

Watching diagnoses develop: Eye movements reveal symptom processing during diagnostic reasoning

Agnes Scholz¹ · Josef F. Krems² · Georg Jahn²

© Psychonomic Society, Inc. 2017

Abstract Finding a probable explanation for observed symptoms is a highly complex task that draws on information retrieval from memory. Recent research suggests that observed symptoms are interpreted in a way that maximizes coherence for a single likely explanation. This becomes particularly clear if symptom sequences support more than one explanation. However, there are no existing process data available that allow coherence maximization to be traced in ambiguous diagnostic situations, where critical information has to be retrieved from memory. In this experiment, we applied memory indexing, an eye-tracking method that affords rich time-course information concerning memory-based cognitive processing during higher order thinking, to reveal symptom processing and the preferred interpretation of symptom sequences. Participants first learned information about causes and symptoms presented in spatial frames. Gaze allocation to emptied spatial frames during symptom processing and during the diagnostic response reflected the subjective status of hypotheses held in memory and the preferred interpretation of ambiguous symptoms. Memory indexing traced how the diagnostic decision developed and revealed instances of hypothesis change and biases in symptom processing. Memory indexing thus

provided direct online evidence for coherence maximization in processing ambiguous information.

Keywords Eye movements · Process tracing · Memory indexing · Diagnostic reasoning · Coherence maximization

Diagnostic reasoning involves finding a probable explanation for a set of observations (Johnson & Krems, 2001; Meder, Mayrhofer, & Waldmann, 2014; Patel, Arocha, & Zhang, 2005). A physician, for example, is required to find the most likely cause for a patient's symptoms. Usually, symptoms are reported sequentially and have to be evaluated based on knowledge stored in long-term memory (Mehlhorn, Taatgen, Lebiere, & Krems, 2011; Thomas, Dougherty, Sprenger, & Harbison, 2008). Symptom information can be sufficient to determine a single explanation, but often the available information supports more than one hypothesis (McKenzie, 1998) and is thus ambiguous (Holyoak & Simon, 1999). An ambiguous case elicits differing final diagnoses from different diagnosticians. Each single diagnostician may adhere to an initial hypothesis or adopt an alternative. In this study, we applied eye tracking to investigate memory processes (memory indexing) during diagnostic reasoning to reveal coherence maximizing in symptom processing.

Previous research has shown that symptom processing in memory is biased toward the hypothesis supported by symptoms presented early in the sequence (Baumann, Krems, & Ritter, 2010; Bussemeyer & Townsend, 1993; Lange, Thomas, & Davelaar, 2012; Rebitschek, Bocklisch, Scholz, Krems, & Jahn, 2015; Rebitschek, Scholz, Bocklisch, Krems, & Jahn, 2012; Weber, Böckenholt, Hilton, & Wallace, 1993), especially if the response is given after all the symptom information has been received (end-of-sequence response mode; Hogarth & Einhorn, 1992). These findings on

Electronic supplementary material The online version of this article (doi:10.3758/s13423-017-1294-8) contains supplementary material, which is available to authorized users.

✉ Agnes Scholz
agnespscholz@gmail.com

¹ Department of Psychology, University of Zurich, Zurich, Switzerland

² Department of Psychology, Technische Universität Chemnitz, Chemnitz, Germany

the so-called diagnosis momentum (Croskerry, 2003) constitute instances of confirmation bias (Nickerson, 1998) and can be interpreted as a reasoner's tendency to strive for a coherent interpretation (Glöckner, Betsch, & Schindler, 2010; Holyoak & Simon, 1999; Mehlhorn & Jahn, 2009; Kostopoulou, Russo, Keenan, Delaney, & Douiri, 2012; Wang, Johnson, & Zhang, 2006). The coherence effect is closely related to research on information distortion (DeKay, Stone, & Sorenson, 2011; Hagmayer & Kostopoulou, 2013; Russo, Medvec, & Meloy, 1996; Strickland & Keil, 2011). Incoherent representations are transformed into coherent representations through information distortion to maximize coherence. Coherence can also be achieved by biased information processing maximizing the belief in one hypothesis while decreasing the belief in alternatives. Maximizing coherence often favors the initially leading hypothesis, yet it can strengthen an alternative hypothesis if stronger evidence for this alternative has accumulated and a hypothesis change takes place.

Coherence maximization has been studied by analyzing the outcome of the reasoning process. For instance, symptom sequences with equal support for multiple hypotheses can provide evidence for coherence maximizing in unequal proportions of diagnoses. Thus, the probability that a certain disease has caused a patient's symptoms given equal support for this disease and an alternative (and equal base rates) is .5 (maximally ambiguous). Deviations of diagnosis proportions from .5 indicate biased symptom processing to increase coherence in a diagnostic decision. In previous studies with maximally ambiguous sequences, the initial hypothesis was chosen as the final diagnosis with a proportion higher than .5 (Rebitschek, Bocklisch, et al., 2015).

Coherence maximization can be described by parallel constraint-satisfaction models (Glöckner & Betsch, 2008; McClelland & Rumelhart 1981; Read, Vanman, & Miller, 1997; Simon, Snow, & Read, 2004; Simon, Stenstrom, & Read, 2015; Thagard, 1989). Theories of coherence maximization are grounded in cognitive consistency theories. At the heart of cognitive consistency lies the Gestaltian principle that human cognition is affected by mutual interaction among constituent elements of a cognitive representation. In parallel constraint satisfaction models, the reasoning task is represented by a network, in which symptoms and diagnoses are interconnected by excitatory and inhibitory links representing positive and negative relations between symptoms and diagnoses. Bidirectional activation and inhibition settles the network in a stable and thus coherent state favoring either one or the other diagnosis.

Despite their merits, it is difficult to use these models to clarify the underlying cognitive processes (Amaya, 2015; Mehlhorn & Jahn, 2009; Rumelhart, Smolensky, McClelland, & Hinton, 1986) that lead to the observed biases in symptom processing. However, this clarification is

necessary to enable understanding of how coherence maximizing lends weight to one of two competing hypotheses, and to clarify how the coherence maximizing process can result in the selection of a less supported diagnosis. One important means of clarifying the cognitive processes is the collection of process data to inform and enhance process models of coherence-based diagnostic reasoning.

Process tracing methods, such as verbal protocols, information boards, or Mouselab allow the study of information processing prior to and during the response (for overviews, see Glaholt & Reingold, 2011; Schulte-Mecklenbeck, Kühberger, & Ranyard, 2011). However, memory-based reasoning processes usually cannot be observed because most of the time cognition proceeds without systematic accompanying overt actions. Recent research on the looking-at-nothing phenomenon and the visual-world paradigm has shown that *eye movements* are applicable to the study of real-time retrieval processes (e.g., Hoover & Richardson, 2008; Johansson, Holsanova, Dewhurst, & Holmqvist, 2012; Johansson, Holsanova, & Holmqvist, 2006; Martarelli, Mast, & Hartmann, 2017; Richardson & Kirkham, 2004; Richardson & Spivey, 2000; Spivey & Geng, 2001) and language processing (Alloppenna, Magnuson, & Tanenhaus, 1998; Altmann, 2004; Altmann & Kamide, 2007, 2009; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Extending these results, *memory indexing* has been developed as a process measure to study higher level cognitive tasks (Renkewitz & Jahn, 2010, 2012) and has been successfully applied to study reasoning and decision making (Jahn & Braatz, 2014; Platzer, Bröder, & Heck, 2014; Scholz, von Helversen, & Rieskamp, 2015). Inferring memory-based processing by observing eye movements is possible because reactivating information that is linked to a location reestablishes a spatial index that leads the gaze to the relevant location (Huettig, Olivers, & Hartsuiker, 2011; Johansson & Johansson, 2014; Scholz, Mehlhorn, & Krems, 2016; Spivey & Dale, 2011).

Jahn and Braatz (2014) applied memory indexing to study sequential diagnostic reasoning. Participants were told to imagine they were physicians trying to identify the chemical with which a worker in a chemical plant had been affected during an accident (chemical accident task; Mehlhorn et al., 2011). Information concerning the symptoms and the chemicals that could potentially elicit such symptoms were learned during a preceding learning phase. Symptom classes and the chemicals (possible diagnoses) were associated to spatial locations on a computer screen. During reasoning trials, the spatial locations that previously contained information during the learning phase were empty, and symptoms were presented auditorily in sequence. Eye movements were recorded during reasoning trials. Gaze allocation to emptied screen locations revealed the changing activation status of hypotheses over the course of a reasoning trial and indicated how symptoms were interpreted. For example, in trials with

early symptoms supporting a hypothesis that had to be changed to arrive at the correct diagnosis, fixation proportions were highest for the initial hypothesis first and highest for the correct hypothesis later. In the study by Jahn and Braatz (2014), most symptom sequences had a single correct diagnosis.

Present study

In the present study, we focused on exploring memory processes during sequential diagnostic reasoning with ambiguous symptom sequences to extend previous findings concerning eye movements during decision making and diagnostic reasoning and to test process assumptions about coherence maximization. In everyday life, people are regularly faced with complex, ambiguous situations that nonetheless call for a decision (e.g., Holyoak & Simon, 1999). Studying ambiguity allows one to specify how conflicting information is integrated and therefore presents a strong case of testing process assumptions about coherence maximization. Ambiguity results when two or more hypotheses are supported by the symptom sequence, and there is no single correct diagnosis at the end after all symptom information has been presented. We use two hypothetical examples to illustrate this next.

First, consider the symptom sequence *a-ab-ab-b*. In this sequence, *a* denotes a symptom that is caused by a Chemical A but by none of the other chemicals in question and thus strongly supports Chemical A as a candidate diagnosis; *ab* denotes a symptom that is caused by Chemical A and by Chemical B and thus supports two Chemicals A and B; and *b* supports only Chemical B. In the sequence *a-ab-ab-b*, two hypotheses are supported equally by the set of symptoms. A second example of an ambiguous symptom set would be an *a-bd-a-ab* sequence. This sequence contains support for the Hypotheses A, B, and D, but with a clear ordering: Hypothesis A is supported by three symptoms (two of which are not caused by any other chemical), Hypothesis B is supported by two symptoms, and Hypothesis D is supported by a single symptom only. Note that, like in the example before, the first symptom elicits A as the leading hypothesis. The second symptom, however, does not support A and suggests B or D instead. Thus, B and D may be added to the set of considered hypotheses and could become strengthened by coherence maximizing in processing later symptoms, such that the final diagnosis could be B although the sequence provides superior support for A.

We tested ambiguous symptom sequences of this kind to explore coherence maximization during diagnostic reasoning by applying the memory indexing method. Recent research has shown that biased information processing and information distortion can increase or decrease the belief in a hypothesis and explain diagnostic preferences beyond mere retrieval

processes. In the framework of parallel constraint satisfaction models, these processes are implemented by bidirectional associations between symptoms and hypotheses that settle a network toward a coherent explanation of given information. Following this line of research, we aimed to demonstrate that eye movements could trace the changing activations resulting from the mutual interactions between symptom information and diagnoses held in memory.

In a first set of analyses, we clarified the broader relation between eye movements and the outcome of the reasoning process; that is, the diagnostic response (Hypothesis 1). Based on the literature review, we wanted to replicate previous findings on the relation between eye movements and complex thinking processes (Hypotheses 2 and 5). This first set of analyses aimed to further strengthen our methodological approach and can be seen as testing preconditions for our second set of analyses. In the second set, we tested more specific hypotheses on the effects of coherence maximization during diagnostic reasoning (integrated probability matching, Hypothesis 3; hypothesis changes, Hypothesis 4). Next, we outline all hypotheses in more detail.

Hypothesis 1: Gaze behavior and diagnostic response

We assume that if gaze data indeed reflect memory retrieval in ambiguous diagnostic situations, eye gaze should correspond to the outcome of the reasoning process. In decision making, it has been shown that the preferred option was gazed at longer (Stewart, Hermens, & Matthews, 2015; for an overview, see Orquin & Mueller Loose, 2013). Additionally, recent research on diagnostic reasoning has shown that eye movements can reflect symptom integration in memory and that eye movements can indicate the diagnostic response (Jahn & Braatz, 2014). Based on these findings, we assume that gaze duration to an alternative during diagnostic reasoning and processing of an ambiguous symptom sequence should predict how likely this alternative is to be chosen at the end of the reasoning process.

Hypothesis 2: Location matching

When the first symptom establishes a single leading hypothesis, gaze data following the presentation of this symptom should reflect which hypothesis it supports and thus the correct retrieval of the symptom location from memory. Consequently, if only one hypothesis is supported by the first symptom, this hypothesis should be gazed at longer than any other hypothesis. Such momentary probability matching has only once been shown in diagnostic reasoning (Jahn & Braatz, 2014). It would be in line with previous findings on the

looking-at-nothing phenomenon (e.g., Richardson & Spivey, 2000).

Hypothesis 3: Integrated probability matching

If eye movements can trace coherence maximization during sequential diagnostic reasoning, eye movements during later symptom presentations should reveal the integration of symptom information (see Renkewitz & Jahn, 2012; Jahn & Braatz, 2014; Scholz et al., 2015). For instance, if a later symptom supports two alternatives, gaze duration should be longer toward the leading hypothesis. Alternatively, if gaze behavior merely reflects retrieval processes without revealing symptom integration, when being presented with a symptom that is equally strongly associated with two hypotheses, participants should look at both diagnoses for about the same duration. Recent findings on gaze allocation during diagnostic reasoning suggest that eye movements reflect integrated probability matching and thus reasoning instead of just memory retrieval. However, there has been no statistical test of this hypothesis.

Hypothesis 4: Hypothesis change

When being presented with an ambiguous symptom sequence, a person's belief can change from the leading to an alternative hypothesis, when enough evidence for an alternative hypothesis has accumulated. Coherence maximization can affect this symptom integration process. For instance, coherence maximization can lead to participants not changing their belief by distorting information supporting an alternative hypothesis, leading them to respond with the initial hypothesis. Assuming that memory-indexing gaze data reveal a participant's currently-preferred hypothesis, if the proportion of fixations toward a hypothesis stays about the same throughout the symptom sequence, this would indicate that a hypothesis change is unlikely to have occurred. By contrast, if there is a change in the proportion of fixations to the leading hypothesis over the symptom sequence, this would likely suggest the occurrence of a hypothesis change. Thus, if eye movements reflect processes of coherence maximization, differences in fixation proportions between the beginning and the end of a symptom sequence should predict the hypothesis change.

Coherence maximization can also affect information processing after a hypothesis change has taken place. Biased information processing can strengthen the alternative hypothesis even if no further evidence supporting this hypothesis is presented (Holyoak & Simon, 1999). If the memory indexing gaze data are able to reveal such biases in information processing, we should observe fixations that are unrelated to the current symptom. That means, the most fixated hypothesis

could be the alternative hypothesis and not the hypothesis that is supported by the symptom sequence.

Hypothesis 5: Response matching

In decision making it has been shown that people choose the option for which the most evidence has been accumulated (Busemeyer & Townsend, 1993, Krajbich, Armel, & Rangel, 2010). Further, fixation durations get longer for the option that is finally chosen (gaze-cascade, e.g., Fiedler & Glöckner, 2012; Glaholt & Reingold, 2011; Shimojo, Simion, Shimojo, & Scheier, 2003). Congruently and in line with previous findings, we expect that fixations directed toward a participant's final diagnosis will increase toward the end of the reasoning trial and will be at the highest proportion during the response interval. Table 1 provides an overview of the tested gaze hypotheses and the main results.

Method

The study consisted of a learning phase followed by a reasoning phase. The reasoning task required participants to determine the most likely cause of a patient's symptoms. In the learning phase, participants first learned how symptoms are assigned to symptom classes, and then how symptom classes relate to chemicals. Participants were informed that the patients in need of diagnosis were workers employed in a chemical plant, that their symptoms were caused by one of the processed chemicals, and that each patient was affected by only one of the listed chemicals (the chemical list was exhaustive with mutually exclusive explanations). Associations between symptom classes and chemicals were established by presenting symptom classes in rectangular frames in the screen quadrants that each represented one chemical (see Fig. 1). During reasoning, symptoms were presented auditorily while participants observed the emptied rectangular frames. Eye movements were recorded throughout the reasoning phase, and the diagnostic decision was collected at the end of each reasoning trial.

Participants

Of the 34 participants, for whom calibration of the eye tracker succeeded to an accuracy of at least 2° of visual angle, two participants were excluded because eye-tracking accuracy decreased during the experiment. The final 32 participants were all students from Chemnitz University of Technology (21 female, 11 male), with a mean age of 22.4 years (ranging from 19 to 39 years). All had normal or corrected-to-normal vision.

Table 1 Study hypotheses and results on memory indexing gaze behavior

<p>Hypothesis 1: Gaze behavior and diagnostic response Gaze behavior can predict the diagnostic response</p>	<p>Fixation proportions toward the A chemical are a significant predictor for the A response. Thus, the longer participants gaze toward the A chemical during the four symptom intervals, the higher the A response proportion. (<i>Confirmed</i>)</p>
<p>Hypothesis 2: Location matching Gaze data following the presentation of the first symptom reflect which hypothesis this symptom supports</p>	<p>Participants fixate the chemical being supported by the first symptom much longer than chance level would predict, thus corroborating our hypothesis on location matching. (<i>Confirmed</i>)</p>
<p>Hypothesis 3: Integrated probability matching Eye movements during later symptom presentations reveal the integration of symptom information beyond mere memory retrieval.</p>	<p>When listening to a symptom that is associated with two hypotheses, participants gaze longer toward the hypothesis that received more support during the sequence of presented symptoms. (<i>Confirmed</i>)</p>
<p>Hypothesis 4: Hypothesis change The change in gaze durations toward an alternative predicts a hypothesis change. After a hypothesis change, fixations were unrelated to the current symptom.</p>	<p>The difference in A-fixation proportions from the first to the last two symptom intervals predicts the response (A vs. not A). After a hypothesis change, participants fixated most on the alternative hypothesis and not on the hypothesis supported by the symptom sequence. (<i>Confirmed</i>)</p>
<p>Hypothesis 5: Response matching Fixations directed toward a participant's final diagnosis increase toward the end of the reasoning trial and will be at the highest proportion during the response interval.</p>	<p>Fixation proportions increased for the chosen hypothesis from the third symptom interval until the response interval. When giving the response, participants gazed longer toward the chosen hypothesis than chance level would predict. (<i>Confirmed</i>)</p>

Apparatus

Participants were seated at a distance of 63 cm in front of a 22-in. computer screen (1680 × 1050 pixels). Stimuli were

presented via E-Prime 2.0. Auditory recordings were presented through headphones and responses were given on a standard keyboard. An SMI RED remote eye tracker sampled data from the right eye at 120 Hz. Gaze data were recorded with

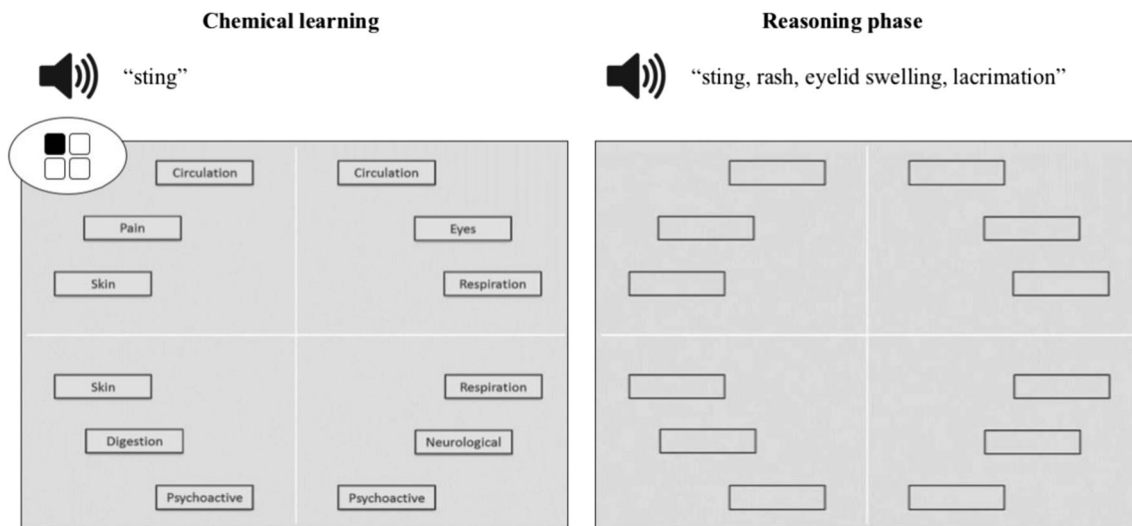


Fig. 1 *Left:* Spatial arrangement of the four chemicals and the symptom classes that each chemical could cause as it was presented during learning. Each of the four screen quadrants represented one chemical and each chemical consisted of three symptom classes. When being tested during learning, participants listened to single symptoms while the arrangement was emptied (as shown on the right) and had to indicate the corresponding chemical. For example, hearing “sting” they were supposed to indicate the top left quadrant because sting belonged to the pain class of symptoms. Participants indicated their response by the corresponding top left response key. *Right:* Emptied spatial arrangement

shown during the reasoning phase. Participants listened to four symptoms and had to indicate which chemical most likely caused the symptoms. The sequence *sting, rash, eyelid swelling, and lacrimation* is an example of an a-ac-b-b sequence. In the example, the *top left* chemical is in the A role supported by *sting* (pain) and *rash* (skin); the *top right* chemical is in the B role supported by *eyelid swelling* (eyes) and *lacrimation* (eyes); the *bottom left* chemical is in the C role supported only by *rash* (skin). The D chemical was not supported by symptoms presented in this sequence and was located diagonally to the A chemical (see main text for more information)

iView X 2.5 following 5-point calibration and analyzed with BeGaze 2.3. Fixation detection used a dispersion threshold of 80 pixels and a duration threshold of 100 ms. For the statistical analyses we used the R language (R Core Team, 2016) and JASP 0.8.0.0 (JASP Team, 2016).

Material

The four chemicals were assigned to screen quadrants (see Fig. 1, left). Each quadrant enclosed three rectangular frames, which contained the three symptom classes that the respective chemical could cause. For example, the chemical at the top left in Fig. 1 triggered symptoms derived from the symptom classes circulation, pain, and skin. One symptom class was unique (pain for the top left chemical) and two symptom classes were shared with other chemicals. Table 2 lists all eight symptom classes and symptoms.

Frames containing symptom classes were arranged in a circle. The distance between the center of the screen and the center of each rectangle was 12.2° of visual angle. The four symptom classes that were uniquely caused by a chemical were presented in the center of the respective quadrant (e.g., the symptom class pain in the center of the top left quadrant in Fig. 1 is located between the symptom classes circulation and skin). The symptom classes that were triggered by two chemicals featured in two quadrants and were presented in two neighboring frames of the circle (e.g., circulation in Fig. 1 is located top right and top left).

Symptoms from symptom classes that were associated with one chemical are denoted with a single small letter (*a*, *b*, *c*, or *d*). Symptoms from symptom classes that were associated with two chemicals are denoted with two small letters (e.g., symptom *ab* can be caused by Chemical A and Chemical B).

A single trial in the reasoning phase consisted of four symptoms presented auditorily; for example, sting, rash, eyelid swelling, and lacrimation (Fig. 1, right). In this example, sting (belonging to the pain class) supported the top left chemical; rash (skin) supported the top left and the bottom left

chemicals, and eyelid swelling (eyes) and lacrimation (eyes) supported the top right chemical. The chemical that was assumed to have an advantage in participants' diagnostic reasoning is the chemical in the *A* role (henceforth called *A* chemical). The advantage may have been due to (1) the chemical being supported by more symptoms than alternative chemicals, or (2) it having received equal support like alternatives but benefited from being supported by the first symptom, or (3) the chemical being supported by an equal number of symptoms but by more diagnostic symptoms or symptoms from more than one symptom category. The competing alternative chemical in this study is referred to as the chemical in the *B* role (henceforth called *B* chemical), with further competitors referred to as *C* and *D* chemicals. Note that the chemical roles changed from trial to trial. Thus, the eye symptom could support a chemical in the *A* role in one trial but support a chemical in the *C* role in another trial.

Sixteen symptom sequences were constructed that contained support for two or three hypotheses and consisted of symptoms that supported either one or two hypotheses. A subset of nine sequences shown in Table 3 was selected to demonstrate how memory indexing tracks the subjective status of hypotheses and provides information about coherence maximization. In all of the selected sequences, the first symptom established a single leading hypothesis. The development of a coherent explanation can most clearly be observed when the first symptom supported one hypothesis. The remaining seven sequences mainly differed from the selected sequences in the order of symptom presentation and in the first symptom supporting two hypotheses (*A* and *C* or *A* and *B*). The full set of sequences and a discussion of order effects on response proportions are included in the [Supplemental Materials](#).

The nine selected sequences varied in the number of consecutive symptoms that supported the *A* hypothesis from the beginning of the sequence onward (see Table 3). Sequence 1 in Table 3 started with three symptoms supporting *A* (*a-ab-ab-b*). Sequence 2 started with two symptoms supporting *A* (*a-ac-b-b*). Sequences 3 and 4 started with a single *a* symptom (*a-bd-bd-a* and *a-bd-a-ab*). Sequence 5 started with one

Table 2 Symptom classes and symptoms (originally in German)

Symptom class	Symptom	Symptom
Eyes (<i>Augen</i>)	Eyelid swelling (<i>Lidschwellung</i>)	Lacrimation (<i>Tränenfluss</i>)
Respiration (<i>Atemwege</i>)	Difficulty breathing (<i>Erstickungsgefühl</i>)	Cough (<i>Husten</i>)
Neurological (<i>Nervensystem</i>)	Speech disorder (<i>Sprachstörung</i>)	Paralysis (<i>Lähmung</i>)
Circulation (<i>Kreislauf</i>)	Sweating (<i>Schwitzen</i>)	Fainting (<i>Ohnmacht</i>)
Pain (<i>Schmerzen</i>)	Twinge (<i>Stechen</i>)	Sting (<i>Brennen</i>)
Skin (<i>Haut</i>)	Rash (<i>Ausschlag</i>)	Acid burn (<i>Verätzung</i>)
Digestion (<i>Verdauung</i>)	Vomiting (<i>Erbrechen</i>)	Diarrhoea (<i>Durchfall</i>)
Psychoactive (<i>Psychoaktiv</i>)	Aggression (<i>Aggressivität</i>)	Anxiety (<i>Angstzustände</i>)

Table 3 Mean response proportions, standard deviations, and within-subjects 95% confidence intervals (Morey, 2008) for nine symptom sequences

No.	Sequence	Response A		Response B		Response D	
		<i>M</i> (<i>SD</i>)	95% CI	<i>M</i> (<i>SD</i>)	95% CI	<i>M</i> (<i>SD</i>)	95% CI
1	<i>a-ab-ab-b</i>	0.54 (0.23)	[0.46, 0.63]	0.44 (0.23)	[0.35, 0.52]		
2	<i>a-ac-b-b</i>	0.40 (0.39)	[0.29, 0.51]	0.58 (0.30)	[0.46, 0.69]		
3	<i>a-bd-bd-a</i>	0.54 (0.28)	[0.44, 0.64]	0.31 (0.23)	[0.23, 0.39]	0.12 (0.15)	[0.06, 0.17]
4	<i>a-bd-a-ab</i>	0.77 (0.19)	[0.69, 0.84]	0.18 (0.21)	[0.10, 0.26]		
5	<i>b-ab-ac-ac</i>	0.57 (0.26)	[0.47, 0.66]	0.25 (0.28)	[0.15, 0.35]		
6	<i>b-b-ac-a</i>	0.40 (0.34)	[0.27, 0.52]	0.53 (0.36)	[0.40, 0.67]		
7	<i>b-b-a-ac</i>	0.36 (0.32)	[0.24, 0.47]	0.59 (0.33)	[0.47, 0.71]		
8	<i>a-b-ab-ac</i>	0.73 (0.24)	[0.65, 0.82]	0.21 (0.21)	[0.13, 0.29]		
9	<i>a-d-ab-ac</i>	0.63 (0.27)	[0.54, 0.73]			0.18 (0.20)	[0.10, 0.25]

Italics mark the consecutive symptoms (three, two, one, or zero) that supported the *A* hypothesis from the beginning of the sequence onward

symptom supporting *B* (*b-ab-ac-ac*). Sequences 6 and 7 started with two symptoms supporting *B* (*b-b-ac-a* and *b-b-a-ac*). Sequences 8 and 9 again started with one symptom supporting *A* and only differed in the second symptom either supporting *B* or *D*.

For each symptom sequence, each of the four chemicals appeared once in the *A* role. This was possible due to the symmetrical symptom class patterns of the chemicals. All possible assignments of symptoms (e.g., lacrimation) to symptom sequences (e.g., *a-ab-ab-a*) were constructed with the restriction that no single symptom occurred twice in the same symptom sequence (e.g., lacrimation occurred only once within the sequence *a-ab-ab-a*). Each sequence was tested four times per participant, with each chemical assuming the *A* role, resulting in 64 (16 sequences \times 4 chemicals) trials per participant.

Procedure

Participants first learned about symptoms and the eight symptom classes that they belonged to. Learning of symptoms and symptom classes proceeded by categorizing single symptoms in one of eight symptom classes (see e.g., Jahn & Braatz, 2014; Rebitschek, Krems, & Jahn, 2015) and continued until all symptoms had been answered correctly once in sequence. Learning about symptoms and symptom classes took 11 min on average ($SD = 10$ min).

In the next phase, participants learned about the four chemicals. They studied the spatial layout as shown in the left half of Fig. 1. During test trials, participants saw only the emptied spatial frames (Fig. 1, right), and single symptoms were presented auditorily. Participants were not explicitly instructed to look at the spatial frames, neither during learning nor during the reasoning phase. They responded by indicating which chemical could have caused the presented symptom by pressing one of four keys on a number block of a keyboard. The keys matched the spatial

positions of the chemicals (e.g., number 1 indicated the chemical at the bottom left). Feedback was provided auditorily and visually (see Jahn & Braatz, 2014). Learning lasted until participants assigned 95% of all symptoms correctly. Learning which symptom classes could be caused by which chemicals took 10 minutes on average ($SD = 9$ min).

Each reasoning trial was initiated by the participant by pressing the space bar. The next slide showed the emptied rectangular frames (Fig. 1, right) and participants were auditorily presented with a sequence of four symptoms. Each symptom presentation lasted 1,000 ms followed by a delay of 2,000 ms. After the fourth symptom and the delay, the response interval started. Participants indicated their diagnosis using the same keys as practiced during learning. Response time was not restricted. On average participants took 2,750 ms ($SD = 2367$ ms) to respond.

After solving three practice trials at the beginning of the reasoning phase, the eye tracker was calibrated. Participants then worked through 64 reasoning trials which took on average 21 minutes ($SD = 3$ min).

Results

Mean response proportions and mean fixation proportions based on fixation durations are reported for the subset of nine sequences (for an overview of the sequences see, Table 3, second column). Response data for all tested sequences are presented in the [Supplemental Materials](#).

Diagnostic response

Diagnostic responses were recorded after the sequence of four symptoms had been presented (end-of-sequence response mode). Participants chose one of the four chemicals (*A*, *B*,

C, or *D* chemical) as the most likely cause of the presented symptoms.

Participants almost always chose one of the contending hypotheses, choosing a chemical that was not supported by the symptom sequence in only 37 trials (1.8% of all trials). In 38 trials (1.9% of all trials), they chose the diagnosis that was only weakly supported by a single symptom when this symptom also pointed to a more supported chemical (e.g., *C* response after *a-ac-b-b*). These cases were excluded from further analysis. Table 3 shows response proportions for the nine sequences and separately for each response.

The *A* response proportions were the highest for Sequences 4, 8, and 9 in Table 3, in which *A* received superior support. Unsurprisingly, people most frequently chose *A* for these sequences. When multiple hypotheses were supported by two symptoms each (Table 3, Sequences 2, 3, 6, and 7), participants more often chose the hypothesis supported by two symptoms from the same symptom class (*b* and *b* or *a* and *a*) rather than selecting a competing hypothesis supported by symptoms that (singly or both) were associated with two chemicals (*a* and *ac* or *bd* and *bd*): Symptoms supporting only one hypothesis (highly diagnostic) were thus evaluated as stronger evidence than symptoms supporting two hypotheses. See [Supplemental Materials](#) for a more detailed discussion of this finding.

Memory indexing gaze behavior

To analyze gaze behavior, we first computed the proportion of trial duration per trial for which no gaze data had been recorded. Trials were discarded if more than 40% of gaze data were missing (4.9% of all trials; see Renkewitz & Jahn, 2012). For one participant, more than 40% of gaze data were missing in every trial, leaving a sample of 31 participants for these analyses. Four areas of interest (AOIs) were defined corresponding to the four quadrants representing the four chemicals. The AOIs were denoted *A*, *B*, *C*, and *D* according to the four chemical roles (remember that quadrants' roles differed from trial to trial). The center of the screen (a circular area around the center of the screen with a diameter of 5.1° of visual angle) was not included in the analysis. Figure 2 shows plots of mean fixation proportions, aggregated over trials and participants, across the five time intervals for the first five symptom sequences that were presented in Table 3. These five sequences were representatives of each class of items (see Fig. S1). Plots on memory indexing gaze data of Sequences 6 to 9 of Table 3 are included in the [Supplemental Materials](#). To show differences in symptom processing resulting in one or the other diagnosis (coherence maximizing), there are separate plots for trials with *A*, *B* and *D* responses (left, middle, and right

Fig. 2 Mean proportions of fixation times in each interval that fell upon the *A*, *B*, *C*, or *D* quadrants for four ambiguous symptom sequences with two contending hypotheses (*A* responses left column, *B* responses middle column) and one ambiguous sequence with three contending hypotheses (additionally *D* responses right column). The number of participants shows how many participants responded at least once with the *A*, *B*, or *D* response. *X*-axis labels show the five symptom intervals with the respective symptoms. *Error bars* represent within-subjects 95% CIs (Color figure online)

column, respectively). Over all responses and sequences, these plots show that symptoms are interpreted and integrated with previous symptoms after presentation. Gaze allocation toward the chemical quadrants measured by fixation proportions differs markedly for the same sequence depending on the finally chosen diagnosis, even after two or three symptoms. In trials with *A* responses, the *A*-fixation proportion dropped when the earlier symptoms supported an alternative hypothesis. Similarly, in trials with *B*-responses fixation proportions for *B* increased the earlier a *B*-supporting symptom was presented, leading to a hypothesis change if the sequence started with a (see top to bottom ordering of sequences in Fig. 2). In the sequence *a-bd-bd-a* (Fig. 2.3), a third hypothesis *D* was as supported as *B*. In trials with *D* responses, the most fixated quadrant shifted from *A* to *D*.

The following analyses focus on showing that gaze data can predict responses, the generation of a leading hypothesis, integration of symptom information, and biased symptom processing to maximize coherence that either favors the leading hypothesis or results in a hypothesis change. Finally, we analyze gaze data during the response interval.

Hypothesis 1: Gaze behavior and diagnostic response To show a link between memory indexing gaze data and the outcome of the reasoning process, we applied linear mixed-effects logistic regression modeling (Bates, Maechler, Bolker, & Walker, 2015). Therefore, we first computed fixation proportions toward the *A* chemical over the four symptom presentations and related them to a binary coding of the diagnostic response, that is, deciding for or against the *A* chemical.¹ Mixed-effects modeling with by-subject and by-item random intercepts and a fixed effect for *A*-fixation proportions predicted the *A* responses significantly better than a model consisting only of by-subject and by-item intercepts as obtained by a chi-square likelihood ratio test of model 1 against the null model, $AIC_{\text{null}} = 1395$, $AIC_{\text{model 1}} = 1324$, $\chi^2(1) = 72.7$, $p < .001$,

¹ In the sequence *b-ab-ac-ac* (Fig. 2.5), the leading hypothesis after the first symptom is *B*. To be able to analyze the data over all five sequences, for all analyses, the sequence *b-ab-ac-ac* was recoded by reversing the *A* and *B* roles so that the *b* symptom became an *a* symptom and the *ac* symptoms became *bd* symptoms, resulting in the sequence *a-ab-bd-bd*. Similarly, the sequence *b-b-ac-a* was recoded to *a-a-bd-b* and the sequence *b-b-a-ac* to *a-a-b-bd* (see Table 3, Sequences 6 and 7).

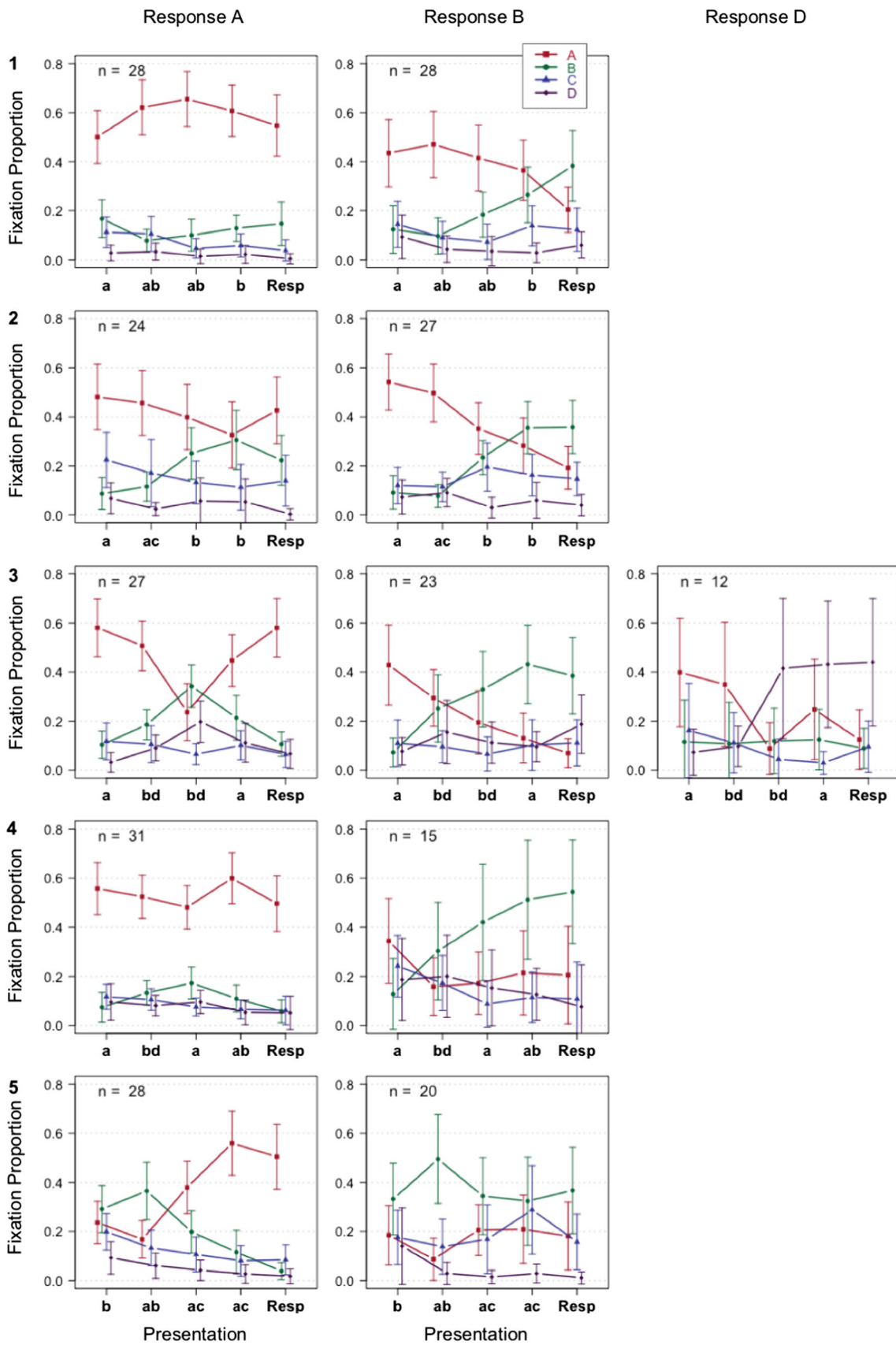


Table 4 Coefficients of mixed-effects logistic regression and z statistics testing A -fixation proportions over all four symptom presentations (Model 1) and A -fixation proportions plus the proportional change in fixations from the first to the last two symptom intervals (Model 2) as predictors of the final A response

	Model 1			Model 2		
	b [95% CI]	z	p	b [95% CI]	z	p
Intercept	-0.54 [-1.02, -0.05]			-0.46 [-0.94, 0.03]		
A -fixation proportions	1.90 [1.45, 2.36]	8.2	<.001	1.96 [1.50, 2.43]	8.3	<.001
Diff. A -fixation proportions				-0.83 [-1.21, -0.45]	-4.3	<.001

Nagelkerke's $R^2 = 9.0$, $N = 1,073$. Additionally, the fixed effect A -fixation proportions significantly predicted the final choice as revealed by the Wald-statistic (also known as z statistic) testing whether the fixed-effect coefficient significantly differed from zero (see Table 4, Model 1). Each increase in A -fixation proportions by 0.1 increased the odds for an A response by 20.9%.

Hypothesis 2: Location matching In the first symptom interval, fixation proportions should reflect how much the first symptom supported each individual hypothesis (momentary probability matching). The first four symptom sequences (see Figs. 2.1–2.4) began with an a symptom. Accordingly, the A quadrant in the first interval should be fixated on longer than the other three spatial areas B , C , and D . Likewise, in the symptom sequence commencing with a b symptom (see Fig. 2.5), B should be fixated on longer than A . Given four possible diagnoses, fixation proportions toward the chemical supported by symptoms during the first symptom interval should differ significantly from the chance level of .25. As expected and confirmed by a one-sample t test, during the first symptom interval, participants gazed at the chemical supported by the first symptom ($M = 0.44$, $SD = 0.19$) longer than predicted by chance, $t(30) = 5.6$, $p < .001$, 95% CI [0.37, 0.51], $d = 1.0$.

Hypothesis 3: Integrated probability matching If eye movements can demonstrate the integration of symptom information beyond mere symptom retrieval, then when listening to a symptom supporting two chemicals, participants should gaze longer toward the more supported chemical. For instance, when listening to “sweating” that is associated with two chemicals, such as A and B (see Fig. 1), participants should look longer toward the A than the B chemical when A is the leading hypothesis. Alternatively, if it is merely retrieval that automatically guides the eyes to all associated spatial locations, when listening to “sweating,” the A and B chemical should be looked at for about the same duration.

In all sequences presented in Fig. 2, a single hypothesis (the A chemical) was established as leading hypothesis, followed by symptoms supporting an alternative hypothesis B^1 . Following the hypothesis on integrated likelihood matching, fixation durations should be longer for the A chemical than for the B chemical when listening to an ab symptom. In order to test this, fixation durations² were aggregated for all sequences and participants for the ab symptoms. In cases with two ab symptoms in one sequence (e.g., Sequence 1 in Fig. 2), we aggregated fixation durations for the two respective intervals. A paired t test supports the hypothesis on integrated probability matching: $M_A = 891.1$ ms, $SD_A = 514.6$ ms, $M_B = 385.8$ ms, $SD_B = 220.9$ ms, $t(30) = 6.84$, $p < .001$, 95% CI [354.4, 656.1], $d = 1.23$. That is, participants looked longer toward the chemical that received more support during the sequence of presented symptoms. Consequently, the null hypothesis that eye movements merely show retrieval processes should be rejected.

Hypothesis 4: Hypothesis change To test whether a change in fixation proportions can predict the dichotomized diagnostic response (A or not A), we ran a second analysis of fixation proportions with a mixed-effects logistic regression model. In this model we included the change in fixation proportions as well as the A -fixation proportions as predictors for the dichotomized diagnostic response. To arrive at a measure for the change in fixation proportions, we first computed two A -fixation proportions: one for the first two symptom intervals computed from fixation durations during the first and second symptom presentations, and another for the last two symptom intervals computed from fixation durations during the third and fourth symptom presentations. Second, we subtracted the A -fixation proportions for the last two intervals from the A -fixation proportions for the first two intervals. If the

² For this analysis, we used fixation durations because fixation proportions toward the A chemical diminish with an increase in fixation proportions toward B . Similarly, fixation proportions toward B diminish with an increase in fixation proportions toward A . Thus, fixation proportions toward different chemicals are not independent of each other.

resulting difference in *A*-fixation proportions (first minus last two symptoms) has a value greater than zero, this means that a participant's orientation toward the *A* chemical was stronger in the first two symptom intervals than during the last two symptom intervals. By contrast, a value smaller than zero indicates that a participant's orientation toward the *A*-chemical increased from the first two to the last two symptom intervals. A value around zero means that *A*-fixation proportions were similar during the first two and last two symptom intervals. Mixed-effects modeling showed that a model with by-subject and by-item random intercepts, and fixed effects for the *A*-fixation proportions and the difference in *A*-fixation proportions, predicted the *A* responses significantly better (chi-square likelihood ratio test) than a model consisting of *A*-fixation proportions as a single fixed effect and by-subject and by-item intercepts, $AIC_{\text{model 1}} = 1324$, $AIC_{\text{model 2}} = 1307$, $\chi^2(2) = 19.06$, $p < .001$, Nagelkerke's $R^2 = 2.5$, $N = 1,073$. The difference in *A*-fixation proportions significantly predicted the final choice as tested with the Wald-statistic (see Table 4, Model 2). Each increase in the early-minus-late-difference in *A*-fixation proportions by 0.1 decreased the odds for an *A* response by 8.0%.

In addition, following a hypothesis change, fixation proportions changed away from the presented symptom information (fixations unrelated to the current symptom). This became visible in cases where participants changed their belief away from the leading *A* hypothesis following the presentation of an inconsistent *bd* symptom (Sequence 3: *a-bd-bd-a* and Sequence 4: *a-bd-a-ab* in Fig. 2). When listening to a *bd* symptom, participants gazed longer toward the *B* or *D* chemical than toward the *A* chemical. In the *a* symptom interval following the presentation of the *bd* symptom, fixation proportions significantly increased for the diagnosis that was chosen—not just if the final diagnosis was *A* but also if the final diagnosis was *B* or *D*. In order to test the reliability of this pattern, we compared mean fixation proportions toward the chosen diagnoses for the *bd*-symptom interval ($M = 0.35$, $SD = 0.33$) with the immediately following *a*-symptom interval ($M = 0.51$, $SD = 0.32$), $t(29) = -2.95$, $p = .006$, 95% CI [-0.28, -0.05], $d = 0.6$. In the immediately following *a*-symptom interval and when responding with *B* or *D* (mean fixation proportion for *B* and *D*: $M = 0.60$, $SD = 0.35$), the *A* hypothesis was almost never gazed at ($M = 0.11$, $SD = 0.18$), $t(20) = 4.91$, $p < .001$, 95% CI [0.28, 0.69], $d = 1.1$. Thus, participants finally choosing *B* or *D* only infrequently looked at the location of the *A* chemical even when an *a* symptom was presented. Instead, they looked at the location of the chemical they believed in.

Hypothesis 5: Response matching To determine whether fixation proportions directed toward a participant's final diagnosis (e.g., *A*-fixation proportions when choosing the *A* diagnosis) increased toward the end of the reasoning trial, a

repeated-measures ANOVA comparing fixation proportions between the third ($M_3 = 0.33$, $SD_3 = 0.12$), fourth ($M_4 = 0.37$, $SD_4 = 0.13$), and response ($M_{\text{resp}} = 0.44$, $SD_{\text{resp}} = 0.18$) intervals for the chosen diagnosis was conducted. The test revealed a significant increase in fixation proportions approaching the response interval, Greenhouse–Geisser corrected, $F(2, 41.7) = 15.96$, $p < .001$, $\eta_p^2 = .35$. Furthermore, during the response interval itself, participants' fixation proportions were the highest for the chosen diagnosis (see Fig. 2) as confirmed by a one-sample *t* test comparing fixation proportions to the chance level of .25, $t(30) = 5.89$, $p < .001$, 95% CI [0.38, 0.51], $d = 1.1$.

Discussion

In everyday life, humans have to cope with ambiguous, uncertain situations. This is particularly evident when people have to find an explanation for a set of inconclusive observations. How do people cope with ambiguity in such challenging instances of diagnostic reasoning? Outcome data suggest that people strive for a coherent interpretation of observations (Glöckner et al., 2010; Holyoak & Simon, 1999; Mehlhorn & Jahn, 2009; Kostopoulou et al., 2012; Wang et al., 2006). A coherent interpretation can be achieved through biased information processing and information distortion. Observing such processes directly had not been done before, because methods were missing that could reveal the changing activation status of hypotheses over the course of a reasoning trial. We tested coherence maximization during diagnostic reasoning using memory indexing—a new method that is based on observing eye movements while participants solve memory-based, higher level cognitive tasks (Jahn & Braatz, 2014; Renkewitz & Jahn, 2010, 2012; Scholz et al., 2015). This study provides evidence that eye movements reflect the tendency to maximize coherence in diagnostic reasoning. The current experiment showed these effects with symptom sequences that were highly ambiguous and supported the initial hypothesis with more or fewer symptoms in a row.

At the beginning of a reasoning trial, gaze behavior reflected the momentary probability of hypotheses given the presented symptom information (location matching, Hypothesis 2), replicating previous findings on the looking-at-nothing behavior (e.g., Richardson & Spivey, 2000). Eye movements, however, did not only reflect (automatic) retrieval processes initiated by hearing an auditorily presented symptom. Instead, eye movements reflected the tendency to strive for a coherent interpretation of symptom information. This became evident during later symptom presentations, in which eye movements were predominantly directed to locations of symptom interpretations consistent with the leading hypothesis and not to all locations that were associated with the

presented symptom (integrated probability matching, Hypothesis 3; fixations unrelated to the current symptom, Hypothesis 4). This finding is in line with previous research demonstrating symptom integration with varying symptom strengths (Jahn & Braatz, 2014, see also Altmann & Kamide, 2007, 2009; Scholz et al., 2015, for similar interpretations of their results). In this study, the location of the symptom classes (small rectangular areas within quadrants) coincided with the chemical locations (the quadrants). Therefore, it is difficult to quantify the amounts of retrieval versus processing of information held in memory and their relation to the resulting fixation duration. All information was learned by heart and tested equally often, which should keep the retrieval effort and time about constant. Thus, the observed differences in fixation proportions can be attributed to differences in information processing. Still, future research is needed to quantify to what extent eye movements reflect retrieval and processing of information held in memory. Initial attempts to disentangle these processes with eye movement measures exist, but thus far these have led to differing results (Glaholt & Reingold, 2011; Horstmann, Ahlgrimm, & Glöckner, 2009; Klichowicz, Scholz, Strehlau, & Krems, 2016).

A change in fixation proportions from the first to the last two symptom intervals can predict a hypothesis change (Hypothesis 4). Thus, by studying eye movements we can directly observe *whether* coherence maximization leads to belief revision during symptom presentation. Here, we compared the first to the last two symptom intervals. This was a simplification, because earlier or later hypothesis changes were possible depending on when during the symptom presentation strong evidence for an alternative hypothesis was presented. Incorporating information on when during the symptom presentation a hypothesis change becomes likely may increase the predictive power of the model of hypothesis change. However, to enable this a more detailed understanding of the timing of the belief updating process and its relation to the execution of eye movements is required. For instance, in the sequence *a-ab-ab-b*, the fourth symptom presented strong evidence for the alternative *B* hypothesis. Nonetheless, even participants who eventually chose the *B* chemical gazed longer toward the *A* than the *B* chemical when considering this piece of information. This result may have been due to the leading *A* hypothesis and *ab* symptoms being initially interpreted as support for *A*, or it may be an artifact of gaze allocation being slower than the memory updating process. The results provide only a first step in studying hypothesis changes by applying memory indexing. Future research is needed to clarify the exact timing between eye movements and belief updating processes, and thus make more specific predictions about hypothesis change.

By separating the fourth symptom interval and the response interval, the gaze cascade effect could be observed more

clearly in this study than in previous experiments. When giving their response, participants fixated longest toward the chosen hypothesis (response matching, Hypothesis 5). It has been argued that the higher fixation duration toward the chosen option demonstrates that eye movements can influence preference judgments (see Shimojo et al., 2003, but see Glaholt & Reingold, 2011). Indeed, manipulating eye movements can lead to better retrieval performance (Johansson & Johansson, 2014; Scholz et al., 2016) and guiding the eyes toward salient cue information can influence the decision strategy (Platzer et al., 2014). Eye movements can thus be both cause and consequence of memory retrieval (Ferreira, Apel, & Henderson, 2008; Richardson, Altmann, Spivey, & Hoover, 2009), and they have been shown to update information processing in memory (Spivey & Dale, 2011). However, when gaze is not guided by a salient event in the visual world, eye movements do not alter the processing of information in memory (Altmann & Kamide, 2007, 2009; Hoover & Richardson, 2008; Richardson & Kirkham, 2004; Richardson & Spivey, 2000; Scholz et al., 2015).

The use of ambiguous symptom sequences in this study resulted in varying responses to the same sequence of symptoms. Although participants were presented with the same symptom sequences, their interpretation differed depending on their subjective evaluation of symptom information and this prompted different final diagnoses. This result conforms to research showing that identical patterns of observed events can lead to different outcomes depending on the reasoners' current causal beliefs (Hayes, Hawkins, Newell, Pasqualino, & Rehder, 2014; Meder et al., 2014). The analysis of gaze behavior by response clearly showed that the final response developed via a process of biased symptom processing and information distortion. For instance, in the *a-ab-ab-b* sequence, the bias toward the initially leading hypothesis was clearly reflected in response proportions. Gaze behavior revealed how this advantage of the leading *A* hypothesis developed, but additionally it showed how the hypothesis change developed in trials in which the competing *B* diagnosis was chosen. By directly tracing biased symptom processing unobtrusively, memory indexing provides strong evidence for theories postulating coherence maximizing through biased information processing and information distortion (Kostopoulou et al., 2012; Russo et al., 1996; Wang et al., 2006).

Bridging two lines of research, on eye movements to emptied spatial locations and on diagnostic reasoning, this study revealed to some degree the processing of ambiguous symptom information and allowed deep insights into the nature and the timing of the process of explanation. As our memory indexing results demonstrate, tracing cognitive processes in highly complex tasks is crucial for a better understanding of higher cognition, and informs process models of reasoning and decision making.

Acknowledgments This research was supported by the Swiss National Science Foundation (SNF) Grant PP00P1_157432 to the first author and German Research Foundation (DFG) Grants KR 1057/17-1 and JA 1761/7-1 to the second and third authors. The authors would like to thank Ricarda Fröde and Claudia Dietzel for their help in conducting the experiment, and Bettina von Helversen, Peter Shepherdson, Yvonne Oberholzer, and Tibor Petzoldt for helpful comments on an earlier version of the manuscript.

References

- Alloppenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*, 419–439. doi:10.1006/jmla.1997.2558
- Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The ‘blank screen paradigm’. *Cognition*, *93*, 79–87. doi:10.1016/j.cognition.2004.02.005
- Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*, 502–518. doi:10.1016/j.jml.2006.12.004
- Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*, *111*, 55–71. doi:10.1016/j.cognition.2008.12.005
- Amaya, A. (2015). *The tapestry of reason: An inquiry into the nature of coherence and its role in legal argument*. Oxford: Hart.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. doi:10.18637/jss.v067.i01
- Baumann, M. R. K., Krems, J. F., & Ritter, F. E. (2010). Learning from examples does not prevent order effects in belief revision. *Thinking and Reasoning*, *16*, 98–130. doi:10.1080/13546783.2010.484211
- Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*, 432–459. doi:10.1037/0033-295X.100.3.432
- R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available from <https://www.R-project.org/>.
- Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine*, *78*, 775–780. doi:10.1097/00001888-200308000-00003
- DeKay, M. L., Stone, E. R., & Sorenson, C. M. (2011). Sizing up information distortion: Quantifying its effect on the subjective values of choice options. *Psychonomic Bulletin & Review*, *19*, 349–356. doi:10.3758/s13423-011-0184-8
- Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends in Cognitive Sciences*, *12*, 405–410. doi:10.1016/j.tics.2008.07.007
- Fiedler, S., & Glöckner, A. (2012). The dynamics of decision making in risky choice: An eye-tracking analysis. *Frontiers in Psychology*, *3*, 1–18. doi:10.3389/fpsyg.2012.00335
- 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*, 125–146. doi:10.1037/a0020692
- 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*, 215–228.
- Glöckner, A., Betsch, T., & Schindler, N. (2010). Coherence shifts in probabilistic inference tasks. *Journal of Behavioral Decision Making*, *23*(5), 439–462.
- Hagmayer, Y., & Kostopoulou, O. (2013). A parallel constraint satisfaction model of information distortion in diagnostic reasoning. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the cognitive science society* (pp. 531–536). Austin: Cognitive Science Society.
- Hayes, B. K., Hawkins, G. E., Newll, B. R., Pasqualino, M., & Rehder, B. (2014). The role of causal models in multiple judgments under uncertainty. *Cognition*, *133*, 611–620. doi:10.1016/j.cognition.2014.08.011
- Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24*, 1–55. doi:10.1016/0010-0285(92)90002-J
- Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General*, *128*, 3–31. doi:10.1037/0096-3445.128.1.3
- Hoover, M. A., & Richardson, D. C. (2008). When facts go down the rabbit hole: Contrasting features and objecthood as indexes to memory. *Cognition*, *108*, 533–542. doi:10.1016/j.cognition.2008.02.011
- 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*, 335–354. <http://papers.ssrn.com/abstract=1393729>.
- Huetting, F., Olivers, C. N. L., & Hartsuiker, R. J. (2011). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica*, *137*, 138–150. doi:10.1016/j.actpsy.2010.07.013
- Jahn, G., & Braatz, J. (2014). Memory indexing of sequential symptom processing in diagnostic reasoning. *Cognitive Psychology*, *68*, 59–97. doi:10.1016/j.cogpsych.2013.11.002
- JASP Team. (2016). JASP (Version 0.8.0.0)[Computer software]. Available from <https://jasp-stats.org/>
- Johansson, R., Holsanova, J., Dewhurst, R., & Holmqvist, K. (2012). Eye movements during scene recollection have a functional role, but they are not reinstatements of those produced during encoding. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 1289–1314. doi:10.1037/a0026585
- 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*, 1053–1079. doi:10.1207/s15516709cog0000
- Johansson, R., & Johansson, M. (2014). Look here, eye movements play a functional role in memory retrieval. *Psychological Science*, *25*, 236–242. doi:10.1177/0956797613498260
- Johnson, T. R., & Krems, J. F. (2001). Use of current explanations in multicausal abductive reasoning. *Cognitive Science*, *25*, 903–939. doi:10.1207/s15516709cog2506_2
- Klichowicz, A., Scholz, A., Strehlau, S., & Krems, J. F. (2016). Differentiating between encoding and processing during sequential diagnostic reasoning: An eye-tracking study. In D. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th annual conference of the cognitive science society* (pp. 129–134). Austin: Cognitive Science Society.
- Kostopoulou, O., Russo, J. E., Keenan, G., Delaney, B. C., & Douiri, A. (2012). Information distortion in physicians’ diagnostic judgments. *Medical Decision Making*, *32*, 831–839. doi:10.1177/0272989X12447241
- Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*, 1292–1298. doi:10.1038/nn.2635
- 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*, 1–16. doi:10.3389/fpsyg.2012.00215

- Martarelli, C. S., Mast, F. W., & Hartmann, M. (2017). Time in the eye of the beholder: Gaze position reveals spatial-temporal associations during encoding and memory retrieval of future and past. *Memory & Cognition*, *45*, 40–48. doi:10.3758/s13421-016-0639-2.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive model of context effects in letter perception. Part 1. An account of basic findings. *Psychological Review*, *88*, 375–407.
- McKenzie, C. R. M. (1998). Taking into account the strength of an alternative hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 771–792. doi:10.1037/0278-7393.24.3.771
- Meder, B., Mayrhofer, R., & Waldmann, M. R. (2014). Structure induction in diagnostic causal reasoning. *Psychological Review*, *121*, 277–301. doi:10.1037/a0035944
- Mehlhorn, K., & Jahn, G. (2009). Modeling sequential information integration with parallel constraint satisfaction. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the cognitive science society* (pp. 2469–2474). Austin: Cognitive Science Society.
- 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*, 1391–1411. doi:10.1037/a0023920
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau. In *Tutorials in Quantitative Methods for Psychology*, *4*, 61–64.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175–220. doi:10.1037/1089-2680.2.2.175
- Orquin, J., & Mueller Loose, S. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, *144*, 190–206. doi:10.1016/j.actpsy.2013.06.003
- Patel, V. L., Arocha, J. F., & Zhang, J. (2005). Thinking and reasoning in medicine. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 727–750). New York: Cambridge University Press.
- Platzer, C., Brüder, A., & Heck, D. W. (2014). Deciding with the eye: How the visually manipulated accessibility of information in memory influences decision behavior. *Memory & Cognition*, *42*, 595–608. doi:10.3758/s13421-013-0380-z
- Read, S. J., Vanman, E. J., & Miller, L. C. (1997). Connectionism, parallel constraint satisfaction processes, and gestalt principles: (Re)introducing cognitive dynamics to social psychology. *Personality and Social Psychology Review*, *1*, 26–53. doi:10.1207/s15327957pspr0101_3.
- Rebitschek, F., Bocklisch, F., Scholz, A., Krems, J. F., & Jahn, G. (2015). Biased processing of ambiguous symptoms favors the initially leading hypothesis in sequential diagnostic reasoning. *Experimental Psychology*, *62*, 287–305. doi:10.1027/1618-3169/a000298
- Rebitschek, F., Krems, J. F., & Jahn, G. (2015). Memory activation of multiple hypotheses in sequential diagnostic reasoning. *Journal of Cognitive Psychology*, *6*, 780–796. doi:10.1080/20445911.2015.1026825.
- Rebitschek, F., Scholz, A., Bocklisch, F., Krems, J. F., & Jahn, G. (2012). Order effects in diagnostic reasoning with four candidate hypotheses. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society* (pp. 905–910). Austin: Cognitive Science Society.
- Renkewitz, F., & Jahn, G. (2010). Tracking memory search for cue information. In A. Glöckner & C. Wittmann (Eds.), *Foundations for tracing intuition: Challenges and methods* (pp. 199–218). New York: Psychology Press.
- 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*, 1622–1639. doi: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*, 235–236. doi: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*, 46–62. doi:10.1037/0096-3445.133.1.46
- Richardson, D. C., & Spivey, M. J. (2000). Representation, space and hollywood squares: Looking at things that aren't there anymore. *Cognition*, *76*, 269–295. doi:10.1016/S0010-0277(00)00084-6
- Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986). Schemata and sequential thought processes in PDP models. In J. L. McClelland, D. E. Rumelhart, & The PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*, (Vol. 2, pp. 7–57). Cambridge, MA: MIT Press.
- Russo, J. E., Medvec, V. H., & Meloy, M. G. (1996). The distortion of information during decisions. *Organizational Behavior and Human Decision Processes*, *66*, 102–110. doi:10.1006/obhd.1996.0041
- Scholz, A., Mehlhorn, K., & Krems, J. F. (2016). Listen up, eye movements play a role in verbal memory retrieval. *Psychological Research*, *80*, 149–158. doi:10.1007/s00426-014-0639-4.
- Scholz, A., von Helversen, B., & Rieskamp, J. (2015). Eye movements reveal memory processes during similarity- and rule-based decision making. *Cognition*, *136*, 228–246. doi:10.1016/j.cognition.2014.11.019
- Schulte-Mecklenbeck, M., Kühberger, A., & Ranyard, R. (2011). The role of process data in the development and testing of process models of judgment and decision making. *Judgment and Decision Making*, *6*, 733–739.
- Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, *6*, 1317–1322. doi:10.1038/nn1150
- Simon, D., Snow, C. J., & Read, S. J. (2004). The redux of cognitive consistency theories: Evidence judgments by constraint satisfaction. *Journal of Personality and Social Psychology*, *86*, 814–837. doi:10.1037/0022-3514.86.6.814.
- Simon, D., Stenstrom, D. M., & Read, S. J. (2015). The coherence effect: Blending cold and hot cognitions. *Journal of Personality and Social Psychology*, *109*, 369–394. doi:10.1037/pspa0000029.
- Spivey, M. J., & Dale, R. (2011). Eye movements both reveal and influence problem solving. In S. P. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 551–562). New York: Oxford University Press.
- Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, *65*, 235–241. doi:10.1007/s004260100059
- Stewart, N., Hermens, F., & Matthews, W. J. (2015). Eye movements in risky choice. *Journal of Behavioral Decision Making*, *29*, 116–136. doi:10.1002/bdm.1854.
- Strickland, B., & Keil, F. (2011). Event completion: Event based inferences distort memory in a matter of seconds. *Cognition*, *121*, 409–415. doi:10.1016/j.cognition.2011.04.007
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634. doi:10.1126/science.7777863
- Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, *12*, 435–467. doi:10.1017/S0140525X00057046.
- Thomas, R. P., Dougherty, M. R., Sprenger, A. M., & Harbison, J. I. (2008). Diagnostic hypothesis generation and human judgment.

Psychological Review, 115, 155–185. doi:[10.1037/0033-295X.115.1.155](https://doi.org/10.1037/0033-295X.115.1.155)

Wang, H., Johnson, T. R., & Zhang, J. (2006). The order effect in human abductive reasoning: An empirical and computational study. *Journal of Experimental & Theoretical Artificial Intelligence*, 18, 215–247. doi:[10.1080/09528130600558141](https://doi.org/10.1080/09528130600558141)

Weber, E. U., Böckenholt, U., Hilton, D. J., & Wallace, B. (1993). Determinants of diagnostic hypothesis generation: Effects of information, base rates, and experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 1151–1164. doi:[10.1037/0278-7393.19.5.1151](https://doi.org/10.1037/0278-7393.19.5.1151)