Supplemental Material

Diagnostic Response

Coherence maximization can be demonstrated by testing ambiguous symptom sequences with varying support for competing hypotheses. However, another kind of ambiguity was introduced in the tested sequences that allows studying effects of *symptom diversity*.

If two alternatives are supported by the same number of symptoms, the supporting symptoms for one hypothesis can be more *diagnostic*, whereas for the other hypothesis, they can be more *diverse*. Symptoms are maximally diagnostic for a hypothesized cause if they are only linked to this cause. For instance, if "eye symptoms" support only one out of several possible hypotheses, eye symptoms (e.g., "eyelid swelling" and "lacrimation") are maximally diagnostic. These symptoms, however, are not diverse. Two symptoms are diverse if they are derived from two different symptom classes linked to a hypothesis (e.g., "lacrimation" and "cough" derived from the symptom classes "eyes" and "respiration"). Diverse symptoms can provide stronger support than non-diverse symptoms. This has been explained by similarity coverage (Osherson, Smith, Wilkie, López, & Shafir, 1990) and causal diversity (Kim & Keil, 2003; Kim, Yopchick, & de Kwaadsteniet, 2008). Ambiguous symptom sets consisting of equal numbers of maximally diagnostic but non-diverse symptoms supporting one hypothesis, and less diagnostic but diverse symptoms supporting another hypothesis, thus pit diagnosticity against diversity.

We assumed that hypotheses supported by diverse symptoms derived from two symptom classes may have been selected to a greater degree than hypotheses supported by symptoms from one highly diagnostic symptom class.

Additionally, we assumed that the same set of symptoms presented in different orders may elicit differing final diagnoses (Baumann et al., 2010; Bergus, Chapman, Levy, Ely, & Oppliger, 1998; Hogarth & Einhorn, 1992; Trueblood & Busemeyer, 2011).

Figure S1 shows proportions of diagnostic responses for all 16 sequences. Figure S1.A, S1.B, and S1.D show the proportions of A- and B-responses for ten sequences with two contending hypotheses. Figure S1.C and S1.E show response proportions for six sequences with three contending hypotheses.

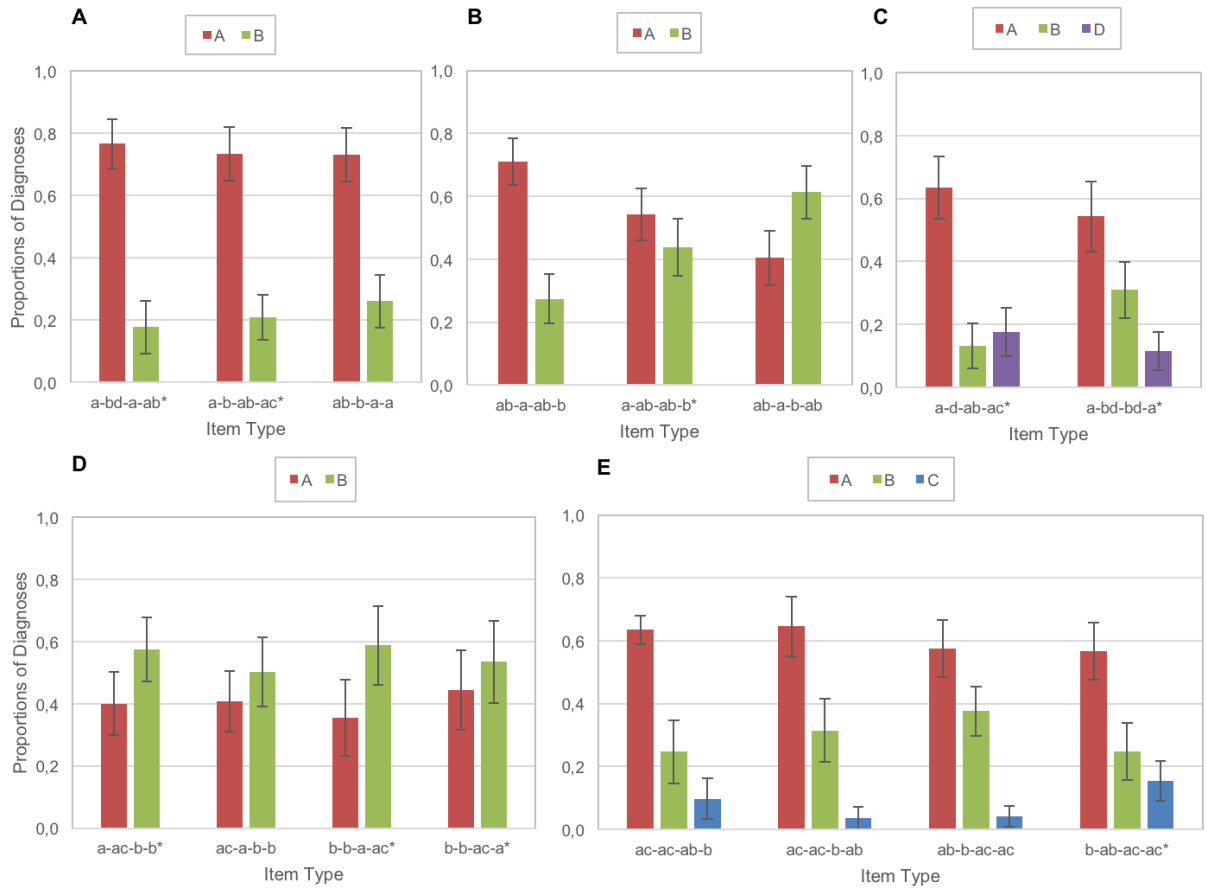

Figure S1. Mean proportions of diagnoses for each symptom sequence with either two or three contending hypotheses. (A) Sequences with *superior support for the A-diagnosis*. (B) Sequences with *equal support for A and B* and the same set of symptoms; only the symptom order is varied. (C) Sequences with *three contending hypotheses* and inconsistent symptoms supporting the D-diagnosis (d or twice bd). (D) Sequences that pit *diagnosticity against diversity* and consist of the same set of symptoms; diverse support for A and maximally diagnostic support for B. (E). Sequences with *three contending hypotheses*; three symptoms supporting the A-diagnosis; all of which also support an alternative diagnosis (B and C). Error

bars represent within-subjects 95% CIs. Sequences marked with a star have been presented in the paper (diagnostic responses and gaze data).

When A was the only hypothesis supported by three symptoms (Figure S1.A and S1.E), the A-response proportion was the highest. Unsurprisingly people most frequently chose the hypothesis that received the most support.

For the sequence ab-a-b-ab (Figure S1.B), in which the b-symptom occurred before the third symptom supporting A, the B-response proportion was higher than the A-response proportion. This reflects a strong order effect across the sequences in Figure S1.B because all symptoms provide equal support for A and B (maximally ambiguous) and differ only in symptom order.

If two or three hypotheses were supported by two symptoms each (Figure S1.D, and right sequence in Figure S1.C) participants more often chose the hypothesis supported by two maximally diagnostic symptoms from the same symptom class (b and b or a and a) rather than selecting a competing hypothesis supported by diverse symptoms that (in part or both) were associated with two chemicals (a and ac or bd and bd): In all four symptom sequences consisting of the symptoms ac, a, b, and b (Figure S1.D), participants chose B as opposed to A on a slightly more frequent basis. In the sequence a-bd-bd-a (Figure S1.C) participants chose A more often than B or D. Diagnosticity was thus often evaluated as stronger than diversity.

In our study, response proportions indicated that participants opted for responses supported by diagnosticity slightly more often, and gaze behavior showed that a second symptom from the same symptom class (b after b) was viewed as additional evidence. This response tendency cannot be compared to a normative model because critical information for computing a normative decision (causal structure, causal strengths, base rates, alternative causation, and presence of unstated symptoms) was implicit and unspecified. It is quite likely

that participants would rather choose the diversely supported hypothesis if critical information favoring diversity had been provided (Rebitschek, Krems, & Jahn, 2016).

Memory Indexing Gaze Data

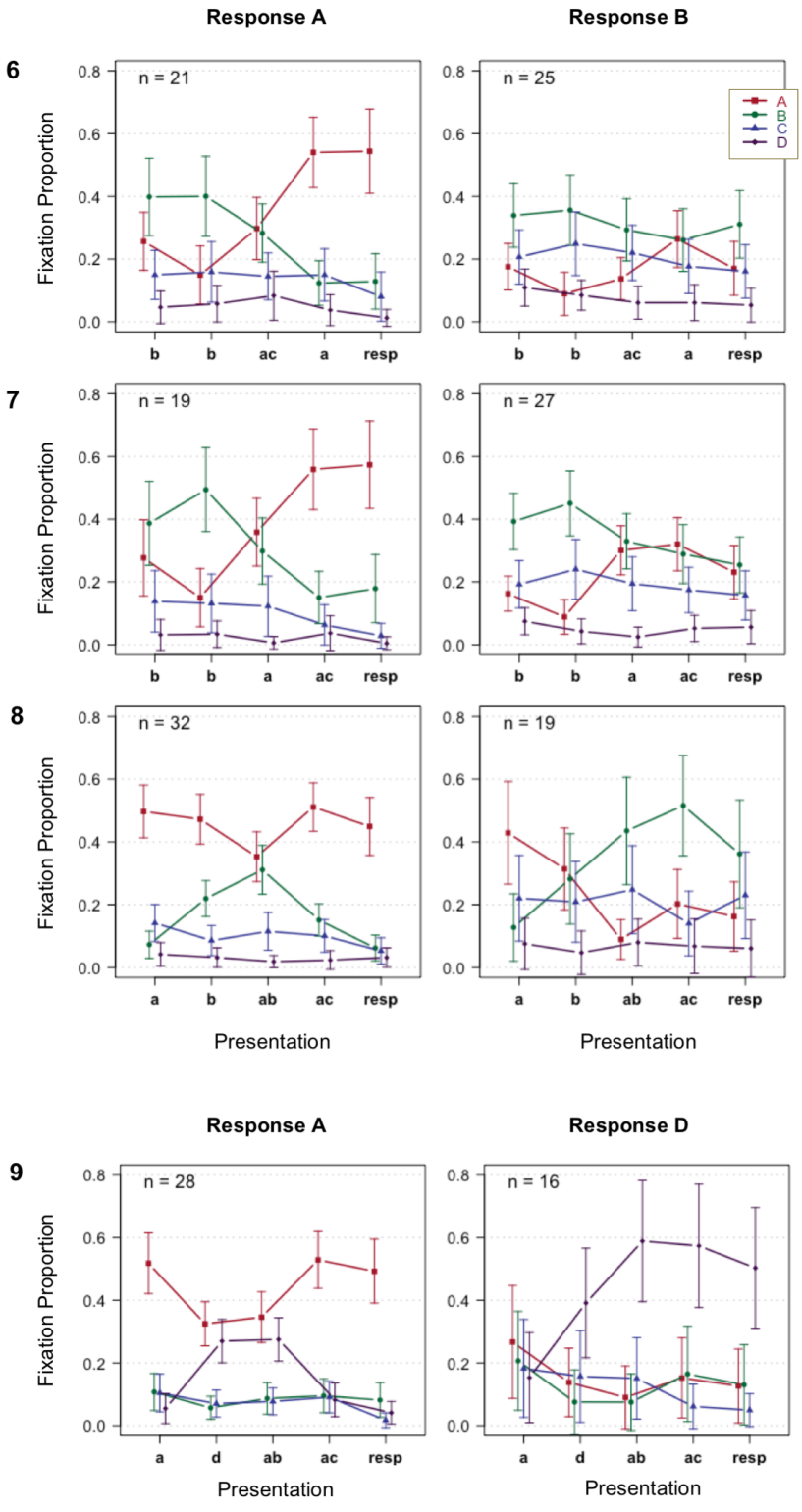

Figure S2. Mean proportion of fixation times in each interval that fell upon the A-, B-, C-, or D-quadrants for four ambiguous symptom sequences with two contending hypotheses (for sequences 6, 7, and 8: A-responses left column, B-responses right column, for sequence 9: A-

responses left column, D-responses right column). The number of participants shows how many participants responded at least once with the A-, B-, or D-response. X-axis labels show the five symptom intervals with the respective symptoms. Error bars represent within-subjects 95% CIs.

Literature

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

Kim, N. S., & Keil, F. C. (2003). From symptoms to causes: Diversity effects in diagnostic reasoning. *Memory & Cognition*, *31*, 155–165. doi: 10.3758/BF03196090

Kim, N. S., Yopchick, J. E., & de Kwaadsteniet, L. (2008). Causal diversity effects in information seeking. *Psychonomic Bulletin & Review*, *15*, 81–88. doi:10.3758/PBR.15.1.81

Osherson, D., Smith, E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, *97*, 185–200. doi: 10.1037/0033-295X.97.2.185

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

Trueblood, J. S., & Busemeyer, J. R. (2011). A quantum probability account of order effects in inference. *Cognitive Science*, *35*, 1518–1552. doi:10.1111/j.1551-6709.2011.01197.x