Baumann, M. R. K., Krems, J. F., & Ritter, F. E. (2010). Learning from examples does not prevent order effects in belief revision. *Thinking & Reasoning. 16*(2). 98-130.

Short title: Experience Does Not Prevent Order Effects

Learning from Examples Does Not Prevent Order Effects in Belief Revision

Martin R.K. Baumann, Josef F. Krems, and Frank E. Ritter

31 march 2010

Contact information:

Dr. Martin Baumann

Institute of Transportation Systems

German Aerospace Center / Deutsches Zentrum für Luft- und Raumfahrt e.V.

Lilienthalplatz 7

38108 Braunschweig

Germany

Tel.: +49 (0)531 - 295-3429   Fax: +49 (0)531 - 295-3402

Email: martin.baumann@dlr.de

Word count: 12743

Abstract

A common finding is that information order influences belief revision (e.g., Hogarth & Einhorn, 1992). We tested personal experience as a possible mitigator. In three experiments participants experienced the probabilistic relationship between pieces of information and object category through a series of trials where they assigned objects (planes) into one of two possible categories (hostile or commercial) given two sequentially presented pieces of probabilistic information (route and ID), and then they had to indicate their belief about the object category before feedback. The results generally confirm the predictions from the Hogarth and Einhorn model. Participants showed a recency effect in their belief revision. Extending previous model evaluations the results indicate that the model predictions also hold for classification decisions, and for pieces of information that vary in their diagnostic values. Personal experience does not appear to prevent order effects in classification decisions based on sequentially presented pieces of information and in belief revision.

Key words: belief revision; opinion revision; order effects; learning from examples; probability learning;

Acknowledgements

We thank Ben Newell, Hongbin Wang, Anke Popken, Frank Renkewitz, and two anonymous reviewers for their valuable comments and helpful suggestions on previous versions of this manuscript.

On July, 3, 1988, Captain Rogers had to make a far reaching decision within a few minutes. He had to decide if an aircraft approaching the USS Vincennes, the ship under his command in the Persian Gulf, was hostile or friendly. He had to form his decision on the grounds of just a few pieces of evidence, including altitude, route, and response on warning. As we all know, Captain Rogers made the wrong decision. He shot down the aircraft, which turned out to be a commercial flight (see Thagard, 1992, p. 138ff for more details). From a cognitive point of view Captain Rogers had to evaluate different hypotheses about the plane's nature and intention in the light of available and newly incoming evidence. He was engaged in a process called belief revision (Wang, 1993).

A key feature of belief revision is its sequential nature. In many cases, like in Captain Rogers' situation, information is received a piece at a time and has to be integrated into an evolving impression. This is very common in many situations of diagnostic reasoning. For example, in medical diagnosis, where symptoms are normally analyzed piece after piece by a physician; also in technical domains like troubleshooting, information regarding a fault is analyzed through a sequence of pieces of evidence to locate the cause of a malfunction. And debugging computer programs—another domain where diagnostic reasoning plays an important role—normally is a sequential process, where new information becomes available step-by-step.

In classification tasks, therefore, the modification of an individual belief, whether an object belongs to a certain category, can be viewed as a process of evidence integration. In this integration, the current belief level is adjusted as a result of analyzing a series of subsequent pieces of information.

Order effects in belief revision

A major research question arises given the sequential nature of belief revision: to what degree does the sequence itself, the manner in which the single pieces of evidence follow each other, influence how people adjust their belief strength? This would be the case if the same evidence has different effects on belief strength depending on it being processed at the beginning, the middle, or the end of a sequence. From a normative point of view (e.g., Bayesian) it should not matter if a piece of evidence A is processed and then evidence B, or the other way round. However, people seem to tend to adjust their beliefs according to the order of data presentation. This is called an order effect according to definitions by Hogarth and Einhorn (1992) and others (Langley, 1995; Ritter, Nerb, Lehtinen, & O'Shea, 2007).

Order effects with regard to belief revision have been found in many areas (see Ritter et al., 2007 for an overview) including impression formation (Asch, 1946), attitude modification (Friedrich & Smith, 1998), deductive reasoning (Johnson-Laird & Steedman, 1978), causal inference (Hogarth & Einhorn, 1992), and diagnostic reasoning (Johnson & Krems, 2001). Where order effects occur, reasoners are paying more attention to the order of the information in a sequence than to the combined content of the information in the sequence. Whereas this might be appropriate when the order of information presentation conveys relevant information in itself, such as when there is reason to believe that the first pieces of information are already outdated when the last ones are perceived, it is not appropriate and represents a cognitive bias when this order conveys no relevant information, such as when a reasoner has to classify an object or person according to the features of an object or person (e.g., a patient), and the order with which these features are considered is determined arbitrarily by the reasoner. Meanwhile, several theoretical models have been proposed to explain why order effects occur or why they do not occur. On one hand, belief revision is described as an on-line anchoring-and-adjustment process in which the current belief, the anchor, is adjusted by new evidence. Hogarth and Einhorn´s (1992) model is the most prominent example for this theoretical position. On the other hand, there are experience-based models explaining biases in decision making by memory-based processes, such as UECHO proposed by Wang, Johnson, and Zhang (2006).

Hogarth and Einhorn's (1992) model of belief revision

Hogarth and Einhorn (1992) suggested a mathematical model to describe the process of belief revision as an anchoring-and-adjustment process that predicts order effects. The adjustment process depends on the direction of the impact of the evidence and on the level of the anchor. If negative evidence is presented, the impact of the evidence is proportional to the current level of belief: the more a person already believes in a hypothesis the greater the impact of the negative evidence piece. This means that strong anchors are weakened more by means of the same evidence than weak anchors. If positive evidence is presented the adjustment weight is inversely proportional to the current anchor. The less a person believes in a hypothesis the greater the impact of the positive piece of evidence. This establishes a contrast effect between the current anchor and new pieces of evidence. The impact of new evidence is larger the greater the difference between anchor and evidence. This contrast effect is the main cause for order effects according to the model.

An example may illustrate this effect. Assume that a person has a neutral belief about a hypothesis at the start of the experiment. In condition A she first receives a piece of positive evidence and then a piece of negative evidence. In condition B the order is reversed. In condition A the belief in the hypothesis is first raised after the positive evidence. Therefore, the impact of the negative piece of evidence in condition A is high as it is contrasted to a high belief. In condition B the belief strength is neutral at the time the negative evidence is presented and its impacts is therefore smaller on the belief adjustment. The analogous argumentation holds true for the positive piece of evidence in both conditions. Because of this greater impact of the last piece of evidence, the result is a recency effect. This example also makes clear that the recency effect prediction holds true only for so called inconsistent sequences of information that contain both positive and negative pieces of evidence. If only consistent information (positive or negative) is given in a sequence no order effect is predicted as each piece of evidence in a consistent sequence is weighted according to the same principle. If only positive evidence is presented each piece of evidence is weighted inversely proportional to the current belief level. If only negative pieces of evidence are presented each piece is weighted proportional to the current belief level.

Tests of the Hogarth and Einhorn model

If one compares results from different studies that have been run one can see confirmation as well as contradictions to these predictions. Johnson (1995) found the predicted recency effect in the domain of auditing. Similar results were obtained by Adelman, Tolcott, and Bresnick (1993) with trained army air defense personnel and by Highhouse and Gallo (1997) examining personnel decision making.

On the other hand there are also contradictions and inconsistencies. Tubbs, Gaeth, Levin, and Van Osdol (1993) report a recency effect, but not only – as predicted by the model – for inconsistent but also for consistent sequences. Chapman, Bergus, and Elstein (1996) found a recency effect in a clinical judgment task with physicians both as predicted if the belief is adjusted after each piece of evidence (Step-by-Step) but also in contradiction to Hogarth and Einhorn (1992) if the belief is adjusted only once after all pieces of information have been presented (End-of-Sequence, EoS). Depending on the content of the sequence Adelman and Bresnick (1992) found both a recency effect and no order effect at all for inconsistent sequences processed in a step-by-step manner in an experiment with Patriot air defense officers. In a detailed replication of this study Adelman, Bresnick, Black, Marvin, and Sak (1996) found a primacy effect in early processing steps as well in later ones, although they presented inconsistent sequences of information in a step-by-step manner. Plach (1998) did not find an order effect at all. To summarize, the general picture on order effects appears mixed.

Order effects in belief revision and task experience

One possible reason for these inconsistencies is that in most of these studies the actual participants' experience with the task structure and their familiarity with the pieces of evidence presented in the trials was not known or controlled. For example, in one of Hogarth and Einhorn's (1992) experiments participants were obliged to rate the probability that a new coaching program is responsible for the improvement of a baseball player´s hitting rate. But participants could not ground their decision on their own experience about how often a change in a coaching program and an improvement in hitting rate occur together. But it was shown previously in other reasoning domains that basing decisions on experience can improve reasoning quite substantially. For example, studies on diagnostic reasoning showed that experience-based learning of the relationship between evidence and related hypotheses can lead to more accurate performance (e. g., Christensen-Szalanski & Beach, 1982). Current work on how people make decisions from experience when they have to learn probabilities and outcomes by observation has suggested that presenting this information as frequencies through examples facilitates processing probabilistic information and eliminating errors (e.g., Edgell et al., 2004; Newell & Rakow, 2007).

With regard to order effects Wang, Johnson, and Zhang (2006) recently found that in a categorical decision task the recency effect decreases and disappears as experience with the task increases when the experience was acquired by processing a series of examples. They ascribe this effect to two main causes. First, by learning from examples the participants' "… beliefs are gradually tuned to the statistical structure of the environment" (p. 222). Second, by learning from examples the participants' confidence in their beliefs increases making them less susceptible to changes due to new pieces of information. There are also studies on belief revision in realistic environments with subject-matter experts as participants that show experience reduces order effects (e.g., Trotman & Wright, 1996).

These results suggest that it might be important when studying belief revision that the participants possess knowledge about the statistical structure of the task acquired most naturally by encountering a series of examples. This might be because learning the statistical structure of the task seems to take place automatically through experiencing examples (Hasher & Zacks, 1984; Sedlmeier, 2002) and allows retrieving the relevant information from memory instead of having to combine pieces of evidence never encountered before in this combination. But this seems not be the whole story either. In an early study Adelman, Tolcott, and Bresnick (1993) found order-effects also with domain experts. Also Zhang, Johnson, and Wang (1998) using a similar paradigm as Wang et al. (2006) report that participants learnt the relevant base rates and conditional probabilities in a classification task quite well. Nevertheless, the participants showed a recency effect both in classification decisions and in a belief revision task when inconsistent sequences of information were presented in accordance with Hogarth and Einhorn (1992). So obviously, even experience-based learning does not prevent order effects in reasoning in all cases.

Goal and outline of this research project

Thus, this paper examines in more detail the relationship between learning the statistical structure of the task by experiencing a series of examples and order effects in belief revision tasks. In each of the experiments the participants completed a learning phase consisting of a series of classification trials to learn the probabilistic relationship between pieces of information about an object and the category membership of this object. The objects were planes approaching a naval ship that had to be classified as either commercial or hostile. The two types of information were information about the plane's route and the plane's answer to the request for identification (ID). Each type of information had two possible values: one favoring the commercial category, one the hostile category.

This learning phase was followed by a belief revision block with several belief revision trials where participants had to rate the probability of the category membership of an object—the plane—after each of two sequentially presented pieces of information about the plane: one about its route and one about its answer to the ID request. Categories and pieces of information were the same as in the earlier classification trials.

This combination of learning phase and belief revision task made it possible to test whether direct experience with the underlying statistical task structure prevents order effects in a following belief revision task. For this test it is not necessary that participants learn the correct probabilities but only to test whether they come to the same final estimation in the belief revision task irrespective of whether this final estimation is correct or not. The same paradigm was used in the experiments of Zhang et al. (1998) and Wang et al. (2006).

In all experiments the trials of the belief revision block were designed in such a way that the participants had to adjust their belief after each piece of information so that, according to Hogarth and Einhorn (1992), a recency effect should occur. The general question was whether the predicted recency effect disappears if the participants acquire knowledge about the probabilistic relationship between pieces of evidence and category membership from a set of examples.

In the first experiment we investigated whether experience with parts of the statistical task structure is sufficient to prevent order effects in a belief revision task. The participants were presented with a series of classification trials in the learning phase where only a single piece of information about the to be classified object was presented on each trial. The aim was that participants acquire knowledge about the relationship between each single piece of information and category membership. In the following belief revision block they were presented with belief revision trials, each one consisting of a sequence of two inconsistent pieces of evidence, and they had to rate their belief after each piece of evidence. Therefore, in this experiment participants had to integrate their partial knowledge—knowledge only about these single relationships—in the belief revision task and could not rely on knowledge about the relationship of combinations of pieces of evidence and category membership.

In the second experiment we investigated whether order effects in a belief revision task disappear if participants experienced the statistical relationships directly that were asked for in the belief revision task. The classification trials of the learning phase now consisted of sequences of two pieces of evidence and the participants had to classify the object after these two pieces. Therefore, the participants had the opportunity to learn the relationship between combinations of pieces of evidence and the object's category membership. Presenting two pieces of information in the classification trials of the learning phase allowed us to examine the occurrence of order effects in the classification decisions. In the following belief revision block participants were presented with belief revision trials, with each trial consisting of a sequence of inconsistent pieces of evidence, and they had to rate their belief after each piece of evidence as in Experiment 1.

In the third experiment we examined order effects in consistent sequences. For this it was necessary to vary the diagnostic value of the two types of information. One type of information gave a strong indication of the object's category, the other type only a weak indication. With this variation it was possible to detect potential order effects also for sequences of two consistent pieces of information. In the learning phase participants were presented with sequences of two pieces of information and had to classify the object as in Experiment 2. In the following belief revision block participants were again presented with belief revision trials each consisting of a sequence of two pieces of evidence, and they had to rate their belief after each piece of evidence. But unlike the first two experiments, in Experiment 3 trials with sequences of consistent pieces of evidence were also presented in the belief revision block. Table 1 gives an overview of the research goals and key manipulations of the three experiments.

Insert Table 1 about here

Experiment 1

Experiment 1 investigated whether people show an order effect if they experience the relationship between single evidence pieces and category membership from a set of examples. In the experiments of Hogarth and Einhorn (1992) participants did not have this experience as they did not have the experience on the statistical relationship between combinations of pieces of evidence and category. In previous experiments examining the relationship between experience and order effects participants encountered always those two pieces of information in the classification trials of the learning phase (e.g., Wang et al., 2006; Zhang et al., 1998) that were also presented in the belief revision task. Therefore, in these experiments participants experienced exactly that relationship directly that was relevant for their *final* belief estimation in the belief revision task. They did not experience the relationships that were relevant for their *intermediate* belief estimation when they had received the first of two pieces of evidence.

Thus, Experiment 1 explored whether order effects disappear if participants had directly experienced the statistical relationship of *single* pieces of evidence and category membership of objects before, but had not experienced directly the statistical relationship between *combinations* of evidence pieces. In this experiment the participants needed to combine the knowledge about the impact of each single piece of evidence in the belief revision task and could not rely on information stored in memory about the impact of the presented combination of evidence pieces. Studies in Bayesian reasoning show that reasoning can be greatly improved if participants are presented exactly with this kind of information and not only if the directly experienced conditional probabilities were asked for but also if the base rate was asked for (Christensen-Szalanski & Beach, 1982), that is, if the acquired knowledge has to be used to estimate a related variable.

Method

Participants. Participants were 20 students from Chemnitz University of Technology who were compensated with extra course credit. The average age was 20.35 (*SD* = 1.5) years. 15 were female.

Materials and Instruction. The cover story for all three experiments was that the participants were to imagine being a captain on a naval ship located in an area with both commercial airplanes and hostile military airplanes. In the learning phase of Experiment 1 participants were presented with a series of classification trials. In each of the classification trials a plane approached the ship and the participants had to decide on the basis of *one* piece of information which of the two categories the approaching plane belonged. This piece of information was either about the plane's route (R) or about the plane's answer to the request for identification (ID). The plane could fly on a commercial route (R+), favoring the commercial plane hypothesis, or it could not (R-), favoring the hostile military hypothesis. Similarly, the plane could identify itself as a commercial plane (ID+), favoring the commercial hypothesis, or it could not answer at all (ID-), favoring the hostile military plane hypothesis. As each of the four pieces of information was associated with each plane category with a certain conditional probability the participants had to make their decisions under uncertainty (the probabilities are shown in Table 2). After they made their decision participants received feedback about the "true" nature of the plane. Figure 1 shows the principal procedure of a classification trial.

Insert Table 2 and Figure 1 about here

In the trials of the belief revision block participants were presented sequentially with three pieces of information – the base information that a plane is approaching the ship, a piece of information about the plane's route, and a piece of information about the plane's identification request answer (ID). It was made clear to the participants in the introduction to the belief revision block that the planes they had to evaluate in the belief revision block trials were drawn from the same population as the planes in the classification trials. The pieces of information about the plane's route and its identification were always inconsistent. That is, on each trial one piece of information after the base information favored the hostile military plane hypothesis and the other one contradicted this hypothesis. Participants received these pieces of information in one session in the order of Route-ID, and the same pieces of information in the other session in the order of ID-Route. The sequences of information in the belief revision trials are shown in Table 3. After each piece of information, including the base information that a plane was approaching the ship, the participants had to rate the probability that the plane was hostile given the currently available pieces of information on a scale from 0 to 100. After this the screen was cleared and the next piece of information was presented. The principal procedure of a belief revision trial is shown in Figure 2.

Insert Table 3 and Figure 2 about here

Procedure. Each participant completed two sessions. Each session was divided into two parts. The first part consisted of a learning phase in which 40 classification trials were presented. In half of the classification trials of each session information about the route of the plane was displayed, in the other half of the trials information about the plane's identification was displayed. Trials with route or identification information were mixed randomly. Half of the planes in each session were commercial planes; half were hostile military planes. The absolute number of trials with a given combination of presented information and plane category in the learning phase of one session is shown in Table 2 as are the resulting base rates and conditional probabilities. In total participants performed 80 classification trials across the two sessions.

In the second part – after learning – participants had to perform the belief revision block where they had to indicate their belief that an approaching plane was hostile given two successively presented pieces of information. In each session they had to perform two belief revision trials. The two sessions differed only in terms of the order of route and ID information in the trials of the belief revision block. The two types of sessions were balanced across participants. The learning phases of the two sessions were identical in terms of the frequencies of combinations of pieces of information and plane category.

After the belief revision block participants were presented with each single piece of information (R+, R-, ID+, ID-, and “plane approaching”) in random order. They had to give a rating from -100 to 100 of how strongly each piece of information on its own supported the hostile plane hypothesis. This was performed as a manipulation check to test whether each single piece of evidence was interpreted in the correct direction, namely R+ and ID+ as supporting the commercial plane hypothesis, R- and ID- as supporting the hostile plane hypothesis, and ‘plane approaching’ as supporting neither.

This check was performed in all experiments reported in this study. It turned out that these five pieces of information were interpreted as intended in all experiments. The average ratings for R+ and ID+ were always significantly below 0 indicating support for the commercial hypothesis, R- and ID- were always significantly above 0 indicating support for the hostile hypothesis (smallest t-statistic for R+ with $M = -17.40$, $SD = 46.04$ in Experiment 1, $t(19) = -2.39$, $p = .03$). The ratings for the information "plane approaching" never differed significantly from 0.

Design. The participants' classifications in the learning phase were analyzed according a 2 (session) x 4 (information) within-subjects design. The first factor represents whether the classifications were made in the first or in the second session and the second factor the kind of information presented in the classification trials of the learning phase R+, R-, ID+, and ID-.

For the analysis of the ratings in the trials of the belief revision block a 2 (order of sessions) x 2 (evidence combination) x 2 (order of information) mixed design was realized. Order of sessions, that is, first the route-ID order session and then the ID-route order session or vice versa was manipulated as a between-subjects factor. The factor evidence combination with two possible combinations, R+ID- and R-ID+, and the third factor, order of information, were manipulated as within-subjects factors.

Results

Learning phase. The decisions for the 40 trials by each participant in each session were transformed into observed base rates and observed conditional probabilities of the answer 'commercial' given a piece of evidence, p('commercial' | evidence). These were averaged across the 20 participants. The results are shown in Figure 3 together with the corresponding theoretical values from Table 2 (labeled "Bayes").

The base rate for the answer ‘commercial’ computed for the two sessions did not differ significantly from the theoretical value, $t(19) = -0.44$, $p = .66$. Neither did the base rate differ when computed for the sessions separately, greatest $t(19) = -1.09$, $p = .23$. The observed conditional probabilities of the answer ‘commercial’ given a piece of information were generally more extreme than the theoretical values. That is, the observed conditional probabilities of the answer ‘commercial’ given R+ and ID+ were greater than .8, and they were less than .2 given each of the two pieces of evidence contradicting the commercial hypothesis R- and ID-. With only one exception, these deviations were significant both if the data of the learning phases of each session were analyzed separately or together, smallest $t(19) = 2.15$, $p = .045$. Only for R- trials in the first session did the observed conditional probability not differ significantly from the theoretical value of .2, $t(19) = .98$, $p = .34$.

Insert Figure 3 about here

A 2x4 analysis of variance (ANOVA) for the observed conditional probabilities comparing the within-subjects variables session (first or second) and piece of information (R+, R-, ID+, ID-) revealed a highly significant main effect of piece of information, $F(3, 57) = 355.68$, $p < .001$, $\eta_P^2 = .95$, indicating that participants clearly reacted differently to the different pieces of information. Also, the interaction between session and piece of information was significant, $F(3, 57) = 2.93$, $p = .041$, $\eta_P^2 = .13$. This was because the observed conditional probabilities for the answer “commercial” became more extreme in the second session compared to the first session.

Belief revision block. Figure 4 shows the average ratings of the probability of the plane being hostile that the participants gave after each piece of information presented during a trial sequence. Each line represents the mean ratings for one of the four possible sequences. The ratings show a clear recency effect as predicted by Hogarth and Einhorn (1992). For both inconsistent evidence combinations the participants estimated the probability of the plane being hostile greater if the final piece of information favored the hostile hypothesis than if the final piece of information contradicted the hostile hypothesis.

This was confirmed in the statistical analysis for which the differences between the final rating and the rating after the baseline information were computed to correct for baseline differences. For these differences a 2x2x2 ANOVA was conducted comparing the between-subjects variables order of sessions (route-ID order in the belief revision block in the first and ID-route in the second session or vice versa) and the within-subjects variables order of information (route-ID or ID-route) in the belief revision trial and evidence combination (R+ID- or R-ID+) in the belief revision trial. It revealed only a significant interaction between evidence combination and order of information, $F(1, 18) = 7.62$, $p = .013$, $\eta_P^2 = .30$, because the final ratings were less than the baseline rating when the last piece of information contradicted the hostile hypothesis, and they were greater than the baseline when the last piece of information confirmed the hostile hypothesis. This pattern of results confirms the recency effect predicted by Hogarth and Einhorn (1992). All other main effects or interactions were not significant, $F < 1$.

Insert Figure 4 about here

Discussion

This experiment showed that participants could acquire base-rate information quite properly from examples. The overall relative frequencies of hostile and friendly classification decisions in the learning phase did not differ significantly from the theoretical value. However, the observed conditional probabilities computed from the classification decisions in the learning phase were all more extreme than the theoretical values that defined the probability distribution.

As predicted by Hogarth and Einhorn's model (1992) the ratings of class membership in the belief evaluation task showed a clear recency effect, although the participants were presented with a series of trials representing the probabilistic relationship between category membership and each single piece of information in the learning phase before. The results from the learning phase show that participants acquired knowledge about the probabilistic relationship between each piece of evidence and the plane's category membership, even though it was not perfectly correct. The results of the belief revision trials show that the participants could not integrate their knowledge about the significance of each single piece of evidence in such a way that order effects were prevented. Therefore, knowledge about the significance of single pieces of evidence acquired by a series of examples does not seem to be sufficient to eliminate order effects in this belief revision task.

Experiment 2

Because experiencing the relationship between single pieces of information and the category membership of an object did not extinguish the recency effect in belief revision in Experiment 1, the major question of Experiment 2 was whether the recency effect would disappear if participants experience the relationship between *combinations* of different pieces of information and category membership through examples. In this case they directly experience the relationship that is later asked for in the belief revision trials. In a previous experiment using a similar procedure Zhang et al. (1998) found that this did not prevent order effects. In contrast to the procedure used by Zhang et al. in which 50 trials were used, Experiment 2 gave participants further classification trials, 100, to acquire knowledge about the probabilistic relationship between evidence and plane category. Furthermore, the participants were presented with the two possible inconsistent evidence combinations in the belief revision block, not only with one as in Zhang et al. (1998).

Method

Participants. 40 undergraduate students from Chemnitz University of Technology participated in the experiment for course credit. They were randomly assigned to one of four groups (10 participants per group). The average age was 21.9 (*SD* = 2.5) years. 27 were female.

Materials, procedure, design. The same cover story was used as in Experiment 1. In Experiment 2 participants performed only one session. As in Experiment 1 this session was divided into two parts: a learning phase with a series of classification trials and a belief revision block. In the classification trials the participants in Experiment 2 received *two* pieces of information (route and ID) about the plane after the message that a plane was approaching (see Figure 5). The participants then had to decide about the plane's identity and received feedback about the plane's "real" identity thereafter. In total each participant performed 100 classification trials divided into two blocks with 50 classification trials each. All four possible combinations of the pieces of evidence (R+ID+; R+ID-; R-ID+; R-ID-) were presented in the classification trials. The frequency of the two plane categories and of the four evidence combinations was the same in both blocks. The relevant probabilities and frequencies are shown in Table 4. The order of route and ID information in the classification trials was constant across all classification trials, that is, each participant received the pieces of information in the classification trials only in one order, either Route-ID or ID-Route.

Insert Figure 5 and Table 4 about here

In the belief revision block trials participants were presented sequentially with three pieces of information as in Experiment 1 (see Figure 2), and they had to rate the probability of the plane being hostile after each piece of information on a scale between 0 and 100. The route and ID information were always inconsistent. In contrast with Experiment 1, participants were presented with each of the two possible inconsistent evidence combinations (R+ID- and R-ID+) only once. That is, they saw the two possible combinations either in the order Route-ID or in the order ID-Route.

For half of the participants the order of the information in the classification trials was Route-ID, for the other half it was ID-Route. Also, the order of information in the belief revision trials was Route-ID for half of the participants and ID-Route for the other half. This yielded four experimental groups, each with ten participants. In two groups the order of information in the learning phase and in the belief revision block was the same, in the other two groups it was different. Consequently, order of information was manipulated as a between-subjects factor in this experiment, replicating Experiment 1 of Zhang et al. (1998). This allowed us to test two effects. The first effect is of a more extended learning phase compared to Zhang et al.. The second test is whether knowledge acquired in the learning phase using one feature order (e.g., route then ID) can be applied in the belief revision task in the opposite order (e.g., ID then route).

For the learning phase this yielded a 2x2x4 mixed factorial design with order of information (Route-ID or ID-Route) as a between-subjects factor and block of trials (first and second block) and evidence pattern (R+ID+, R-ID+, R+ID-, R-ID-) presented in a trial as within-subjects factor. The belief revision block consisted of a 2x2x2 mixed factorial design with order of information in the learning phase and order of information in the belief evaluation task as between-subjects factors and evidence pattern (R+ID-, R-ID+) as within-subjects factor.

Results

Learning phase. The decisions for each block of 50 trials by each participant were transformed into observed base rates and observed conditional probabilities of the answer "commercial" given an evidence pattern, which were then averaged across the 20 participants for the route-ID order and across the 20 participants for the ID-route order. The observed base rate differed for both route-ID and ID-route orders of information neither in the first block nor in the second block significantly from the theoretical value of .5, largest $t(19) = -0.96$, $p = .35$.

The observed conditional probabilities (see Figure 6) for trials with consistent sequences were both for the first and the second block and for both orders of information more extreme than the respective theoretical values. For the pattern R+ID+ this deviation was significant for both orders of information in the second block and for the order route-ID in the first block, smallest $t(19) = 8.94$, $p < .001$. For the pattern R-ID-this deviation was significant only in the first block for the order route-ID, $t(19) = -5.11$, $p < .001$.

For inconsistent sequences the deviations from the theoretical value of .5 follow a recency effect pattern. If the last piece of information in a sequence favored the hostile hypothesis the observed conditional probability is greater than .5. If the last piece of information favored the commercial hypothesis the observed conditional probability is less than .5. Only the sequence ID-R+ in the first block (in the ID-R+ order of the R+ID-pattern) does not show this pattern. Both for the first and the second block of trials these deviations are significant only if the information is presented in the order route-ID, smallest $t(19) = -2.10$, $p = .049$.

Insert Figure 6 about here

A comparison of the observed conditional probabilities for different orders of information and evidence patterns confirms the recency effect for inconsistent sequences. In both blocks and for both inconsistent evidence patterns the observed conditional probabilities for the answer “commercial” were greater if the last piece of information favored the commercial hypothesis than if it favored the hostile hypothesis. This was statistically confirmed by a 2x2x2 ANOVA for the inconsistent sequences where the between-subjects variable order of information (route before or after ID) and the within-subjects variables block (first and second block of trials) and evidence pattern (R+ID-, R-ID+) were compared. It revealed a significant interaction between order of information and evidence pattern, $F(1, 38) = 11.97$, $p < .001$, $\eta_P^2 = .24$. This is due to the recency effect pattern described above. Neither the main effect of block nor any interactions involving this term reached significance, $F < 1$.

For the consistent sequences a 2x2x2 ANOVA with the between-subjects variable order of information (route before or after ID) and the within-subjects variables block of trials (first or second block of trials) and evidence pattern (R+ID+, R-ID-) was computed. Only the main effect of evidence pattern was significant, $F(1, 38) = 7{,}262.23$, $p < .001$, $\eta_P^2 = .9$, confirming Hogarth and Einhorn’s (1992) prediction of no order effect for consistent sequences.

Belief revision block. The results of the belief revision block collapsed over the two orders of information in the learning phase are shown in Figure 7. The ratings show a recency effect: for each of the two inconsistent sequences the probability of the plane being hostile was rated higher if the last piece of information favored the hostile hypothesis than if the last piece of information favored the commercial hypothesis.

To correct for baseline differences, the differences between the rating after the last piece of information and after the baseline information were computed. A 2x2x2 ANOVA for these differences was conducted for the between-subjects factors order of information in the learning phase (route before or after ID) and order of information in the belief evaluation task (route before or after ID), and the within-subjects factor evidence pattern (R+ID- or R-ID+). The effect of evidence pattern was highly significant, $F(1, 36) = 17.99$, $p < .001$, $\eta_P^2 = .27$. If the evidence pattern R+ID- was presented the mean difference was 20.48 (*SD* = 28.25). If the evidence pattern R-ID+ was presented the mean difference was -6.45 (*SD* = 34.06). The interaction between order of information in the belief revision block and evidence pattern was marginally significant, $F(1, 36) = 3.89$, $p = .056$, $\eta_P^2 = .07$, confirming the recency effect pattern mentioned above. Neither the main effect of order of information in the training phase, $F(1, 36) = 1.24$, $p = .27$, $\eta_P^2 = .03$, nor any interactions involving this term were significant, $F < 1$.

Insert Figure 7 about here

Discussion

The results of the learning phase confirm the predictions of the model of Hogarth and Einhorn (1992). A clear recency effect was observed after inconsistent sequences in the classification trials of the learning phase replicating the recency effect Zhang et al. (1998) found in classification decisions. The recency effect of our Experiment 2 was found both in the first 50 trials and in the second 50 trials of the learning phase. More experience on the task did not reduce the order effect. Therefore, the recency effect cannot be attributed to a lack of task experience, at least over this period of time.

As in the first experiment the ratings in the belief revision block trials after inconsistent sequences also show a recency effect pattern that falls just short of significance. Therefore, it can be concluded that even if participants experience the same sequences of information in the learning phase and are trained on the relationship between combinations of pieces of information and category membership they still show an order effect in the probability ratings.

One may criticize that this order effect is caused by the knowledge the participants acquired during the learning phase. First, they could have associated each specific sequence of information as a whole with a certain level of belief in the relevant hypothesis. Then, they reproduced this knowledge in the evaluation task replicating the pattern of results of the learning phase in the belief evaluation task and producing a recency effect. This would imply that they evaluated each sequence in the belief evaluation task as a single unit, not considering each piece of information contained in a sequence on its own.

Two arguments speak against this assumption. First, the results of the manipulation check showed that the participants clearly acquired the correct knowledge about the significance of each single piece of evidence, although they were trained only on combinations of pieces of evidence. Second, the ratings after the second item in the belief evaluation task clearly followed the expected direction. That is, participants were able to evaluate the impact of a single piece of information on their current belief and were able to adjust their belief accordingly. They did not wait with the adjustment until they saw all pieces of evidence, which one could expect according to the above argument. Thus, it seems plausible to assume that the participants evaluated each piece of information in the belief evaluation task following an step-by-step process. According to Hogarth and Einhorn (1992) the order effect then is caused by the contrast effect between the to-be-adjusted anchor and the evaluated piece of evidence. The results of this experiment, together with those of Experiment 1, and those of Johnson (1995) and Adelman et al. (1993), indicate that experience on the task does not influence this mechanism.

Experiment 3

The results of Experiments 1 and 2 show that in accordance with the predictions of Hogarth and Einhorn (1992) a recency effect occurred both for decisions in classification trials (Experiment 2) and for the answers in a belief revision block (Experiments 1 and 2) when the participants were presented with inconsistent sequences. These two experiments did not address the question of order effects for sequences of *consistent* pieces of evidence, however. As explained above, the model of Hogarth and Einhorn (1992) predicts no order effect for consistent sequences in belief revision tasks. However, Tubbs et al. (1993) compared order effects for consistent and inconsistent sequences of information and found, contradictory to the model predictions, a recency effect for both consistent and inconsistent sequences. That is, the order of consistent pieces of evidence with different predictive validity will also have a recency effect. But Tubbs et al. did not give their participants the opportunity to experience the statistical nature of their task structure directly by a series of experiences with the regularities in the stimuli before having them to complete a belief revision task. Therefore, a third experiment with a broader focus was conducted. Experiment 3 examined whether order effects occurred both for consistent and inconsistent sequences of evidence both for decisions in classification trials and for answers in a belief revision task and whether such effects are affected by the direct experience of the relationship between pieces of evidence and object category.

To investigate order effects in consistent sequences it is necessary to use pieces of information that differ in their diagnostic power. These differentially diagnostic pieces of information can be generated by setting the conditional probability of object category, given some piece of information, in such a way that for one type of information (e.g., route), the conditional probabilities of plane type given the different values of this information type vary strongly across these values (i.e., flying or not flying on a commercial route). This type of information is called the strong evidence. For the other type of information the different values of this information type (e.g., answering or not answering to ID request) lead to similar conditional probabilities of plane category given each piece of evidence. This is called the weak evidence.

This manipulation allowed us to combine strong and weak pieces of evidence in different orders across sequences. A recency effect would be found if the respective hypothesis is more strongly believed when the strong evidence is presented as the last piece of information than when the weak evidence is presented as the last one. A primacy effect would be found if the respective hypothesis is more strongly believed when the strong evidence is presented first than when the weak evidence is presented first.

Method

Participants. 40 undergraduate students at Chemnitz University of Technology participated. Participants received course credit. The average age was 20.57 (*SD* = 1.91) years. 24 were female.

Materials, procedure, design. The same cover story was used as in the previous experiments. The experimental session was divided into two parts: a learning phase with a series of classification trials and a belief revision block. The basic procedure of the classification trials of the learning phase of Experiment 3 was the same as in Experiment 2. Participants received two pieces of information (route and ID) about the plane after the message that a plane was approaching and then had to decide whether the plane was a commercial or a hostile military plane. All possible combinations of pieces of evidence were shown in the classification trials (R+ID+, R+ID-, R-ID+, R-ID-). The conditional probabilities for the two types of planes given the different pieces of evidence and their possible combinations were set in such a way that one type of information provided strong evidence for the classification of the plane and the other one weak evidence.

Table 5 shows the conditional probabilities of hostile and commercial planes given information about the route and the ID. For half of the participants—the strong route group—the information about the route of the plane was a strong indicator of the identity of the plane, and the ID information was only a weak one. Therefore, if in this group the plane flew on a commercial route (R+) and did answer the ID request (ID+) the probability that it was a commercial plane was .94, and if it did not answer the request the probability of being a commercial plane was still .78. If the plane did not fly on a commercial route (R-) but answered the ID request (ID+) the probability of being commercial was only .22 and if it did not fly on a commercial route and did not answer the ID request the probability was only .06. For the other half of the participants— the strong ID group — the information about the response on the ID request was a strong indicator, and the route information had lower predictive power. Here the conditional probabilities of plane type given information about route or ID were just interchanged. From these settings it follows that when an inconsistent evidence pattern was presented (R+ID- or R-ID+) the probability of a plane being commercial was .78 if the strong piece of evidence favored the commercial hypothesis and it was .22 if the strong piece of evidence favored the military hostile hypothesis. An inspection of Table 5 shows how these probabilities were achieved. For example, given the evidence pattern R+ID- in the route strength condition there was a total of 18 trials in one session with this evidence pattern. In 14 trials the plane was commercial, in the other four it was hostile resulting in a conditional probability p('Commercial' | R+, ID-) of .78.

Insert Table 5 about here

In the belief revision block after the learning phase participants had to complete four belief revision trials. In each trial one of the four possible evidence patterns (R+ID+, R+ID-, R-ID+, R-ID-) was presented, that is, both the inconsistent patterns as in Experiment 1 and 2 and now also the consistent patterns. The procedure of each trial was the same as in the previous experiments.

Each participant completed two sessions. Each session consisted of a learning phase and a belief revision block. The learning phase of each session consisted of 104 classification trials and the belief revision block of each session consisted of four belief revision trials. The sessions differed only in the order of route and ID information. In one session the order of information in the learning phase and in the belief revision block was route-ID, in the other session it was ID-route. The order of sessions was balanced across participants.

To summarize, in this experiment the following design was realized. There were two strength conditions manipulated as a between-subjects factor. 20 participants received route as the strong information; 20 participants received ID as the strong information. The order of information (Route-ID or ID-Route) was realized as a within-subjects factor. As a third factor, the evidence pattern was varied as a within-subjects factor (R+ID+, R-ID-, R-ID+, and R+ID-). Thus, the learning phase and the belief revision block consisted of a 2 (strength condition) x 2 (order of information) x 4 (evidence pattern) mixed factorial design.

Results

Learning phase. Observed base rates and observed conditional probabilities for each evidence pattern were calculated from the 208 decisions of each participant for each evidence order and strength condition. These were then averaged over the 20 participants in each strength condition.

The observed base rate of the answer 'commercial' did not differ significantly from the expected theoretical value of .5 both for the route strength condition, $t(19) = -0.49$, $p = .63$, and for the ID strength condition, $t(19) = -0.74$, $p = .46$.

The results for both the theoretical and the observed conditional probabilities for the answer "commercial" are shown in Figure 8. The observed conditional probabilities for trials with consistent sequences were all more extreme than the respective theoretically expected values. For R+ID+ this difference was significant only for the ID strength condition, smallest $t(19) = 6.48$, $p < .001$. Given the evidence pattern R-ID- the observed conditional probabilities were always significantly less than .06, smallest $t(19) = -3.36$, $p = .003$. For inconsistent sequences the observed conditional probabilities converged closer to .5 than the expected values. For all sequences where a strong positive piece of information was combined with a weak negative one, the observed conditional probability was less than the theoretical value of .78. But this deviation was significant only for the sequence R+ID- in the route strength condition, $t(19) = -2.30$, $p = .033$, and for the sequence ID+R- in the ID strength condition, $t(19) = -2.74$, $p = .013$. For all sequences where a strong negative piece of information was combined with a weak positive one, the observed conditional probability was greater than the theoretical value of .22. This deviation is significant for three out of four possible sequences: R-ID+, $t(19) = 2.39$, $p = .027$, ID+R-, $t(19) = 2.79$, $p = .012$ in the route strength condition, R+ID-, $t(19) = 2.318$, $p = .032$ in the ID strength condition.

Given the same evidence pattern, e.g. R+ID-, the observed conditional probability for the answer ‘commercial’ is much higher in the route strength condition than in the ID strength condition. The analogue holds true for the evidence pattern R-ID+ (see Figure 8), confirming that the participants clearly identified route as the strong and ID as the weak information in the route condition and vice versa in the ID condition. This was confirmed by a 2x2x2 ANOVA calculated for the inconsistent sequences separately and comparing the between-subjects variable strength condition (route or ID condition) and the within-subjects variables order of information (Route-ID or ID-Route) and evidence pattern (R+ID-, R-ID+). It revealed a highly significant interaction between strength condition and evidence pattern, $F(1, 38) = 40.11$, $p < .001$, $\eta_P^2 = .51$.

Insert Figure 8 about here

Regarding the order effect in the classification decisions the observed conditional probabilities for the inconsistent evidence patterns R+ID- and R-ID+ in the route strength condition and R-ID+ in the ID strength condition show a recency effect pattern. The probabilities are greater when the last piece of information favored the commercial hypothesis. But neither the main effect of order of information nor any interaction involving this factor reached significance in the ANOVA, showing that the recency effect in the classification decisions was not significant in this experiment.

Also, for the consistent sequences a 2x2x2 ANOVA was conducted for the observed conditional probabilities comparing the between-subjects variable strength condition (route or ID condition) and the within-subjects variables order of information (route before or after ID) and evidence pattern (R+ID+, R-ID-). It revealed a highly significant main effect of strength condition, $F(1, 38) = 10.04$, $p = .003$, $\eta_P^2 = .21$. The observed probability of the answer ‘commercial’ was higher in the ID strength condition, $M_{\underline{ID}} = .50$, $SD_{ID} = .49$, than in the route strength condition, $M_R = .49$, $SD_R = .47$. The main effect of evidence pattern was also highly significant, $F(1, 38) = 8{,}632.78$, $p < .001$, $\eta_P^2 = .99$. The probability of the answer ‘commercial’ given the evidence pattern R+ID+ was .97, $SD = .058$, and it was .02, $SD = .032$ given the evidence pattern R-ID-. No other effects reached significance. Especially, neither the main effect of order of information nor any interaction involving this factor reached significance indicating that the participants did not show any order effect for consistent sequences of information in their categorical decisions.

Belief revision block. Figure 9 shows the participants’ average ratings of the probability of the plane being hostile after each piece of information for consistent and inconsistent sequences and for both strength conditions. As for the learning phase consistent and inconsistent sequences are reported and analyzed separately. The ratings after consistent sequences do not show any order effect as predicted by Hogarth and Einhorn (1992).

Insert Figure 9 about here

For the statistical analysis the difference between the final rating and the rating after the baseline information was computed for each participant to correct for baseline differences. Using these differences as dependent variable a 2x2x2 ANOVA was conducted comparing the between-subjects variable strength condition (route or ID) and the within-subjects variables order of information (route-ID or ID-route) and evidence pattern (R+ID+ or R-ID-). The only effect that was found significant was the main effect of evidence pattern, $F(1, 38) = 337.33$, $p < .001$, $\eta_P^2 = .90$. The mean difference between last rating and base rating was 38.83 (*SD* = 16.38) for the pattern R-ID-, and –34.91 (*SD* = 21.93) for the pattern R+ID+.

Besides one exception, the ratings for all inconsistent sequences show the pattern characteristic for a recency effect. Given the same evidence pattern the hostile hypothesis is regarded more probable if the last piece of evidence is in favor of this hypothesis. Only the ratings involving the evidence pattern R-ID+ in the ID strength condition do not show a recency effect. Again, for the statistical analysis the difference between the final rating and the rating after the baseline information was computed for each participant to correct for baseline differences. A 2x2x2 ANOVA was conducted comparing the between-subjects variable strength condition (route or ID) and the within-subjects variables order of information (route-ID or ID-route) and evidence pattern (R+ID+ or R-ID-). The ANOVA revealed significant main effects of information order, $F(1, 38) = 8.53$, $p = .006$, $\eta_P^2 = .18$, and evidence pattern,

$F(1, 38) = 12.39$, $p = .001$, $\eta_P^2 = .25$. The mean difference between the ratings after the last and the first piece of evidence was 15.94 (*SD* = 25.38) for the order route-ID, and 6.45 (*SD* = 24.03) for the order ID-route. The mean difference for the evidence pattern R+ID- was 16.71 (*SD* = 25.18), and 5.68 (*SD* = 23.91) for the pattern R-ID+. Most important, the two-way interaction between order of information and evidence pattern was also significant, $F(1, 38) = 7.91$, $p = .008$, $\eta_P^2 = .17$, statistically confirming the recency effect pattern described above. Given the evidence pattern R+ID- the mean difference between the last and first rating was 26.63 (*SD* = 18.99) if the piece of information favoring the hostile hypothesis was presented last (ID-), and 6.80 (*SD* = 26.87) if the piece of information favoring the commercial hypothesis (R+) was presented last. A similar pattern was found for the evidence pattern R-ID+. The mean difference was 5.25 (*SD* = 26.65) if the last information (ID+) favored the commercial hypothesis, and it was 6.10 (*SD* = 21.15) if the last piece of information (R-) favored the hostile hypothesis.

Discussion

As in the previous experiments Experiment 3 showed that experiencing the probabilistic relationship between pieces of evidence and object category by examples did not eliminate order effects in the probability ratings of a belief revision task. The pattern of results we found in the probability ratings in general supports the predictions of Hogarth and Einhorn's (1992) model. There was a recency effect for inconsistent sequences of evidence as found in Experiments 1 and 2 and no order effect for consistent sequences. But the recency effect for inconsistent sequences was less robust as in the previous experiments, as only in three of four possible sequences the effect was found. Additionally, the classification decisions after inconsistent sequences in the learning phase trials show a recency effect pattern in three of four possible cases but this effect did not reach significance as opposed to Experiment 2 and in Zhang et al. (1998).

We can only speculate about the reasons. But there are two possible explanations for the decrease in the size of the order effect in general in this experiment, that go beyond some unsystematic variations for single evidence combinations, such as the R-ID+ sequence in the ID strength condition. The first possible explanation is the introduction of weak and strong evidence in this experiment. The strong evidence in Experiment 3 had about the same diagnostic value as the evidence in Experiment 2. For example, P('Commercial Plane' | 'Plane is on commercial route') was .88 in the route condition of Experiment 3 and .80 in Experiment 2. But P('Commercial' | 'Plane does not respond') was .31 in the route condition of Experiment 3 and .2 in Experiment 2. When combined in an inconsistent sequence Hogarth and Einhorn's (1992) model predicts a smaller order effect given the values of Experiment 3. Given the variance in the participants' ratings this could be the reason for a less robust recency effect in Experiment 3 in the belief revision trials. The same argument can be applied to the failure of finding an order effect in the classification trial decisions. For a decision to be made, the belief strength has to be transformed into a decision. Given the inter- and intrapersonal variance in the belief revision process a reduced order effect has a reduced chance to show up in the classification decisions.

The second possible explanation is based on Wang et al.'s (2006) model of belief revision. In their model they assume that the more experience reasoners acquire with the statistical nature of the task the better is their knowledge tuned to this nature and the more confident they are in their knowledge. According to Wang et al. this results in changes to the sensitivity to new knowledge leading to a decrease and even to the extinction of recency effects. In Experiment 3 the participants completed a learning phase that consisted of more than twice as many classification trials than in the Experiments 1 and 2. Therefore, they had much more opportunity to tune their knowledge to the probabilistic structure of the task than in the previous experiments.

General Discussion

Three experiments were conducted to examine the relationship between experience and order effects in belief revision. The predictions from Hogarth and Einhorn's (1992) belief-adjustment model were evaluated in situations where participants could acquire knowledge about the underlying probabilistic task structure by a series of classification trials before giving ratings of category membership in a belief revision task. In Experiment 1 participants could experience the relationship between single pieces of evidence and the category membership of objects and then had to apply this experience in a belief revision task where they were presented with sequences of inconsistent pieces of evidence.

In Experiment 2 participants could experience the relationship between combinations of pieces of evidence and category membership and were presented with the subset of inconsistent combinations in a belief revision task. In this experiment both types of evidence provided the information with the same diagnosticity validity about category membership.

In Experiment 3 participants experienced the relationship between combinations of pieces of evidence as in Experiment 2, but here the two types of evidence provided differently strong information about category membership. All three experiments demonstrate that order effects were not prevented by the direct experience of the probabilistic relationship between pieces of evidence and category membership. In accordance with the predictions of Hogarth and Einhorn (1992) we found recency effects for inconsistent sequences both for the classification decisions in the learning phase and for the ratings in the belief revision task. Additionally, and also in accordance with Hogarth and Einhorn, we found no order effect for consistent sequences both for the classification decisions and for the ratings in the belief revision task.

A conclusion to draw from the reported experiments is that direct experience with the probabilistic structure underlying a task (a possible mitigator of order effects) does not appear to improve reasoning performance in belief revision tasks. But other studies demonstrate an improvement of reasoning performance after such experience (e.g., Christensen-Szalanski & Beach, 1982; Wang et al., 2006). Christensen-Szalanski and Beach (1982) presented their participants with a series of trials, each trial representing a patient that either suffered from a certain disease or not and showed a certain symptom or not. Consequently, the participants directly experienced the probabilistic relationship between symptom and disease in this training procedure. They found that such a training procedure made the base rate fallacy in a diagnostic reasoning task disappear. They explained this with the participants' experience with the underlying task structure. Why did the experience with the underlying task structure in the experiments described here not prevent the occurrence of order effects? The answer to this question offered by the results described in this paper is that order effects in diagnostic reasoning tasks seem not to be based on memory retrieval processes. Both in the experiments of Christensen-Szalanski and Beach (1982) and in the experiments presented here participants acquired knowledge about the statistical structure of the task by encountering a series of examples. In both studies participants experienced the probabilistic relationship between object category (ill or not in Christensen-Szalanski and Beach; hostile or commercial plane type in our study) by a series of examples. This knowledge could then directly be applied in the diagnostic reasoning task of Christensen-Szalanski and Beach when they asked their participants for the probability that a patient suffers from the disease given the symptom. In this case the necessary information simply had to be retrieved from memory. To decide about an object’s category in the classification trials of the learning phase and to perform the ratings in the belief revision tasks of our experiments the necessary information (i.e., the conditional probabilities of plane type given a certain evidence) had not only to be retrieved from memory but also to be integrated by the belief revision process. Therefore, this integration process seems to be the cause of the order effect found in our experiment. The results presented here support the assumption of the model of Hogarth and Einhorn (1992) that one major characteristic of this integration process is that new evidence is put into contrast to the current belief strength leading to a weighting of pieces of evidence in inconsistent sequences that depends on the order of evidence presentation.

How does this relate to Wang et al.’s (2006) results? Their results indicate that the integration process is modified by the reasoners’ experience with the task structure resulting in a reduction and finally an extinction of the recency effect for inconsistent sequences such as those used in our experiments. They postulate that with increasing experience the reasoners’ beliefs become better tuned to the task structure and the reasoners’ confidence in their beliefs increases. This leads to the reasoners’ reduction of sensitivity to new pieces of information, decreasing the recency effect. Why was this not the case in our experiments? Despite some similarities in the experimental procedure of their and our experiments, there are also major differences. Of major importance could be that Wang et al.’s participants had to perform the belief revision task several times during the experimental session with each evidence pattern presented in both orders in each block of belief revision. This might have caused a deeper self-reflection of their beliefs increasing the confidence in their beliefs. In contrast to this, the participants of the experiments described in this paper performed only two blocks of belief revision trials during an experiment. In each block each evidence pattern was presented only once. Therefore, there was much less necessity and opportunity to explicitly think about one's beliefs in these experiments, supposedly delaying the process of confidence growth. We can only speculate here because confidence was not controlled or measured in our experiments. But the results of Experiment 3 give some indication that the direct experience of the probabilistic nature of the task structure by a great number of examples indeed might lead to an increase in confidence that in turn leads to the reduction and perhaps extinction of order effects. The role of confidence in belief revision might also be the clue to explain the different results of studies on order effects in belief revision tasks with subject-matter experts (e.g., Trotman & Wright, 1996 vs Adelman, Tolcott, & Bresnick, 1993). Confidence was not measured in these studies. Therefore, it seems worthwhile in future work to examine the effect of confidence in one's beliefs. Studying the relationship of confidence to one's sensitivity to new pieces of evidence in particular and to order effects in general, both in controlled laboratory tasks and more applied domains may provide further insights into belief revision.

References

Adelman, L., & Bresnick, T. A. (1992). Examining the effects of information sequence on Patriot air defense officers′ judgments. *Organizational Behavior and Human Decision Processes, 53*, 204-228.

Adelman, L., Bresnick, T. A., Black, P. K., Marvin, F. F., & Sak, S. G. (1996). Research with Patriot air defense officers: Examining information order effects. *Human Factors, 38*(2), 250-261.

Adelman, L., Tolcott, M. A., & Bresnick, T. A. (1993). Examining the effects of information order on expert judgment. *Organizational Behavior and Human Decision Processes, 56*, 348-369.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41*, 258-290.

Chapman, G. B., Bergus, G. R., & Elstein, A. S. (1996). Order of information affects clinical judgment. *Journal of Behavioral Decision Making, 9*(3), 201-213.

Christensen-Szalanski, J. J. J., & Beach, L. R. (1982). Experience and the base-rate fallacy. *Organizational Behavior And Human Performance, 29*, 270-278.

Edgell, St. E., Harbison, J. I., Neace, W., Nahinsky, I. D., & Lajoie, A. S. (2004). What is Learned from Experience in a Probabilistic Environment? *Journal of Behavioral Decision Making, 17*, 213-229.

Friedrich, J., & Smith, P. (1998). Suppressive influence of weak arguments in mixed-quality messages: An exploration of mechanisms via argument rating, pretesting, and order effects. *Basic & Applied Social Psychology, 20*(4), 293-304.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist, 39*(12), 1372-1388.

Highhouse, S., & Gallo, A. (1997). Order effects in personnel decision making. *Human Performance, 10*(1), 31-46.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology, 24*, 1-55.

Johnson, E., N. (1995). Effects of information order, group assistance, and experience on auditors` sequential belief revision. *Journal of Economic Psychology, 16*, 137-160.

Johnson, T., & Krems, J. (2001). Use of current explanations in multicausal abductive reasoning. *Cognitive science, 25*, 903-939.

Johnson-Laird, P. N., & Steedman, M. (1978). The psychology of syllogisms. *Cognitive Psychology, 10*(1), 64-99.

Langley, P. (1995). Order effects in incremental learning. In P. Reimann & H. Spada (Eds.), *Learning in humans and machines* (pp. 154-167). Kidlington, UK: Pergamon.

Newell, B., & Rakow, T. (2007). The role of experience in decisions from description. *Psychonomic Bulletin & Review, 14(*6), 1133-1139.

Plach, M. (1998). *Prozesse der Urteilsrevision: Eine kognitive Modellierung der Verarbeitung unsicheren Wissens.* Wiesbaden: DUV.

Ritter, F. E., Nerb, J., Lehtinen, E., & O'Shea, T. M. (Eds.). (2007). *In order to learn: How the sequences of topics affect learning*. New York, NY: Oxford University Press.

Sedlmeier, P. (2002). Associative learning and frequency judgments: The PASS model. In P. Sedlmeier & T. Betsch (Eds.), *Frequency processing and cognition* (pp. 137-152). Oxford: Oxford University Press.

Thagard, P. (1992). *Conceptual revolutions*. Princeton, NJ: Princeton University Press.

Trotman, K. T., & Wright, A. (1996). Recency effects: Task complexity, decision mode, and task-specific experience. *Behavioral Research in Accounting, 8*, 175-193.

Tubbs, R. M., Gaeth, G. J., Levin, I. P., & Osdol, L. A. V. (1993). Order effects in belief updating with consistent and inconsistent evidence. *Journal of Behavioral Decision Making, 6*, 257-269.

Wang, H., Johnson, T. R., & Zhang, J. (2006). The order effect in human abductive reasoning: an empirical and computational study. *Journal of Experimental & Theoretical Artificial Intelligence, 18*(2), 215-247.

Wang, P. (1993). Belief revision in probability theory. In *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence* (pp. 519-526). San Mateo, CA: Morgan Kaufmann Publisher.

Zhang, J., Johnson, T. R., & Wang, H. (1998). The relation between order effects and frequency learning in tactical decision making. *Thinking and Reasoning, 4*(2), 123-145.

Tables

Table 1

Description of the goals and the key manipulations of each experiment.

Exp.	Research Question	Key manipulation	Learning phase	Belief revision task
1	Do order effects disappear if part of the statistical task structure was directly experienced before?	Presentation of only single pieces of information and category type in learning phase	Single piece of information	Inconsistent sequences with two pieces of information
2	Do order effects disappear if the complete task structure was experienced directly before?	Presentation of sequences with two pieces of information in learning phase	Consistent and inconsistent sequences of two pieces of information	Inconsistent sequences with two pieces of information
3	Do order effects exist for sequences with only consistent information?	High and low informative types of information	Consistent and inconsistent sequences of two pieces of information	Consistent and inconsistent sequences of two pieces of information

Table 2

Conditional probability of plane type given each piece of information in one session of Experiment 1

Piece of evidence	Plane type	Number of trials	P(plane type | evidence)
Plane flying on commercial route (R+)	Commercial	8	.8
Plane not flying on commercial route (R-)	Commercial	2	.2
Plane answering (ID+)	Commercial	8	.8
Plane not answering (ID-)	Commercial	2	.2
Plane flying on commercial route (R+)	Hostile	2	.2
Plane not flying on commercial route (R-)	Hostile	8	.8
Plane answering (ID+)	Hostile	2	.2
Plane not answering (ID-)	Hostile	8	.8

Table 3

Sequence of pieces of information in the belief revision trials of Experiment 1.

Sequence Type	Base information	First piece of information	Second piece of information
R+/ID-	Plane approaching	On commercial route	Does not answer
R-/ID+	Plane approaching	Not on commercial route	Identifies itself as commercial
ID-/R+	Plane approaching	Does not answer	On commercial route
ID+/R-	Plane approaching	Identifies itself as commercial	Not on commercial route

Table 4

Conditional probabilities and corresponding absolute number of trials in one block (in brackets) for the plane type "commercial" given the different evidence combinations in Experiment 2.

Evidence		T = 'Commercial'			T = 'Hostile'			
A	B	P(T | A)	P(T | B)	P(T | A,B)	P(T | A)	P(T | B)	P(T | A,B)	Total trials with AB
R+	ID+	.80 (20)	.80 (20)	.94 (16)	.20 (5)	.20 (5)	.06 (1)	17
R+	ID-	.80 (20)	.20 (5)	.50 (4)	.20 (5)	.80 (20)	.50 (4)	8
R-	ID+	.20 (5)	.80 (20)	.50 (4)	.80 (20)	.20 (5)	.50 (4)	8
R-	ID-	.20 (5)	.20 (5)	.06 (1)	.80 (20)	.80 (20)	.94 (16)	17

Note: T stands for Type of plane

Table 5

Conditional probabilities and corresponding absolute number of trials for one session (in brackets) for the plane types given the different evidence combinations in the route condition (first line in each evidence row) and ID condition (second line in each evidence row) in Experiment 3.

Evidence		T = 'Commercial'			T = 'Hostile'			
A	B	P(T | A)	P(T | B)	P(T | A,B)	P(T | A)	P(T | B)	P(T | A,B	Total trials with A,B
R+	ID+	.88 (46)	.69 (36)	.94 (32)	.12 (6)	.31 (16)	.06 (2)	34
		.69 (36)	.88 (46)	.94 (32)	.31 (16)	.12 (6)	.06 (2)	
R+	ID-	.88 (46)	.31 (16)	.78 (14)	.12 (6)	.69 (36)	.22 (4)	18
		.69 (36)	.12 (6)	.22 (4)	.31 (16)	.88 (46)	.78 (14)	
R-	ID+	.12 (6)	.69 (36)	.22 (4)	.88 (46)	.31 (16)	.78 (14)	18
		.31 (16)	.88 (46)	.78 (14)	.69 (36)	.12 (6)	.22 (4)	
R-	ID-	.12 (6)	.31 (16)	.06 (2)	.88 (46)	.69 (36)	.94 (32)	34
		.31 (16)	.12 (6)	.06 (2)	.69 (36)	.88 (46)	.94 (32)	

Note: T stands for Type of plane.

Figure Captions

Figure 1. Procedure of classification trial in Experiment 1.

Figure 2. Procedure of belief revision trial in all experiments.

Figure 3. Observed conditional probabilities and base rates for the answer "commercial" in the learning phase of Experiment 1; error bars represent SEM (R+: on commercial route, R-: off commercial route; ID+: answers ID request; ID-: no answer to ID request).

Figure 4. Probability rating of the plane being hostile after each piece of information in test phase of Experiment 1*;* error bars represent SEM (R+: on commercial route, R-: off commercial route; ID+: answers ID request; ID-: no answer to ID request).

Figure 5. Procedure of classification trial in Experiments 2 and 3.

Figure 6. Observed conditional probabilities for the answer 'commercial' for the first and second block of the learning phase of Experiment 2; error bars represent SEM (R+: on commercial route, R-: off commercial route; ID+: answers ID request; ID-: no answer to ID request).

Figure 7. Probability rating of the plane being hostile after each piece of information in Experiment 2; error bars represent SEM (R+: on commercial route, R-: off commercial route; ID+: answers ID request; ID-: no answer to ID request).

Figure 8. Observed conditional probabilities for the answer 'commercial' in the learning phase of Experiment 3; error bars represent SEM (R+: on commercial route, R-: off commercial route; ID+: answers ID request; ID-: no answer to ID request).

Figure 9. Probability ratings of the plane being hostile after each piece of information in Experiment 3; error bars represent SEM (R+: on commercial route, R-: off commercial route; ID+: answers ID request; ID-: no answer to ID request).

Figure 1

Figure 2

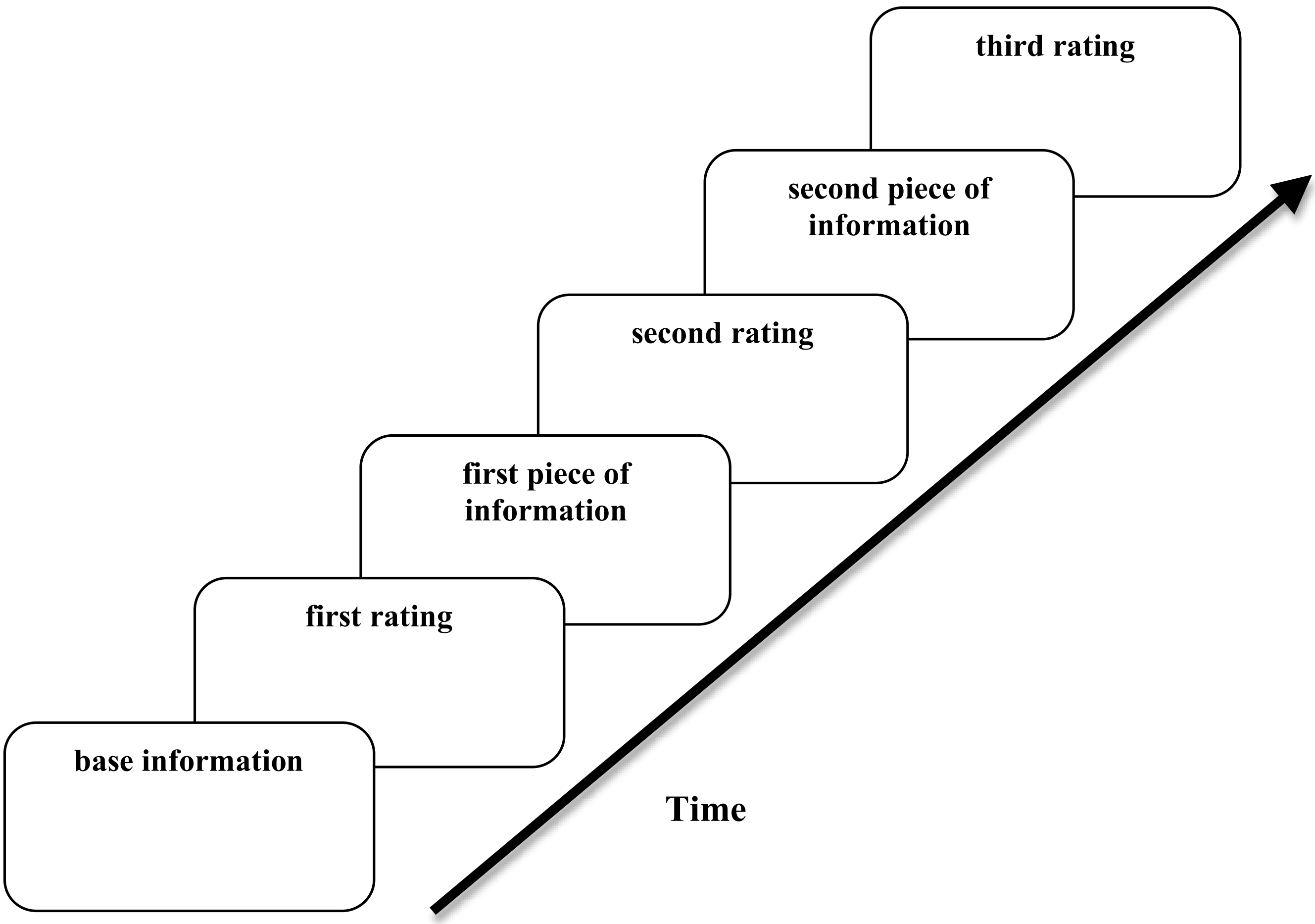

third rating
second piece of information
second rating
first piece of information
first rating
base information
Time

Figure 3

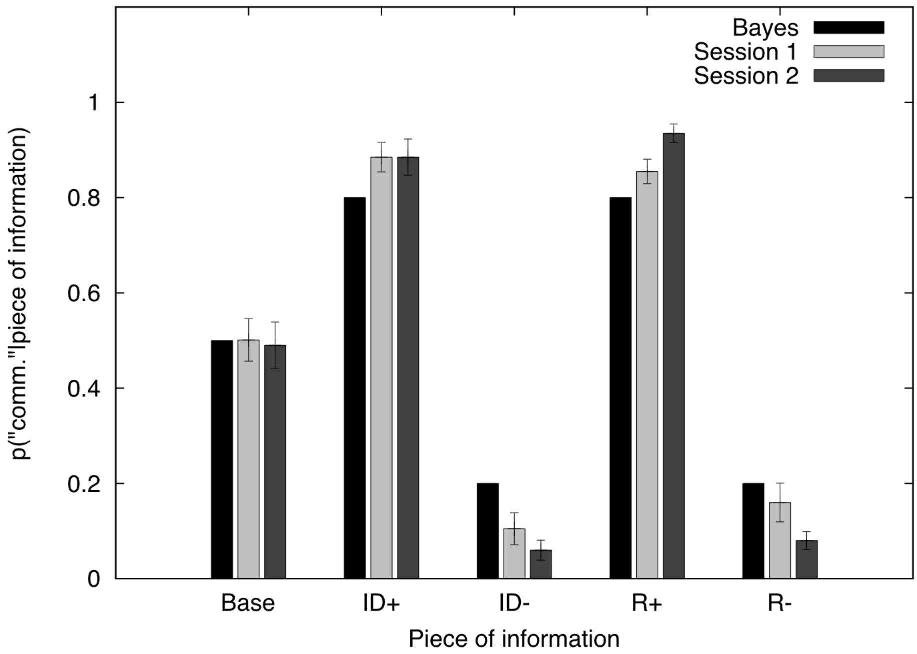

Bayes
Session 1
Session 2
p("comm."|piece of information)
1
0.8
0.6
0.4
0.2
0
Base
ID+
ID-
R+
R-
Piece of information

Figure 4

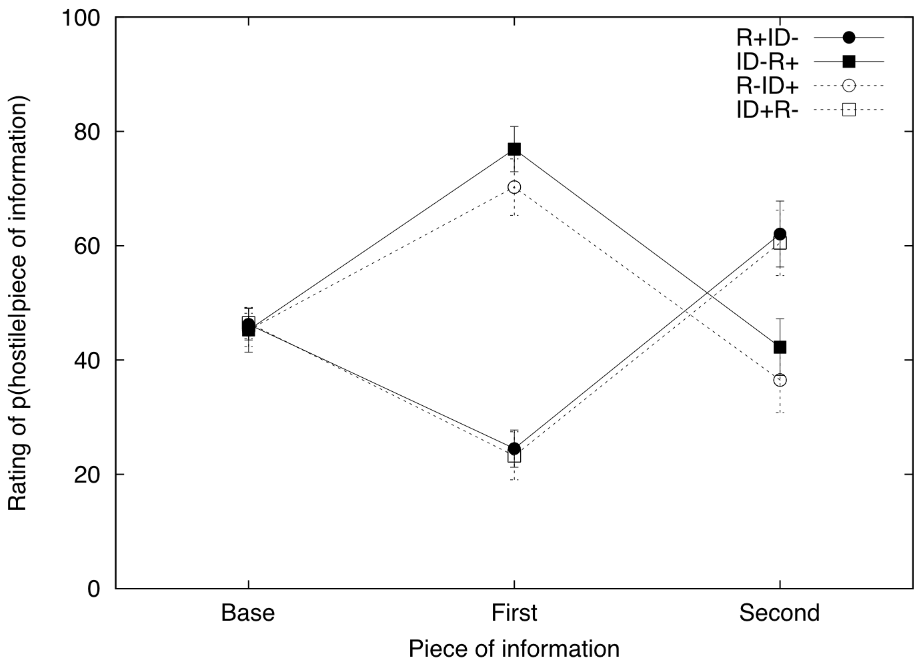

R+ID-
ID-R+
R-ID+
ID+R-
Rating of p(hostile|piece of information)
0
20
40
60
80
100
Base
First
Second
Piece of information

Figure 5

Figure 6

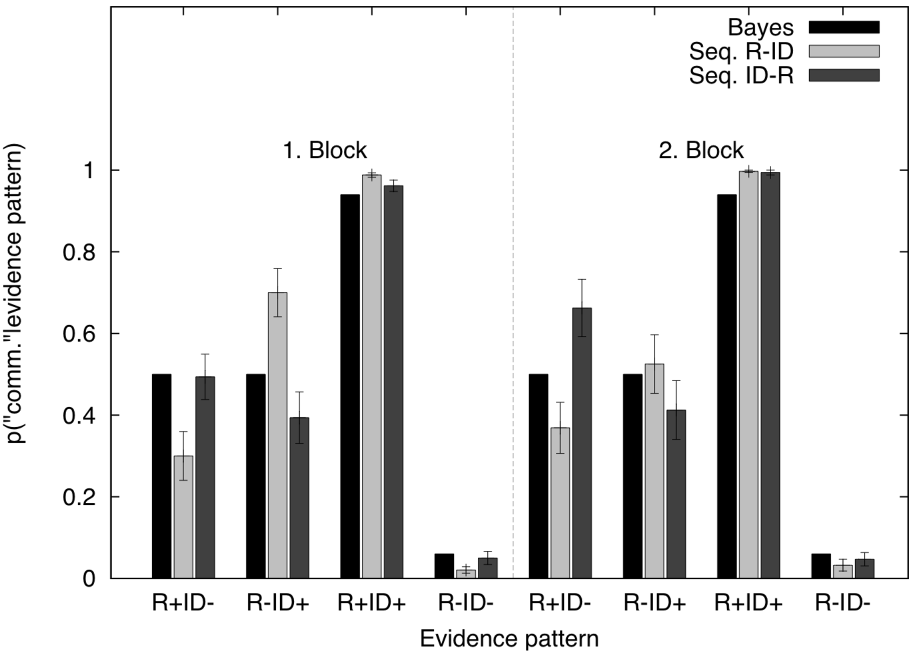

Bayes
Seq. R-ID
Seq. ID-R
1. Block
2. Block
p("comm."|evidence pattern)
1
0.8
0.6
0.4
0.2
0
R+ID-
R-ID+
R+ID+
R-ID-
R+ID-
R-ID+
R+ID+
R-ID-
Evidence pattern

Figure 7

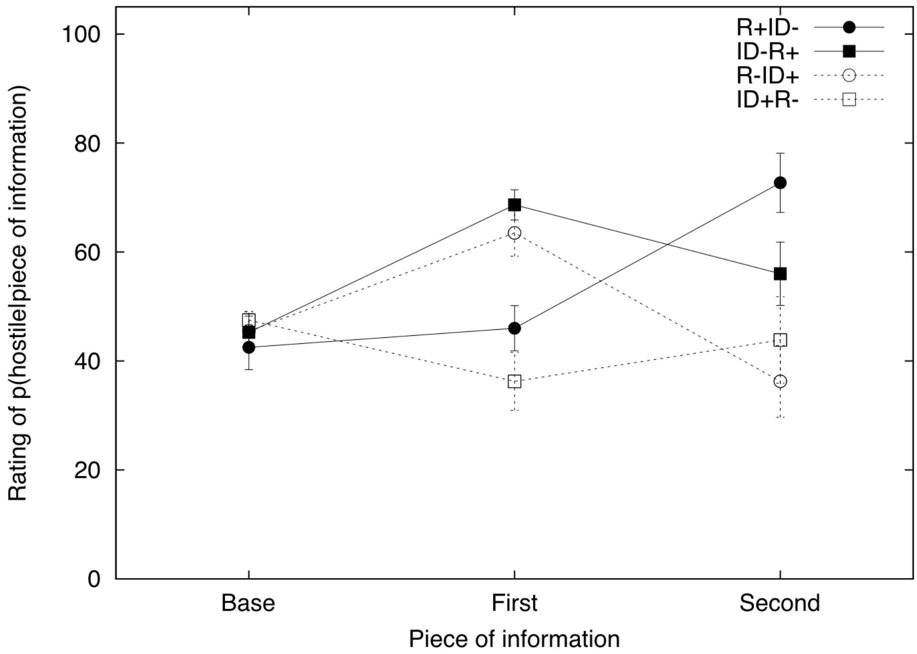

R+ID-
ID-R+
R-ID+
ID+R-
Rating of p(hostile|piece of information)
100
80
60
40
20
0
Base
First
Second
Piece of information

Figure 8

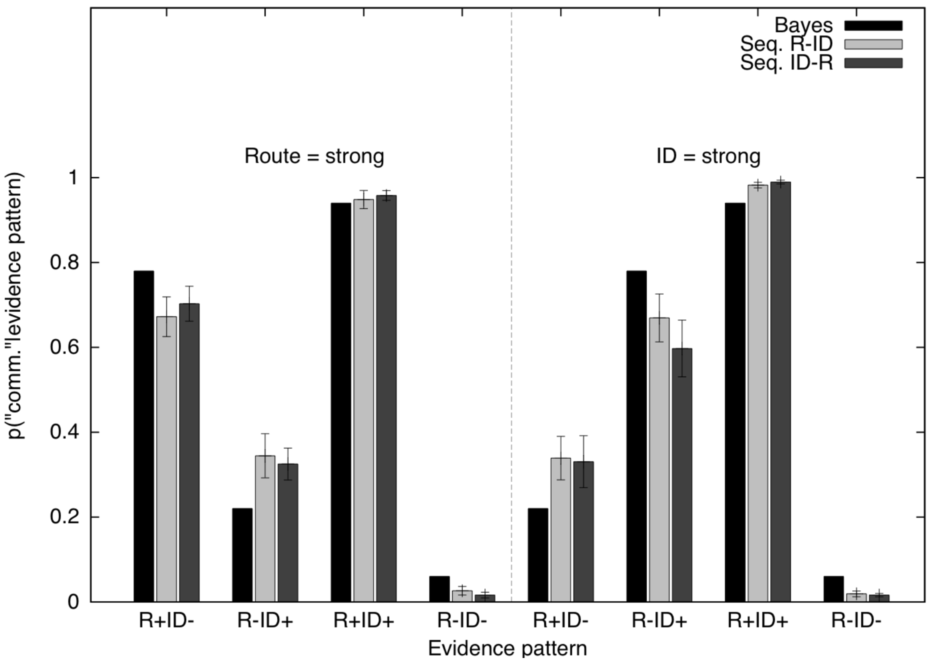
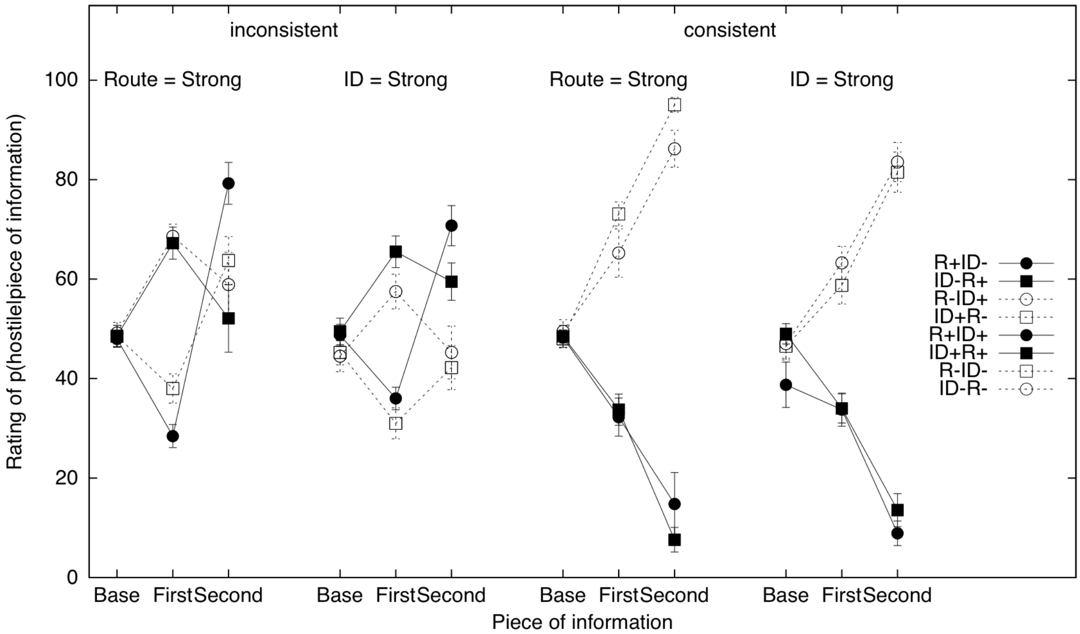

Bayes
Seq. R-ID
Seq. ID-R
Route = strong
ID = strong
p("comm."|evidence pattern)
0
0.2
0.4
0.6
0.8
1
R+ID-
R-ID+
R+ID+
R-ID-
R+ID-
R-ID+
R+ID+
R-ID-
Evidence pattern

Figure 9


inconsistent
consistent
Route = Strong
ID = Strong
Route = Strong
ID = Strong
Rating of p(hostile|piece of information)
0
20
40
60
80
100
Base
First
Second
Base
First
Second
Base
First
Second
Base
First
Second
Piece of information
R+ID-
ID-R+
R-ID+
ID+R-
R+ID+
ID+R+
R-ID-
ID-R-

Biographical sketches

Martin Baumann is team leader at the Institute for Transportation Systems of the German Aerospace Center studying drivers' information processing while driving.

Josef F. Krems is a professor of psychology at the Chemnitz University of Technology.

Frank Ritter is on the faculty of the College of IST, an multidisciplinary academic unit at Penn State to study how people process information using technology.