Tracing Current Explanations in Memory: A Process Analysis Based on Eye-Tracking

Journal:	*Quarterly Journal of Experimental Psychology*
Manuscript ID	QJE-STD-19-188.R2
Manuscript Type:	Standard Article
Date Submitted by the Author:	13-Mar-2020
Complete List of Authors:	Klichowicz, Anja; Chemnitz University of Technology, Strehlau, Sascha; Chemnitz University of Technology Baumann, Martin; Department of Psychology, University of Ulm, Germany Krems, Josef; Chemnitz University of Technology Rosner, Agnes; University of Zurich Department of Psychology
Keywords:	sequential abductive reasoning, eye movements, memory indexing, process tracing, mental representation

DOI: 10.1177/1747021820922509

Tracing Current Explanations in Memory: A Process Analysis Based on Eye-Tracking

Anja Klichowicz [a], Sascha Strehlau [a], Martin R. K. Baumann[b], Josef F. Krems [a], and Agnes Rosner[c]

[a]Department of Psychology, Chemnitz University of Technology, Germany; [b]Department of Psychology, University of Ulm, Germany; [c]Department of Psychology, University of Zurich, Switzerland

Correspondence concerning this article should be addressed to Anja Klichowicz, Department of Psychology, Chemnitz University of Technology. E-mail: anja.klichowicz@psychologie.tu-chemnitz.de

DOI: 10.1177/1747021820922509

Abstract

Sequential abductive reasoning is the process of finding the best explanation for a set of observations. Explanations can be multicausal and require the retrieval of previously found ones from memory. The theory of abductive reasoning (TAR) allows detailed predictions on what information is stored and retrieved from memory during reasoning. In the research to date, however, these predictions have never been directly tested. In the present study, we tested process assumptions such as the construction of a mental representation from TAR using memory indexing, an eye-tracking method that makes it possible to trace the retrieval of explanations currently held in working memory. Gaze analysis revealed that participants encode the presented evidence (i.e., observations) together with possible explanations into memory. When new observations are presented, the previously presented evidence and explanations are retrieved. Observations that are not explained immediately are encoded as abstractly explained. Abstract explanations enter a refinement process in which they become concrete before they enter the situation model. With the memory indexing method, we were able to assess the process of information retrieval in abductive reasoning, which was previously believed to be unobservable. We discuss the results in the light of TAR and other current theories on the diagnostic reasoning process.

Keywords: sequential abductive reasoning, eye movements, memory indexing, process tracing, mental representation

Introduction

The belief that "nihil sine causa" ("there is nothing on earth without a cause", Job 5:6) is deeply rooted in human experience. Yet often, outcomes have multiple causes. Digestive difficulties, for instance, may be caused by both a lactose intolerance as well as a soy allergy. Research on diagnostic and abductive reasoning examines how people generate and evaluate observations to find the best causal explanation (Johnson & Krems, 2001). In diagnostic reasoning, explanations are retrieved from a predefined set held in long-term memory, whereas in abductive reasoning, explanations are inferred from a given case—for instance, existing observations—based on a rule (Peng & Reggia, 1990). Note that the term "observation" describes the evidence that is available to the reasoner, but does not say whether that evidence has actually been examined ("observed") by the reasoner. In this study, we concentrate on the application of rules to find the best causal explanation and therefore generally use the term abductive reasoning, but occasionally use the terms diagnostic and abductive reasoning interchangeably.

Research on abductive reasoning is complex. This is due, first, to the high number of alternative explanations that exist for any given set of observations (multicausality). Second, different pieces of evidence often do not become available at once but only over time (e.g., we listen to one piece of information after the other; we first look at one piece of information before shifting our gaze to the next visuospatial location due to limited cognitive capacities). This is why research on diagnostic reasoning studies how a number of sequentially presented observations are integrated over the course of the diagnostic process (Jahn & Braatz, 2014; Johnson & Krems, 2001).

The sequential integration of information enables the retrieval of critical information from memory in order to generate and evaluate the best causal explanation. Recent research on this subject has shown that this generation and evaluation process depends on the activation of information from long-term memory (Mehlhorn, Taatgen, Lebiere, & Krems,

DOI: 10.1177/1747021820922509

2011; Rebitschek, Krems, & Jahn, 2016; Thomas, Dougherty, Sprenger, & Harbison, 2008). In a study by Mehlhorn et al. (2011), participants first learned associations between medical symptoms and potential chemicals that caused them (and therefore act as their explanations). Second, participants were asked to diagnose which chemical accounted for a number of sequentially presented medical symptoms by retrieving the before learned associations from long-term memory. The availability of potential explanations (chemicals) in memory was measured with an implicit probe reaction time task, which is based on the idea that explanations evaluated as highly plausible lead to faster responses than implausible alternatives. For instance, if a chemical A is associated with a symptom, then asking participants if probe A is the name of a chemical after the symptom has been presented should lead to a faster response than presenting an unrelated letter that is not associated with the symptom or no chemical at all (such as E). The behavioral results reported by Mehlhorn et al. (2011) are best described with a cognitive model in which all observations (symptoms) stored in working memory have the same potential to activate explanations from long-term memory. In that study, it should be noted that the final diagnosis consisted of only one explanation.

As retrieval from memory is essential for successful reasoning, the goal of our study was to observe *what* information is retrieved from memory in order to find a multicausal explanation for sequentially presented observations. We studied a case in which only the rules are learned and the participants infer explanations from presented observations (i.e., abductive reasoning); that is, only rules are retrieved from long-term memory. This makes it likely that explanations and observations will be stored in working memory for further processing over the course of the trial. Finally, instead of using an implicit probe reaction time task, we used eye movements as an implicit measure of which observations and explanations a reasoner retrieves from memory. In the following, we first describe current theories on abductive reasoning and then describe the eye-tracking method we used.

DOI: 10.1177/1747021820922509

Current Explanations of Abductive Reasoning

Current theories on the process of abductive reasoning often focus on testing whether and to what extent humans deviate from normative solutions to reasoning problems—that is, to what extent they reason without taking memory retrieval into account (e.g., Johnson-Laird, Byrne, & Schaeken, 1992; Johnson-Laird, 1980). In the revised version of the famous mental models theory, Johnson-Laird, Khemlani, and Goodwin, (2015) argued that as a person encounters new information, they generate all possible interpretations of an argument. In the case of abductive reasoning, this would mean that all possible explanations are generated from the set of observations, which poses a huge challenge to working memory. In a similar vein, Bayesian and related probabilistic approaches assume that with every piece of new information, the reasoner updates the likelihood of each explanation. Whereas mental models and Bayesian approaches describe diagnostic reasoning behavior relatively well (e.g., Meder & Mayrhofer, 2017), they are strictly normative approaches that cannot explain how memory influences the reasoning process.

In contrast, process theories focus on describing the actual operations needed to perform a reasoning task. For instance, the HyGene model (Thomas et al., 2008) successfully describes how explanations are generated as well as how those explanations serve as the basis for probability judgments concerning an explanation and hypothesis testing in order to find the most likely one (Lange, Thomas, & Davelaar, 2012). In HyGene, explanations are generated by activating past instances stored in long-term memory. Selection among generated explanations can further be described with the help of spreading activation models, in which the current explanation receives the highest level of activation (Baumann, Mehlhorn, & Bocklisch, 2007; Mehlhorn et al., 2011). However, recent process models have examined only how individuals infer a single cause for a set of observations.

DOI: 10.1177/1747021820922509

One theory that has focused on how people infer a multicausal explanation on the basis of sequentially presented observations is the theory of abductive reasoning (TAR, Johnson & Krems, 2001). It is a comprehensive process theory on abductive reasoning for multicausal explanations that allows detailed predictions about information storage and retrieval in the process of abductive reasoning. In the following, we describe the main assumptions of TAR in more detail.

1. **Current explanation in the situation model**. TAR assumes that observations and explanations are stored in memory in a single situation model that represents the *current explanation* of the observations made thus far. The current explanation is therefore the overall explanation that takes all observations and explanations into account that have been obtained to date. As is the case with most real-world problems, observations are presented sequentially. Each new piece of information therefore runs through a comprehension process in order to be related to previous observations. If an explanation can be inferred, observations and explanations are integrated into the situation model. Integration, in this case, refers to "chunking" observations and their explanations and then assigning the chunks to slots in memory.
2. **Types of explanation—concrete vs. abstract**. TAR differentiates between several classes of explanations: A *concrete* explanation is able to explain one or more observations. For instance, experiencing stomach pain after drinking cow's milk suggests possible lactose intolerance. If an observation can be explained only by a set of related concrete explanations, TAR refers to this set as an *abstract* explanation. For instance, without additional tests, it is not clear whether the symptom of chronic fatigue is best explained by liver disease, depression, or

DOI: 10.1177/1747021820922509

something else entirely. In this study, we use only two classes of explanations, *concrete* and *abstract,* to describe how observations are explained.[1]

3. **Consistency checks and explanation refinement**. Each new observation leads to the generation of an explanation, which is then tested in terms of its implications for the current explanation over all observations. This process is called consistency checking and includes the following steps. First, the current explanation (consisting of all explanations found thus far) is retrieved. Second, the current observation is evaluated according to its fit with the current explanation. Third, if the current observation can be explained by a single concrete explanation, that explanation becomes part of the current explanation, but if it can only be explained by a set of concrete explanations, an abstract explanation is constructed. Abstract explanations thus allow the current explanation to represent a range of possible explanations that must then go through a *refinement process*. The refinement process can be described in the following way. Each possible concrete explanation associated with the abstract explanation is evaluated in terms of its overlap with the current explanation. One of three cases can occur: (a) If only one concrete explanation in the abstract set overlaps with the current explanation, it is integrated and the observation is "concretely explained"; (b) If the observation can be explained by a pre-existing concrete explanation, no additional explanation is stored in the situation model; (c) If the intersection contains more than one concrete explanation, an abstract explanation is produced and integrated into the situation model as a set of concrete explanations. In the latter case, additional information must then be collected.

[1] TAR proposes a third class of explanations (*disjunctive* explanations, which contradict each other; see Johnson & Krems, 2001) which, for simplification purposes, we do not address further in this study.

DOI: 10.1177/1747021820922509

Tracing the Reasoning Process

Observing memory processes in multicausal abductive reasoning is challenging, since the processes of interest are hidden from view. Previous research focused on examining the outcomes of the reasoning process (e.g., response times and choices) and was only able to deduce reasoning processes indirectly (e.g., Mehlhorn et al., 2011). Recent research has shown that eye movements can be used to observe memory retrieval during diagnostic reasoning (Jahn & Braatz, 2014; Scholz, Krems, & Jahn, 2017). This so-called memory indexing method (Jahn & Braatz, 2014; Renkewitz & Jahn, 2012) is a process measure based on measuring eye movements. It utilizes the looking-at-nothing phenomenon, which describes the behavior in which, when retrieving information that was previously encoded at a spatial position, a person's gaze returns to the previous location of the information, even when the information in question is no longer displayed (Laeng & Teodorescu, 2002; Richardson & Kirkham, 2004; Scholz, Klichowicz, & Krems, 2017; Spivey & Geng, 2001; for an overview see Ferreira, Apel, & Henderson, 2008; Richardson, Altmann, Spivey, & Hoover, 2009). When encoding information, a spatial index is stored along with the semantic memory content (e.g., Pylyshyn, 1998; Richardson & Kirkham, 2004). When the stored information is probed, the associated spatial index is reactivated as well. Activating the spatial index calls upon the oculomotor system, which leads to an eye movement toward the associated, but now vacant spatial location. Jahn and Braatz (2014) applied memory indexing to study sequential diagnostic reasoning. In their study, participants were asked to diagnose which of four chemicals a worker in a chemical plant had encountered during an accident in the factory (chemical-accident task; see also Mehlhorn et al., 2011). Participants first encoded causal information on chemicals and their symptoms in a learning phase in which each symptom class and possible diagnoses (chemicals) were presented in rectangles on the screen. During reasoning trials, participants then listened to an auditory presentation of the symptoms in sequence while viewing the empty rectangles on the computer screen. Participants looked at the empty rectangles when listening to new symptoms and when inferring the explanation. That is, the higher the likelihood of an explanation, the more participants looked toward the spatial location associated with this explanation. Given this evidence, studying eye movements based on the memory indexing procedure appears to be a promising means of studying memory retrieval during the process of finding the best explanation for a set of observed symptoms, that is, the processes of abductive reasoning.

Present Study

The goal of this study is to test process assumptions derived from TAR regarding memory retrieval during abductive reasoning in the case of a multicausal explanation. To achieve this goal, we use the memory indexing method, which allows one to trace memory retrieval via eye movements. We combine the memory indexing method with a reasoning task that was developed to test assumptions of TAR and that has been shown to accurately examine the abductive reasoning process: the so-called “black box task” (BBX; Johnson & Krems, 2001). BBX is a visuospatial reasoning task in which participants must locate a number of atoms hidden in a box by shooting light rays into the box and observing where the rays exit the box (see Figure 2). The entrance and exit locations of the light rays represent the observations. The combination of exit and entry locations are the basis for inferring a path of the light ray through the black box. This path is generated by the interactions of the light rays with the hidden atoms as defined by a small number of predefined rules. Assumptions regarding the locations of the hidden atoms represent the explanations. Consequently, in the black box task, explanations must be inferred from a set of predefined rules. Even though deductive problems also involve the inference of a special case based on a common rule, our task is, according to Johnson and Krems (2001), clearly abductive since it involves finding a hidden state of a system. That is, the reasoner enters input into a system and derives its state based on observing the corresponding output. In our case, the entry position of the rays form the input,

DOI: 10.1177/1747021820922509

and the exit positions form the output. The hidden state is described by the atoms, which determine the path the rays take through the black box.

Further, the BBX method enables researchers to trace the generation of *causal* explanations (e.g., by studying consistency checking and the refinement process) rather than the retrieval of learned associations or past instances stored in memory (e.g., Klahr & Dunbar, 1988; Thomas et al., 2008). This makes the task unique in comparison to previously studied paradigms like the medical diagnoses task (Jahn & Braatz, 2014; Scholz, Krems, & Jahn, 2017; Mehlhorn et al., 2011). The visuospatial nature of the task makes it especially well suited to eye-tracking methods (see Hedge & Leonards, 2013, or Spivey & Geng, 2001 for studying eye movements "to nothing" in a visuospatial memory task). Furthermore, as will be explained in more detail below, it is possible to experimentally manipulate the case that a *single* explanation might explain more than one observation, whereas *several* explanations might be needed to explain one observation.

The focus of this study was on investigating the use of the current explanation stored in memory. Therefore, previously seen observations and set atoms were removed from the screen as soon as a new observation was requested. This procedure differs from those used in previous studies, in which all seen observations and set atom locations were visible to the reasoner throughout the trial (e.g., Johnson & Krems, 2001). Even though this method can reveal what information is retrieved during the reasoning process, one caveat is that up to now, there has been no way to differentiate retrieval and inference processes in peoples' eye movements. For instance, it is not possible to disentangle whether participants actually retrieve explanation locations or infer them from retrieved observation locations. However, the retrieval of information (either only the observation location, in order to infer the atom location, or the explanation and observation locations) is the most important assumed mechanism in TAR, and up to now, it has not been possible to test it directly. Similar problems have been raised with regard to differentiating between encoding and retrieval

DOI: 10.1177/1747021820922509

during decision-making (Glaholt & Reingold, 2011; Horstmann, Ahlgrimm, & Glöckner, 2009). Still, this is the first study to use memory indexing in multicausal abductive reasoning. One of its primary goals is to show that eye tracking provides a method to reveal which information is used during multicausal reasoning. In the following, we outline our hypotheses concerning memory indexing gaze behavior in the abductive reasoning task.

Hypothesis 1: Information Stored in the Situation Model

TAR states that all observations and explanations are encoded into a situation model. With each new observation, individuals attempt to comprehend the information and check the new explanation generated to explain the observation for consistency with the situation model. Concerning the memory indexing gaze behavior, we therefore expect that participants will look toward the current observation (because it is visible on the screen) as well as the atom location that must be inferred from the observation (which, in the case of a concrete explanation, can be inferred by applying learned rules).

Furthermore, TAR states that the derived explanations must be checked for consistency with the current explanation. Therefore, TAR assumes that all previous observations and atom locations are retrieved. Consequently, we expect participants' gaze to return to the previous observation locations and previously inferred atom locations (concrete explanations), even though previous observations are no longer visible in the visual array and must, therefore, be retrieved from memory. In the same vein, already inferred atoms must be retrieved from memory and should, therefore, be revisited.

Additionally, we expect that as more information is included in the situation model, eye gaze is also directed toward an increasing number of locations in the black box as more information has to be retrieved.

Hypothesis 2: Types of Explanation—Concrete vs. Abstract

As explained above, abstract explanations are the representation of a set of concrete explanations for an observation that cannot be explained immediately. The abstract

DOI: 10.1177/1747021820922509

explanation must then enter a refinement process to become a concrete one that is coherent with the situation model. This additional process requires additional processing time.

Therefore, we first expect participants to look more toward the set of concrete explanations representing the abstract explanation than toward any irrelevant field, as all concrete explanations belonging to the abstract explanation that overlap with previous knowledge become part of the situation model.

Because abstract explanations enter the refinement process, which represents an additional step in TAR, we expect participants to look more toward abstractly explained observations than toward concretely explained observations as participants try to further concretize the corresponding abstract explanation. This should also result in increased response times (increased time before a new observation is obtained) for abstractly explained observations in comparison to concretely explained observations.

Method

Participants

The sample size was chosen similarly to the only two other studies we know of that have used memory indexing to study sequential diagnostic reasoning (Jahn & Braatz, 2014, Scholz, Krems, & Jahn, 2017). Of the 29 participants for whom eye tracker calibration succeeded with an accuracy of at least 2° of visual angle, one participant had to be excluded due to a drop in eye tracking accuracy during the experiment. The final 28 participants were all students from Chemnitz University of Technology (20 female, 8 male) with a mean age of 22.3 years (*SD* = 3). All had normal or corrected to normal vision.

Apparatus

Participants were seated at a distance of 63 cm from the front of a 22-inch computer screen (1680 x 1050 pixels). Stimuli were presented via E-Prime 2.0. Participants provided responses using a standard keyboard and computer mouse. An SMI RED remote eye tracker sampled

DOI: 10.1177/1747021820922509

data from the right eye at 120 Hz during the reasoning task. Gaze data were recorded with iView X 2.5 following five-point calibration. Event detection and heat map analyses were performed with BeGaze 3.0. Fixation detection used a dispersion threshold of 100 pixels and a duration threshold of 80 ms. Data were analyzed using IBM SPSS 24, Microsoft Excel 2016, and JASP0.8.

Material

The BBX-task consisted of a 10 x 10 grid with a size of 25.92 x 26.27° of visual angle (1015 x 1029 pixels). In this grid, participants were asked to locate hidden atoms by observing where light rays enter and exit the box. Participants did not see the path of the light rays. They only saw the entrance and exit positions of each light ray within the BBX, which were indicated by a number appearing in the outer rectangles of the BBX (see Figure 1). These observation locations marked the observations that had to be explained according to TAR. As shown in Figure 1, each atom has a field of influence (a circle around the atom). When a light ray hits that field, both interact according to a set of predefined rules. When the light ray does not hit a field of influence, it goes straight through the box. If the ray hits the field of influence at an angle, it is reflected 90° away from the atom, resulting in a so-called L-pattern. The observation is concretely explained. If the ray hits the atom directly, it is absorbed and does not exit the BBX. Combinations of two reflections can result in a Z- or a U-pattern. In the following, the observations will be called L-pattern, absorption pattern, Z-pattern, and U-pattern. However, the actual observations are still visible in the observation locations of the ray. The names of the observation patterns describe the path the ray would be most likely to take based on the observation locations, and are used to describe the observation for a better understanding of the participant's task. Note that an immediate concrete explanation of absorption, U-pattern, and Z-pattern might not be possible since the observations indicate only the row or column of the atom location. To locate the exact position, at least one other

DOI: 10.1177/1747021820922509

observation must be concretely explained, which can be presented either before (for an example, see Figure 2A) or after the absorption, Z-pattern, or U-pattern (for an example, see Figure 2B). If the absorption, U-pattern, or Z-pattern cannot be explained with the help of previously observed observations or previously set atoms, the atom's location is abstractly explained and, therefore, results in an abstract explanation.

--- Insert Figure 1 about here ---

After each new observation (i.e., entry/exit location of a ray), participants were asked to indicate the location of the atoms based on the information drawn from the light rays they shot into the box. Participants were asked to place on the grid the minimum number of atom markers necessary to explain the ray pattern of the respective trial. That is, not every new observation called for a new atom as it was also possible that an observation was already explained by previously set atoms. On the other hand, two atoms could be necessary in order to explain an observation (especially in case of U-patterns and Z-patterns). This underlines the multicausal nature of the task, as a number of atoms were needed to explain the observations throughout a trial. As a result, the number of atoms differed between 2 and 4 depending on the observation pattern in the trial. Furthermore, participants were able to proceed to the next observation without placing an atom, as sometimes participants did not yet have all of the information needed to place the atom (e.g., in the case of an unexplained absorption) or observations were explained by previously set atoms (e.g., in case of an explained absorption). Currently placed atoms remained visible until participants acquired the next observation information. As soon as a new observation was presented, all previously seen observations and atoms set by the participants disappeared. During a single trial, participants watched between 4 and six rays in sequential order. Each ray was indicated by a number at the entrance and exit position in the BBX (Figure 2A). Previous observation locations as well

DOI: 10.1177/1747021820922509

as the locations of previously set atoms disappeared when a new observation was presented. Participants had to remember on average three already placed atoms per trial in order to place the atom corresponding to the last observation. In this way, the paradigm allowed us to maximize the possible influence of information stored and retrieved in memory on abductive reasoning.

Of the 48 experimental trials presented to each participant, each ray could immediately be concretely explained in 28 trials (trials consisting of only concrete explanations). Twenty trials contained rays that could only be explained once subsequent light rays were observed (abstract explanation). The test set contained 25 trials with four rays, 13 trials with five rays, and 10 trials with six rays. Each trial could involve zero, one, or two abstract explanations.

The participant decided when to observe the next light ray by pressing the space bar. The number of remaining observations to be observed in each trial was displayed by a digit at the upper left corner of the grid. An example of the trial is displayed in Figure 2A.

--- Insert Figure 2 about here ---

Procedure

The experiment began with an initial instruction phase in which participants learned the rules of the BBX, followed by two training phases. During the first training phase, both observation locations and atoms remained visible throughout the trial—thus, they did not have to be stored in memory. In the second training phase, participants completed the trial under test conditions. That is, only the locations at which the current light ray entered and exited the BBX and the counter in the upper left corner were displayed. Once participants used the computer mouse to indicate the possible position of the atom, the location of both the atom and its field of influence were displayed on the computer screen. Participants could then press

DOI: 10.1177/1747021820922509

the space bar to shoot a new light ray into the BBX. Each training phase consisted of seven trials, which included feedback regarding the true position of the atoms at the end of the trial. That is, atoms that were placed correctly appeared in green, whereas incorrectly placed atoms appeared in red. Atoms that participants failed to place at all were displayed in blue.

Following eye tracker calibration, participants completed 48 test trials that followed the same procedure as the second training phase, but without feedback. Participants needed 53 to 127 minutes to complete the entire experiment (M = 85.2 min; SD = 21.1). Due to the large number of trials and high complexity of the task, participants were allowed to take a five-minute break every 30 minutes, after which the experiment was continued with a recalibration of the eye tracker. Each calibration procedure lasted until an accuracy of at least 2° visual angle was reached.

Results

Performance

Participants solved, on average, 78.7% of the trials correctly. There was no difference in accuracy between trials with only concretely explained (CE) observations (M_{CE} = 80.2%, SD = 10.4) and trials including abstractly explained (AE) observations (M_{AE} = 76.2%, SD = 15.7, $t(27) = 1.45$, $p = .16$, $d = 0.27$, $BF_{01} = 1.97$), indicating that all of the trials in the task were of comparable difficulty. However, there was a difference in accuracy between trials with four (M_4 = 85.1%, SD = 11.5), five, (M_5 = 69.6%, SD = 14.3) and six (M_6 = 69.4%, SD = 19.7) observations (F (1.71, 46.25) = 15.1, $p < .001$, $\eta_p^2 = 0.36$, $BF_{10} > 1000$). This difference is due to significantly better performance in trials with four light rays as compared to trials with five or six rays (both ps = <.001). Bonferroni pairwise comparisons reveal no difference between trials with five and six rays (p = 1).

Participants required an average of 38 seconds (SD = 15 s) to complete a trial. There was no difference in the time needed per trial between trials consisting of only concretely

DOI: 10.1177/1747021820922509

explained observations (M_{CE} = 37.04 s, *SD* = 15.82 s) and trials including abstractly explained observations (M_{AE} = 39.34 s, *SD* = 17.12 s, $t(27) = -1.08$, $p = .29$, $d = -0.20$, $BF_{01} = 2.95$). There was a difference in the time needed to solve trials consisting of four (M_4 = 35.19 s, *SD* = 16.38 s), five (M_5 = 39.81 s, *SD* = 16.10 s), and six (M_6 = 46.40 s, *SD* = 21.35 s) observations ($F(1.60, 43.30) = 9.17$, $p = .001$, $\eta_p^2 = 0.25$, $BF_{10} = 1000$).

We only analyzed correctly solved trials, as we formulated no hypotheses regarding participants' reasoning behavior when they made errors. For this reason and to provide a better understanding, we report results collapsed over trials with four, five, and six observations. However, when controlling the number of observations presented per trial, the analysis yields exactly the same pattern of results (see Appendix 1).

Gaze Analyses

Each square of the grid and the grey border of the black box were treated as areas of interest (AOIs), resulting in 100 separate AOIs with a size of 2.64° x 2.64° of visual angle (102 x 102 pixels) each. We coded the AOIs where the rays entered and exited the BBX and hit an atom's field of influence as well as the AOIs where atoms should be placed if participants correctly considered all the observations presented thus far (see Figure 3). For the analyses, both the field containing the atom and the field containing the location at which the respective light ray hit the atom's field of influence were combined into a single atom AOI. Additionally, we coded the entire row or column in which an abstractly explained atom could currently be located as well as all fields between observation locations and atoms for all patterns that could result in an abstract explanation (absorption, U-pattern, Z-pattern).

We coded the observation currently being seen as the current observation and the currently inferred explanation location as the current atom. Note that current observation and explanation locations are still visible on the screen. For each observation within a trial sequence, we summed up fixation durations to the previous and current observation locations, and the previous, current, and irrelevant atom locations (see Figure 2). This procedure

DOI: 10.1177/1747021820922509

allowed us to ensure that fixations were not counted twice as "current" and "previous" fixations. Next, we aggregated these values across the four to six observations presented within one trial sequence for the analyses of Hypothesis 1 and across abstract and concrete explanations within one trial sequence for the analyses of Hypothesis 2. Note that previous explanation and observation locations were no longer present in the visual array and had to be retrieved from memory.

--- Insert Figure 3 about here ---

As dependent measures, we analyzed fixation times, fixation counts, and strike rates. Fixation times are the sum of all fixation durations to an explanation in milliseconds (e.g., an atom location) or an observation (both entry and exit positions of the ray). Fixation counts were calculated congruently, as the sum of the number of fixations (e.g., to the observation locations) of relevant AOIs. As a third dependent variable, we introduced the strike rate, which was the total number of AOIs looked at. Even if an AOI was looked at only once, it was included in the index. As fixation time, fixation count, and strike rate show very similar result patterns, we only report results on fixation times in this section. The analysis of fixation counts and strike rates can be found in the Appendix. The time participants spent looking at the grid, from the onset of the observation information until the participant requested a new observation, will henceforth be referred to as viewing time.

Fixation times to the atoms as well as the observation locations of the light rays were compared to irrelevant AOIs, which were chosen randomly from all irrelevant AOIs for each trial. Irrelevant AOIs were those that contained neither an observation, nor an atom, nor a field of influence throughout the trial. This was to ensure that fixations to previous atom locations differ from randomly hitting the AOI with the eyes while looking at the black box grid. The number of irrelevant AOIs included in the analyses corresponds to the number of

DOI: 10.1177/1747021820922509

relevant AOIs to which it was compared. Therefore, the number of randomly sampled irrelevant AOI varies over different analyses. However, observation positions are always compared to irrelevant AOIs of the grey border and atom locations, and abstract explanations are compared to a corresponding number of irrelevant AOI located in the white grid.

Hypothesis 1: Information Stored in the Situation Model

Based on the assumptions of TAR, we expected, first, that to make an inference, current observations and atom locations are looked at for a longer period of time than previous observation and atom locations or irrelevant AOIs. Second, we assumed that to check for consistency with previously seen and inferred observations and atoms, participants should also look longer at previous atom and observation locations than at irrelevant areas of the grid. To test the two hypotheses, we conducted a 3 (information type: current, previous, irrelevant locations) x 2 (location: atom, observation) repeated measures ANOVA. As sphericity was not given, a Greenhouse-Geisser correction was used.

Additionally, we expected a growing situation model, as more information is included with each new observation. This should result in a growing number of AOIs that participants look at throughout each trial. To test this, we calculated separate repeated measures ANOVAs for the total factor number of seen observations for set sizes 4, 5, and 6.

Current and previous information. We found a main effect for the information type $F(1.11, 30{,}04) = 148.70$, $p < .001$, $\eta_p^2 = 0.85$, $BF_{10} > 1000$). As shown in Table 1, Bonferroni post hoc comparisons revealed that participants directed their gazes toward the current observation and atom location significantly longer than toward previous observation and atom locations ($p < .001$, $d = 1.76$) and the irrelevant AOIs ($p < .001$, $d = 1.68$). Also, participants looked significantly longer at previous locations than at irrelevant areas ($p < .001$, $d = 0.79$).

DOI: 10.1177/1747021820922509

Note that these previous locations were not visible in the visual array but, instead, had to be retrieved from memory in order to make a correct inference.

We found a main effect for the factor atom vs. observation locations (F(1.00, 27.00) = 118.00, $p < .001$, $\eta_p^2 = 0.81$, $BF_{10} > 1000$) as well as a significant interaction between the factors information type and locations (F(1.24, 33.86) = 124.6, $p < .001$, $\eta_p^2 = 0.82$, $BF_{10} >$ 1000). As becomes evident in Figure 4, participants looked for a shorter period of time at observation than at atom locations. This decline in fixation durations was steeper for atoms than for observations, reflecting the fact that people spent by far the longest amount of time looking at the current atom location. It should be noted that a paired sample t-test reveals that the difference between previous observation locations and irrelevant observation AOIs is also significant ($t(27) = 6.09$, $p < .001$, $d = 1.15$, $BF_{10} > 1000$).

The same result pattern can be found when analyzing fixation counts and strike rates (see Appendix 2).

--- Insert Figure 4 about here ---

Expansion of the situation model. As more information is included in the situation model, eye gaze should also be directed toward an increasing number of locations in the black box grid, as more information has to be retrieved. As only the current observation is visible on the screen, eye movements have to be driven by the retrieval of previous information. Therefore, gazes toward other areas within the BBX reflect the updating of the situation model.

We tested this assumption with separate repeated measures ANOVAs for the factor number of seen observations for set sizes 4, 5, and 6. The Greenhouse-Geisser corrected repeated measures ANOVA indicates that in line with the hypothesis, the total number of

DOI: 10.1177/1747021820922509

AOIs gazed at[2] increases with an increasing number of presented observations for trials with four, five, and six observations ($F_4(1.26,34.13) = 32.65$, $p < .001$, $\eta_p^2 = 0.55$, $BF_{10} > 1000$, $F_5(3.03,81.91) = 5.49$, $p = .002$, $\eta_p^2 = 0.17$. $BF_{10} = 56.70$, $F_6(3.88,104.79) = 5.84$, $p < .001$, $\eta_p^2 = 0.18$, $BF_{10} = 325.77$). Figure 5 illustrates this finding in the heat map of one and the same trial consisting of four observations that was solved correctly by all 28 participants. It shows that as more information must be stored in the situation model, participants direct their eyes toward an increasing number of locations in the BBX grid.

In sum, the results are in line with the assumptions of Hypothesis 1.

--- Insert Figure 5 about here ---

Hypothesis 2: Different Types of Explanations—Concrete vs. Abstract

According to TAR, observations for which there is not a single concrete explanation result in an abstract explanation that needs to go through a refinement process in order to become a concrete explanation. To test this assumption, we performed a set of three analyses. First, concerning fixation durations to AOIs in the black box grid, we tested whether participants look longer for AOIs of the *abstract explanations* in comparison to irrelevant AOIs, as this is an indicator of abstract explanations being represented in the current explanation. Second, as evidence of the refinement process, we tested whether participants look longer for *observation locations* of abstractly as opposed to concretely explained observations. Third, again as a measure of refinement, we tested whether viewing times increased (in terms of the time

[2] We chose to include all AOIs of the grid in this analysis rather than only relevant atom AOIs for the following reasons: First, in line with results on Hypothesis 1, showing that participants almost never looked at completely irrelevant AOIs, we also assumed in this analysis that participants only looked at task-relevant areas (see also Hayhoe, Bensinger, & Ballard, 1998). Second, research on looking-at-nothing behavior has shown that when retrieving information from memory, eye movement patterns correspond to the patterns seen during the encoding stage, but at a smaller spatial scale, a phenomenon known as shrinkage of gaze patterns (e.g., Johansson, Holsanova, & Holmqvist, 2006). Analysing all AOIs reduces the chance to miss relevant fixations. It should also be noted that previous atom and observation locations were no longer visible when a new observation was presented. In any case, these locations had to be retrieved from memory.

DOI: 10.1177/1747021820922509

before a new observation is obtaines) for abstractly explained observations in comparison to concretely explained observations.

For the first analysis, we compared summed fixation durations to AOIs that were part of the abstract explanation (i.e., squares of rows and columns associated with the abstract explanation; see dotted AOIs in Figure 6), with summed fixation durations for an exact same number of randomly chosen irrelevant AOIs of the grid. A Wilcoxon-signed rank test for data aggregated over absorptions, U-patterns and Z-patterns (i.e., patterns resulting in an abstract explanation) revealed longer fixation durations for locations that were part of the abstract explanation ($W = 363$, $p <.001$, $r = 0.79$, $BF_{10} = 65.83$, for fixation count and strike measures see Appendix 3). This result is in line with the assumption that abstract explanations are part of the current explanation.

--- Insert Figure 6 about here ---

For the second analysis, testing whether participants looked for a longer period of time at observation locations of abstractly as opposed to concretely explained observations, we conducted a repeated measures ANOVA analyzing summed fixation durations for observation locations for the factors explanation type (concrete, abstract) and information type (current, previous, irrelevant)[3]. The Greenhouse-Geisser corrected ANOVA revealed a main effect for the factor explanation type (abstract/concrete atoms; $F(1.00, 26.00) = 4.79$, $p = .04$, $\eta_p^2 = 0.16$, $BF_{10} = 0.24$). That is, in line with our hypothesis, participants spent a longer period of time looking at observation locations of abstractly explained as opposed to concretely explained observations. However, there is also a strong effect of information type ($F(1.01, 26.13) = 38.25$, $p < .004$, $\eta_p^2 = 0.60$, $BF_{10} > 1000$) as well as an interaction between

[3] Due to the large differences in fixation durations between current and previous observations observed in the analyses of Hypothesis 1, we added information type to this analysis as well.

DOI: 10.1177/1747021820922509

information type and explanation type (F(1.03, 26.76) = 5.47, p = .03, $\eta_p^2 = 0.17$, BF_{10} > 1000, see also Figure 7). That is, the difference in fixation durations to observation locations mainly occurs for the current observation location, but not for previous observation locations (Bonferroni pairwise comparison p < .001, d = 1.56). This is still in line with Hypothesis 2, which states that people spend more time on the currently abstractly explained observations than the currently concretely explained observations (Table 1; for fixation counts and strike rates, see Appendix 4).

--- Insert Figure 7 about here ---

In the third analysis, we tested whether viewing times increased when a concrete explanation was not possible (Table 1). However, this proved not to be the case. For Z-patterns and U-patterns, viewing times were longer for concretely explained as opposed to abstractly explained observations. Participants directed their gazes at absorptions at an approximately equal rate. Even though the time for encoding abstractly explained observations rose, as indicated by the memory indexing behavior toward the observation locations (second analysis), the results for viewing times indicate that participants may interrupt observation processing as soon as they reach the conclusion that an observation cannot be concretely explained with the current explanation. We calculated a repeated measures ANOVA for viewing times, which included the observation number as well as explanation type as a factor and collapsed data over the three patterns. Note that observation number only included the second, third, fourth, and fifth observation, as there was no explained absorption, U-pattern, or Z-pattern during the first and no unexplained observation during the last observation possible. The factor explanation type included collapsed data on absorption, U-pattern, and Z-pattern. Furthermore, data was collapsed over trials with four, five, and six observations. The ANOVA revealed two main effects and no interaction.

DOI: 10.1177/1747021820922509

Participants took significantly longer for explained as opposed to unexplained observations, $F(1.00, 18.00) = 6.06$, $p = .02$, $\eta_p^2 = 0.25$, $BF_{10} = 0.19$. The viewing times between observation numbers differed significantly, $F(1.44, 25.84) = 0.74$, $p = .04$, $\eta_p^2 = 0.19$, $BF_{01} = 0.37$, with the fourth observation taking longest. However, Bonferroni comparisons show that there is only a significant difference between the second and third observation ($p = .04$), with the third observation taking less time. All other comparisons are not statistically different (all ps > .05). As it might be possible that viewing times for explained observations are increased because they occurred later in the trial, resulting in a more complex situation model that has to be retrieved, we also looked at the interaction between observation number and explanation type. The ANOVA reveals a null effect for the interaction between observation number and explanation type ($F(1.14, 20.58) = 0.74$, $p = .42$, $\eta_p^2 = 0.04$, $BF_{01} = 6.20$), indicating that participants take longer for explained observations, independent of their presentation time within a trial. In any case, these findings are not in line with Hypothesis 2 or the predictions of TAR. We will return to this point later in the discussion section.

--- Insert Table 1 about here ---

Discussion

This study investigated what information is stored in and retrieved from memory during multicausal sequential abductive reasoning when the causal structure is defined by a set of rules. The theory of abductive reasoning (TAR, Johnson & Krems, 2001) makes it possible to derive detailed hypotheses about what information is stored in and retrieved from memory. Its main assumptions are that the reasoner builds a model describing the current situation and attempts to find a coherent explanation for all observations. If explanations cannot be provided right away, they go through a refinement process. To test these assumptions, we

DOI: 10.1177/1747021820922509

applied a new method—memory indexing—that is based on measuring eye movements to trace the activation of information stored in memory (e.g., Jahn & Braatz, 2014).

Information Stored in the Situation Model

We found that participants gazed back toward all observations and locations where they inferred the position of an atom, which served as the explanations in this study. That is, memory indexing gaze data provided support for the assumption that observations and inferred explanations are stored in a representation in memory that can have the form of a situation model in line with the assumptions of TAR.

Directly comparing the number of eye gazes toward the explanations and observations revealed that participants were more likely to gaze back toward the explanations than toward the observations. Although not directly proposed by TAR, this finding can be explained by the theory. That is, in contrast to observations that do not change, explanations can change during the process of checking a new explanation for consistency with previously set explanations. This process is reflected in overall longer gaze durations toward explanations than observations.

Furthermore, according to TAR, as more observations are presented, an increasing amount of information is stored in the situation model. In line with this prediction, we found that as more information is included in the situation model, gaze behavior is directed toward an increasing number of locations in the BBX grid.

Concretely and Abstractly Explained Observations

TAR assumes that when observations result in an abstract explanation—that is, when they cannot be immediately concretely explained—they must enter a refinement process. The reasoner must then search for intersections between explanations. The explanations can become concrete only if a single intersection is found. If more than one intersection is found, a set of concrete explanations forming an abstract explanation is stored as an explanation in

DOI: 10.1177/1747021820922509

the situation model. Our data supports this assumption only partially. In line with the assumption of a refinement process, participants direct their gazes at the observation locations of abstractly explained observations more extensively than the locations of concretely explained observations. Furthermore, participants spent longer looking toward locations of the abstract explanations than toward locations that were irrelevant for the explanation.

Based on TAR, we assumed that the time needed to respond to an observation is longer for abstractly explained than concretely explained observations. This claim was not supported by our experimental results. Especially in the case of U- and Z-patterns, participants did not even require half the time for abstractly explained observations as they did for concretely explained observations. This result holds even when controlling for the point in time during a trial at which a concrete explanation could be inferred (remember that as more observations are seen, the situation model becomes more complex, possibly increasing the processing load). We would argue that responding to a concretely explained observation resulting from a U- or Z-pattern might take more time because information on several seen observations has to be integrated to place the atoms. In contrast, when noticing that an observation cannot be concretely explained, participants tend to keep the unexplained observation in mind and immediately request new observation information in order to explain it.

Taken together, the results presented here show that participants take longer to encode the observation location but do not engage in abductive reasoning once they have realized that an observation cannot be explained right away. When they realize this, they may engage in a deliberate reasoning strategy that prevents them from further processing steps as described in the TAR theory.

TAR and Current Theories on Diagnostic Reasoning

In summary, our results support the following main assumptions of TAR: 1) A process of

DOI: 10.1177/1747021820922509

comprehension takes place for every new piece of information in which the new observation is interpreted in light of existing knowledge. 2) Consistency checking occurs such that previous explanations in the situation model are retrieved with every new observation. 3) There exist two classes of explanations, and abstractly explained observations must enter a refinement process.

TAR is particularly useful in studying multicausal abductive reasoning problems in which the causal structure is not learned but instead results from applying a set of rules. However, since TAR was first introduced, other theoretical approaches for understanding how abductive reasoning can be explained have been proposed in the literature. In the following, we compare their assumptions with TAR and discuss how our results fit these approaches.

HyGene (Thomas et al., 2008), like TAR, bases the evaluation of existing explanations on new observations. Furthermore, both theories acknowledge the role of memory for storing observations and explanations. In contrast to TAR, HyGene assumes that not all explanations are represented as the number of explanations reaches a certain limit. Our data show that participants are able to maintain observations and explanations of up to six observations in memory. The maximum number of explanations that can be represented in memory is a topic for future research. Furthermore, both theories assume that explanations are chosen from the set of all those available. In TAR, this occurs when an observation cannot be explained immediately, which results in an abstract explanation. According to HyGene, the reasoner constructs a number of explanations based on experience (retrieval from long-term memory) and transfers the ones with the strongest relation to semantic long-term memory into working memory. According to HyGene, which explanations are transferred seems to depend more on individual experience than on the explanations' explanatory value. It may be interesting to study whether HyGene can be applied to the situation in which an explanation does not have to be retrieved from long-term memory but can be inferred from learned rules. Although HyGene allows a number of explanations to be represented in working memory, it results in

DOI: 10.1177/1747021820922509

only one explanation for observed information. As stated earlier, the reasoning process in TAR can also result in a combination of more than one explanation. HyGene is a valuable model for describing explanation generation and evaluation. However, in order to answer research questions concerning multicausal reasoning, it must be extended to include the possibility of multiple explanations.

The consistency check postulated by TAR is closely related to the main assumption of the construction-integration approach (Kintsch, 1998), which states that only knowledge compatible with the current representation remains activated. Incompatible knowledge is excluded from the current network (Baumann et al., 2007). However, the construction-integration approach goes beyond the assumptions of TAR by attempting to differentiate between inhibition and decay. Whereas Baumann et al. (2007) assume that explanations that are not relevant for the current task can decay, Mehlhorn, et al. (2011) state that observations inhibit irrelevant explanations in memory. TAR makes no asumption on interference mechanisms, and only assumes that abstract explanations with no overlap to existing knowledge do not enter the situation model. The question of decay and inhibitition of irrelevant explanations is a topic for future research. Furthermore, whereas the construction-integration approach assumes consistency-checking in favor of the leading explanation and, therefore, only one explanation for an observation pattern (Kintsch & Dijk, 1978), again, one of TAR's greatest advantages is that it allows a combination of explanations (atoms).

In summary, TAR appears to be the most specialized theory for the case of abductive reasoning, which was the focus of this study (multicausality, causal structure definded by rules in comparison to learned associations). By applying a new process-tracing mehod—memory indexing—we demonstrated that the main assumptions of TAR hold true, thereby showing that this theory is still worth investigating as a competitor to other process theories on abductive reasoning. However, future research should examine whether other theories may also be able to explain behavior in the case of multicausal abductive reasoning.

DOI: 10.1177/1747021820922509

Tracing memory processes

Examining where individuals direct their gaze during thinking and reasoning has been shown to provide deep insights into the underlying cognitive processes (Jahn & Braatz, 2014; Renkewitz & Jahn, 2012; Rosner & Helversen, 2019; Scholz, Krems, & Jahn, 2017; Scholz, von Helversen, & Rieskamp, 2015). In this study, we combined the memory indexing method with a visuospatial reasoning task consisting of 100 small grids on a computer screen. Even with this task, memory indexing provided useful insights into retrieval processes that take place during abductive reasoning. It allowed us to trace which information is retrieved from memory during abductive reasoning and was sensitive to differences in the status of explanation (i.e., concretely versus abstractly explained observations).

Our research raises the interesting question of whether participants actually retrieve explanation locations or simply infer the retrieved observation locations. In this study, observation locations had to be retrieved as they disappeared when new information was presented. TAR merely states that new information is evaluated in light of existing knowledge; it would therefore be interesting to further investigate when the existing knowledge is retrieved from memory. Analyzing the time course of gazes at previous observation and explanation locations might provide insights into the retrieval dynamics of abductive reasoning.

In this study, we observed large differences in fixation durations between currently presented information and previous observations and explanations no longer visible on the screen. This may be the case, because different processes are reflected in participants' eye movement behavior. Eye movements toward currently visible information may reflect encoding and comprehension of the visually presented pieces of information, spatial reasoning, and the motor response. Eye movements toward previous information locations demonstrate the retrieval of observations and explanations. So far, the processes that take place simultaneously while looking at currently visible information cannot be disentangled.

DOI: 10.1177/1747021820922509

One key finding of this study is the fact that participants look longer at previous (but currently invisible) locations containing explanations than at irrelevant areas. This demonstrates memory retrieval in order to comprehend new observations as well as consistency checking with current explanations forming the situation model. However, this might also be due to a sample of irrelevant AOI that does not mirror the distribution of atom and observation AOI. We admit that both distributions differ, as it is not possible to entirely match every occupied AOI with an irrelevant AOI that unites the same spatial features (e.g., distance to the center of the BBX). Still, we are confident the central fixation bias (e.g., Tseng, Carmi, Cameron, Munoz, & Itti, 2009) plays only a minor role in our data. The central fixation bias describes the fact that peoples eye movements are biased towards the center when viewing a natural scene (Tseng et al., 2009). According to Tseng et al. (2009) this is mainly driven by (1) the experience that photographer usually arrange interesting objects in the center of their work (photographer bias) and (2) by viewing strategies as people learn that information sampling starting in the middle of a scene is a most effective use of attentional resources. As the photographer bias is mostly supported in natural scenes, we do not assume that it can be generalized to our artificial material. Second, participants learn very early in the experiment that rays of light enter the BBX from the border. That is, even in the practice trials, participants are aware of the fact that observations have to be gathered at the edges of the BBX making viewing strategies that are biased towards the center unlikely.

Further, it would be interesting to disentangle comprehension and consistency checking in eye movements to previous information locations to study the influence of memory on the reasoning process in more detail. Research already investigated the extent to which different eye movement measures (e.g., number of fixations, fixation durations) can be used to differentiate between encoding and processing (Glaholt & Reingold, 2011; Horstmann et al., 2009; Klichowicz, Scholz, Strehlau, & Krems, 2016), however, with mixed results. Another approach would be to better understand the interaction of both top-down-driven

DOI: 10.1177/1747021820922509

(memory retrieval) and bottom-up-driven factors (information visible on screen) on memory indexing gaze behavior during abductive reasoning. Here, one could manipulate the amount of information that has to be retrieved from memory and test how this affects memory indexing and reasoning behavior.

Conclusion

Studying eye movements based on the memory indexing method (e.g., Renkewitz & Jahn, 2012; Jahn & Braatz, 2014) can be used to explore memory retrieval during multicausal abductive reasoning. We found evidence of retrieval processes during multicausal abductive reasoning that were in line with the assumption of the theory of abductive reasoning (Johnson & Krems, 2001) As all observation and explanation information are kept in and retrieved from working memory in order to engage in information comprehension (that is, integration into a mental representation), consistency checking, and refinement. Our data even expands assumption made by TAR as we found that explanation information receive much more attention than already explained observations. We conclude that this theory makes a useful contribution to explaining the abductive reasoning processes.

Supplementary Material

The Supplementary Material is available at: qjep.sagepub.com

Acknowledgments

Parts of the data were presented at the 38th annual meeting of the Cognitive Science Society (August 2016, Philadelphia, PA). Agnes Rosner gratefully acknowledges the support of the Swiss National Science Foundation (grant 157432). Furthermore, the authors would like to thank Daniela Eileen Lippoldt for her help in preparing the data and Paul Walther for his help

DOI: 10.1177/1747021820922509

programming the experiment and two anonymous reviewers for their valuable comments on a previous version of this paper.

DOI: 10.1177/1747021820922509

References

Baumann, M. R. K., Mehlhorn, K., & Bocklisch, F. (2007). The activation of hypotheses during abductive reasoning. *Proceedings of the 29th Annual Cognitive Science Society*, 803–808. Retrieved from http://csjarchive.cogsci.rpi.edu/proceedings/2007/docs/p803.pdf

Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends in Cognitive Sciences*, *12*(11), 405–410. https://doi.org/10.1016/j.tics.2008.07.007

Glaholt, M. G., & Reingold, E. M. (2011). Eye movement monitoring as a process tracing methodology in decision making research. *Journal of Neuroscience, Psychology, and Economics*, *4*(2), 125–146. https://doi.org/10.1037/a0020692

Hayhoe, M. M., Bensinger, D. G., & Ballard, D. H. (1998). Task constraints in visual working memory. *Vision Research*, *38*(1), 125–137. https://doi.org/10.1016/S0042-6989(97)00116-8

Hedge, C., & Leonards, U. (2013). Using eye movements to explore switch costs in working memory. *Journal of Vision*, *13*(4), 18–18. https://doi.org/10.1167/13.4.18

Horstmann, N., Ahlgrimm, A., & Glöckner, A. (2009). How distinct are intuition and deliberation ? An eye-tracking analysis of instruction-induced decision modes. *Judgment and Decision Making*, *4*(5), 335–354. https://doi.org/10.2139/ssrn.1393729

Jahn, G., & Braatz, J. (2014). Memory indexing of sequential symptom processing in diagnostic reasoning. *Cognitive Psychology*, *68*, 59–97. https://doi.org/10.1016/j.cogpsych.2013.11.002

Johansson, R., Holsanova, J., & Holmqvist, K. (2006). Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and in complete darkness. *Cognitive Science*, *30*(6), 1053–1079. https://doi.org/10.1207/s15516709cog0000

Johnson-Laird, P. N. (1980). Mental Models in Cognitive Science. *Cognitive Science*, *4*(1), 71–115. https://doi.org/10.1016/S0364-0213(81)80005-5

Johnson-Laird, P. N., Byrne, R. M., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review*, *99*(3), 418–439. https://doi.org/10.1037/0033-295X.99.3.418

Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015). Logic, probability, and

DOI: 10.1177/1747021820922509

human reasoning. *Trends in Cognitive Sciences*, *19*(4), 201–214. https://doi.org/10.1016/j.tics.2015.02.006

Johnson, T. R., & Krems, J. F. (2001). Use of current explanations of multicausal abductive reasoning. *Cognitive Science*, *25*, 903–939. https://doi.org/10.1016/S0364-0213(01)00059-3

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* New York, NY: Cambridge University Press.

Kintsch, W., & Dijk, T. A. Van. (1978). Psychological Review. *Psychological Review*, *85*(5), 363–394.

Klahr, D., & Dunbar, K. (1988). Dual Space Search During Scientific Reasoning. *Cognitive Science*, *12*, 1–48.

Klichowicz, A., Scholz, A., Strehlau, S., & Krems, J. F. (2016). Differentiating between Encoding and Processing during Diagnostic Reasoning: An Eye tracking study. *38th Annual Conference of the Cognitive Science Society*, 129–134.

Laeng, B., & Teodorescu, D. S. (2002). Eye scanpaths during visual imagery reenact those of perception of the same visual scene. *Cognitive Science*, *26*(2), 207–231. https://doi.org/10.1207/s15516709cog2602

Lange, N. D., Thomas, R. P., & Davelaar, E. J. (2012). Temporal dynamics of hypothesis generation: The influences of data serial order, data consistency, and elicitation timing. *Frontiers in Psychology*, *3*(JUN), 1–16. https://doi.org/10.3389/fpsyg.2012.00215

Meder, B., & Mayrhofer, R. (2017). Diagnostic causal reasoning with verbal information. *Cognitive Psychology*, *96*, 54–84. https://doi.org/10.1016/j.cogpsych.2017.05.002

Mehlhorn, K., Taatgen, N. A., Lebiere, C., & Krems, J. F. (2011). Memory activation and the availability of explanations in sequential diagnostic reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1391–1411. https://doi.org/10.1037/a0023920

Peng, Y., & Reggia, J. A. (1990). *Abductive inference models for diagnostic problem solving.* New York: Springer-Verlag.

Pylyshyn, Z. (1998). *The Role of Visual Indexes in Spatial Vision and Imagery ∗ Backgound*. 1–10.

Rebitschek, F. G., Krems, J. F., & Jahn, G. (2016). The diversity effect in diagnostic reasoning. *Memory & Cognition*, *44*(5), 789–805. https://doi.org/10.3758/s13421-016-0592-0

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. https://doi.org/10.1037/a0028073

Richardson, D. C., Altmann, G. T. M., Spivey, M. J., & Hoover, M. a. (2009). Much ado about eye movements to nothing: a response to Ferreira et al.: taking a new look at looking at nothing. *Trends in Cognitive Sciences*, *13*(6), 235–236. https://doi.org/10.1016/j.tics.2009.02.006

Richardson, D. C., & Kirkham, N. Z. (2004). Multimodal events and moving locations: eye movements of adults and 6-month-olds reveal dynamic spatial indexing. *Journal of Experimental Psychology. General*, *133*(1), 46–62. https://doi.org/10.1037/0096-3445.133.1.46

Rosner, A., & Helversen, B. Von. (2019). Memory shapes judgments : Tracing how memory biases judgments by inducing the retrieval of exemplars. *Cognition*, *190*(November 2018), 165–169. https://doi.org/10.1016/j.cognition.2019.05.004

Scholz, A., Klichowicz, A., & Krems, J. F. (2017). Covert shifts of attention can account for the functional role of "eye movements to nothing." *Memory and Cognition*, 1–14. https://doi.org/10.3758/s13421-017-0760-x

Scholz, A., Krems, J. F., & Jahn, G. (2017). Watching diagnoses develop: Eye movements reveal symptom processing during diagnostic reasoning. *Psychonomic Bulletin & Review*, 1–15. https://doi.org/10.3758/s13423-017-1294-8

Scholz, A., von Helversen, B., & Rieskamp, J. (2015). Eye movements reveal memory processes during similarity- and rule-based decision making. *Cognition*, *136*, 228–246. https://doi.org/10.1016/j.cognition.2014.11.019

Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: eye movements to absent objects. *Psychological Research*, *65*(4), 235–241.

Thomas, R. P., Dougherty, M. R., Sprenger, A. M., & Harbison, J. I. (2008). Diagnostic hypothesis generation and human judgment. *Psychological Review*, *115*(1), 155–185. https://doi.org/10.1037/0033-295X.115.1.155

Tseng, P. H., Carmi, R., Cameron, I. G. M., Munoz, D. P., & Itti, L. (2009). Quantifying

DOI: 10.1177/1747021820922509

center bias of observers in free viewing of dynamic natural scenes. *Journal of Vision*, *9*(7), 1–16. https://doi.org/10.1167/9.7.4

DOI: 10.1177/1747021820922509

Figure Captions

Figure 1. Rules of the black box task (BBX). A light ray entering the BBX can take the following paths, depending on where the ray hits the hidden atom: 1 = straight through, 2 = L-Pattern, 3 = absorption, 4 = U-pattern, 5 = Z-pattern.

Figure 2. (A) Example of the trial of the BBX in which the third observation is concretely explained by the atom set to explain the first (i.e., the dotted circle in the third frame). (B) Exemplary trial of the BBX with an abstractly explained observation at the beginning, which can only be concretely explained once the third observation is presented. Prior to observing the third observation, an abstract explanation is found for the first, illustrated by the dotted circles.

Figure 3. AOI definitions used to analyze the data. (vertical stripes) = atom ; (horizontal stripes) = observation location; (dots) = possible atom locations of abstractly explained observations (middle) or AOIs between observation locations and the atom for patterns that could have potentially resulted in an abstract explanation but that were already explained by a previous observation (right).

Figure 4. Mean summed fixation durations to the atom locations (grey) and the observation locations (black). Fixation durations toward the current atom and observation locations are compared to previous atom and observation locations and to irrelevant AOIs. Error bars represent one standard error of the mean.

Figure 5. Heat maps of one trial with four observations for all 28 participants. The picture on the left displays the first atom, the picture on the right, the last. Participants only saw the current atom. Areas on the screen that received more attention are highlighted in warmer

DOI: 10.1177/1747021820922509

colors (i.e., in red). All previous atoms (illustrated with dotted lines) were invisible and only added in this figure to demonstrate participants' gaze behavior to absent but previous atoms that had to be retrieved to make a correct inference.

Figure 6. Heat maps with gaze data from all 28 participants for an abstractly explained absorption (left), an abstractly explained U-pattern (middle), and an abstractly explained Z-pattern (right), indicating that participants show eye movements within the rows and columns in which atoms may be located. Note that atoms were defined as the AOI containing the center of the atom and the AOI where the ray hit the field of influence of the atom. As can be seen, for U-patterns and Z-patterns, participants gazed longer at the AOI where the ray hits the field of influence (indicated by warmer colors). Possible atom locations are visualized by dotted atoms.

Figure 7. Mean summed fixation times toward the observation locations for concretely explained (grey) and abstractly explained (black) observations for current and previous observations and irrelevant AOIs. Error bars represent one standard error of the mean.

DOI: 10.1177/1747021820922509

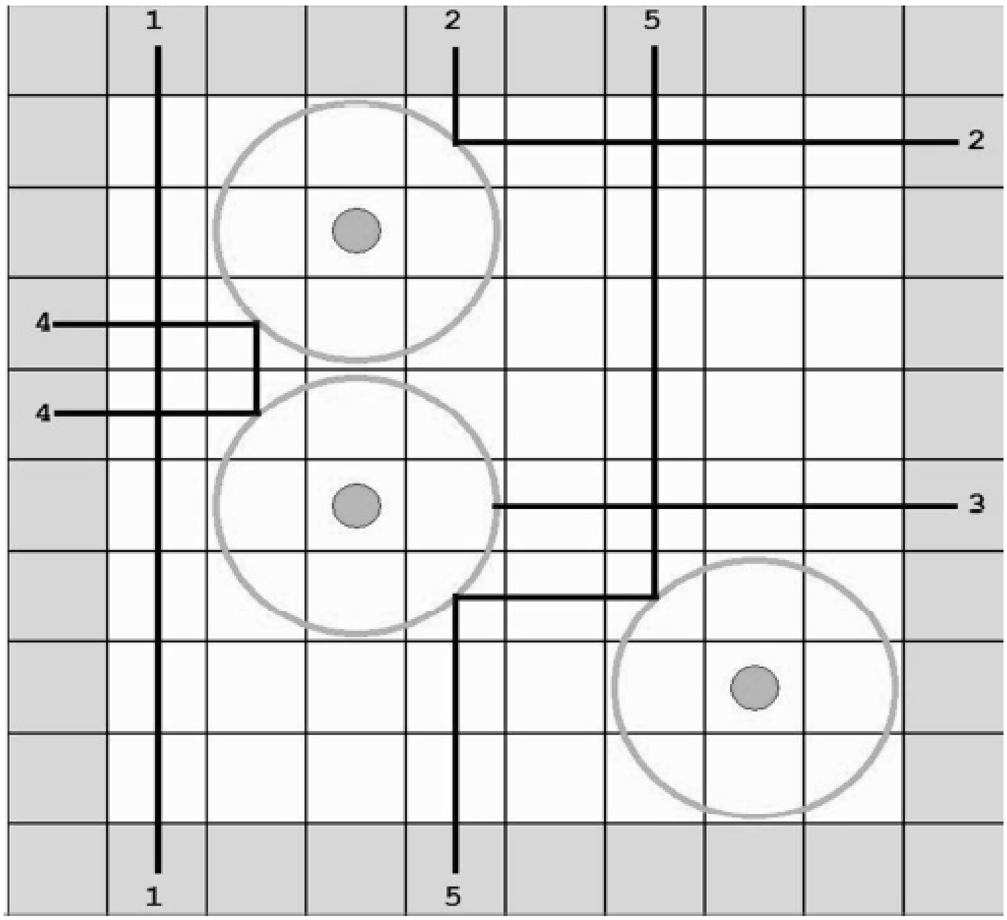
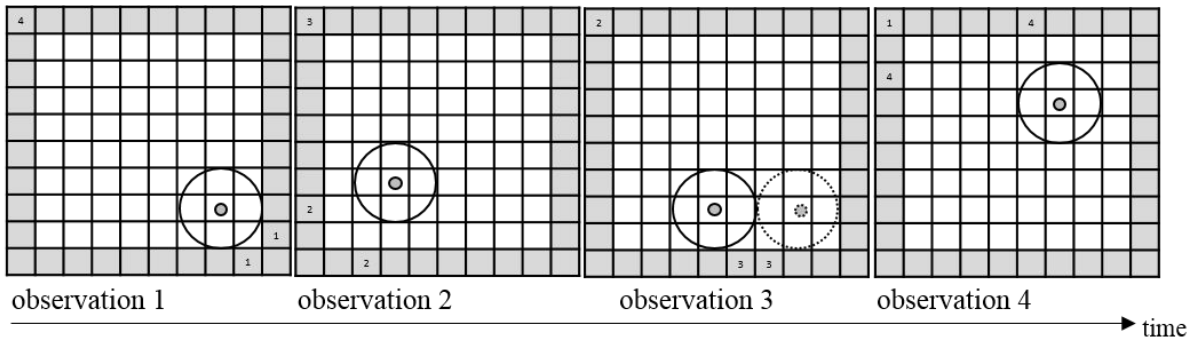


Figure 1. Rules of the black box task (BBX). A light ray entering the BBX can take the following paths, depending on where the ray hits the hidden atom: 1 = straight through, 2 = L-Pattern, 3 = absorption, 4 = U-pattern, 5 = Z-pattern.

DOI: 10.1177/1747021820922509

A



B

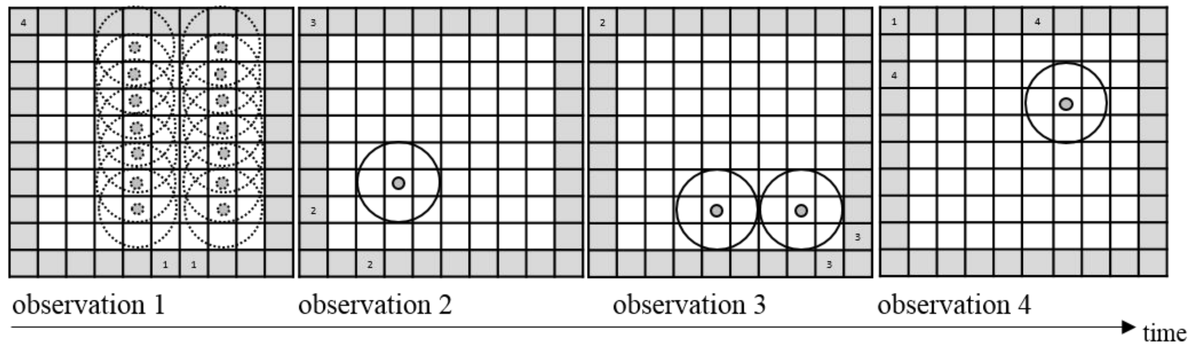


Figure 2. (A) Example of the trial of the BBX in which the third observation is concretely explained by the atom set to explain the first (i.e., the dotted circle in the third frame). (B) Exemplary trial of the BBX with an abstractly explained observation at the beginning, which can only be concretely explained once the third observation is presented. Prior to observing the third observation, an abstract explanation is found for the first, illustrated by the dotted circles.

DOI: 10.1177/1747021820922509

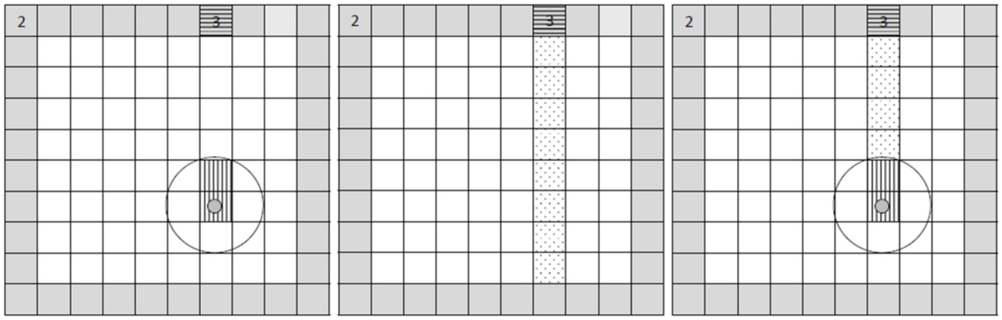


Figure 3. AOI definitions used to analyze the data. Vertical stripes = atom ; Horizontal stripes = observation location; Dots = possible atom locations of abstractly explained observations (middle) or AOIs between observation locations and the atom for patterns that could have potentially resulted in an abstract explanation but that were already explained by a previous observation (right).

DOI: 10.1177/1747021820922509

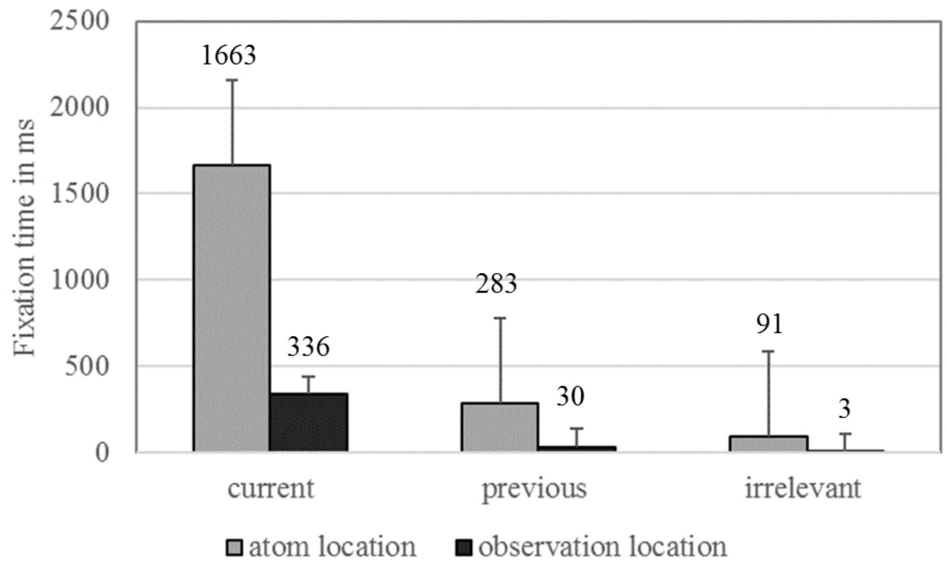


Figure 4. Mean summed fixation durations to the atom locations (grey) and the observation locations (black). Fixation durations toward the current atom and observation locations are compared to previous atom and observation locations and to irrelevant AOIs. Error bars represent one standard error of the mean.

DOI: 10.1177/1747021820922509

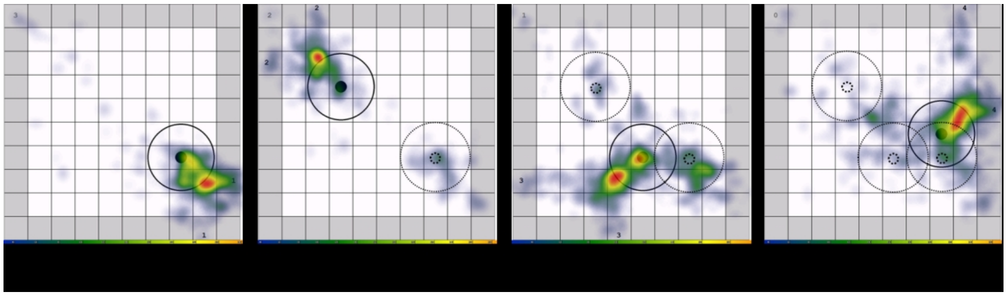


Figure 5. Heat maps of one trial with four observations for all 28 participants. The picture on the left displays the first atom, the picture on the right, the last. Participants only saw the current atom. Areas on the screen that received more attention are highlighted in warmer colors (i.e., in red). All previous atoms (illustrated with dotted lines) were invisible and only added in this figure to demonstrate participants' gaze behavior to absent but previous atoms that had to be retrieved to make a correct inference.

DOI: 10.1177/1747021820922509

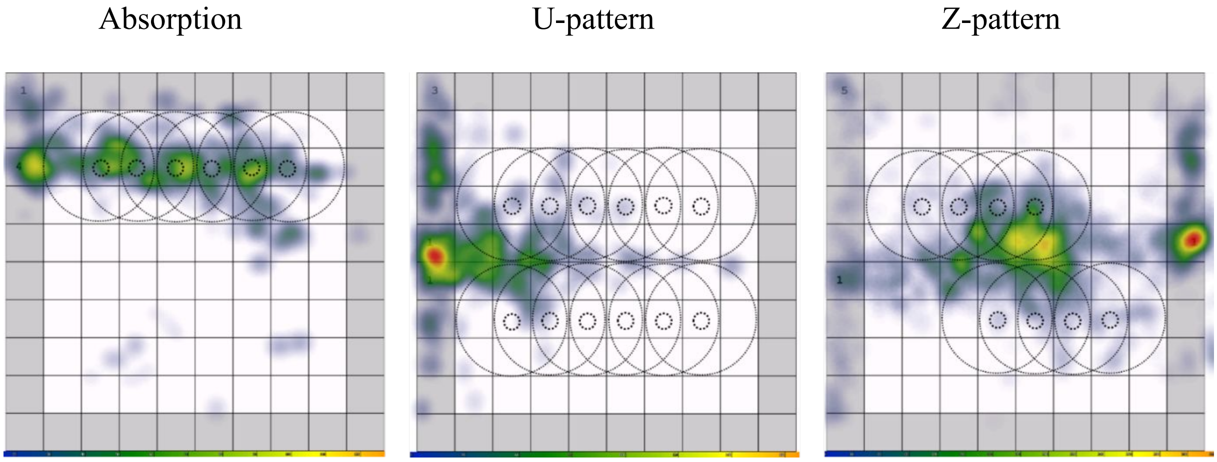


Figure 6. Heat maps with gaze data from all 28 participants for an abstractly explained absorption (left), an abstractly explained U-pattern (middle), and an abstractly explained Z-pattern (right), indicating that participants show eye movements within the rows and columns in which atoms may be located. Note that atoms were defined as the AOI containing the center of the atom and the AOI where the ray hit the field of influence of the atom. As can be seen, for U-patterns and Z-patterns, participants gazed longer at the AOI where the ray hits the field of influence (indicated by warmer colors). Possible atom locations are visualized by dotted atoms.

DOI: 10.1177/1747021820922509

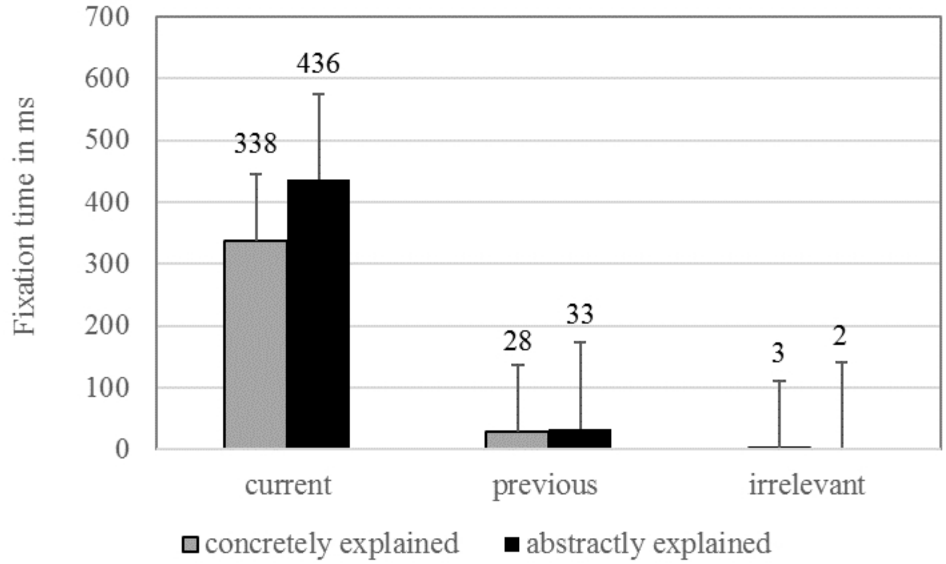


Figure 7. Mean summed fixation times toward the observation locations for concretely explained (grey) and abstractly explained (black) observations for current and previous observations and irrelevant AOIs. Error bars represent one standard error of the mean.

DOI: 10.1177/1747021820922509

Table 1. Means (*M*) and standard deviations (*SD*) of fixation time in s and viewing time in s toward the observation locations of the light rays for concretely explained and abstractly explained observations.

	Fixation time		Viewing time	
	concrete	abstract	concrete	abstract
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*
Absorption	0.30 (0.27)	0.35 (0.30)	4.45 (1.81)	4.62 (1.88)
U-pattern	0.59 (0.58)	0.83 (0.51)	12.26 (5.27)	5.17 (2.37)
Z-pattern	0.47 (0.55)	1.59 (3.34)	104.23 (162.83)	16.36 (12.67)

DOI: 10.1177/1747021820922509