluding the merely typical (but not necessary) features. This second comparison process consumes additional time, and hence, positive instances of a category should be categorized faster the more characteristic features they share with the category (because this makes the second step unnecessary). *Negative* instances, however, should be correctly classified faster the *fewer* characteristics they share with the concept (e.g. “a whale is a bird” is denied quicker than “a bat is a bird”). These quite complex predictions have been observed, thus corroborating the feature comparison model (Rips et al., 1973).[7] A second example of how response time data have been used to validate cognitive models is described in Textbox 3.3.

Evaluation: A precise cognitive theory or model should ideally make predictions about the (relative) duration of processes or tasks. Hence, as the above examples have shown, response times can yield valuable information to test theories. Some early approaches to measure process durations like Donders’ (1868) and S. Sternberg’s (1966; 1969) methods rely on strict seriality assumptions which are sometimes questioned and hard to justify since processes may operate in parallel (e.g. Ellis & Humphreys, 1999). In addition, the subtraction method often makes unrealistic demands for task construction. As the paradigm case of the DDM has shown, response times may also be a good indicator of ability, task ease, bias, and motivation if analyzed in the context of a model (see Donkin & Brown, 2018). Currently, promising general approaches are being developed that combine outcome-based measurement models (see Section 4.1) with response time data (Heck and Erdfelder, 2016), and more general approaches try to

[7] Corroborating a theory does not “verify” it. There may be even better theories that can explain the same data and make new predictions beyond the corroborated model.

tackle the question as to whether processes operate in parallel and whether they are self-terminating or exhaustive. Finally, for many applications in logical reasoning and problem solving, response times are simply a good indicator of task difficulty in addition to solution rates. Since they are easy to obtain in computerized experiments, this additional source of information should always be recorded.

Textbox 3.3: Validating outcome-based strategy classification with response time data

In Textbox 3.2, we described how Bröder and Schiffer (2003) classified people as using the decision strategies TTB, WADD, TALLY or GUESS based on the decision outcomes in a set of diagnostic tasks. Bröder and Gaissmaier (2007) reasoned that if the classification really reflected the processes assumed by the strategies, one should expect a specific response time pattern for each group classified as using this strategy. Specifically, when people use TTB, they should need more time the more cues they have to retrieve from memory. Remember that TTB searches cues in the order of decreasing validity and stops search as soon as a discriminating cue is found. Hence, for TTB, we expect increasing response times with the position of the most valid discriminating cue. Since WADD and TALLY retrieve all four cues anyway, they should not show such an increase in response times, at least a much smaller one. WADD should generally take more time than TALLY since it also weighs the cue information with validity which TALLY does not require. Finally, GUESSing should be quickest altogether not showing systematic variations with cue position. As Figure 3.7 shows, the predictions were largely confirmed. Hence, the response time analysis lent additional credibility to the classification procedure that was initially based on decision outcomes alone.

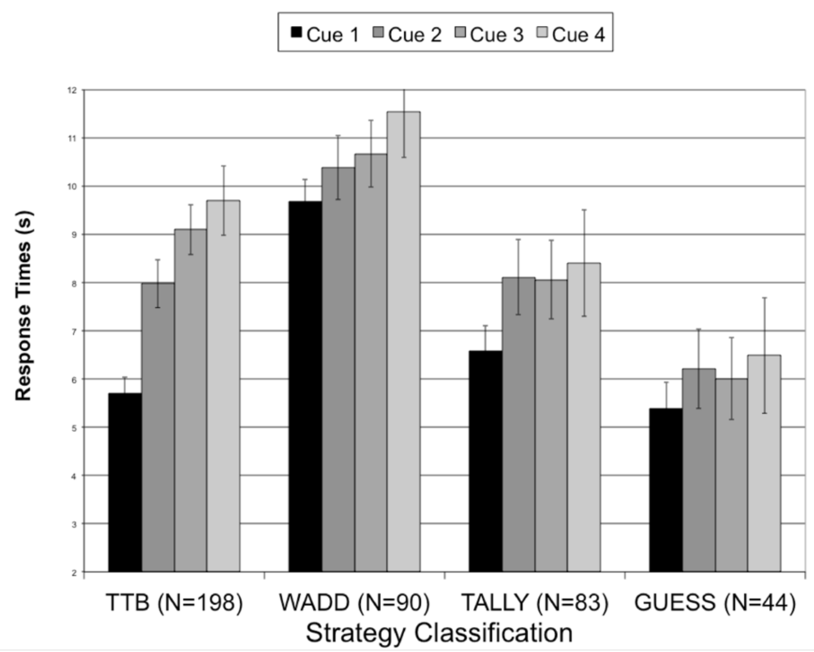

Figure 3.8: Results of the response time analysis by Bröder and Gaissmaier (2007). (See Text for details). ©Springer Nature. Reprinted with permission.

3.4.2.2 Monitoring Solution Steps and Information Search

With the rise of information processing models of thinking, problem solving research shifted to a type of sequential tasks that allowed the researcher to monitor directly the intermediate steps participants took to solve the problem. A famous example is the "Tower of Hanoi" problem in which three (or more) discs of different sizes are stacked on one of three pegs. The person's task is to move the discs to the third peg according to two rules: first, never put a larger disc on top of a smaller one, and second, only move one disc at a time (see Chapter 9, "Problem Solving", Figure 9.5). A second famous example is the "hobbits-and-orcs" problem where a boat with only two seats can be used to transfer 3 hobbits and 3 orcs across a river following the rule that there must not be more orcs than hobbits on any side of the river at any time. Participants' solution steps can be filmed, protocoled, or assessed by accompanying think-aloud protocols. These kinds of tasks helped to diagnose the general strategies people use and where these heuristics may lead to impasses, for example (Thomas, 1974).

Whereas this research strategy using sequential tasks with "observable steps" has proven fruitful, it is very restricted in scope. A somewhat more generally applicable approach is to monitor the *information search* prior to a problem solution or decision. In this paradigm, decision-relevant information is hidden from the subject's view and has to be actively uncovered or asked for. We will illustrate both a structured version in an *information board* and an unstructured *open questioning* paradigm.

Information search board. The first applications of this method actually used information cards hidden in envelopes and laid out on a table or pinned to a board (e.g. Payne, 1976). With the advent of computerized experimenting, a so-called "Mouse-Lab" version was first published by Payne et al. (1988) which presents information boxes on a screen that can be uncovered by just clicking it with the computer mouse. This methodology is often used

Dish name	Camembert and noodles	Fish sticks with mashed potatoes
		
Price	3,00 €	3,60€
Calories	661 kcal	504 kcal
Protein	31,5 g	24,6 g
Fat	34,5 g	19, 2 g
Carbohydrate	53,7 g	56, 4 g
Cholesterol	146 mg	42 mg
Sodium	1311 mg	937 mg
	Choose this dish	Choose this dish

Figure 3.7: Example of a hypothetical MouseLab layout similar to the one used in the study by Schulte-Mecklenbeck et al. (2013) where participants could choose from two different meal options. All cells of the table were closed, and participants could acquire information by clicking on the cells. They are opened here only for illustration.

to investigate multi-attribute decisions, and it has been developed in the meantime also for use in Web-based studies (e.g., Willemsen & Johnson, 2019).

Figure 3.8 shows a typical display from Schulte-Mecklenbeck et al. (2013) in which the decision *options* are arranged in columns, whereas the *attributes* are arranged in rows. In this study, the participant had to choose between meals offered in a virtual canteen, each of which was described by the same set of attributes (price, calories, different nutrients). You may be familiar with these kinds of matrices from consumer reports, for example, in which several products are compared on various attributes. In an information board study, all information is initially hidden, and the decision maker can uncover information she desires (sometimes incurring some search costs) and finally make a decision. The information may remain visible after clicking, or it may disappear again if the cursor leaves the respective box. The latter procedure more heavily taxes working memory. As you can imagine, this procedure yields a wealth of information about the search, such as the search *sequence*, the *amount* of information searched, and the *time* spent inspecting each piece of information. Payne et al. (1988) have collected various measures derivable from these data that are believed to reflect aspects of the decision strategy (see Textbox 3.4), in particular if decision making tends to ignore information and focuses on comparing options on important attributes ("noncompensatory" decision making) or whether the strategy tends to use all information and compares overall evaluations of the options ("compensatory" strategies). Willemsen and Johnson (2019) report new developments to visualize aspects of the search process in this paradigm.

Unstructured open questioning formats. The information board technique described in the previous section contains pre-structured information which may create some experimental demands in suggesting which kinds of information the experimenter deems relevant. This allows the inferring of the *relative* importance people put on attributes but not whether they find them important in the first place. Huber, Wider, and Huber (1997) therefore developed a technique with quasi-realistic decision scenarios. After reading the scenarios (e.g. about the problem of saving an endangered turtle species), participants could ask for any further information they wished, receiving answers from a large set of predefined information. This procedure has shown repeatedly that participants tend to ignore probability information (Huber et al., 1997) and that they ask for information on how to eliminate risks (Huber, Bär & Huber, 2009).

Evaluation: Observing the steps involved in thinking by monitoring corresponding behavior is one possibility to more "closely" follow thinking processes. Monitoring stepwise problem solving is restricted to a very specific type of tasks, however. Another possibility is to register the information search processes prior to a decision or action, for example via MouseLab. As we have seen, this can yield a wealth of data that may inform us about the strategies people use. As a caveat, it should be noted that information *search* is not necessarily indicative of how the information is *integrated* (see Bröder, 2000), both may be quite different processes governed by different rules (Glöckner & Betsch, 2008). For example, one may look up all relevant information (seemingly indicating compensatory decision making), but decide to ignore most of it (leading to noncompensatory integration). Or one can decide in a compensatory manner without exhaustive search (if the remaining information could not reverse a decision anyway). Researchers do not always distinguish between search and integration, which may lead to misunderstandings in theory testing (Lohse & E. J. Johnson, 1996). Hence, to apply the methodology, it must be clear which part of cognition is under scrutiny. Finally, the active information search paradigm by Huber et al. (1997) has the advantage of not suggesting experimental demands to the study participants but it is a rather explorative method for generating instead of testing cognitive theories.

Textbox 3.4: MouseLab Decision Strategy Indicators

Payne, Bettman, and E. J. Johnson (1988) and Payne (1976) derived various measures from the search sequences and inspection times of information in MouseLab, for example the *strategy index* SI (sometimes also called search index or PATTERN) which codes the relative amount of option-wise search (i.e. moving within options to new attributes) versus attribute-wise search (comparing different options on the same attribute). Option-wise search is thought to indicate so-called *compensatory strategies* that use all information and compare overall evaluations of the options (examples are WADD or TALLY in the previous textboxes), whereas attribute-wise search is believed to reflect *noncompensatory strategies* that ignore information (such as TTB in the previous textboxes). If n_o is the number of search transitions within an option to a different attribute and n_a the number of transitions within an attribute to another option (transitions switching both option and attribute are ignored), the search index can be computed as

$$SI = \frac{n_o - n_a}{n_o + n_a},$$

and it varies from -1 to +1, reflecting pure attribute- and option-wise search respectively. Böckenholt and Hynan (1994) proposed a modified version of the index for asymmetric options x attributes tables as in Figure 3.8. The following table contains further measures and their interpretation.

Measure	**Definition**	**higher values indicate...**
Strategy Index SI	see text	compensatory
ACQ	number of acquisitions in trial	compensatory
TPERACQ	time per acquisition	compensatory
PTMI	percentage of time spent inspecting most important attribute	noncompensatory
VAR-ATTR	variance of times spent on different attributes	noncompensatory
VAR-ALTER	variance of times spent on different options	noncompensatory

3.4.2.3 Tracking of Eye Movements

A method which has gained popularity in recent years involves the registration of eye movements while thinking, based on the assumption that a person's momentary attention and focus of processing is reflected by his or her fixation on a stimulus. While early **eye tracking** devices were expensive and intrusive by requiring people to have their head fixated (for example by biting a board) or to wear heavy helmets with cameras and contact lenses, new (and cheaper) devices allow for the remote monitoring of eye movements by use of infrared light reflected from the cornea, either in front of a computer screen or even in more natural environments (see Ball, 2014, and Russo, 2019, for brief introductions). Eye-tracking has been used extensively in research on reading and language comprehension, but it is also becoming increasingly popular in decision research and research on thinking (see Orquin & Loose, 2013). For example, by using an open information board, tracking the gaze sequence may yield similar information as with a MouseLab procedure.

The motor activity of the eyes is composed mainly of *saccades*, which are quick movements during which no information is registered, and *fixations* which are brief resting periods during which the viewer registers visual stimulus information (e.g., Holmqvist et al., 2011). Consequently, the sequence, number, average duration and cumulative duration of fixations are of main interest to researchers.

For *explorative* (hypothesis-generating) research, several methods for visualizing the gaze behavior of participants exist. *Heatmaps* color-code the frequency of fixations to certain parts of the stimulus, and *scanpaths* contain additional information about the sequence and the duration of the fixations (see Figure 3.9 for examples of the same data presented as a heatmap or a scanpath). These visualizations are often used in applied research settings like usability and consumer research in order to optimize displays and ads.

In hypothesis-*testing* research, the stimulus display is typically arranged in a way that important parts are clearly separated into **areas of interest (AOI)** that contain different aspects of the problem. For example, Figure 3.10 (left) shows a display with five letters, four of which build an anagram (= scrambled word puzzle) with a four-letter solution, the fifth letter being a superfluous distractor. The letters are widely distributed across the screen for an error-free detection of the stimulus a person is looking at a specific moment.

Often, processing hypotheses can be formulated in a way that different problem aspects are expected to receive more attention than others which can be tested by comparing the number or duration of fixations at the respective AOIs. I will describe a research example from problem solving research. To test whether people acquire solution knowledge even *before* they have a conscious insight into the correct solution, Ellis, Glaholt, and Reingold (2011) used anagram problems like the one depicted in Figure 3.10 and monitored eye movements during problem solution. The anagrams consisted of five letters, one of which was not part of the four-letter solution word. Participants were instructed to press a button as soon as they had found the solution word, and in Experiment 1b additionally stated whether the solution “popped up” in a sudden “aha” experience. Ellis et al. (2011) tested the hypothesis that participants would accumulate knowledge prior to finding the solution even if the solution appeared suddenly in their consciousness. This should be reflected in

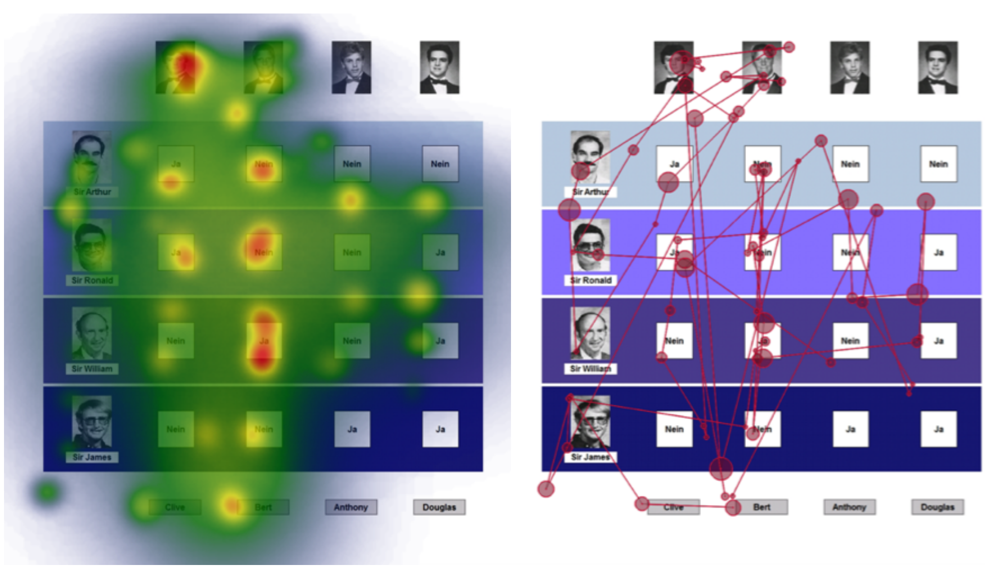

Figure 3.9: Heatmap and scanpath representation of the same eye tracking data of a person in a decision trial. In this task, the options (columns) were card players, and participants had to predict their success based on advice of experts (rows). In this trial, the participant focuses on the two leftmost options in a predominantly option-wise manner. (Data from Ettlin & Bröder, 2015, Experiment 4).

decreasing attention to the distractor relative to the solution letters. In fact, there was a significant tendency to ignore the distractor letter on average 2.5 s before particpants announced they had found the solution, confirming the hypothesis of knowledge accumulation before conscious insight.

Evaluation: The tracking of eye movements has become cheaper, more user-friendly and less intrusive in recent years. Holmqvist et al. (2011) give an extensive overview of theory and application. As we have seen, eye-tracking data can reveal a lot about the sequence of processing and the allocation of attention while thinking, and it can be used both in an explorative and a hypothesis-testing fashion. The latter requires experimental setups with theoretically defined AOIs for which gaze durations and frequencies can be compared. Furthermore, important extensions are under development such as the *memory indexing* method developed by Renkewitz and Jahn (2012). This ingenious idea is based on the "looking-at-nothing" effect first investigated by Richardson and Spivey (2000), demonstrating that during memory retrieval, people tend to look at the location (on a computer screen, for instance) where they learned that information. Basically, this method therefore allows the monitoring of sequences of hidden memory processes by analyzing gaze data! A study by Scholz, Krems, and Jahn (2017) on (hy-

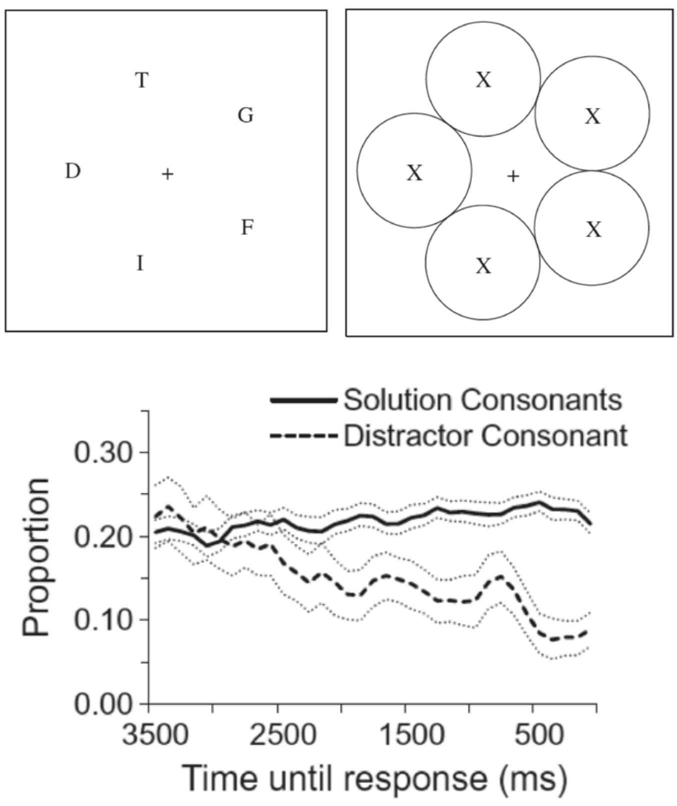

Figure 3.10: Top left: Anagram setup of Ellis et al. (2011), one letter does not belong to the four-letter solution word. Top right: Areas of interest (AOIs) from which fixations to the letters are recorded (not visible to participants). Bottom: Mean proportion of time looked at solution letters and distractor prior to solution in Experiment 1b. ©Elsevier. Reprinted with permission.

pothetical) medical diagnoses not only replicated the looking-at-nothing effect but also showed that the gaze behavior reflects the diagnosis currently most active in working memory, and it also allows the prediction of participants' final decisions. Also, new software methods allow to change displays contingent on gaze behavior "on the fly" (e.g. Franco-Watkins & J. G. Johnson, 2011), thus opening new possibilities for experiments.

There are a few downsides to the eye-tracking method, however: first, the connection between visual attention and gaze direction is not always as close as assumed since spatial attention can also be directed to locations without moving the eyes. Second, many other factors (like salience or reading routines) influence our gaze behavior, thus data are often quite noisy, and it is not always easy to separate meaningful data from the unsystematic variation. Third, depending on the quality of the equipment used, often several participants have to be excluded (e.g. those wearing glasses or contact lenses). Finally, at the moment of writing, explorative rather than theory-testing applications seem to prevail in the literature which may of course change in the future.

3.4.2.4 Response Dynamics

A recent development pioneered by Spivey, Grosjean, and Knoblich (2005) uses the characteristics of the motor behavior (specifically, participants' hand movements) during a decision response to draw conclusions about internal thinking processes and their dynamics. Since most experiments use the computer mouse as the input device, this methodology has been christened *mouse-tracking*, although other devices have been used to record participants' hand movements as well (e.g., the Nintendo Wii Remote, a handle, or motion capture systems). One assumption is that the decision dynamically evolves during the mouse movement, and its trajectory may therefore reflect the extent to which a decision conflict is present (Stillerman & Freeman, 2019). In a typical setup, each trial presents two choice options in the upper left and right corners of the computer screen. The participant has to initiate a trial by clicking on a start button that is typically placed in the neutral middle at the lower end of the screen (cf. Figure 3.11) upon which the decision-critical information is presented (either immediately, after a delay, or following an initial upwards movement; see Scherbaum & Kieslich, 2018, for a discussion about the different starting procedures and their consequences for mouse-tracking data). During the (sometimes speeded) response, the participant will then choose one option by clicking it while her mouse movements are continuously recorded. If the decision maker feels a conflict between both options, the mouse path will probably not be totally straight, but it will be "drawn" a bit to the competing alternative. Several measures can be derived to quantify this deviation, the simplest is the "maximum absolute deviation (MAD)" of the curved trajectory from the straight line leading to the chosen option. Figure 3.11 shows a typical display investigating the "Simon effect" along with visualized raw data as well as average trajectories from data published by Scherbaum et al (2010).

Although it is quite new, the method has been applied to a variety of domains, such as categorization tasks (animals, gender, race), spoken word recognition, risky decision making, word and sentence comprehension, truth judgments, social cognition and more (see Freeman et al., 2011; Freeman, 2018). It provides a sensitive measure of conflict between response options. Furthermore, the exact analysis of the temporal dynamics in the trajectories (including speed and acceleration metrics) can even provide information about *when* the conflict arises, which can signify whether a specific piece of information is processed earlier or later in the decision process (Dshemuchadse, Scherbaum, & Goschke, 2013; Sullivan, Hutcherson, Harris, & Rangel, 2015). For example, Sullivan et al. (2015) had their participants choose between food items they had rated before on healthiness and taste. Independent of which food was chosen in a trial, the mouse trajectory was influenced by the taste difference earlier than by the healthiness information, indicating that the initial preference tendency is driven by pleasure, whereas health considerations come into play somewhat later in the decision process.

Evaluation: The way in which participants move the mouse to choose an option is an unobtrusive

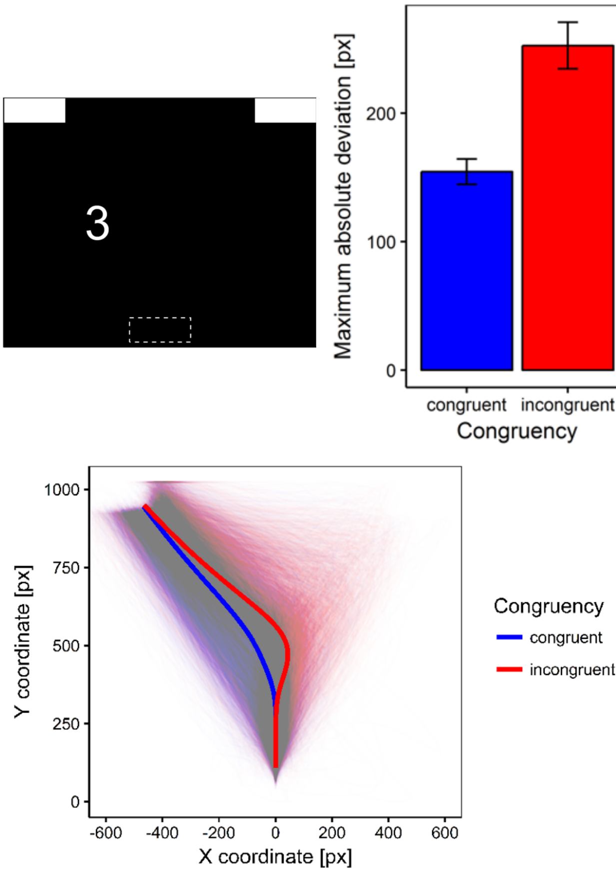

Figure 3.11: Top left: Exemplary mouse-tracking setup of Experiment 2 by Scherbaum et al. (2010) to investigate the Simon effect. Participants had to click a start button at the bottom center of the screen (dashed lines), when moving the cursor upwards, a number x appeared, and participants had to click left or right for x<5 and x>5 respectively. The presentation side of the number varied, creating congruent (x<5 left or x>5 right) vs. incongruent (x<5 right or x>5 left) trials. Top right: The summary mean absolute deviation of mouse trajectories demonstrates the Simon effect with greater average deviation for incongruent trials. Bottom: Individual and average (thick lines) mouse trajectories for congruent and incongruent trials (note that all trajectories were flipped to the left and only correct trials were analyzed).

method for revealing conflicting response tendencies. As the food choice example shows, even quite detailed information about the time course of processing can be gathered. Furthermore, easy-to-use implementation and analysis software has been developed, for example, the mousetrap plugin for creating mouse-tracking experiments in the free and open-source graphical experiment builder OpenSesame (Kieslich & Henninger, 2017) and the mousetrap R package for analyzing and visualizing mouse-tracking data (Wulff, Haslbeck, Kieslich, Henninger, & Schulte-Mecklenbeck, 2019). As a relatively novel method, mouse-tracking faces a number of challenges. Many aspects of the design of mouse-tracking studies (e.g. the starting procedure and mouse sensitivity settings) require careful consideration to reduce the amount of noise in the data and to ensure that the decision process takes place during (and not before) the movement (e.g. Scherbaum & Kieslich, 2018). Also, averaged trajectories may be misleading and suggest a smooth curve when in fact, they are averaged across different types of trajectories in different trials (Wulff et al., 2019). Finally, it is currently unknown whether cognitive conflicts always influence response dynamics and therefore how to interpret the *absence* of trajectory effects.

3.4.3 Computer Simulations

Beginning with Newell, Shaw, and Simon's (1958) work on a computer program later called the "General Problem Solver" (although it was rather limited in its abilities), cognitive scientists have attempted to formulate their theories in precise formal terms and to translate them into computer programs. The aim is to *simulate* human performance in cognitive tasks, including typical errors and fallacies or shortcomings in memory etc. Computer versions of theories are also termed **computational models** (Farrell & Lewandowsky, 2018). The scope of such models ranges from very specific theories about certain tasks to broad overarching "cognitive architectures" (e.g. ACT-R by Anderson et al., 2004) that entail many empirically informed constraints for modeling and predicting human behavior.

The advantages of formalizing theories and cognitive processes in such a way are manifold: first, the precision of the theory typically has to be increased. Whereas verbal theories are often quite vague, an implementation in the computer demands precise concepts. Second, such a formalization may reveal inconsistencies in the theory that would have gone unnoticed without formalizing it. Third, in addition to just predicting qualitative "effects" (e.g. the existence of group differences), precise models may even give quantitative predictions about effect sizes. Hence, in addition to the experimental tools researchers use to observe people's behavior, matching it with computer simulations can reveal a lot about the validity of cognitive theories. We refer the interested reader to Farrell and Lewandowsky (2018) for an excellent introduction to cognitive modeling.

3.4.4 Neuroscientific Methods

Since all our cognitive functions including thinking depend on *brain functions*, an ultimate understanding of cognition will have to include knowledge about these functions. The traditional approach of **neuropsychology** gains many insights into the localization of cognitive functions in the human cortex by carefully assessing cognitive impairments caused by specific brain injuries. These investigations have inspired the view that the brain's architecture is largely *modular* with certain modules being responsible for specific abilities.

In recent decades, brain imaging methods—mostly functional magnetic resonance imaging **(fMRI)**—have dramatically increased our knowledge about the brain structures involved in diverse cognitive tasks including thinking, although enthusiastic claims that fMRI can "watch the brain while thinking" are quite overstated (see Satel & Lilienfeld, 2013, for a critique). Basically, the standard fMRI method can contrast the metabolic activity pattern in the brain during a task with the activity pattern in another (control) task, and the regions with the greatest activity differences are probably involved in the processes that differ between the tasks. Hence, the experimental logic is quite similar to Donders' (1868) subtraction method for response times, and the better the tasks are chosen, the more meaningful the interpretation of the activation differences. In the last few years, complex statistical methods called *connectivity analysis* have also been developed which give very detailed information about the path and time course of activation that spreads through the brain during specific tasks (see Bastos & Schoffelen, 2016, for a review).

A wealth of knowledge about brain structures involved in various cognitive activities has been accumulated in the meantime, and a deeper treatment of neuroscientific methods is beyond the scope of this chapter. For the interested reader, I highly recommend Ward (2015) and Purves et al. (2013), for introductions into cognitive neuroscience.

3.5 Conclusion

The behaviorists believed that investigating thoughts and consciousness would require introspection and verbal reports which are subjective and notoriously unreliable. Hence, they believed the mind to evade serious scientific investigation. As this chapter has shown, cognitive psychologists have proven this aspect of behaviorism to be blatantly wrong. Numerous innovative techniques that rely on objective data were developed that shed light on the proverbial "black box" of the mind. As recent devlopments like response dynamics and eye tracking show, this development of clever methods is still going on, and it will without doubt help to reveal more fascinating insights into cognition in the future.

Acknowledgements

I would like to thank Pascal Kieslich, Sophie Scharf, and Yury Shevchenko for helpful comments on a draft of this chapter.

Summary

How can theories about unobservable events like cognitive processes be tested and evaluated empirically? Since the method of introspection (self-observation) was criticized very early on for various reasons, cognitive scientists have developed a large toolbox of other methods that yield more objective data for testing theories about cognition. The idea behind this is that cognitive processes like retrieving a memory or solving a logical puzzle lead to observable consequences in behavior. The easiest methods just measure the outcome of a process, e.g. whether an item is solved or not. Depending on how precise the theory is, this can provide surprisingly detailed information about cognition. For example, items may be chosen in a way that different processes predict different solution patterns across these items which may allow the inferring of a strategy. Another set of methods tries to tackle the underlying processes more closely, for example by dissecting response times or by monitoring information uptake with information boards or eye movement analyses. Also, movements during response generation can reveal conflicting response tendencies. Finally, theories about thinking and cognition can profit very much from computer simulations and of course neuroscientific research that investigates the neural underpinnings of the processes.

Review Questions

1. Why is it important to have objective measurements or observations in science, meaning that in principle, different observers would come to similar conclusions about the observed phenomena?
2. Explain in your own words how it is possible to draw conclusions about latent cognitive processes or strategies by observing overt behavior.
3. If you employed an information board setup with an option-by-attributes matrix for a decision problem, but displayed all information freely from the beginning, which information could you extract from the scanpath of fixations if you monitor the gaze movements of a person?
4. Looking at Figure 3.11, what problems would you expect with the mouse-tracking procedure, both theoretically and in practice?

Hot Topic: Single or multiple mechanisms in decision making?

Arndt Bröder

My research in the last two decades has been greatly inspired by research on "adaptive decision making" showing that people flexibly adapt their decision behavior to changing environmental demands, such as time pressure, memory retrieval demands, or payoff structures. The predominant view has been that we can choose from a large repertoire of qualitatively different strategies and heuristics that we employ under appropriate circumstances (see Textboxes 3.3. and 3.4). This idea of a strategy "toolbox" was especially promoted by Gigerenzer et al. (1999) and stimulated a lot of research. After developing valid methods for *diagnosing* these strategies in a valid manner (see Textbox 3.3), my further research investigated under which circumstances these strategies and simple heuristics are applied (see Bröder, 2012, for an overview). However, there are also critics of the toolbox metaphor, claiming that we might rather use a *single* mechanism for deciding, such as the evidence accumulation model described in Section 3.4.2 (Figure 3.5), and widening or narrowing the gap between decision thresholds may just *mimick* the use of different strategies, although people just change a parameter in a universal strategy. Both views are notoriously hard to differentiate empirically. In a series of elegant studies, my doctoral student Anke Söllner showed that indeed the evidence accumulation view is more plausible than the multiple heuristics view to describe information acquisition (Söllner & Bröder, 2016). Recent joint work with colleagues favoring another "unified strategy" approach based on coherence-maximization principles also showed that predicions from this theory appear to explain search behavior better than the multiple strategies view (Jekel, Glöckner & Bröder, 2018). The debate which metaphor is more appropriate will probably continue for a while (see Marewski, Bröder & Glöckner, 2018), but I always try to respect Konrad Lorenz' advice: "It is a good morning exercise for a research scientist to discard a pet hypothesis every day before breakfast. It keeps him young." [a]

References

Bröder, A. (2012). The quest for take-the-best: Insights and outlooks from experimental research. In P. M. Todd, G. Gigerenzer, & The ABC Research Group (Eds.), *Ecological rationality: Intelligence in the world.* (pp. 216–240). New York: Oxford University Press. doi:10.1093/acprof:oso/9780195315448.003.0067

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Jekel, M., Glöckner, A., & Bröder, A. (2018). A new and unique prediction for cue-search in a parallel-constraint satisfaction network model: The attraction search effect. *Psychological Review*, *125*(5), 744–768. doi:10.1037/rev0000107

Marewski, J. N., Bröder, A., & Glöckner, A. (2018). Some metatheoretical reflections on adaptive decision making and the strategy selection problem. *Journal of Behavioral Decision Making*, *31*(2), 181–198. doi:10.1002/bdm.2075

Söllner, A., & Bröder, A. (2016). Toolbox or adjustable spanner? A critical comparison of two metaphors for adaptive decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(2), 215–237. doi:10.1037/xlm0000162

[a] Lorenz, K. (1966/2002). *On aggression*. London: Routledge. (p. 9)

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An Integrated Theory of the Mind. *Psychological Review*, *111*(4), 1036–1060. doi:10.1037/0033-295X.111.4.1036

Arnold, N. R., Bröder, A., & Bayen, U. J. (2015). Empirical validation of the diffusion model for recognition memory and a comparison of parameter-estimation methods. *Psychological Research*, *79*(5), 882–898. doi:10.1007/s00426-014-0608-y

Ball, L. J. (2014). Eye-tracking and reasoning: What your eyes tell about your inferences. In W. De Neys, M. Osman, W. De Neys (Ed), & M. Osman (Ed) (Hrsg.), *New approaches in reasoning research*. (pp. 51–69). New York, NY, US: Psychology Press. doi:10.4324/9781315879857

Bassok, M., & Novick, L. R. (2012). Problem solving. In K. J. Holyoak & R. G. Morrison (Hrsg.), *The Oxford handbook of thinking and reasoning.* (pp. 413–432). New York, NY, US: Oxford University Press. doi:10.1093/oxfordhb/9780199734689.001.0001

Bastos, A. M., & Schoffelen, J.-M. (2016). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in Systems Neuroscience, 9*. doi:10.3389/fnsys.2015.00175

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*(1), 57–86. doi:10.3758/BF03210812

Böckenholt, U., & Hynan, L. S. (1994). Caveats on a process-tracing measure and a remedy. *Journal of Behavioral Decision Making*, *7*(2), 103–117. doi:10.1002/bdm.3960070203

Bröder, A. (2000). A methodological comment on behavioral decision research. *Psychologische Beiträge*, *42*(4), 645–662.

Bröder, A. (2010). Outcome-based strategy classification. In A. Glöckner & C. Witteman (Hrsg.), *Foundations for tracing intuition: Challenges and methods*. (pp. 61–82). New York, NY US: Psychology Press. doi:10.4324/9780203861936

Bröder, A., & Gaissmaier, W. (2007). Sequential processing of cues in memory-based multiattribute decisions. *Psychonomic Bulletin & Review*, *14*(5), 895–900. doi:10.3758/bf03194118

Bröder, A., & Schiffer, S. (2003). Take The Best versus simultaneous feature matching: Probabilistic inferences from memory and effects of reprensentation format. *Journal of Experimental Psychology: General*, *132*(2), 277–293. doi:10.1037/0096-3445.132.2.277

Bühler, K. (1908). Antwort auf die von W. Wundt erhobenen Einwände gegen die Methode der Selbstbeobachtung an experimentell erzeugten Erlebnissen [Response to W. Wundt's Objections against the Method of Self-Observation of Experimentally Created Experiences]. *Archiv für die Gesamte Psychologie*, *12*, 93–123.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*(3), 432–459. doi:10.1037/0033-295X.100.3.432

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, *3*(3), 472–517. doi:10.1016/0010-0285(72)90019-9

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, *11*(6), 671–684. doi:10.1016/S0022-5371(72)80001-X

Donders, F. C. (1868). Die Schnelligkeit psychischer Prozesse. *Archiv für Anatomie, Physiologie und wissenschaftliche Medizin*, (657–681).

Donkin, C., & Brown, S. D. (2018). Response Times and Decision-Making. In J.T. Wixted (Ed.). *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*. Wiley online library. doi:10.1002/9781119170174.epcn509

Dshemuchadse, M., Scherbaum, S., & Goschke, T. (2013). How decisions emerge: Action dynamics in intertemporal decision making. *Journal of Experimental Psychology: General*, *142*(1), 93–100. doi:10.1037/a0028499

Ellis, R., & Humphreys, G. W. (1999). *Connectionist psychology: A text with readings*. Hove, England: Psychology Press.

Ellis, J. J., Glaholt, M. G., & Reingold, E. M. (2011). Eye movements reveal solution knowledge prior to insight. *Consciousness and Cogni-*

tion: An International Journal, *20*(3), 768–776. doi:10.1016/j.concog.2010.12.007

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie/Journal of Psychology*, *217*(3), 108–124. doi:10.1027/0044-3409.217.3.108

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data., Rev. ed.* Cambridge, MA, US: The MIT Press. doi:10.1097/00005053-198511000-00011

Ericsson, K. A., & Moxley, J. (2019). Thinking aloud during superior performance on tasks involving decision making. In M. Schulte-Mecklenbeck, A. Kühberger & J. G. Johnson (Eds.). *A Handbook of Process Tracing Methods* (2nd Edition). London: Routlege.

Ettlin, F., & Bröder, A. (2015). Perceptual grouping does not affect multi-attribute decision making if no processing costs are involved. *Acta Psychologica*, *157*, 30–43. doi:10.1016/j.actpsy.2015.02.002

Farrell, S., & Lewandowsky, S. (2018). *Computational modeling of cognition and behavior*. New York, NY: Cambridge University Press. doi:10.1017/9781316272503

Franco-Watkins, A. M., & Johnson, J. G. (2011). Decision moving window: Using interactive eye tracking to examine decision processes. *Behavior Research Methods*, *43*(3), 853–863. doi:10.3758/s13428-011-0083-y

Freeman, J. B. (2018). Doing psychological science by hand. *Current Directions in Psychological Science*, *27*(5), 315–323. doi:10.1177/0963721417746793

Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology*, *2*, 59. doi:10.3389/fpsyg.2011.00059

Funke, J. (2003). *Problemlösendes Denken [Problem-solving thinking]*. Stuttgart: Kohlhammer.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669. doi:10.1037//0033-295x.103.4.650

Glöckner, A., & Betsch, T. (2008). Modeling option and strategy choices with connectionist networks: Towards an integrative model of automatic and deliberate decision making. *Judgment and Decision Making*, *3*(3), 215–228. doi:10.2139/ssrn.1090866

Hattori, M., Sloman, S. A., & Orita, R. (2013). Effects of subliminal hints on insight problem solving. *Psychonomic Bulletin & Review*, *20*(4), 790–797. doi:10.3758/s13423-013-0389-0

Heck, D. W., & Erdfelder, E. (2016). Extending multinomial processing tree models to measure the relative speed of cognitive processes. *Psychonomic Bulletin & Review*, *23*(5), 1440–1465. doi:10.3758/s13423-016-1025-6

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Huber, O., Bär, A. S., & Huber, O. W. (2009). Justification pressure in risky decision making: Search for risk defusing operators. *Acta Psychologica*, *130*(1), 17–24. doi:10.1016/j.actpsy.2008.09.009

Huber, O., Wider, R., & Huber, O. W. (1997). Active information search and complete information presentation in naturalistic risky decision tasks. *Acta Psychologica*, *95*(1), 15–29. doi:10.1016/S0001-6918(96)00028-5

Jekel, M., Fiedler, S., & Glöckner, A. (2011). Diagnostic task selection for strategy classification in judgment and decision making: Theory, validation, and implementation in R. *Judgment and Decision Making*, *6*(8), 782–799.

Johnson, E. J., Schulte-Mecklenbeck, M., & Willemsen, M. C. (2008). Process models deserve process data: Comment on Brandstätter, Gigerenzer, and Hertwig (2006). *Psychological Review*, *115*(1), 263–272. doi:10.1037/0033-295X.115.1.263

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods*, *49*(5), 1652–1667. doi:10.3758/s13428-017-0900-z

Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review*, *107*(4), 852–884. doi:10.1037/0033-295X.107.4.852

Knoblich, G., & Wartenberg, F. (1998). Unbemerkte Lösungshinweise begünstigen Veränderungen der Problemrepräsentation. [Unnoticed hints facilitate representational change in problem solving]. *Zeitschrift für Psychologie mit Zeitschrift für angewandte Psychologie*,

Lohse, G. L., & Johnson, E. J. (1996). A comparison of two process tracing methods for choice tasks. *Orga-*

nizational Behavior and Human Decision Processes, *68*(1), 28–43. doi:10.1006/obhd.1996.0087

Maier, N. R. F. (1931). Reasoning in humans. II. The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology*, *12*(2), 181–194. doi:10.1037/h0071361

Massen, C., & Bredenkamp, J. (2005). Die Wundt-Bühler-Kontroverse aus der Sicht der heutigen kognitiven Psychologie [The Wundt-Bühler-Controversy from the View of Modern Cognitive Psychology]. *Zeitschrift für Psychologie*, *213*(2), 109–114. doi:10.1026/0044-3409.213.2.109

Mata, A., Ferreira, M. B., Voss, A., & Kollei, T. (2017). Seeing the conflict: An attentional account of reasoning errors. *Psychonomic Bulletin & Review*, *24*(6), 1980–1986. doi:10.3758/s13423-017-1234-7

Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review*, *65*(3), 151–166. doi:10.1037/h0048495

Newen, A. (2017). What are cognitive processes? An example-based approach. *Synthese*, *194*(11), 4251–4268. doi:10.1007/s11229-015-0812-3

Orquin, J. L., & Loose, S. M. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, *144*(1), 190–206. doi:10.1016/j.actpsy.2013.06.003

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior & Human Performance*, *16*(2), 366–387. doi:10.1016/0030-5073(76)90022-2

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534–552. doi:10.1037//0278-7393.14.3.534

Purves, D., Cabeza, R., Huettel, S. A., LaBar, K. S., Platt, M. L., Woldorff, M. G., & Brannon, E. M. (2013). *Principles of cognitive neuroscience., 2nd ed.* Sunderland, MA, US: Sinauer Associates. doi:10.1086/670558

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. doi:10.1037/0033-295X.85.2.59

Ratcliff, R., & Smith, P. (2015). Modeling simple decisions and applications using a diffusion model. In J. R. Busemeyer, Z. Wang, J. T. Townsend, A. Eidels, J. R. Busemeyer (Ed), Z. Wang (Ed), ... A. Eidels (Ed) (Hrsg.), *The Oxford handbook of computational and mathematical psychology*. (pp. 35–62). New York, NY, US: Oxford University Press. doi:10.1093/oxfordhb/9780199957996.001.0001

Reisen, N., Hoffrage, U., & Mast, F. W. (2008). Identifying decision strategies in a consumer choice situation. *Judgment and Decision Making*, *3*(8), 31–31.

Renkewitz, F., & Jahn, G. (2012). Memory indexing: A novel method for tracing memory processes in complex cognitive tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1622–1639. doi:10.1037/a0028073

Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition*, *76*(3), 269–295. doi:10.1016/s0010-0277(00)00084-6

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning & Verbal Behavior*, *12*(1), 1–20. doi:10.1016/S0022-5371(73)80056-8

Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition*, 17(6), 759–769. doi:10.3758/BF03202637

Russo, J. E. (2019). Eye fixations as a process trace. In M. Schulte-Mecklenbeck, A. Kühberger & J. G. Johnson (Eds.). *A Handbook of Process Tracing Methods* (2nd Edition). London: Routlege.

Satel, S., & Lilienfeld, S. O. (2013). *Brainwashed: The seductive appeal of mindless neuroscience*. New York, NY, US: Basic Books.

Scherbaum, S., & Kieslich, P. J. (2018). Stuck at the starting line: How the starting procedure influences mouse-tracking data. *Behavior Research Methods*, 50(5), 2097–2110. https://doi.org/10.3758/s13428-017-0977-4

Scherbaum, S., Dshemuchadse, M., Fischer, R., & Goschke, T. (2010). How decisions evolve: The temporal dynamics of action selection. *Cognition*, *115*(3), 407–416. doi:10.1016/j.cognition.2010.02.004

Scholz, A., Krems, J. F., & Jahn, G. (2017). Watching diagnoses develop: Eye movements reveal symptom processing during diagnostic reasoning. *Psychonomic Bulletin & Review*, *24*(5), 1398–1412. doi:10.3758/s13423-017-1294-8

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychol-*

ogy: General, *122*(2), 166–183. doi:10.1037/0096-3445.122.2.166

Sullivan, N. J., Hutcherson, C., Harris, A., & Rangel, A. (2015). Dietary self-control is related to the speed with which health and taste attributes are processed. *Psychological Science*, 26(2), 122–134. doi:10.1177/0956797614559543

Schulte-Mecklenbeck, M., Johnson, J. G., Böckenholt, U., Goldstein, D. G., Russo, J. E., Sullivan, N. J., & Willemsen, M. C. (2017). Process-Tracing Methods in Decision Making: On Growing Up in the 70s. *Current Directions in Psychological Science*, *26*(5), 442–450. doi:10.1177/0963721417708229

Schulte-Mecklenbeck, M., Kühberger, A., & Ranyard, R. (2011). *A handbook of process tracing methods for decision research: A critical review and user's guide*. (M. Schulte-Mecklenbeck, A. Kühberger, & R. Ranyard, Hrsg.). New York, NY, US: Psychology Press.

Schulte-Mecklenbeck, M., Sohn, M., de Bellis, E., Martin, N., & Hertwig, R. (2013). A lack of appetite for information and computation. Simple heuristics in food choice. *Appetite*, *71*, 242–251. doi:10.1016/j.appet.2013.08.008

Schulze, R. (1909). *Aus der Werkstatt der experimentellen Psychologie und Pädagogik*. [From the workshop of experimental psychology and pedagogy]. Leipzig, Germany: Voigtländer.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, *81*(3), 214–241. doi:10.1037/h0036351

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *102*(29), 10393–10398. doi:10.1073/pnas.0503903102

Sternberg, S. (1966). High-speed scanning in human memory. *Science*, *153*(3736), 652–654. doi:10.1126/science.153.3736.652

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, Amsterdam*, *30*, 276–315. doi:10.1016/0001-6918(69)90055-9

Stillerman, B. J., & Freeman, J. B. (2019). Mouse-tracking to understand real-time dynamics of social cognition. In M. Schulte-Mecklenbeck, A. Kühberger & J. G. Johnson (Eds.). *A Handbook of Process Tracing Methods* (2nd Edition). London: Routlege.

Thomas Jr., J. C. (1974). An analysis of behavior in the hobbits-orcs problem. *Cognitive Psychology*, *6*(2), 257–269. doi:10.1016/0010-0285(74)90013-9

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, *32*(7), 1206–1220. doi:10.3758/BF03196893

Ward, J. (2015). *The student's guide to cognitive neuroscience., 3rd ed.* New York, NY, US: Psychology Press. doi:10.4324/9781315742397

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*(2), 158–177. doi:10.1037/h0074428

Willemsen, M. C., & Johnson, E. C. (2019). (Re)Visiting the decision factory: Observing cognition with MouselabWEB. In M. Schulte-Mecklenbeck, A. Kühberger & J. G. Johnson (Eds.). *A Handbook of Process Tracing Methods* (2nd Edition). London: Routlege.

Wulff, D. U., Haslbeck, J. M. B., Kieslich, P. J., Henninger, F., & Schulte-Mecklenbeck, M. (2019). Mouse-tracking: Detecting Types in Movement Trajectories. In M. Schulte-Mecklenbeck, A. Kühberger & J. G. Johnson (Eds.). *A Handbook of Process Tracing Methods* (2nd Edition). London: Routlege.

Wundt, W. (1907). Über Ausfrageexperimente und über die Methoden zur Psychologie des Denkens [About Interrogation Experiments and about the Methods of a Psychology of Thinking]. *Psychologische Studien*, *3*, 301–360.

Wundt, W. (1908). Kritische Nachlese zur Ausfragemethode [Critical Gleanings on the Interrogation Method]. *Archiv für die Gesamte Psychologie*, *11*, 445–459.

Glossary

area of interest (AOI) Region on a display that is predefined as theoretically interesting when fixated by a participant during eye tracking. 42

computational model A theory or theoretical assumptions about cognitive processes cast in precise mathematical terms to predict and/or explain empirical phenomena. 46

correlation study Empirical study in which variables that are expected to covary are meaured or observed in order to test whether they actually covary. This typically does not allow for causal interpretations of their connection. 27

dependent variable (DV) The variable of interest, the value of which we try to explain by other (independent) variables. 27

experiment The independent variable (IV) is actively manipulated by an experimenter to observe its impact on the dependent variable (DV). If participants are randomly assigned to different levels of the IV, the effects of the IV on the DV can be interpreted causally. 28

eye tracking A method for monitoring people's visual information intake by recording their eye movements during the inspection of a display. 41

fMRI Functional Magnetic resonance Imaging is a non-invasive method to measure the regional blood flow in the brain during task processing. Contrasting the activity during different tasks allows to infer the brain regions crucially involved in the processes that differ between tasks. 46

independent variable (IV) The IV is a variable that is theoretically assumed to influedce another variable of interest (the dependent variable). 27

indicator An indicator variable ist an observable variable (e.g. the number of test items solved) that reflects a theoretically interesting unobservable variable (e.g. intelligence). 28

introspection The method of observing one's own cognitive processes during a task and to report on them. This method has been criticized very early as subjective and error-prone. 28

methodological behaviorism A methodological position that keeps the behaviorist conviction to base empirical data solely on objectively observable behavior. However, in contrast to radical behaviorism, the position does not deny that unobservable processes (like cognitive processes) exist, and hypotheses about them can be tested by relying objective data. 29

neuropsychology Classical method of precisely documenting cognitive impaiments caused by circumscribed bran damage in order to localize brain functions. 46

operationalization The process of translating a latent variable (e.g. memory strength) into an empirically observable variable (e.g. number of recalled items). 27

reactivity A problem that may arise if the assessment method of a process changes the process itself. 29

thinking-aloud-method Verbalization of all thoughts during a problem solution. This method may give a researcher hints on the nature of strategies used. 29