Research Article

Biased Processing of Ambiguous Symptoms Favors the Initially Leading Hypothesis in Sequential Diagnostic Reasoning

Felix G. Rebitschek,[1] Franziska Bocklisch,[2] Agnes Scholz,[2] Josef F. Krems,[2] and Georg Jahn[1]

[1]Department of Psychology, University of Greifswald, Germany, [2]Department of Psychology, Technische Universität Chemnitz, Germany

Abstract. In sequential diagnostic reasoning, observed pieces of evidence activate hypotheses in memory and are integrated to reach a final diagnosis. The order of evidence can influence diagnostic reasoning. This article examines the processing of ambiguous evidence underlying order effects if multiple hypotheses are activated. In five experiments with a quasi-medical scenario, participants dealt with symptom sequences supporting multiple diagnoses. The symptom order, the response mode (end-of-sequence, step-by-step), and the consistency of evidence were manipulated. A primacy order effect occurred with both response modes suggesting that ambiguous pieces of evidence were distorted toward the hypothesis that strongly corresponded with the first piece. The primacy effect was partially counteracted by stepwise belief ratings, which strengthened the weight of recent evidence and promoted switching to an alternative diagnosis. We conclude that once hypotheses are generated, the interplay of coherence-oriented information distortion and memory-dependent analytic processes propagates into distinct order effects in diagnoses.

Keywords: order effects, diagnostic reasoning, information distortion, probabilistic inference, evidence integration, belief updating

In diagnostic reasoning, belief updating, and impression formation, the order in which information is processed can affect the result of information integration (e.g., Anderson, 1981; Asch, 1946; Wang, Johnson, & Zhang, 2006; for a review about order effects, see Hogarth & Einhorn, 1992). Order effects in diagnostic reasoning about sequentially observed evidence are of particular interest in research on medical decision making (Bergus, Chapman, Levy, Ely, & Oppliger, 1998; Chapman, Bergus, & Elstein, 1996; Cunnington, Turnbull, Regehr, Marriott, & Norman, 1997; Kostopoulou, Mousoulis, & Delaney, 2009) and are most perspicuous if diagnoses differ depending on the order in which the same set of symptoms is encountered. Here we aim to examine how biased processing of equivocal symptom sequences favors early diagnostic hypotheses resulting in a primacy order effect, and how to counteract this bias in diagnoses. The experiments in this article presented diagnostic problems with four hypothetical causes. Each symptom was consistent with more than one cause and thus ambiguous. Sequences of ambiguous symptoms that in combination provided equal support for two causes are referred to as equivocal sequences in the following. The response procedure and the consistency of symptom sequences were varied across experiments.

The experiments put the participant in the role of a physician who observes symptoms a patient suffers from. The first observed symptom triggers the first hypothesis or set of hypotheses about the potential cause (Patel, Arocha, & Zhang, 2005; Thomas, Dougherty, Sprenger, & Harbison, 2008). These initial hypotheses have to be updated and evaluated as subsequent symptoms are observed (Weber, Böckenholt, Hilton, & Wallace, 1993). Subsequent symptoms can be consistent with initial hypotheses or they can be inconsistent with all of them and generate new hypotheses from memory. After four symptoms, the participant has to select a final diagnosis from multiple hypotheses.

Bayesian belief updating as a normative standard integrates observations irrespective of symptom order (Shanteau, 1972), but human reasoning already about a single hypothetical cause can be affected by the order of information. Depending on task characteristics, reasoners may produce primacy as well as recency order effects. Such order effects were addressed by the theory of information integration (Anderson, 1981) and by the belief-adjustment model (Hogarth & Einhorn, 1992), but the consideration of multiple hypothetical causes in parallel that is focused in this article is not captured in these theories. Considering

multiple candidate hypotheses is more challenging than evaluating a single candidate (Bylander, Allemang, Tanner, & Josephson, 1991), may exceed the cognitive capacity of the diagnostician (Johnson & Krems, 2001), and, presumably, provokes stronger order effects.

Early evidence in a sequence can have a disproportionately strong impact on judgments. Such *primacy order effects* have been obtained for impression formation (e.g., Anderson & Hubert, 1963), for contingency judgments (e.g., Yates & Curley, 1986), and for subjective probability revision (e.g., Peterson & DuCharme, 1967). In sequential diagnostic reasoning, a primacy effect favors the initial hypothesis that has been triggered by the first piece of evidence. This can be indicated by a higher than warranted confidence in this hypothesis and an increased probability of choosing this hypothesis as the final diagnosis.

Primacy effects can be explained by biased predecisional processing of evidence (Bond, Carlson, Meloy, Russo, & Tanner, 2007) that is encountered subsequent to an initially formed preference for an option (for a review see Brownstein, 2003). Such predecisional distortion of information (Russo, Medvec, & Meloy, 1996) shall ensure that beliefs are consistent (Russo, Carlson, Meloy, & Yong, 2008). Encountered information is processed with the goal to achieve coherence with the emerging decision (Bond et al., 2007; Simon, Snow, & Read, 2004). Hence, the interpretation of new evidence can be distorted in favor of an explanation that already emerged as preferable (Holyoak & Simon, 1999; Simon et al., 2004).

Information distortion was shown to produce primacy effects for stepwise as well as single-judgment response procedures in single-option tasks (Bond et al., 2007; but see Hogarth & Einhorn, 1992) and for the final choice in multi-option tasks (Carlson, Meloy, & Russo, 2006). Similar primacy effects were observed in diagnostic reasoning (Kostopoulou, Russo, Keenan, Delaney, & Douiri, 2012). However, studies on sequential diagnostic reasoning using response procedures with a single final judgment (end-of-sequence, EoS) and with multiple candidate hypotheses are rather inconclusive with respect to order effects (no order effect, Sprenger & Dougherty, 2012; primacy or recency effects, Lange, Thomas, Buttaccio, Illingworth, & Davelaar, 2013). To clarify conditions for order effects in diagnostic reasoning, the present experiments systematically vary factors that may affect order effects.

First, we aim to contribute novel evidence for a primacy effect in diagnostic reasoning using an EoS procedure with four contending candidate hypotheses and ambiguous pieces of evidence. Distortion of evidence increases if its diagnosticity declines (Russo, Meloy, & Medvec, 1998) and if potential causes are equally attractive (cf. Brownstein, 2003). Thus, we present sequences including ambiguous symptoms that support two candidate hypotheses at once (one strongly, one weakly). Moreover, total evidence equally supports two out of four candidate hypotheses in equivocal sequences. If ambiguous symptoms are distorted toward the higher activated hypothesis, this early leading hypothesis (initial hypothesis) is expected to be favored as the final diagnosis over an equally supported alternative (primacy effect). Distortion that can propagate into a primacy effect is described, for instance, in the probabilistic constraint satisfaction model by Hagmayer and Kostopoulou (2013). According to their model a need for coherence drives processes distorting validity of ambiguous pieces of evidence during sequential reasoning.

An alternative approach is offered by the extended computational model of hypothesis generation, HyGene, that comprises sequential symptom processing, the generation of hypotheses, and order effects in diagnostic reasoning (Lange et al., 2013; Thomas et al., 2008). HyGene ascribes primacy as well as recency effects to the working memory dynamics that are modeled with parameters for recurrent activation of sequentially presented symptoms and their mutual inhibition (Lange et al., 2013). However, this model lacks a mechanism of how a set of multiple contending hypotheses is maintained and sequentially updated while symptoms are processed. Thus, distorted processing of ambiguous evidence, which produces a primacy effect as we expect it here, is not captured in the present version of the extended HyGene model.

Second, it is highly relevant, how to counteract such order effects in diagnostic reasoning (Croskerry, 2003; Curley, Young, Kingry, & Yates, 1988). Deliberate strategies can help to surmount reasoning biases (for an overview, see Lilienfeld, Ammirati, & Landfield, 2009) such that two hypotheses can be considered impartially (e.g., McKenzie, 1998). For instance, decreasing overconfidence in a hypothesis is known to debias judgments (Koriat, Lichtenstein, & Fischhoff, 1980). Likewise, considering the opposite (Lord, Lepper, & Preston, 1984; Mumma & Wilson, 1995) and considering not necessarily opposite but plausible alternative hypotheses (Anderson & Sechler, 1986; Hirt & Markman, 1995) are promising strategies for counteracting a bias such as the primacy effect. The strategy of considering alternative hypotheses may be applicable by means of a procedure that prompts the distinct consideration of multiple alternative causes. Hence, step-by-step ratings of all hypothetical causes after each symptom may counteract the primacy effect.

Regarding single-belief updating, primacy effects for a brief sequence of evidence that was fully consistent with the initial belief were shown to diminish if the step-by-step procedure (SbS) was used instead of an EoS procedure (belief-adjustment model, Hogarth & Einhorn, 1992). Thus, for the present study, this stepwise procedure is adapted to encourage the full and repeated consideration of all candidate hypotheses. Participants are prompted after each piece of evidence to rate each hypothetical alternative's probability of being the cause given the observed evidence. This procedure ensures that knowledge about all hypothetical causes is repeatedly generated from long-term memory and encourages evaluative processes that may counteract distorting interpretations of evidence. Hence, we assume that the SbS procedure including repeated belief ratings may reduce a primacy effect, although this is not true of stepwise procedures in general (e.g., Bond et al., 2007).

There is a second way in which the step-by-step procedure can attenuate a primacy effect. As shown in several studies with multiple contending hypotheses and comparable stepwise response procedures (e.g., stepwise lists of candidates, Sprenger & Dougherty, 2012), the step-by-step procedure can induce recency effects. When multiple hypothetical causes are repeatedly rated, considerable time elapses between the early pieces of evidence and the final diagnosis. This may reduce the memory activation of early evidence in favor of later evidence, resulting in a recency effect (cf. Lange et al., 2013). Early evidential information may even be lost due to working memory's capacity limitations (Sprenger & Dougherty, 2012). If stepwise ratings of all hypothetical causes exceed cognitive capacities for holding symptoms and updating hypotheses, early information can be edged out of working memory. Then, a gradual shift to the hypothesis that is triggered by the last diagnostic symptom in a sequence is expected, independent from information distortion. Hence, given a SbS procedure, such a recency shift in diagnoses may ensue in combination with a reduced primacy effect.

Up to this point, the focus laid on diagnostic reasoning about evidence (symptoms) that is consistent with two contending candidate hypotheses. But in real-world settings, inconsistent evidence is not uncommon. Hence, the third issue examined in the experiments concerns how diagnostic reasoning with four candidate hypotheses is altered if evidence is encountered that conflicts with established hypotheses. Research on single-belief updating distinguishes consistent and mixed sequences (Hogarth & Einhorn, 1992). Consistent sequences contain only pieces of evidence that can be subsumed under the initially supported candidate hypothesis, whereas mixed sequences include inconsistent pieces that cannot be reconciled with this candidate. In the present case of multiple candidate hypotheses for equivocal sequences, a sequence is *consistent* if all evidence can be subsumed under two equally supported candidate hypotheses. Despite consistency, participants likely experience unpleasant dissonance because they are prompted for a single choice in the face of equivocal evidence. Presumably, they aim to resolve this dissonance by distorting evidence (cf. Holyoak & Simon, 1999; Russo et al., 2008).

If such equivocal yet consistent sequences are extended with disconfirming evidence (*mixed sequences*) that only supports deviating alternatives, dissonance may be further increased (Brownstein, 2003). In contrast to confirming and non-diagnostic pieces of evidence, disconfirming evidence obviously cannot be distorted to support a leading hypothesis. Instead, information distortion probably weakens this disconfirming piece of evidence and, thus, minimizes the chance of the deviating alternative (Carlson et al., 2006; Russo et al., 1996, 1998).

A conflicting piece of evidence may increase dissonance and it may instigate analytic reasoning (e.g., Klaczynski, 2000; Maheswaran & Chaiken, 1991). Inconsistent evidence receives increased attention (Neuberg & Fiske, 1987) during reasoning, with a special pattern of cortical activation as known from error detection and conflict resolution (cf. Barbey & Barsalou, 2009). The conflict provoked by evidence that cannot be directly reconciled with considered candidate hypotheses may prompt a deepened analysis (Chinn & Brewer, 1993; Hastie & Kumar, 1979) to achieve an inconsistency resolution (e.g., Wyer & Srull, 1989). Conflicting evidence is processed more extensively (Wyer & Gordon, 1982; but see Bodenhausen & Lichtenstein, 1987) to be refuted, rejected, or reinterpreted for rationalizing one's (motivated, Klaczynski, 2000) reasoning. Such elaborative processing is effortful (Wyer & Srull, 1989).

In impression formation, the elaborative processes form relations to associated information for reconciling information that is inconsistent with the established concept. Cognitive processes underlying the evaluation of evidence increase associative linkages between inconsistent information and already established information (presumably processed in working memory, Wyer & Srull, 1986). Likewise, pieces of diagnostic evidence that were distorted earlier in the sequence might be consulted again and possibly reevaluated for integration, which in turn might debias diagnostic judgment. Hence, if disconfirming evidence prompts a deepened analysis, mixed sequences are expected to reveal a reduced primacy effect compared with consistent sequences.

A Quasi-Medical Diagnostic Reasoning Task

We use a diagnostic reasoning task embedded in a quasi-medical scenario. Participants diagnose patients whose sequentially presented symptoms are caused by one of four chemicals (Baumann, Mehlhorn, & Bocklisch, 2007; Jahn & Braatz, 2014; Mehlhorn & Jahn, 2009; Mehlhorn, Taatgen, Lebiere, & Krems, 2011). Prior to the diagnostic task, participants acquire knowledge about fictional chemicals and the symptoms they can cause. Each diagnostic trial consists of four sequentially presented symptoms a patient suffers from.

We manipulate the order of symptoms, the response mode (step-by-step, end-of-sequence), and the consistency of evidence (consistent, mixed sequences). Because we are interested in the updating of multiple hypotheses, we use sequences that remain consistent with multiple candidate hypotheses. Moreover, we use equivocal sequences that equally support two candidate hypotheses. Such a sequence presents an early symptom pointing strongly to one candidate hypothesis (establishing the initially leading hypothesis) and weakly to another candidate hypothesis. Later in the sequence, another symptom points strongly to the previously weakly supported candidate hypothesis and weakly to the previously strongly supported candidate hypothesis (AB-sequences in Table 3).

Regarding equivocal sequences, the forced choice procedure in the present experiments collects a single diagnosis and likely captures the highest activated hypothesis at the end of a symptom sequence. The proportions of final diagnoses across equivalent trials could turn out balanced between two equally supported hypotheses (both 50%)

without an advantage of the initial hypothesis. Differing proportions, however, indicate order effects. The proportion could be higher for the hypothesis that was supported more strongly by the initial symptom (predominant primacy effect) or for the hypothesis that was supported more strongly by a later symptom (predominant recency effect).

We expect order effects depending on the response procedure and the consistency of evidence. The end-of-sequence procedure should promote distortion of consistent evidence toward the initially highly activated hypothesis (primacy effect). The step-by-step procedure that makes all candidate hypotheses repeatedly salient should reduce the primacy effect. Yet, stepwise analytic reasoning is limited by working memory capacity and hence, early evidence could be forgotten or less attended than later evidence if capacity is exceeded. Thus, there may even develop a recency shift in diagnoses with the SbS procedure. Furthermore, the bias toward the initial hypothesis should be reduced by sequences including inconsistent evidence whose integration might provoke an analysis counteracting biased information integration.

Experiments 1A and 1B include items consistently supporting two candidate hypotheses and employ EoS and SbS procedures, respectively. Experiments 2A and 2B use the same EoS and SbS procedures presenting items with inconsistent evidence. Thus, we systematically cross response procedures with consistency of evidence to vary the primacy effect. A subsequent third experiment replicates Experiments 2A and 2B in a single experiment. The experiments extend the range of findings on diagnostic reasoning because they set the response mode and item type combinations in direct comparison and because tasks with four contending candidate hypotheses and an EoS procedure have not been studied before.

Experiments 1A and 1B – Two Contending Hypotheses

The diagnostic reasoning items in Experiments 1A and 1B contain symptoms consistent with two diagnostic candidate hypotheses. The two candidate hypotheses were either equally supported after four symptoms or one had stronger support. Participants' diagnostic judgments should reflect the differing level of support. For the items with equal support (equivocal AB-sequences), we expected a primacy effect for the EoS procedure; for the SbS procedure, we expected a reduced primacy effect and a stronger weight of recent evidence.

Method

Participants

Seventy-nine undergraduate students (58 female; mean age 22.9, SD = 2.8) took part in the experiments. Experiment 1A was conducted at the University of Greifswald with 40 participants and Experiment 1B was conducted at the TU Chemnitz with 39 participants.

Table 1. Symptom categories and symptoms (with the German terms as used in the experiment)

Symptom category	Symptom
Eye (Augen)	Eyelid swelling (Lidschwellung)
	Lacrimation (Tränenfluss)
Respiration (Atemwege)	Cough (Hustenreiz)
	Difficult breathing (Erstickungsgefühl)
Skin (Haut)	Acid burn (Verätzung)
	Rash (Ausschlag)
Neurological (Nervensystem)	Paralysis (Lähmung)
	Speech disorder (Sprachstörung)
Pain (Schmerzen)	Twinge (Stechen)
	Sting (Brennen)
Circulatory problems (Kreislauf)	Sweating (Schwitzen)
	Swoon (Ohnmacht)

Note. Original material was in German.

Design

Response mode (EoS vs. SbS) was varied between-subjects. Participants in Experiment 1A worked through the experimental trials with an end-of-sequence response procedure, whereas participants in Experiment 1B were prompted to state their current belief in each hypothetical cause subsequent to each presented symptom (step-by-step procedure). Furthermore, the design included the within-subjects factor item type (AAB, AB, and ABB) varying the support for candidate hypotheses A and B in symptom sequences.

Material

Learning Material

There were six symptom categories and four chemicals. Each symptom category included two symptoms. For example, *Eyelid Swelling* and *Lacrimation* were elements of the symptom category *Eyes*. Table 1 shows the symptom categories and single symptoms.

There were two gaseous and two fluid chemicals named by single letters. Table 2 shows the four chemicals and the symptom categories they could cause. Each of the chemicals caused symptoms from one symptom category *almost always* and symptoms from two further categories only *occasionally.* There were symptom categories specific for fluid (*Skin*, *Neurological*) or gaseous chemicals (*Eyes*, *Respiration*) and unspecific symptom categories that could be caused by all chemicals (*Pain*, *Circulatory Problems*). Note that symptoms from group-specific categories pointed strongly to one chemical and weakly to another chemical of a group. For example, *Rash* (a *Skin*-symptom) pointed strongly to W and weakly to K.

Table 2. Domain-specific knowledge about chemicals and associated symptom categories participants had to acquire in the learning phase (all experiments)

Group	Chemical	Almost always	Occasionally	Occasionally
Gaseous	R	Eyes	Respiration	Circulatory problems, Pain
	B	Respiration	Eyes	Circulatory problems, Pain
Fluid	W	Skin	Neurological	Circulatory problems, Pain
	K	Neurological	Skin	Circulatory problems, Pain

Note. Original material was in German.

Experimental Material

Within each experimental trial four symptoms were sequentially presented. Three types of sequences with three variants each were used (see Table 3). In all sequences, the presented symptoms were finally consistent with two of the four chemicals. In equivocal sequences (AB sequences), two chemicals were finally equally supported.

In the following, we refer to the chemicals that are strongly suggested at the beginning of a symptom sequence as A-chemicals or A-diagnoses (the initial candidate hypothesis) and to those weakly suggested as B-chemicals or B-diagnoses (the alternative candidate hypothesis). Thus, in a sequence starting with a *Skin*-symptom, W would be the A-diagnosis and K would be the B-diagnosis. The *Skin*-symptom in this case will be referred to as an Ab-symptom because it points strongly to the A-diagnosis and weakly to the B-diagnosis.

Consider a sequence that continues after rash (a *Skin*-symptom pointing strongly to the A-diagnosis and weakly to the B-diagnosis) with sweating (an unspecific *Circulatory Problem*), followed by paralysis and speech disorder that both are *Neurological* symptoms pointing strongly to the B-diagnosis and weakly to the A-diagnosis. The sequence of symptoms in this example is *Ab-x-Ba-Ba* (item type ABB) and can be found in the seventh row of Table 3. Unspecific *Circulatory Problems* and *Pain* symptoms are referred to as x-symptoms.

Table 3 shows all item types with their respective sequences of symptoms. Sequences of item type AAB contained two Ab-symptoms that pointed to the candidate hypothesis strongly supported by the first symptom (A-hypothesis) and one Ba-symptom. Sequences of item type ABB contained two Ba-symptoms that pointed to the candidate hypothesis weakly supported by the first symptom (B-hypothesis) and one Ab-symptom. Sequences of item type AB contained equal support for the A- and the B-hypothesis.

Symptoms were assigned to these sequences such that each of the four chemicals (Table 2) was placed in the role of the A-chemical once for each sequence (Table 3). Thus, 36 combinations of A-chemicals and sequences were presented to each participant partitioned into four blocks. Each block contained all nine sequences and per item type within a block no chemical was repeated as the A-chemical. The order of sequences within a block was randomized

Table 3. Experiments 1A and 1B (above) and Experiments 2A, 2B, and 3 (below): Sequences of symptoms related to first (A) and second (B) or third (C) and fourth (D) diagnosis; x denotes an unspecific symptom

Item type	Sequence	Exemplary sequence
Experiments 1A and 1B		
AB Consistent	Ab-x-x-Ba	Eyelid swelling–Twinge–Sweating–Cough
	Ab-x-Ba-x	
	Ab-Ba-x-x	
AAB Consistent	Ab-Ab-x-Ba	Acid burn–Rash–Sting–Paralysis
	Ab-Ab-Ba-x	
	Ab-x-Ab-Ba	
ABB Consistent	Ab-x-Ba-Ba	Rash–Sweating–Paralysis–Speech Disorder
	Ab-Ba-Ba-x	
	Ab-Ba-x-Ba	
Experiments 2A, 2B, and 3		
AB Consistent	x-Ab-Ba-x	Twinge–Eyelid swelling–Cough–Sweating
	x-Ab-x-Ba	
	x-x-Ab-Ba	
CAB Inconsistent	Cd-Ab-Ba-x	Acid burn–Cough–Eyelid swelling–Sting
	Cd-Ab-x-Ba	
	Cd-x-Ab-Ba	
ABC Inconsistent	Ab-Ba-Cd-x	Paralysis–Rash–Eyelid swelling–Swoon
	Ab-Ba-x-Cd	
	Ab-x-Ba-Cd	

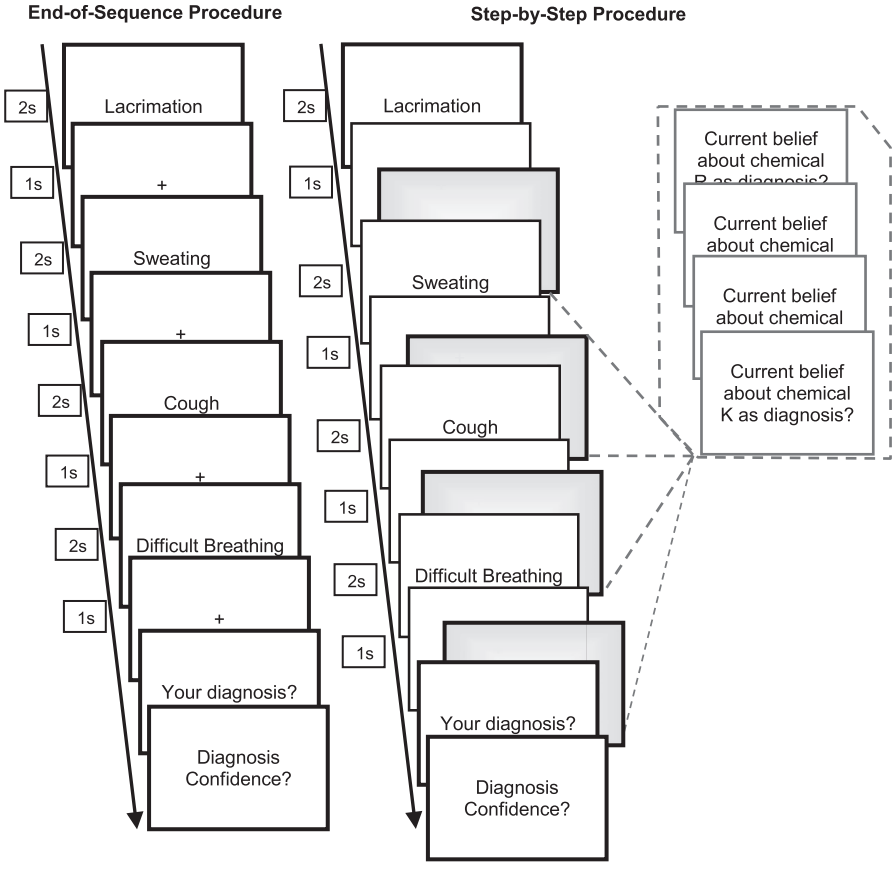

Figure 1. Procedures of exemplary trials in the end-of-sequence-condition (Experiments 1A and 2A, EoS condition of Experiment 3) and in the step-by-step-condition (Experiments 1B and 2B, SbS condition of Experiment 3).

and the order of the four blocks was counterbalanced. At the level of a single trial, the assignment of the actual symptoms was randomly drawn from the complete set of possible symptom combinations excluding repetitions of the same symptom.

Procedure

After the introduction of the cover story, participants were informed that they would learn about the chemicals processed in the plant and the symptom categories that could be caused by the chemicals. Then, symptom categories and the symptoms they subsumed were presented in a table (Table 1). In each trial of the subsequent first learning task, single symptoms (e.g., *Rash*) had to be attributed to symptom categories (e.g., *Skin*) presented as a randomly ordered list. Participants had to respond with the number of the correct category (e.g., *Skin* for *Rash*) and received feedback. The set of 12 single symptoms was repeated in random order until it was once answered without errors.

Next, a table (similar to Table 2) introduced the chemicals, to which group they belonged, and the symptom categories they could cause almost always or occasionally. After participants had studied the table, they started the second learning task training the causal links. Participants saw combinations of a single symptom category with frequency information (e.g., *Pain occasionally*) and had to enter the letter of the corresponding chemical via a standard keyboard (e.g., R). They received feedback. In cases of a wrong response the complete table was shown again until participants continued self-paced. The set of 10 training items was randomly repeated until it was once answered without errors.

After participants had completed learning, the diagnostic reasoning task was explained. They were informed that each trial contained four symptoms that a worker in the chemical plant suffered from. Each worker had come into contact with exactly one of the four chemicals and the task was to decide which chemical most likely had caused the worker's symptoms. Then, they were acquainted with the procedure of the diagnostic task in four training trials. Trials were started self-paced. Participants working through the EoS condition (Experiment 1A) processed trials as shown in the left panel of Figure 1.

Four symptoms were presented sequentially in the center of the screen. Each symptom was visible for 2 s and followed by a fixation cross shown for 1 s. After the last fixation cross, participants were prompted to indicate the chemical that most likely had caused the observed symptoms. The answer was given via the corresponding letter on the keyboard. Finally, participants stated their confidence in the given diagnosis on a 7-point-scale from *very unsure* to *very sure* via number keys.

In the SbS condition (Experiment 1B), participants rated after each symptom for each chemical their belief in the

respective chemical as the cause of the symptoms observed until then (right panel of Figure 1). The instruction read (originally in German):

> "Please enter, how likely (from 0 to 100 per cent) you think chemical … caused the symptoms you observed until now. For how many of 100 similar cases the diagnosis … would be correct? Please enter a number between 0 and 100."

The four chemicals were prompted in random order after each symptom. Following the belief ratings after the fourth symptom, they responded with a final diagnosis and a confidence rating for their diagnosis.

After the four training trials, the 36 experimental trials were presented with a short break after 18 trials. The entire experiment lasted about 45 min.

Results

Diagnoses

C- and D-diagnoses were not supported by specific symptoms and chosen with small rates of 4.8% and 3.6% in Experiments 1A and 1B, respectively. These C- and D-diagnoses were discarded and only A- and B-responses were included in the reported analyses. The proportions of A- and B-diagnoses are shown in Figure 2 for both experiments and separated by item type (see Electronic Supplementary Material, ESM 2). In both experiments, the proportion of A-diagnoses decreased from AAB to AB to ABB items reflecting the decreasing relative support for A by strong symptoms. Polynomial within-subjects contrasts confirmed a significant linear trend with $F(1, 39) = 230.84$, $p < .001$, $\eta_p^2 = .86$ for the EoS condition (Experiment 1A) and with $F(1, 38) = 474.09$, $p < .001$, $\eta_p^2 = .93$ for the SbS condition (Experiment 1B).

For the equivocal AB-items, response proportions deviating from 50% indicate order effects. In the EoS condition, the proportion of A-diagnoses was higher than 50% confirming the expected primacy effect, $t(39) = 4.54$, $p < .001$, $d = 0.72$. In the SbS condition, the proportion of A-diagnoses did not differ significantly from 50%, $t(38) = -0.10$, $p = .924$. This was confirmed with a Bayesian one-sample t-test, using JASP 0.6 (Love et al., 2015) with a default Cauchy prior width $r = 0.707$ ($2^{1/2}/2$ following Morey, Romeijn, & Rouder, 2015; Morey & Rouder, 2014) on the alternative hypothesis. A Bayes factor of $BF_{01} = 5.8$ (with 8.0×10^{-8} error %) indicated that the observed balance of A- and B-diagnoses in the SbS condition is nearly six times more likely under the null hypothesis (substantial evidence, cf. Jarosz & Wiley, 2014). However, the individual response proportions suggest an underlying shift in order effects.

The proportions of participants showing balanced diagnoses (from 5 to 7 A-diagnoses out of 12 diagnoses for AB-items) were similar for the EoS (30.0%) and the SbS condition (33.3%). The overall balanced diagnoses in the SbS condition were due to a remarkable shift toward the B-hypothesis that was supported strongly by the later Ba-symptom, as indicated by the increased proportion of participants, who gave more B- than A-diagnoses (35.9% in the SbS condition vs. only 12.5% in the EoS condition). The shift toward lower proportions of A-diagnoses is visible in the frequency distributions of A-proportions in Figure 3. A comparison of the quartiles of proportions' frequencies confirmed the difference in distributions with $\chi^2(3) = 18.02$, $p < .001$.

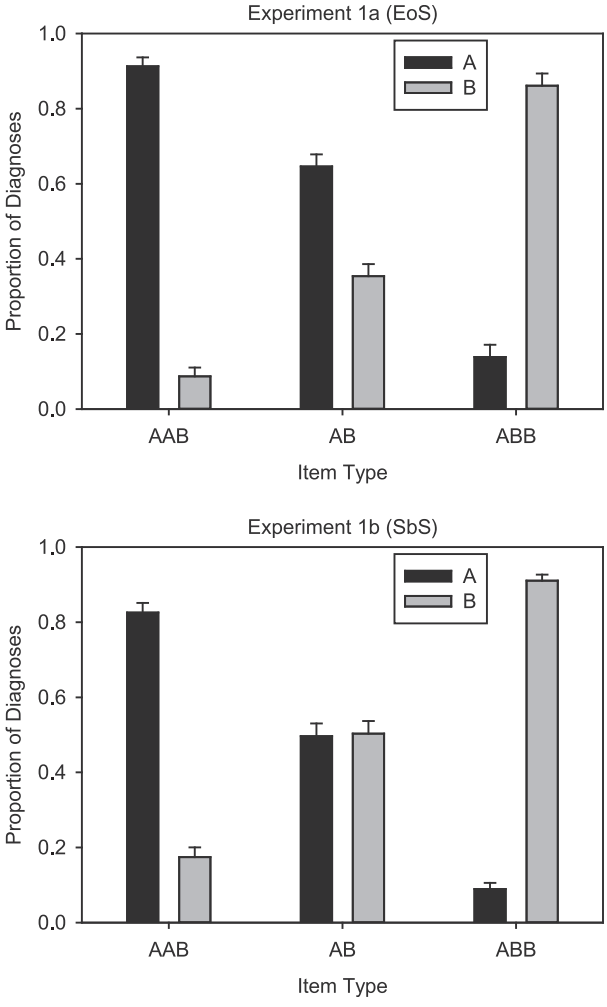

Figure 2. Mean proportions of diagnoses for each item type in Experiments 1A and 1B. Error bars show the standard error of the mean.

Confidence Ratings

Mean confidence ratings for both experiments are depicted in Table 4 (see ESM 1). The confidence ratings were lower for AB-items than for AAB- and ABB-items, which both contained one more specific symptom consistent with A and B.

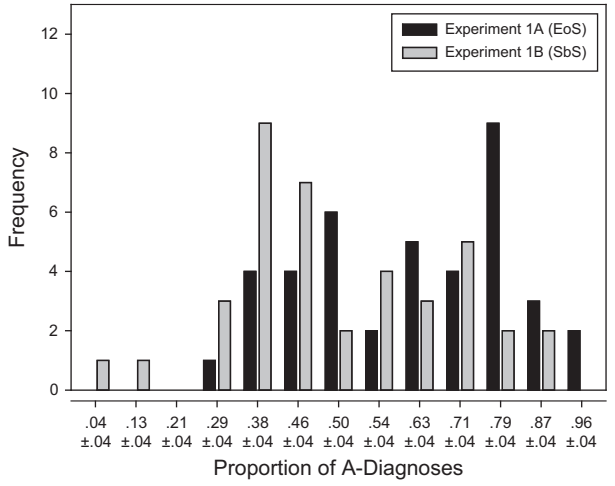

Figure 3. Frequency distribution of A-proportions in individual participants' final diagnoses for AB-items in Experiments 1A (EoS) and 1B (SbS).

Not surprisingly, the confidence in A-diagnoses was higher for AAB-items than for AB-items with *d*s of 1.66 and 1.51 in the EoS and SbS conditions, respectively. Similarly, the confidence in B-diagnoses was higher for ABB-items than for AB-items with *d*s of 1.54 and 1.70, respectively. A stronger influence of the initial hypothesis in the EoS than in the SbS condition is indicated by the higher mean confidence in the rare A-diagnoses for ABB-items; $t(43) = 3.21$, $p = .002$, $d = 0.96$.

Belief Ratings

Sequential belief ratings in the SbS condition reflected mainly the diagnostic value of the symptom presented just before the ratings. This is evident in the mean belief ratings for the three sequences of the AAB item type shown in Table 5.

The ratings for the B-hypothesis after the late Ba-symptom are much higher than for the A-hypothesis although two Ab-symptoms strongly supporting the A-hypothesis had been presented before and participants chose almost always A as the final diagnosis briefly afterwards. Hence, the belief ratings obviously do not reflect the result of symptom integration. Either the participants did not understand the instructions or they did not spend the effort to consider the whole set of observed symptoms. The belief ratings for the AB- and ABB-items similarly reflected the diagnostic value of the just presented symptom and therefore are not further reported.

Table 4. Experiments 1A and 1B: Means of confidence ratings related to the diagnoses (*SD* in parentheses)

Experiment (response mode)	Item type	A	B
1A (End-of-sequence)	AAB	5.66 (0.96)	4.08 (1.80)
	AB	3.77 (1.19)	3.69 (1.29)
	ABB	4.63 (1.11)	5.47 (1.14)
1B (Step-by-step)	AAB	5.10 (1.07)	4.06 (1.34)
	AB	3.49 (0.94)	3.38 (1.06)
	ABB	3.54 (1.16)	5.16 (0.95)

Discussion

The experiments presented diagnostic problems that afforded either sticking with an initial candidate hypothesis (A) or switching to an alternative candidate hypothesis (B). With equal support for A and B in AB-items and the EoS procedure (Experiment 1A), we obtained a clear primacy order effect in final diagnoses favoring the initial candidate hypothesis A. This primacy effect is consistent with earlier findings (e.g., Kostopoulou et al., 2012). Furthermore, it suggests a distorted processing of the symptoms presented subsequently to the generation of an initial hypothesis to achieve coherence with this initial candidate (cf. Bond et al., 2007; Carlson et al., 2006; Hagmayer & Kostopoulou, 2013). Consequently, the initially strongly supported candidate hypothesis was frequently favored over the equally supported contender. This finding of a primacy effect in diagnostic reasoning with an EoS procedure extends previous evidence on information distortion to problems with four candidate hypotheses.

The SbS procedure (Experiment 1B) with stepwise ratings of each hypothetical cause's probability of having caused the presented symptoms increased the proportion of participants who showed a stronger influence of late symptoms. This recency shift presumably reflects an attenuation of the primacy effect and limitations of working memory for early presented symptoms. For SbS procedures with four contending candidate hypotheses, Sprenger and Dougherty (2012, Exp. 1, 2) similarly showed a higher contribution by pieces of evidence that are presented late in a sequence both for hypothesis generation and for probability judgments. Differing from their procedure with ratings of focal hypotheses, the SbS procedure in the present experiments prompted ratings of all hypothetical causes after each symptom's presentation. Consequently, the procedure repeatedly put forward all candidate hypotheses for consideration, it delayed the presentation rate, and it increased the demands for retaining and rehearsing early symptoms. Thus, in our experiment, the SbS procedure attenuated the primacy effect while the stronger influence of late symptoms is probably due to the delayed presentation rate (cf. Lange, Thomas, & Davelaar, 2012; Lange et al., 2013) and the intermittent ratings that created recency effects in memory for presented symptoms.

In the following, further evidence on the propensity to distort information as an explanation of the primacy effect in diagnostic reasoning is provided. In two experiments, novel types of equivocal sequences were presented to confirm the primacy effect (EoS condition) and the recency shift (SbS condition). The sequences included consistent

Table 5. Experiment 1B: Means of belief ratings (exemplary for the sequences of AAB-items) of A- and B-hypotheses, related to the serial position of a symptom (*SD* in parentheses)

Sequence	First symptom	Second symptom	Third symptom	Fourth symptom
Ab-Ab-x-Ba	Ab	Ab	x	Ba
A-Hypothesis	76.41 (17.68)	82.19 (14.35)	41.44 (22.13)	43.39 (21.90)
B-Hypothesis	32.60 (15.05)	34.08 (15.87)	29.83 (13.19)	73.28 (20.33)
Ab-Ab-Ba-x	Ab	Ab	Ba	x
A-Hypothesis	78.38 (15.89)	82.29 (14.29)	41.04 (20.65)	40.65 (22.44)
B-Hypothesis	33.32 (15.41)	33.94 (16.39)	76.14 (18.36)	36.47 (17.07)
Ab-x-Ab-Ba	Ab	x	Ab	Ba
A-Hypothesis	81.13 (14.56)	39.62 (20.77)	78.59 (15.99)	42.09 (20.74)
B-Hypothesis	32.83 (16.03)	29.93 (13.52)	33.35 (17.23)	72.83 (21.19)

and inconsistent evidence. Encountering disconfirming inconsistent evidence is expected to induce a conflict in symptom integration encouraging the reevaluation of the current reasoning state, which might reduce the primacy effect (EoS condition). A recency shift is again expected in the SbS condition.

Experiments 2A and 2B – Two and Four Contending Hypotheses

Besides the equivocal item type AB, which supports two contending candidate hypotheses, Experiments 2A and 2B included items pointing to four candidate hypotheses. These items contain a Cd-symptom strongly suggesting C and weakly suggesting D either at the beginning (CAB) or at the end of the symptom sequence (ABC). The conflict induced by these symptoms that cannot be subsumed under a single explanation was expected to encourage a reevaluating analysis that counteracts biased reasoning. Thus, for the EoS response procedure (Experiment 2A), for ABC- and for CAB-items the primacy effect favoring A-diagnoses was expected to be reduced compared with AB-items. Although CAB-items start with a symptom strongly supporting C, no primacy effect favoring C-diagnoses was expected for CAB-items because of their inferior final support compared with A- and B-diagnoses. To our knowledge, comparable studies about EoS diagnostic reasoning with four contending candidate hypotheses and such mixed support do not exist.

Compared to the EoS condition, the primacy effect was expected to be reduced in the SbS condition for AB-items, and likewise for ABC- and CAB-items, provided a primacy effect favoring A-diagnoses would result in the EoS condition as expected. Furthermore, the SbS procedure should tip proportions of diagnoses toward the late supported candidate hypothesis (a recency effect favoring C-diagnoses in ABC-items compared to CAB-items and a shift increasing B-diagnoses in CAB-items compared to ABC-items) as already shown in Experiment 1B (AB-items).

Method

Participants

Seventy-nine (58 female; mean age 22.5, *SD* = 2.9) undergraduate students took part in the experiments. Experiment 2A was conducted at the University of Greifswald with 40 participants and Experiment 2B was conducted at the TU Chemnitz with 39 participants.

Design

Again, response mode was varied between-subjects (EoS in Experiment 2A vs. SbS in Experiment 2B). The within-subjects factor item type (CAB, AB, and ABC) varied the consistency of evidence (consistent: AB vs. mixed: CAB and ABC) and the position of the deviating symptom supporting two additional candidate hypotheses (early: CAB vs. late: ABC).

Material

Learning Material

The domain specific knowledge participants had to learn was identical to the learning material in Experiments 1A and 1B; for recapitulation, see Tables 1 and 2.

Experimental Material

Again, four symptoms were sequentially presented in each trial. As in the previous experiments, we refer to the chemical that shares specific symptoms with a contender but is the first to be strongly suggested as A-chemical or A-diagnosis and to the contender as B-chemical or B-diagnosis. Three item types (AB, CAB, ABC) encompassing three symptom sequences each were used (see Table 3). The sequences were either equivocal in the sense that presented symptoms were finally and equally consistent with

two of the four chemicals (AB), or they were equivocal with regard to A and B but additionally inconsistent in the sense that a Cd-symptom pointed to two further chemicals (C and D) that the Ab- and Ba-symptoms were not consistent with (CAB and ABC).

For example, an ABC-item could start with a *Skin*-symptom as an Ab-symptom putting W in the role of the A-diagnosis and K in the role of the B-diagnosis. Then eyelid swelling (category *Eyes*) would be a Cd-symptom because it is not consistent with both A and B but strongly suggests the C-diagnosis (chemical R) and weakly the D-diagnosis (chemical B). Note that the sequences for the item type AB (Table 3) differ from the sequences used in the previous experiments and that all three start with an unspecific symptom.

As in the previous experiments, symptoms were assigned to these sequences such that each of the four chemicals was placed in the role of the A-diagnosis once for each sequence. All resulting 36 combinations of A-chemicals and sequences were presented to each participant partitioned into four blocks. Each block contained all nine sequences and per item type within a block no chemical was repeated as the A-chemical. The order of sequences within a block was randomized and the order of the four blocks was counterbalanced. Again, at the level of a single trial the assignment of the actual symptoms was randomly drawn from the complete set of possible symptom combinations excluding repetitions of the same symptom.

Procedure

The procedure was the same as in the previous experiments. Participants acquired domain specific knowledge in two learning tasks. When the learning criteria were reached, they worked through four training trials and then four blocks of experimental trials. In the EoS condition (Experiment 2A), four symptoms were sequentially presented and subsequently a diagnosis and a confidence rating had to be given (left panel of Figure 1). Participants in the SbS condition (Experiment 2B) additionally were prompted after each symptom to rate for each chemical their belief in the respective chemical as the cause of the symptoms observed until then (right panel of Figure 1). The entire experiment lasted about 45 min.

Results

Diagnoses

Proportions of final diagnoses are depicted in Figure 4 (see ESM 4). For the item type AB, A-diagnoses were more frequent than B-diagnoses in the EoS condition but had only a slight advantage in the SbS condition. C- and D-diagnoses were chosen with rates of 9.4% and 6.0% in AB-items of Experiments 2A and 2B, respectively. As in the previous experiments, the analysis of item type AB was restricted

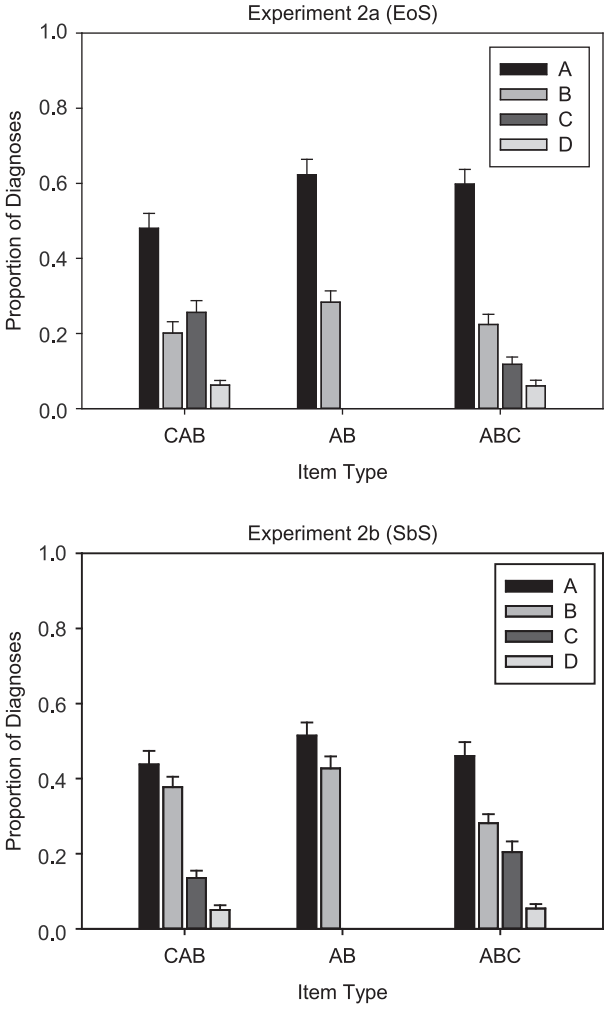

Figure 4. Mean proportions of diagnoses for each item type in Experiments 2A and 2B. Error bars show the standard error of the mean. Proportions for AB-items do not sum to one because proportions of wrong C and D diagnoses are not plotted.

to the A- and B-responses, and the proportion of A-responses was tested against 50%. The primacy effect for the EoS condition was confirmed, $t(39) = 4.61$, $p < .001$, $d = 0.70$, whereas the A-proportion did not differ significantly from 50% in the SbS condition, $t(38) = 0.91$, $p = .368$, $d = 0.15$. Again, a Bayesian one-sample t-test confirmed the latter finding. Using the default Cauchy prior width $r = 0.707$, the Bayesian one-sample t-test with $BF_{01} = 3.9$ (with 5.2×10^{-8} error %) indicated that the observed balance of A- and B-diagnoses in the SbS condition is nearly four times more likely under the null hypothesis.

Again, individual proportions of A-diagnoses for AB-items were examined to identify participants showing either balanced diagnoses (from 5 to 7 A-diagnoses out of 12 diagnoses for AB-items) or order effects. As in the previous experiments, the proportions of participants showing

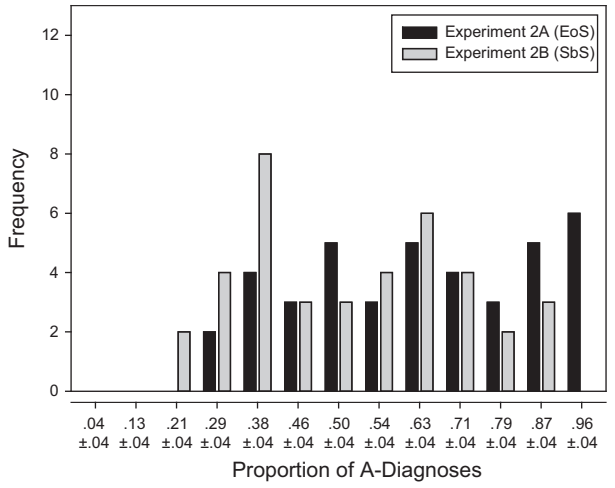

Figure 5. Frequency distribution of A-proportions in individual participants' final diagnoses for AB-items in Experiments 2A (EoS) and 2B (SbS).

balanced diagnoses were similar for the EoS (27.5%) and the SbS condition (25.6%). Furthermore, the remarkable shift in the SbS condition to the hypothesis supported more strongly by a later symptom was confirmed in the proportion of participants who chose predominantly B-diagnoses (35.9% vs. 15.0% for EoS). This shift toward lower A-proportions is also reflected in the frequency distributions plotted in Figure 5. A comparison of proportions' quartiles confirmed the difference in distributions with $\chi^2(3) = 11.35$, $p = .005$.

For the item types CAB and ABC in the EoS condition, A-diagnoses were also clearly more frequent than B-diagnoses. This advantage of A-diagnoses was decreased in the SbS condition, but more so for CAB-items than for ABC-items, which started with an Ab-symptom. As for AB-items, we tested for an order effect between the equally supported alternatives A and B by considering only A- and B-responses (together 100%) and testing the proportion of A-responses against 50%. The advantage of A was confirmed for CAB and ABC in the EoS condition with $t(39) = 4.58$, $p < .001$, $d = 0.72$ and $t(39) = 6.21$, $p < .001$, $d = 0.98$, respectively. Contrary to expectations, the primacy effect for ABC ($d = 0.98$) was larger than for AB ($d = 0.70$).

In the SbS condition, the A-proportion did not differ significantly from 50% for CAB-items, $t(38) = 0.25$, $p = .804$, $d = 0.04$. With a Bayes factor of $BF_{01} = 5.6$ (with 7.8×10^{-8} error %) the observed balance of A- and B-diagnoses is about 5.6 times more likely under the null hypothesis. For ABC-items a primacy effect favoring A over B was confirmed, $t(38) = 2.87$, $p = .007$, $d = 0.46$. Thus, compared with the EoS condition, the primacy effects in both items were reduced, as expected. However, contrary to expectations, the primacy effect for ABC ($d = 0.46$) was larger than for AB ($d = 0.15$).

Somewhat surprisingly, participants selected the C-diagnosis to considerable proportions for item types CAB and ABC, although it was clearly less supported than A or B (one vs. two specific symptoms). The 2 × 2 pattern of C-proportions (dark gray bars in Figure 4) reveals a primacy effect in the EoS condition (CAB higher than ABC) and a recency effect in the SbS condition (ABC higher than CAB). This interaction was confirmed in the corresponding 2 × 2 ANOVA including item type as a within-subjects variable and response condition as a between-subjects variable, with $F(1, 77) = 18.66$, $p < .001$, $\eta_p^2 = 0.20$. The main effects of item type ($p = .16$) and response condition ($p = .62$) were insignificant.

A stronger weight of late symptoms in the SbS condition was also revealed by the higher proportion of B-responses for item type CAB, in which the latest specific symptom points to B, compared with ABC, $t(38) = 3.02$, $p = .004$, $d = 0.63$, and, of course, by higher B-proportions for item types CAB and AB in the SbS- than in the EoS condition.

Confidence Ratings

Mean confidence ratings are depicted in Table 6 (see ESM 3). In each item type, each diagnosis was strongly supported by only one symptom. Accordingly, confidence ratings were more similar to those for AB-items rather than the higher ratings for AAB and ABB items in the previous experiments.

The consistent item type AB overall resulted in higher confidence ratings for A- and B-diagnoses than the inconsistent item types ABC and CAB: In the EoS condition, the confidence in A-diagnoses was higher for AB-items than for ABC- and CAB-items (with *d*s of 0.65 and 0.76, respectively). The confidence in B-diagnoses was higher for AB-items than for ABC-items ($d = 0.50$), but only slightly higher than for CAB-items ($d = 0.35$). In the SbS condition, the confidence in A-diagnoses for AB-items was higher than for ABC-items ($d = 0.35$), but not higher than

Table 6. Experiments 2A and 2B: Means of confidence ratings related to the diagnoses (*SD* in parentheses)

Experiment (response mode)	Item type	A	B	C	D
2A (End-of-sequence)	AB	3.80 (1.30)	3.53 (1.33)	2.33 (1.11)	
	ABC	3.32 (1.29)	3.14 (1.27)	3.13 (1.57)	2.59 (1.29)
	CAB	3.32 (1.26)	3.32 (1.23)	2.91 (1.24)	2.46 (1.26)
2B (Step-by-step)	AB	3.73 (1.33)	3.99 (1.18)	3.51 (1.67)	
	ABC	3.43 (1.18)	3.47 (1.17)	3.07 (1.65)	3.07 (1.53)
	CAB	3.69 (1.13)	3.43 (1.10)	2.98 (1.18)	2.87 (1.51)

for CAB-items ($p > .70$). The confidence in B-diagnoses for AB-items was higher than for ABC- and CAB-items (with *d*s of 0.54 and 0.76, respectively). The less frequent C-diagnoses, in general, received lower confidence ratings than A- and B-diagnoses in the inconsistent item types in line with their weaker support.

Belief Ratings

The sequential belief ratings in the SbS condition (Experiment 2B) as in Experiment 1B were not reasonable ratings of the current support status of each chemical considering all presented symptoms. To keep Experiment 2B as similar as possible to Experiment 1B, belief rating instructions were not changed. Again, the participants seemed to rate predominantly the support provided by the just presented symptom. Thus, the belief ratings did not conform with final diagnoses, cannot be interpreted as intended, and are not further reported.

Discussion

In the EoS condition, the primacy order effect was again obtained for AB-items and likewise for inconsistent ABC- and CAB-items. In the SbS condition, only ABC-items showed a primacy effect. In AB- and CAB-items, the late Ba-symptoms received more weight than in the EoS condition (recency shift), which increased the proportion of B-diagnoses to the detriment of the directly competing A-diagnoses. For ABC-items, the proportion of C-diagnoses was analogously increased by the late Cd-symptom (recency effect), however, this did not prevent a primacy effect favoring A-diagnoses because C was not a direct competitor of A.

Notably, the primacy effect for ABC-items was increased compared with AB-items in the EoS as well as in the SbS condition. Thus, we reject our assumption that in mixed sequences the conflict by inconsistent symptoms might evoke a deepened analysis counteracting order effects. Instead, the initially established A-hypothesis had a larger advantage over the B-hypothesis in ABC-items compared to AB-items. Presumably, a change of the initially leading hypothesis had a similar probability in ABC- and AB-items but could take the form of a switch to B or C in ABC-items. Thus, switches to C reduced the proportion of B-responses in ABC-items and consequently increased the advantage of A over B.

Another unexpected result was the considerable proportion of C-diagnoses given for CAB-items (EoS condition) although C was clearly less supported by specific symptoms than A and B. This suggests that the initial C-hypothesis in several trials was strong enough to distort evidence from opposing symptoms during subsequent symptom integration.

Despite a considerable proportion of C-diagnoses for CAB-items in the EoS condition, the A-diagnosis had a strong advantage over the B-diagnosis even with C triggered by the first symptom. Thus, our novel approach with four candidate hypotheses revealed that evidence can be distorted toward an earlier preferred hypothesis even if this hypothesis is not the initially triggered hypothesis. This was confirmed by the primacy effect obtained for the sequences starting with non-diagnostic symptoms (e.g., x-x-Ab-Ba).

The reported comparisons of response procedures were comparisons of results obtained in separate experiments at different laboratories and thus cannot exclude possible differences between settings and participant populations as alternative explanations of the observed effects of response procedures. Thus, to strengthen our conclusions, we conducted a further experiment at a single laboratory with random assignment of participants from the same population to the response procedures. In the following experiment, presenting the same sequences as in the Experiments 2A and 2B to confirm the primacy effect (EoS condition) and the recency shift (SbS condition), each participant either worked through the EoS or through the SbS condition.

Experiment 3 – Four Contending Hypotheses

Experiment 3 combined the Experiments 2A and 2B with the response procedure as a between-subjects manipulation (EoS vs. SbS). In the light of the results of Experiment 2A, with the EoS procedure, we expected a primacy effect favoring A-diagnoses for ABC-, for CAB-, and for AB-items. Although CAB-items start with a symptom strongly supporting C, a primacy effect favoring A-diagnoses should again be confirmed.

With the SbS procedure, we expected a reduced primacy effect for AB-items, and likewise for ABC- and CAB-items. Furthermore, the SbS procedure should tip proportions of diagnoses toward the candidate hypothesis supported strongly by a later symptom (a recency effect favoring C-diagnoses in ABC-items compared with CAB-items and a shift increasing B-diagnoses in CAB-items compared with ABC-items) as in Experiment 2B.

Method

Participants

Seventy-seven (48 female; mean age 23.3, $SD = 4.0$) undergraduate students took part in the experiment that was conducted at the TU Chemnitz.

Design

Response mode was varied between-subjects (EoS vs. SbS). The within-subjects factor item type varied the consistency

of evidence (consistent: AB vs. mixed: CAB and ABC) and the position of the deviating symptom supporting two additional candidate hypotheses (early: CAB vs. late: ABC).

Material

Learning Material

The domain specific knowledge participants had to learn was identical to the learning material in the foregoing experiments; for recapitulation, see Tables 1 and 2.

Experimental Material

As in the previous experiments, four symptoms were sequentially presented in each trial. The three item types from the Experiments 2A and 2B (AB, CAB, ABC) encompassing three sequences each were used again (see Table 3). The experimental material and its presentation were equal to materials and procedures in Experiments 2A and 2B.

Procedure

The learning procedure was the same as in the previous experiments. Participants acquired domain specific knowledge in two learning tasks. When the learning criteria were reached, they worked through four training trials and then four blocks of experimental trials. In the EoS condition, four symptoms were sequentially presented and subsequently a diagnosis and a confidence rating had to be given (left panel of Figure 1). Participants in the SbS condition additionally were prompted after each symptom to rate for each chemical their belief in the respective chemical as the cause of the symptoms observed until then (right panel of Figure 1). The entire experiment lasted about 60 min.

Results

Diagnoses

Proportions of diagnoses are depicted in Figure 6 (see ESM 6). For item type AB, A-diagnoses were more frequent than B-diagnoses in the EoS condition but not in the SbS condition. C- and D-diagnoses were chosen with rates of 3.0% and 4.4% for AB-items in the EoS condition and the SbS condition, respectively. As in the previous experiments, the analysis of item type AB was restricted to the A- and B-responses, and the proportion of A-responses was tested against 50%. The primacy effect for the EoS condition was confirmed, $t(38) = 4.89$, $p < .001$, $d = 0.78$, whereas the A-proportion did not differ significantly from 50% in the SbS condition, $t(37) = -0.01$, $p = .994$. A Bayesian one-sample t-test (Cauchy prior width $r = 0.707$) confirmed the latter finding with $BF_{01} = 5.7$

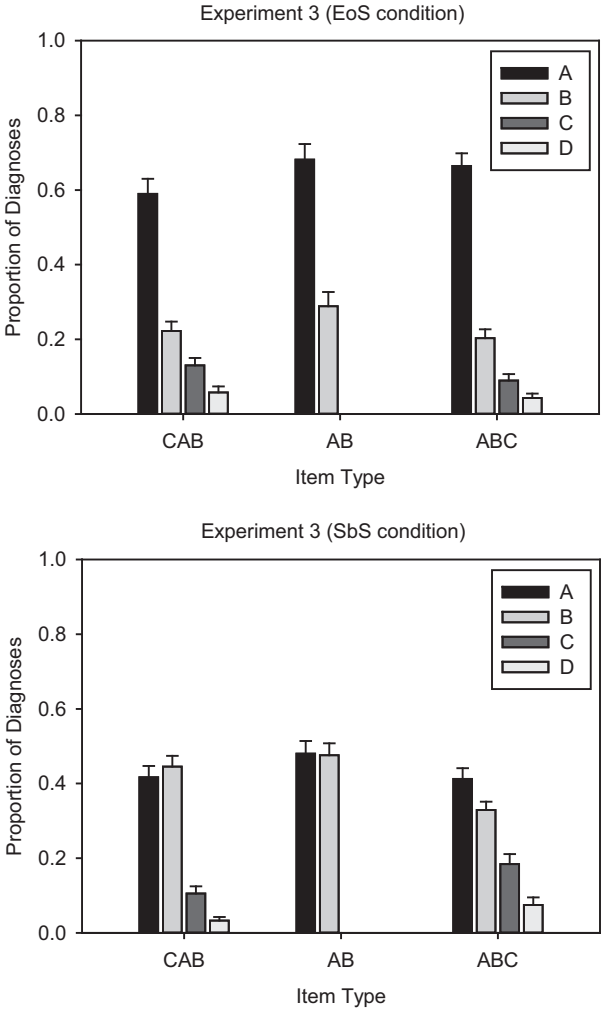

Figure 6. Mean proportions of diagnoses for each item type in Experiment 3. Error bars show the standard error of the mean. Proportions for AB-items do not sum to 1 because proportions of wrong C and D diagnoses are not plotted.

(with 6.0×10^{-8} error %), according to which the observed balance of A- and B-diagnoses in the SbS condition is about 5.7 times more likely under the null hypothesis.

Again, individual proportions of A-diagnoses for AB-items were examined to identify participants showing either balanced diagnoses (from 5 to 7 A-diagnoses out of 12 diagnoses for AB-items) or order effects. The proportion of participants showing balanced diagnoses this time was lower for the EoS (17.9%) than for the SbS condition (42.1%). However, the remarkable recency shift in the SbS condition was confirmed in the proportion of participants who chose predominantly B-diagnoses (31.6% vs. 15.4% for EoS). This shift toward lower A-proportions is also reflected in the frequency distributions plotted in Figure 7. Again, a comparison of proportions' quartiles additionally confirmed the difference in distributions, with $\chi^2(3) = 23.41$, $p < .001$.

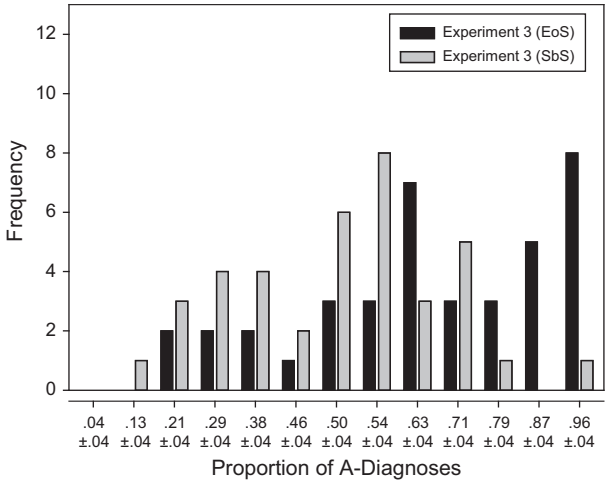

Figure 7. Frequency distribution of A-proportions in individual participants' final diagnoses for AB-items in the EoS condition and the SbS condition of Experiment 3.

For the item types CAB and ABC in the EoS condition, A-diagnoses were also clearly more frequent than B-diagnoses. As for AB-items, we tested for an order effect between the equally supported alternatives A and B by considering only A- and B-responses (together 100%) and testing the proportion of A-responses against 50%. The advantage of A was confirmed for CAB and ABC in the EoS condition with $t(38) = 5.91$, $p < .001$, $d = 0.95$ and $t(38) = 9.10$, $p < .001$, $d = 1.46$, respectively. Conforming the finding from Experiment 2A, the primacy effect for ABC ($d = 1.46$) was larger than for AB ($d = 0.78$).

In the SbS condition, again, the A-proportion did not differ significantly from 50% for CAB-items, $t(37) = -0.56$, $p = .581$, $d = -0.09$, but for ABC-items the expected primacy effect favoring A over B could not be confirmed this time, $t(37) = 1.81$, $p = .078$, $d = 0.29$. A Bayesian one-sided *t*-test, with a substantial Bayes factor of $BF_{01} = 5.0$ (with 4.9×10^{-8} error %), confirmed that the observed balance of A- and B-diagnoses in CAB-items is more likely under the null hypothesis. However, for ABC-items the available evidence was inconclusive with regard to a primacy effect: Neither a Bayes factor of $BF_{01} = 1.3$ (with 4.0×10^{-9} error %) in favor of the null hypothesis nor, given the results from Experiment 2B, a one-sided test for a primacy effect with a Bayes factor of $BF_{10} = 1.5$ (with 2.3×10^{-6} error %) provided conclusive evidence. Thus, compared with the EoS condition, the primacy effect was eliminated in CAB-items, as it was expected for AB-items, and was reduced or eliminated in ABC-items.

Again, participants selected the C-diagnosis to considerable proportions for item types CAB and ABC, although it was clearly less supported than A or B (one vs. two specific symptoms). The 2 × 2 pattern of C-proportions (dark gray bars in Figure 6) indicates a primacy effect in the EoS condition (CAB higher than ABC) and a recency effect in the SbS condition (ABC higher than CAB). This interaction was confirmed in the corresponding 2 × 2 ANOVA including item type as a within-subjects variable and response condition as a between-subjects variable, with $F(1, 75) = 12.61$, $p = .001$, $\eta_p^2 = 0.14$. The main effects of item type ($p = .26$) and response condition ($p = .16$) were insignificant.

The stronger weight of late symptoms in the SbS condition was additionally confirmed by the higher proportion of B-responses for item type CAB, in which the latest specific symptom points to B, compared with ABC, $t(37) = 3.26$, $p = .002$, $d = 0.99$, and by higher B-proportions for item types CAB and AB in the SbS- than in the EoS condition.

Confidence Ratings

Mean confidence ratings are depicted in Table 7 (see ESM 5).

The consistent item type AB overall resulted in higher confidence ratings for A- and B-diagnoses than the inconsistent item types ABC and CAB: In the EoS condition, the confidence in A-diagnoses was higher for AB-items than for ABC- and CAB-items (with *d*s of 0.80 and 0.89, respectively). Likewise, the confidence in B-diagnoses was higher for AB-items than for ABC- and for CAB-items (with $d = 0.66$ and $d = 0.49$, respectively). In the SbS condition, the confidence in A-diagnoses for AB-items was higher than for ABC-items ($d = 0.94$) and for CAB-items ($d = 0.56$). The confidence in B-diagnoses for AB-items was higher than for ABC- and CAB-items (with *d*s of 0.76 and 0.40, respectively). Ratings of less supported C-diagnoses were not clearly different from those of B-diagnoses in ABC-items, neither in the EoS condition ($d = -0.03$) nor in the SbS condition ($d = 0.27$). Regarding CAB-items, ratings of C-diagnoses were lower than those of B-diagnoses in the SbS condition ($d = -0.48$) and in the EoS condition ($d = -0.41$) as well.

Table 7. Experiment 3: Means of confidence ratings related to the diagnoses (*SD* in parentheses)

Response mode	Item type	A	B	C	D
End-of-sequence	AB	4.05 (1.22)	3.67 (1.13)	2.17 (1.00)	
	ABC	3.51 (0.98)	3.12 (1.03)	3.19 (1.44)	2.63 (1.15)
	CAB	3.44 (0.83)	3.23 (1.25)	2.63 (0.92)	2.24 (1.26)
Step-by-step	AB	3.91 (1.10)	3.53 (1.14)	2.54 (1.75)	
	ABC	3.03 (1.09)	2.93 (1.21)	3.07 (1.51)	2.24 (1.24)
	CAB	3.36 (1.19)	3.12 (1.14)	2.48 (1.38)	2.18 (1.65)

Belief Ratings

To keep Experiment 3 as similar as possible to Experiments 1B and 2B, belief rating instructions were not changed. Because the instructions did not prompt reasonable ratings of each chemical's status as in the previous experiments, these ratings are not further reported.

Discussion

The third experiment with a random assignment of participants to the response conditions confirmed the findings of the previous experiments. In the EoS condition, the primacy effect was again obtained for AB-items and for inconsistent ABC- and CAB-items. In contrast, in the SbS condition the primacy effect was attenuated and the late Ba-symptoms in AB- and CAB-items received additional weight, which additionally increased the proportion of B-diagnoses to the detriment of the primacy effect. Moreover, for ABC-items, the proportion of C-diagnoses was increased by the late Cd-symptom.

Furthermore, despite a considerable proportion of C-diagnoses for CAB-items in the EoS condition, the A-diagnosis had a strong advantage over the B-diagnosis even with C triggered by the first symptom. Thus, it was confirmed that evidence can be distorted toward an earlier preferred hypothesis even if this hypothesis is not the initially triggered hypothesis.

Again, the primacy effect for ABC-items was increased compared with AB-items for the EoS condition. Presumably, switches to C have reduced the proportion of B-responses in ABC-items and consequently increased the advantage of A over B.

General Discussion

In diagnostic reasoning, hypotheses explaining observed pieces of evidence are generated from memory and updated until a diagnosis is finally selected from these candidates. The order in which the pieces of evidence are encountered can influence their integration and in turn diagnostic judgments (Weber et al., 1993). The purpose of the reported experiments, in which symptom orders, response modes, and the consistency of symptom sequences were manipulated, was to illuminate reasoning processes that propagate into order effects. These effects were reflected in unequal proportions of diagnoses with equal support (A vs. B), and diagnoses with normatively insufficient support (C-diagnoses).

The main finding is a remarkably strong primacy effect in sequential diagnostic reasoning with four candidate hypotheses and equivocal symptom sequences. The effect was obtained for different sequences: A-diagnoses had an advantage in sequences beginning with a diagnostic symptom (e.g., Ab-x-Ba-x), in sequences beginning with a non-diagnostic symptom (e.g., x-x-Ab-Ba), and across consistent and mixed sequences. A conflict resulting from an inconsistent symptom did not reduce the primacy effect. The effect could only be counteracted partially by a stepwise rating procedure that repeatedly reminded of all hypothetical causes and weakened the memory representation of early relative to late symptoms.

On the one hand, the strong primacy effect in reasoning with multiple candidate hypotheses is in line with Hogarth & Einhorn's (1992) review that reports a primacy effect in 19 out of 27 studies (mostly single belief updating). On the other hand, our finding is opposed to earlier findings given a similar EoS procedure. For instance, Lange and colleagues (2012, 2013) repeatedly showed a recency effect, given a comparable presentation rate. Sprenger and Dougherty (2012) who noticed a weak primacy effect in one out of four of their EoS conditions did not provide evidence for order effects. Lange and colleagues and Sprenger and Dougherty used a different learning procedure. Their participants acquired diagnostic knowledge from a series of cases (learning from experience) whereas our participants studied verbally described causal relationships presented in a table (learning from description). Thus, the difference in findings could be attributed to a difference in learning procedures. However, a recent master's thesis that used the chemical accident paradigm and equivocal sequences replicated the primacy effect for learning from description as well as learning from experience (Gade, 2014). Hence, we presume that high ambiguity favors the primacy effect because it enables information distortion.

Equivocal sequences were purposefully designed to support multiple candidate hypotheses. First, presented symptoms were consistent with both A- and B-diagnoses and did not support one dramatically stronger than the other. Thus, either candidate hypothesis was a reasonable interpretation even of symptoms more strongly supporting the alternative. This ambiguity of each piece of evidence invited information distortion (Russo et al., 1998). Second, A- and B-hypotheses were finally equally supported in equivocal sequences. This equivocal support by a symptom sequence was the equivalent of similarly attractive options in a choice task, which encourage distortion processes (e.g., Tyszka, 1998).

Hence, we conclude that the observed primacy effect is likely the result of information distortion. The subjective diagnostic value of ambiguous symptoms was distorted to establish coherence with the initially generated and preferred hypothesis. Consequently, the initially stronger supported candidate hypothesis was preferred as the final diagnosis over an overall equally supported alternative hypothesis (Kostopoulou et al., 2012). This primacy effect due to the distortion of evidence is in line with the probabilistic constraint satisfaction model of information distortion in diagnostic reasoning (Hagmayer & Kostopoulou, 2013). Furthermore, this finding confirms a study of Jahn and Braatz (2014). They revealed by recording gaze behavior during reasoning how the interpretation of symptoms was biased in favor of the initially leading hypothesis.

An even stronger impact of the initially leading hypothesis is indicated by C-diagnoses for inconsistent CAB-items, which presented initially a single Cd-symptom strongly supporting candidate hypothesis C while

subsequent symptoms (Ab, Ba) jointly pointed to the candidate hypotheses A and B. Although support was stronger for A and B, there was a considerable proportion of C-diagnoses in the end-of-sequence condition. The initially leading hypothesis (C) was, thus, not necessarily reevaluated in the light of inconsistent evidence (Hendrick & Constantini, 1970). The inconsistency was sometimes maintained (cf. attitudinal ambivalence, Priester & Petty, 1996) and despite their stronger support by the symptoms Ab and Ba, the corresponding hypotheses (A and B) were sometimes suppressed in favor of the initially activated hypothesis (C). This did only occur if C was the initially strongly supported candidate hypothesis as reflected in the comparison with ABC-items.

Several alternative explanations for primacy effects in diagnostic reasoning have been proposed, but as we argue in the following, the strong primacy effect in our experiments is best explained by information distortion. According to the support accumulation model (Koehler, White, & Grondin, 2003), diagnostic reasoning can be limited to a focal hypothesis and alternative hypotheses can remain unspecified. However, in the present experiments, the set of contending hypotheses was specified and confidence ratings indicated awareness of alternatives. Symptoms in the end of a sequence can be less accessible and consequently less influential if they are inconsistent with the initial anchor hypothesis (e.g., Mussweiler & Strack, 1999). However, in AB-items the late symptom was consistent with both candidate hypotheses. Furthermore, our experiments exclude a biased search from the environment (e.g., Doherty, Mynatt, Tweney, & Schiavo, 1979) as an explanation of the primacy effect because information search and confirmatory testing (Klayman & Ha, 1987) were not possible. Participants' commitment to the initial hypothesis was possibly heightened because information was selectively exposed during the sequential symptom presentation (Jonas, Schulz-Hardt, Frey, & Thelen, 2001). Yet, this alone cannot explain the primacy effect toward a later triggered hypothesis (A-diagnosis) in CAB-items. Finally, participants could have attributed increased importance to the first symptom (e.g., drawing on a naïve law of primacy, Tulving, 2007). A pragmatic heuristic (e.g., Johnson & Raab, 2003) according to which the symptom mentioned first is of particular importance or predominant could have contributed to primacy order effects. However, this also provides no explanation for an advantage of A over B in the item type CAB that sets in with a symptom strongly suggesting C.

Rather than alternative explanations of primacy effects, the present results imply biased symptom processing that propagates into primacy effects corroborating postulates about information distortion in coherence-based reasoning (e.g., Kostopoulou et al., 2009). Biased processing in sequential reasoning is also consistent with further theories postulating a tendency toward coherence, for instance, construction-integration theory (Kintsch, 1998), which was originally developed in the context of text comprehension and includes mechanisms of parallel-constraint satisfaction (Kunda & Thagard, 1996). Construction-integration theory was adapted for the online integration of medical symptoms and diagnostic hypotheses (Arocha & Patel, 1995). Because an early supported hypothesis frames the integration of later symptoms, this leading hypothesis can remain the highest activated hypothesis that is more likely selected as the final diagnosis. The theory, however, lacks predictions regarding response procedures or the consistency of evidence.

A recent extension (Lange et al., 2013) of the computational model of hypothesis generation HyGene (Thomas et al., 2008) tries to account for order effects in sequential reasoning by mere temporal dynamics of symptom memory. The modeled dynamics of symptom memory (adapted from the context activation model of list recall, Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005) can be tuned to generate recency as well as primacy effects. However, as noted by the authors and as indicated by the present results, not only the retention of observed symptoms should be modeled but also the memory demands and dynamics of retaining and updating multiple hypotheses while processing symptoms. Like Sprenger and Dougherty (2012), we expect that a trade-off between the maintenance of evidence and the maintenance of hypotheses is required. We presume that in addition a mechanism distorting ambiguous evidence in favor of strongly activated hypotheses would account for a strong primacy effect as we showed it.

One should note that the task was designed to increase the likelihood of order effects by the forced choice procedure. Our participants had to select a diagnosis even if they were aware of alternatives and – if possible – would have postponed their decision to request, search, or generate more information. The expectation of forced choice increases the propensity to distort evidence (cf. Brownstein, 2003). This propensity possibly may not manifest with a response option or a rating procedure allowing participants to express awareness of equal support (e.g., Sprenger & Dougherty, 2012). Yet, there are real-world scenarios that demand choices based on limited information similar to the present experiments and provoke order effects, for instance, diagnoses of experienced family practitioners (Chapman et al., 1996). Indeed, information distortion is a remarkably robust phenomenon that occurs in real-world choices (Carlson & Pearo, 2004).

The SbS procedure was shown to partially counteract the primacy effect. The repeated belief ratings promoted the saliency of four contending candidate hypotheses with the intended effect of encouraging their exhaustive consideration for reasoning. Furthermore, the increased demand on retention and rehearsal of the early symptoms due to the delayed presentation and the rating activities probably attenuated the primacy effect additionally. Thus, the SbS procedure highlighted alternative candidate hypotheses and attenuated biased diagnostic reasoning.

The exhaustive belief ratings did not work as a measure of parallel belief updating. Yet, the changing status of multiple hypotheses during sequential symptom processing still would be of considerable interest. This also applies to the EoS conditions, but by definition the status of hypotheses in EoS conditions cannot be inquired via belief ratings. Recent suggestions for indirect process tracing include probe reactions to infer memory activation from response times (Mehlhorn et al., 2011; Rebitschek, Krems, &

Jahn, 2015) and tracking eye movements between locations linked to hypothetical causes and symptoms (Jahn & Braatz, 2014). The latter has yielded converging evidence for primacy effects and biased symptom processing.

In summary, our results extend the research on order effects to reasoning with multiple hypotheses and revealed confirmatory propensities of distorting information as well as capacity-based limitations. Both presumably affect diagnostic reasoning with multiple hypothetical causes even in real-world tasks. Experience does not prevent order effects (Baumann, Krems, & Ritter, 2010) and the risk of biased processing increases with the importance of a task, with the temporal proximity to the decision, and its difficulty (Brownstein, 2003). Particularly diagnostic judgments of equivocal cases are prone to order effects even if alternatives are individually considered.

Acknowledgments

This research was supported by German Research Foundation (DFG) Grants JA 1761/7-1 and KR 1057/17-1. Parts of the data were presented at the 34th Annual Meeting of the Cognitive Science Society (August 2012, Sapporo, Japan).

Electronic Supplementary Material

The electronic supplementary material is available with the online version of the article at http://dx.doi.org/10.1027/1618-3169/a000298

ESM 1. Sav file.
Confidence ratings of Experiment 1.
ESM 2. Sav file.
Diagnosis Experiment 1.
ESM 3. Sav file.
Confidence ratings of Experiment 2.
ESM 4. Sav file.
Diagnosis Experiment 2.
ESM 5. Sav file.
Confidence ratings of Experiment 3.
ESM 6. Sav file.
Diagnosis Experiment 3.

References

Anderson, C. A., & Sechler, E. S. (1986). Effects of explanation and counterexplanation on the development and use of social theories. *Journal of Personality and Social Psychology, 50*, 24–34. doi: 10.1037/0022-3514.50.1.24

Anderson, N. H. (1981). *Foundations of information integration theory.* New York, NY: Academic Press.

Anderson, N. H., & Hubert, S. (1963). Effects of concomitant verbal recall on order effects in personality impression formation. *Journal of Verbal Learning and Behavior, 2*, 379–391. doi: 10.1016/S0022-5371(63)80039-0

Arocha, J. F., & Patel, V. L. (1995). Construction-integration theory and clinical reasoning. In C. A. Weaver III, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 359–381). Hillsdale, NJ: Erlbaum.

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology, 41*, 258–290. doi: 10.1037/h0055756

Barbey, A. K., & Barsalou, L. W. (2009). Reasoning and problem solving: Models. In L. Squire (Ed.), *Encyclopedia of Neuroscience* (Vol. 8, pp. 35–43). Oxford, UK: Academic Press.

Baumann, M., Krems, J., & Ritter, F. (2010). Learning from examples does not prevent order effects in belief revision. *Thinking & Reasoning, 16*, 98–130. doi: 10.1080/13546783.2010.484211

Baumann, M., Mehlhorn, K., & Bocklisch, F. (2007). The activation of hypotheses during abductive reasoning. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 803–808). Austin, TX: Cognitive Science Society.

Bergus, G. R., Chapman, G. B., Levy, B. T., Ely, J. W., & Oppliger, R. A. (1998). Clinical diagnosis and the order of information. *Medical Decision Making, 18*, 412–417. doi: 10.1177/0272989X9801800409

Bodenhausen, G. V., & Lichtenstein, M. (1987). Social stereotypes and information-processing strategies: The impact of task complexity. *Journal of Personality and Social Psychology, 52*, 871–880. doi: 10.1037/0022-3514.52.5.871

Bond, S. D., Carlson, K. A., Meloy, M. G., Russo, J. E., & Tanner, R. J. (2007). Information distortion in the evaluation of a single option. *Organizational Behavior and Human Decision Processes, 102*, 240–254. doi: 10.1016/j.obhdp.2006.04.009

Brownstein, A. (2003). Biased predecision processing. *Psychological Bulletin, 129*, 545–568. doi: 10.1037/0033-2909.129.4.545

Bylander, T., Allemang, D., Tanner, M., & Josephson, J. R. (1991). The computational complexity of abduction. *Artificial Intelligence, 49*, 25–60. doi: 10.1016/0004-3702(91)90005-5

Carlson, K. A., Meloy, M. G., & Russo, J. E. (2006). Leader-driven primacy: Using attribute order to affect consumer choice. *Journal of Consumer Research, 32*, 513–518. doi: 10.1086/500481

Carlson, K. A., & Pearo, L. K. (2004). Limiting predecisional distortion by prior valuation of attribute components. *Organizational Behavior and Human Decision Processes, 94*, 48–59. doi: 10.1016/j.obhdp.2004.02.001

Chapman, G. P., Bergus, G. R., & Elstein, A. S. (1996). Order of information affects clinical judgment. *Journal of Behavioral Decision Making, 9*, 201–211. doi: 10.1002/(SICI)1099-0771(199609)9:3<201::AID-BDM229>3.0.CO;2-J

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research, 63*, 1–49. doi: 10.3102/00346543063001001

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

Cunnington, J. P. W., Turnbull, J., Regehr, G., Marriott, G., & Norman, G. R. (1997). The effect of presentation order in clinical decision making. *Academic Medicine, 72*, S40–S42. doi: 10.1097/00001888-199710001-00014

Curley, S. P., Young, M. J., Kingry, M. J., & Yates, J. F. (1988). Primacy effects in clinical judgments of contingency. *Medical Decision Making, 8*, 216–222. doi: 10.1177/0272989X8800800310

Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review, 112*, 3–42. doi: 10.1037/0033-295X.112.1.3

Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. D. (1979). Pseudodiagnosticity. *Acta Psychologica, 43*, 111–121. doi: 10.1016/0001-6918(79)90017-9

Gade, E.-M. (2014). *The effect of the way how knowledge is acquired on sequential diagnostic reasoning* [Der Effekt der Form des Wissenserwerbs auf sequenzielles diagnostisches Schließen], (Unpublished diploma thesis, master's level). University of Greifswald, Germany.

Hagmayer, Y., & Kostopoulou, O. (2013). A probabilistic 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, TX: Cognitive Science Society.

Hastie, R., & Kumar, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology, 37*, 25–38. doi: 10.1037/0022-3514.37.1.25

Hendrick, C., & Constantini, A. F. (1970). Effects of varying trait inconsistency and response requirements on the primacy effect in impression formation. *Journal of Personality and Social Psychology, 15*, 158–164. doi: 10.1037/h0029203

Hirt, E. R., & Markman, K. D. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of Personality and Social Psychology, 69*, 1069–1086. doi: 10.1037/0022-3514.69.6.1069

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

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

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving, 7*, 2–9. doi: 10.7771/1932-6246.1167

Johnson, J. G., & Raab, M. (2003). Take the first: Option-generation and resulting choices. *Organizational Behavior and Human Decision Processes, 91*, 215–229. doi: 10.1016/S0749-5978(03)00027-X

Johnson, T., & Krems, J. F. (2001). Use of current explanations in multicausal abductive reasoning. *Cognitive Science, 25*, 903–939. doi: 10.1207/s15516709cog2506_2

Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology, 80*, 557–571. doi: 10.1037/0022-3514.80.4.557

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

Klaczynski, P. A. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: A two-process approach to adolescent cognition. *Child Development, 71*, 1347–1366. doi: 10.1111/1467-8624.00232

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211–228. doi: 10.1037/0033-295X.94.2.211

Koehler, D. J., White, C. M., & Grondin, R. (2003). An evidential support accumulation model of subjective probability. *Cognitive Psychology, 46*, 152–197. doi: 10.1016/S0010-0285(02)00515-7

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 107–118. doi: 10.1037/0278-7393.6.2.107

Kostopoulou, O., Mousoulis, C., & Delaney, B. C. (2009). Information search and information distortion in the diagnosis of an ambiguous presentation. *Judgment and Decision Making, 4*, 408–418.

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

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel constraint satisfaction theory. *Psychological Review, 103*, 284–308. doi: 10.1037/0033-295X.103.2.284

Lange, N. D., Thomas, R. P., Buttaccio, D. R., Illingworth, D. A., & Davelaar, E. J. (2013). Working memory dynamics bias the generation of beliefs: The influence of data presentation rate on hypothesis generation. *Psychonomic Bulletin & Review, 20*, 171–176. doi: 10.3758/s13423-012-0316-9

Lange, N. D., Thomas, R. P., & Davelaar, E. J. (2012). Temporal dynamics of hypothesis generation: Influences of data serial order, data consistency, and elicitation timing. *Frontiers in Psychology: Cognitive Science, 3*, 215. doi: 10.3389/fpsyg.2012.00215

Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on Psychological Science, 4*, 390–398. doi: 10.1111/j.1745-6924.2009.01144.x

Lord, C., Lepper, M., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology, 47*, 1231–1243. doi: 10.1037/0022-3514.47.6.1231

Love, J., Selker, R., Verhagen, A. J., Marsman, M., Gronau, Q. F., Jamil, T., … Wagenmakers, E.-J. (2015). JASP (Version 0.6) [Computer Software]. *The JASP Project*. Retrieved from https://jasp-stats.org/download

Maheswaran, D., & Chaiken, S. (1991). Promoting systematic processing in low-motivation settings: Effect of incongruent information on processing and judgment. *Journal of Personality and Social Psychology, 61*, 13–25. doi: 10.1037/0022-3514.61.1.13

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

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. 1391–1411). Austin, TX: 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., Romeijn, J.-W., & Rouder, J. N. (2015). *The philosophy of Bayes factors and the quantification of statistical evidence*. Manuscript submitted for publication. Retrieved from http://pcl.missouri.edu/sites/default/files/paper.pdf

Morey, R. D., & Rouder, J. N. (2014). *BayesFactor 0.9.6*. Comprehensive R Archive Network. Retrieved from http://cran.r-project.org/web/packages/BayesFactor/index.html

Mumma, G. T., & Wilson, S. B. (1995). Procedural debiasing of primacy/anchoring effects in clinical-like judgments.

Journal of Clinical Psychology, 51, 841–853. doi: 10.1002/1097-4679(199511)51:6%3C841::AID-JCLP2270510617%3E3.0.CO;2-K

Mussweiler, T., & Strack, F. (1999). Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology, 35*, 136–164. doi: 10.1006/jesp.1998.1364

Neuberg, S. L., & Fiske, S. T. (1987). Motivational influences on impression formation: Outcome dependency, accuracy-driven attention, and individuating processes. *Journal of Personality and Social Psychology, 53*, 431–444. doi: 10.1037/0022-3514.53.3.431

Patel, V. L., Arocha, J. F., & Zhang, J. (2005). Thinking and reasoning in medicine. In K. J. Holyoak (Ed.), *Cambridge handbook of thinking and reasoning* (pp. 727–750). Cambridge, UK: Cambridge University Press.

Peterson, C. R., & DuCharme, W. M. (1967). A primacy effect in subjective probability revision. *Journal of Experimental Psychology, 73*, 61–65. doi: 10.1037/h0024139

Priester, J. R., & Petty, R. E. (1996). The gradual threshold model of ambivalence: Relating the positive and negative bases of attitudes to subjective ambivalence. *Journal of Personality and Social Psychology, 71*, 431–449. doi: 10.1037/0022-3514.71.3.431

Rebitschek, F. G., Krems, J. F., & Jahn, G. (2015). Memory activation of multiple hypotheses in sequential diagnostic reasoning. *Journal of Cognitive Psychology*, 1–17. (ahead-of-print).

Russo, J. E., Carlson, K. A., Meloy, M. G., & Yong, K. (2008). The goal of consistency as a cause of information distortion. *Journal of Experimental Psychology: General, 137*, 456–470. doi: 10.1037/a0012786

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

Russo, J. E., Meloy, M. G., & Medvec, V. H. (1998). Predecisional distortion of product information. *Journal of Marketing Research, 35*, 438–452. doi: 10.2307/3152163

Shanteau, J. C. (1972). Descriptive versus normative models of sequential inference judgment. *Journal of Experimental Psychology, 93*, 63–83. doi: 10.1037/h0032509

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

Sprenger, A., & Dougherty, M. R. P. (2012). Generating and evaluating options for decision making: The impact of sequentially presented evidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 550–575. doi: 10.1037/a0026036

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

Tulving, E. (2007). On the law of primacy. In M. A. Gluck, J. R. Anderson, & S. M. Kosslyn (Eds.), *Memory and Mind: A Festschrift for Gordon H. Bower* (pp. 31–48). Hillsdale, NJ: Erlbaum.

Tyszka, T. (1998). Two pairs of conflicting motives in decision making. *Organizational Behavior and Human Decision Processes, 74*, 189–211. doi: 10.1006/obhd.1998.2775

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

Wyer, R. S., & Gordon, S. E. (1982). The recall of information about persons and groups. *Journal of Experimental Social Psychology, 18*, 128–164. doi: 10.1016/0022-1031(82)90047-6

Wyer, R. S., & Srull, T. K. (1986). Human cognition in its social context. *Psychological Review, 93*, 322–359. doi: 10.1037/0033-295X.93.3.322

Wyer, R. S., & Srull, T. K. (1989). *Memory and cognition in its social context*. Hillsdale, NJ: Erlbaum.

Yates, J. F., & Curley, S. P. (1986). Contingency judgment: Primacy effects and attention decrement. *Acta Psychologica, 62*, 293–302. doi: 10.1016/0001-6918(86)90092-2

Received March 6, 2014
Revision received March 23, 2015
Accepted March 23, 2015
Published online June 30, 2015

Felix Rebitschek

Harding Center for Risk Literacy
Max Planck Institute for Human Development
14195 Berlin
Germany
Tel. +49 176 244-36511
E-mail rebitschek@mpib-berlin.mpg.de