Changing Explanations in the Face of Anomalous Data in Abductive Reasoning

Martin R. K. Baumann (martin.baumann@phil.tu-chemnitz.de)
Franziska Bocklisch (franziska.bocklisch@phil.tu-chemnitz.de)
Katja Mehlhorn (katja.mehlhorn@phil.tu-chemnitz.de)
Josef F. Krems (krems@phil.tu-chemnitz.de)
Chemnitz University of Technology, Department of Psychology,
Wilhelm-Raabe-Str. 43, 09107 Chemnitz, Germany

Abstract

The integration of anomalous data is an essential subprocess of abductive reasoning. Abductive reasoning is viewed as a comprehension process by which observations are sequentially interpreted and explained in relation to existing knowledge. This emphasizes the importance of the reasoner's knowledge structure also for the reaction to anomalous data. In this paper we investigated the effect of the specificity of anomalous data, that is whether they are related to only one category of causes or to different categories, on how hypotheses are changed to solve an anomaly. The results show that the specificity of observations facilitates the abductive reasoning process, especially in cases where the category of hypotheses must be changed.

Keywords: abductive reasoning; anomaly; fuzzy pattern classification; changing explanations.

Introduction

Abductive reasoning is the process of finding the best explanation for a set of observations. In its simplest form this kind of reasoning can be described as follows: Given knowledge that A causes B, and B is observed, then A is hypothesized as the explanation for B (Josephson & Josephson, 1994). This kind of reasoning is part of many real world tasks such as scientific discovery, medical diagnosis, or software debugging. In these real world tasks there is often not only one observation that has to be explained but a whole set of observations and each of these observations is associated with different possible causes. For example, when a patient complains about headache the physician faces the problem that headache is a common symptom of many diseases. Furthermore, the patient often shows not only one symptom but a set of symptoms that all are associated with different causes and that become known to the physician sequentially. The physician's task is to decide between the different possible causes of the symptoms during the reasoning process and to find the combination of causes that explains all the symptoms best.

The generation of this explanation is in many cases a sequential process that can be viewed as a comprehension process by which observations are sequentially interpreted and integrated into the current explanation (Johnson & Krems, 2001). Hence, after recognizing the initially presented symptoms, the physician will generate an initial explanatory hypothesis that will be used as a context for the interpretation of following symptoms (Johnson & Krems, 2001). Because of the sequential nature of this process it might happen that a new symptom contradicts this initial hypothesis. Such a situation is called an anomaly (Krems & Johnson, 1995). In this situation the physician has to change the current hypothesis to be able to integrate the contradicting symptom into a coherent explanation of all symptoms. There are two ways to do this. The physician can either modify the current explanation so that it explains both the new symptom and the previous ones, or select an alternative explanation for the new symptom that is compatible with the explanation for the previous observations. In our experiment we focused on situations where new observations enforce the reasoner to modify the current explanatory hypothesis as the new contradicting symptom could not be interpreted in a way compatible with the current explanation.

The process of changing from an existing hypothesis to a new one is affected by several factors, such as the availability of alternative hypotheses (e.g., Burbules & Linn, 1988; Johnson & Krems, 2001; Krems & Johnson, 1995) or the entrenchment of the current explanation or the anomalous observation (Keinath & Krems, 1998).

Given the importance of knowledge for real world abductive reasoning tasks, such as medical diagnosis, the structure of the domain knowledge should also affect the process of changing hypotheses. This knowledge is often organized into different levels of abstraction such that higher order concepts form categories under which lower level explanatory hypotheses are subsumed (e.g., Arocha & Patel, 1995). Previous studies on the generation of explanatory hypotheses in the domain of scientific discovery indicate that people need to encounter substantial negative outcomes of their hypothesis evaluations before they start to consider hypotheses from a different category as relevant explanations (Klahr & Dunbar, 1988). Hence, given these results it can be assumed that solving an anomaly is more difficult if it requires to change to a new category of explanatory hypotheses than to switch to a new hypothesis within the same category.

Additionally, the relation of the contradicting observation to the current explanation's category should also be relevant. The observation could be specific and linked specifically to the hypotheses of one category or it could be unspecific and linked to hypotheses of different categories. If an anomaly is caused by a specific observation that contradicts the current category of explanations it should facilitate solving the anomaly. This observation could be used to exclude all hypotheses of the current category from

further consideration at once and to focus on those explanatory hypotheses of that category compatible with the observation. On the other hand, if the anomalous observation is unspecific it is not possible to exclude a category from further consideration as a whole and the participants should show more difficulties in solving the anomaly. Therefore the specificity of the anomalous observation should facilitate the change of hypotheses between categories.

The goal of our experiment was to evaluate the role the specificity of contradicting observations and the requirement to change the category of hypotheses play in solving anomalies in abductive reasoning tasks.

Experiment

The abductive reasoning task

To explore the abductive reasoning process we used an experimental task called "chemical accident". Participants were told the following cover story: "Imagine you are a physician at a chemical plant. After a chemical accident an employee comes to see you as he suffers from several symptoms that are caused by a chemical the employee came in contact with. It is your task to identify the chemical causing the symptoms."

Table 1: Chemical categories, chemicals, and the symptoms each chemical causes.

category	chemical	symptoms
Landin	B	breathlessness cough headache eye irritation
	T	breathlessness cough headache itching
	W	cough eye irritation itching
Amid	Q	redness of the skin chemical burn of the skin eye irritation itching
	M	redness of the skin chemical burn of the skin headache itching
	G	chemical burn of the skin eye irritation headache

As can be seen in Table 1 there were two categories of chemicals called "Landin" and "Amid" each including three chemicals: B, T, W and Q, M, G. Each chemical caused three or four typical symptoms. In total there were seven different symptoms. Each chemical could be identified unambiguously only when all its symptoms were presented. This artificial task with a small number of possible hypotheses was used a) to be able to train participants to a high degree of familiarity with the material, b) to simultaneously avoid potential effects of interindividual differences in knowledge with the use of more realistic abductive reasoning task, and c) to be able to measure the plausibility of all possible hypotheses repeatedly during the abductive reasoning process.

Figure 1 presents the general procedure of a trial. On each trial the symptoms for one hypothetical patient were presented sequentially. These could be either three or four symptoms. After the presentation of each symptom the participant had to rate the plausibility of each of the six chemicals as being the explanation for the symptoms presented so far. This was done on a scale ranging from 1 meaning "very implausible" to 7 meaning "very plausible". The order of the chemicals was randomized in this rating procedure to avoid any order effects in the ratings. After the last symptom the participants had to indicate their decision which chemical caused the symptoms.

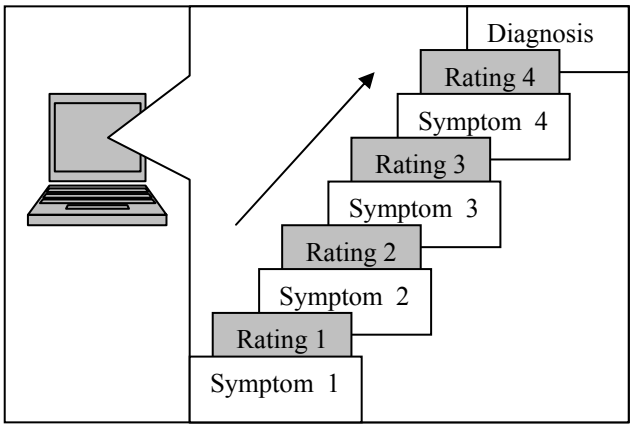

Figure 1: Sequence of events in one trial.

There were specific symptoms that were only caused by the chemicals of one category. For example "breathlessness" and "cough" were characteristic for the "Landin" category whereas "redness of the skin" and "chemical burn of the skin" were specific to the "Amid" category. On the other hand there were unspecific symptoms such as "headache", "eye irritation" and "itching" that could occur in both categories.

The symptoms could appear in a strong form such as "strong headache" or in a slight form such as "slight headache". When a symptom appeared in its strong form it was always caused by a chemical. By contrast when the symptom appeared in its slight form it could be caused by a chemical or any other unrelated circumstance. If, for example, the patient showed "strong redness of skin" this symptom could only be caused by the chemicals Q or M, whereas when the patient showed slight redness of skin this

could be caused either by the chemicals Q and M or in some rare cases by something else, such as a sunburn. The introduction of this feature allowed us to generate trials with anomalous symptoms. Slight symptoms at the beginning could be used to induce a certain hypothesis or group of hypotheses about the chemical probably causing the symptoms. Later in the trial a strong symptom was presented that contradicted the current hypothesis. To solve this anomaly the participant had to disregard the slight symptom presented at the beginning. A new hypothesis had to be generated that could explain both the new symptom and the previous ones except the slight symptom. Slight symptoms were also presented very frequently in consistent trials with no contradicting symptoms where they had to be considered to identify the chemical in the same way as strong symptoms. This should prevent participants from always disregarding slight symptoms.

There were two types of anomalous trials. In one type the anomaly could be solved by switching to a new hypothesis within the same category, for example from chemical T to chemical W of the Amid category. In the other type the anomaly had to be solved by switching between categories, for example from the chemicals B and T of the Landin category to the chemical G of the Amid category.

Table 2 shows the basic structure of each trial type by means of an example for each trial. The abstract structure of the anomalous trials, that is the sequence of specific and unspecific symptoms and the sequence of strong and slight symptoms was the same for all anomalous trials of the respective type.

The top row of Table 2 is an example of a consistent trial without contradictions. The basic principle of this kind of trials is that with each new symptom some of the hypotheses can be rejected until one remains that represents the solution. In this example, after the first presented symptom "strong headache" the chemicals B, T, M and G are plausible hypotheses. The following second symptom "slight breathlessness" allows rejecting two chemicals from the set of possible explanations. This specific symptom points directly to the chemical group "Landin" and only B and T remain as plausible explanations. The third symptom "strong cough" does not yet differentiate between B and T, but after "strong eye irritation" occurring as the last symptom, T can be identified unambiguously as the solution.

The medium row shows an example of an anomalous trial with a hypothesis switch within the category. In this example the specific symptom "slight breathlessness" is presented after the unspecific symptom "strong itching". These two symptoms indicate that the chemical T of the Landin category might have caused the symptoms. With the next symptom "strong eye irritation", the anomaly occurs as this symptom must be caused by a chemical (as it is in its strong form). For this kind of trials the anomalous symptom is always unspecific, such as "eye irritation". Because there is no chemical that causes itching, shortness of breath, and eye irritation, the only way to solve this anomaly is to assume that the "slight breathlessness" symptom is unrelated to the chemical accident. The only relevant symptoms are "itching" and "eye irritation". Therefore T cannot be the solution. After another specific symptom, "strong cough" the chemical W from the same Landin category can be identified as solution.

The bottom row of Table 2 illustrates an anomalous trial with a hypothesis change between categories. As in the within change trials, the anomaly occurs with the third symptom. The important point here is that in these trials the anomalous symptom had always to be a specific symptom, such as "cough". The anomaly can only be resolved by disregarding the second symptom, in this example "slight breathlessness", and changing in this case from the Amid to the Landin category. The last symptom identifies W from the Landin category as the solution.

Table 2: Example trials

trial type	symptom 1	symptom 2	symptom 3	symptom 4
consistent	strong headache	slight breath-lessness	strong cough	strong eye irritation
plausible hypotheses	B T M G	B T	B T	T
anomalous within	strong itching	slight breath-lessness	strong eye irritation	strong cough
specificity	unspecific	specific	unspecific	specific
plausible hypotheses	T W Q M	T	W Q	W
anomalous between	strong itching	slight chemical burn of the skin	strong cough	strong eye irritation
specificity	unspecific	specific	specific	unspecific
plausible hypotheses	T W Q M	Q M	T W	W

Fuzzy pattern classification

For the analysis of the ratings after each symptom presentation the method of fuzzy pattern classification was used (Bocklisch, 1987). This multivariate method is based on Zadeh`s fuzzy set theory (Zadeh, 1965). It is suitable for the analysis and modeling of empirical data. According to the fuzzy pattern classification method, classes are represented by patterns defined in a multidimensional feature space given by a set of relevant features derived from the measured variables. A special potential of this method is also the possibility of parallel processing of several features, e.g. using the plausibility ratings for all chemicals as a vector with six dimensions. The class

specific ranges for the features are described in a fuzzy way. That means that at each point of the feature space a membership value to each class is defined. This value represents the value of truth that an object or observation belongs to the specific class. A class membership function can be built with expert knowledge given rules based on linguistic expressions, e.g. small, medium or big intensity of a phenomenon. A second way is the calculation of a class membership function based on a large or small sample of data. Each data point is described by a membership function. These membership functions are used instead of probability functions.

The result of the classification of an unknown object is a standardized membership value gradually varying between 0 and 1. 0 means that the classified object is not a member of the class. 1 means that the object is a prototypical representative of the class. We used this method to calculate the similarity of each participant's plausibility ratings with an ideal rating behavior. These ideal ratings were defined using the poles of the rating scale. High plausibility of a hypothesis was coded with "7" meaning "very plausible", low plausibility with "1" for "very implausible". Hence, if a chemical could cause the symptoms presented so far in the current trial it was rated with "7", otherwise with "1". Between these two extremes a nonlinear transition is defined by a generalized Aizerman`s potential function (Bocklisch, 1987).

The correspondence of the participants' ratings with the defined ideal rating behaviour was then computed using fuzzy pattern classification. The results of this computation were calculated membership values for each chemical, on each rating point in time and for each single trial. These membership values express the degree of correspondence between the ideal rating classifier and the participants' ratings. High membership values (0.75 to 1) show high correspondence, medium membership values (0.5 to 0.75) show uncertainty and values lower than 0.5 mark low agreement.

Participants, Procedure, Design

11 participants, all undergraduate students at Chemnitz University of Technology took part in this experiment. The abductive reasoning task was presented on a computer. The experiment started with the learning phase where participants acquired the task knowledge displayed in Table 1 followed by a practice phase where the participants performed at least 24 practice trials. This practice phase was repeated until the participants achieved a level of 84% correctly solved practice trials. The data collection phase comprised 36 trials that were presented in random order. Four of the 36 trials were anomalous trials, two requiring a within category change of the hypothesis, two requiring a between change. The dependent variables were percentage of correct chemical identifications at the end of each trial and the plausibility ratings on a seven point scale after each presented symptom.

Results and Discussion

We will first report the results regarding the accuracy in the diagnosis task. The rate of correct diagnoses for the consistent trials was rather high (91.6%) indicating that the participants were able to solve these trials in nearly all cases. In comparison with the consistent trials, the anomalous trials were solved correctly clearly less frequently, as expected. The percentage of correct diagnoses was 47.8% in these trials demonstrating the participants' difficulties to solve the anomalies. This is similar to previous results on anomalous data in abductive reasoning (e.g., Keinath & Krems, 1998; Krems & Johnson, 1995).

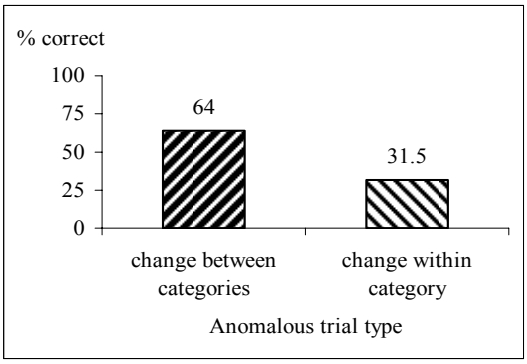

Figure 2: Percentages of correct diagnoses in anomalous trials with within and between category change of hypothesis.

Considering the diagnosis performance in the two types of anomalous trials, the results are contrary to our expectation. Based on the results of Klahr & Dunbar (1988) it was expected that anomalies requiring a change of the hypothesis within the same category should be easier to solve than anomalies requiring a change between categories. But as can be seen in Figure 2, participants identified the correct chemical much more frequently when they had to change the hypothesis between categories of chemicals to solve the anomaly than when they had to change the hypothesis within a category. The percentage of correct diagnoses in the "change between categories" trials is with 64% twice as high as in the "change within category" trials.

The fuzzy pattern classification analysis of the ratings after each symptom sheds some light on the possible reason for this unexpected result. Figure 3 shows the mean membership values of those hypotheses that should be rated as "very plausible" according to the definition of the ideal rating behavior in the two types of anomalous trials after each symptom presentation. The critical membership values are those after the second and the third symptom presentation. These represent the membership values of possible explanations before and after the anomaly.

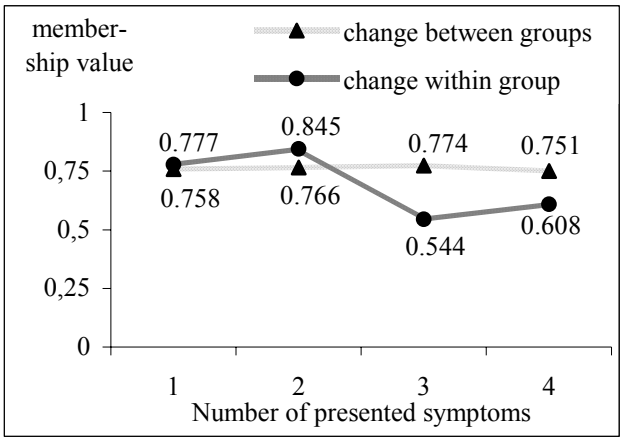

Figure 3: Mean membership values for the ratings of ideally very plausible rated chemicals after each symptom for anomalous trials with hypothesis change between and within category.

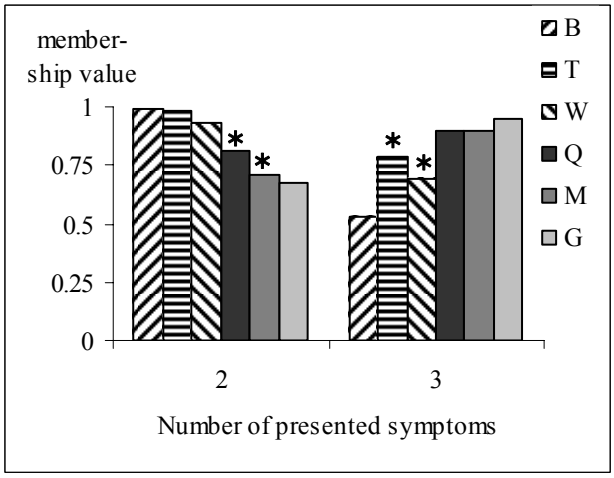

Figure 4: Anomalous trial requiring hypothesis change between categories (dashed bars: Landin, solid bars: Amid; asterisks mark plausible hypotheses).

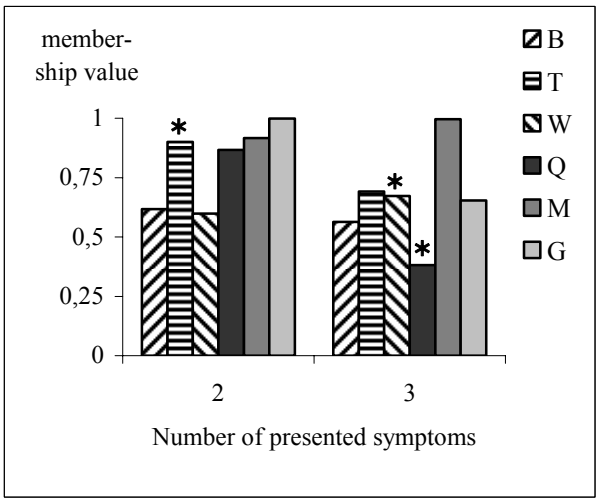

Figure 5: Anomalous trial requiring change within category (dashed bars: Landin, solid bars: Amid; asterisks mark plausible hypotheses).

Before the anomaly occurred, participants followed the ideal rating behavior in both types of anomalous trials quite well. But after the anomaly, membership values for ratings of ideally very plausible hypotheses clearly dropped in those anomalous trials requiring a hypothesis change within the category. This drop was not observed for anomalous trials requiring a hypothesis change between categories. In these trials the membership values for plausible hypotheses remained constant after the anomaly. This indicates that participants switched to the correct hypothesis to a much lesser extent after the anomaly when the new hypothesis was in the same category than when it was in the other category. Even an additional symptom after the anomaly, the fourth symptom in the trial, did not help the participants to identify the correct hypothesis within the category. This resulted in the low diagnosis performance for these trials .

Figures 4 and 5 present these averaged data in more detail as they represent the individual membership values for the different hypotheses ratings before and after the anomalous symptom in two single anomalous trials. Figure 4 presents the data for the anomalous trial requiring the change of hypothesis between categories that was presented also in the bottom row of Table 2. Figure 5 shows the membership values of the ratings of the anomalous trial requiring a hypothesis change within a category that was also presented in the middle row of Table 2. Dashed bars represent the Landin category, solid bars represent the Amid category. Asterisks mark those chemicals that represent plausible hypotheses. As we found the same pattern of membership values for the other anomalous trials of the respective types only the membership values for these trials are presented. Two features of these data shall be emphasized.

First, in the trial that required a hypothesis change between categories participants correctly recognized that the current category Amid became irrelevant after the anomalous symptom “strong cough”. Participants correctly rejected the Amid group and rated all chemicals of this category as implausible consistent with the ideal rating behavior (high membership values for the ratings of chemicals in this category). This indicates that participants recognized the anomalous symptom as a specific feature for the Landin category and were able to use it to reject that category that was not associated with it (Amid). But they were rather unsure which chemical to adopt from the new category as new hypothesis. Both the two plausible (T and W) and the one implausible hypothesis (B) show lower

membership values than the hypotheses in the irrelevant Amid category.

Second, considering the trials that required a hypothesis change within the current category Figure 5 shows that the membership values for all but one hypothesis dropped clearly after the anomalous symptom occurred leading to the low average membership value after symptom three for this kind of trials as presented in Figure 3. For this kind of trials the anomalous symptom was an unspecific symptom, such as "eye irritation" (see middle row in Table 2). It seems that participants were not able to use this symptom as efficiently as the specific symptom in the trials with a hypothesis change between trials. They were not able to reject implausible or to identify the plausible hypotheses after the anomaly.

The importance of the specificity feature of the symptoms is not only evident regarding the reaction to anomalous data. The membership values for the ratings after the second symptom of the within category change trials also indicate the importance of this symptom feature. Even though the specific symptom at the second position in the trial sequence was only presented in its slight form, it lead to the correct rejection of all chemicals of the category that were not associated with this symptom (for the data in Figure 5 it was "breathlessness" as can be seen in the middle row of Table 2). As for the specific symptom in the between category change trials that caused the anomaly, the specific symptom was efficiently used to reject one category but it did not help the participants to identify plausible hypotheses. The membership values for the plausible hypotheses after the second symptom are quite low.

Summary

We view abductive reasoning as a comprehension process This view emphasizes the importance of the domain knowledge structure for task performance, such as whether observations are specifically associated with possible explanations from a certain category or whether they are associated with hypotheses from different categories. This feature should also influence how people solve anomalies encountered in abductive reasoning. The goal of our experiment was to examine whether anomalous observations that are specifically linked to a category of explanatory hypotheses are easier explained than anomalous observations that are linked to explanations from different categories. This is especially important as previous results on scientific discovery indicate that it should be more difficult to solve anomalies that require switching to a hypothesis of a different category than to switch to a new hypothesis within the same category independent of the specificity of the anomalous observation.

To examine this question participants had to perform several trials of an abductive reasoning task where they encountered several observations sequentially. In half of the critical trials with anomalous data the participants had to switch from the current hypothesis to a different hypothesis within the same category of related hypotheses to explain all observations. In the other half of the trials the participants had to switch to a hypothesis from a different category of hypotheses. To test the above predictions, the switch within the category involved an unspecific anomalous observation, whereas the switch between groups involved a specific anomalous observation.

The results indicate that the specificity of an anomalous observation is much more important for finding an explanation for the anomaly than whether it is necessary to switch to new categories of explanations to find an explanation for the anomaly. Specific anomalous observations were used to exclude the incompatible category that was not linked to the observation thus facilitating the reasoning process. This was not possible with the unspecific anomalous observation. But both kinds of observations did not facilitate the activation of the relevant explanatory hypotheses in the remaining category.

Acknowlegements

We thank Georg Jahn and three anonymous reviewers for their helpful comments on an earlier draft of this paper.

References

Arocha, J.A., & Patel, V.L. (1995). Novice diagnostic reasoning in medicine: accounting for clinical evidence. *Journal of Learning Sciences, 4*, 355-384.

Bocklisch, S.F. (1987). *Prozeßanalyse mit unscharfen Verfahren.* VEB Verlag Technik Berlin.

Burbules, N.C. & Linn, M.C. (1988). Response to contradiction:Scientific reasoning during adolescence. *Journal of Educational Psychology, 80,* 67-75.

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

Josephson, J. R., & Josephson, S. G. (1994). *Abductive inference: computation, philosophy, technology.* Cambridge University Press.

Keinath, A., & Krems, J. F. (1998). The Influence of Anomalous Data on Solving Human Abductive Tasks. In L. Magnani, N. Neressian & P. Thagard (Hrsg.), *Philosophica,* Abduction and Scientific Discovery [Special Issue], 61(1), 39-50

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*, 1-48.

Krems, J.F. & Johnson, T.R. (1995). Integration of Anomalous Data in Multicausal Explanations. In J.D. Moore & J.F. Lehmann (Hrsg.), *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (S. 277-282). Mahwah, NJ: Lawrence Erlbaum.

Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control, 8,* 338-353.