RUPRECHT-KARLS-
UNIVERSITÄT
HEIDELBERG

Fakultät für Verhaltens- und Empirische Kulturwissenschaften

Pseudocontingencies – rule based and associative

Dissertation

zur Erlangung des akademischen Grades

doctor philosophiae (Dr. phil.)

vorgelegt dem Rat der Fakultät für Verhaltens- und Empirische Kulturwissenschaften

der Ruprecht-Karls-Universität Heidelberg

von Dipl.-Psych. Florian Kutzner

geboren am 01.03.1978 in München

Gutachter:

1. Prof. Klaus Fiedler (Betreuer)
2. Prof. Thorsten Meiser

Tag des Rigorosums: 01. September 2009

Acknowledgements

Three and a half years have passed since I started this work. Many people supported me in many different ways. I want to express my deep gratefulness:

To my Pseudo-Team. Tobbe was always up to one of countless discussions on work and all what surrounds it. He has become a friend whom I hope to keep. Peter has supported me whenever I needed and showed me how entertaining theory can be. Finally, my supervisor Klaus has not only sharpened my thinking but continuously reaffirms me that psychological research is worth spending the life with.

To Henning, who helped me getting started in Heidelberg and getting through long afternoons. Additionally, I want to thank Henning because he had a perfect timing, twice. He invited me to apply for the position when I had not thought of it yet and he invited me to think about what my thesis was when I had not thought of it yet.

To Michi, who never ceased to give me advice and gave me a great opportunity to cooperate with colleagues outside and from outside Heidelberg.

To the rest of what have been my CRISPies, Christian, Matthias, Cveta, Alice, Theo and the Hiwis, for creating a very pleasant and still critical platform for me to develop.

However, I also want to thank those who were not primarily involved in my work. My parents, Ute and Luitpold, supported me with their love and always gave me the feeling that I am working on the right thing. My friends Oli, Peter, Sebastian, Benjamin, Jan and Hannes listened and softly reminded me that there are many different perspectives.

Finally, I want to thank Livia for being involved in my work and my life. She gives me the love to carry on and never accepts a premature conclusion.

Summary

The present work puts forward a rule-based model for judging the direction of a contingency. A set of "alignment rules" (ARs) is defined, all of which bind frequent observations to frequent observations and infrequent observations to infrequent observations. These rules qualify as possible mechanisms behind pseudocontingencies (PCs, Fiedler, Freytag, & Meiser, 2009). Six experiments, involving social and non-social stimuli, are presented that pit the predictions of the rule-based PCs against associative models for contingency judgments (Van Rooy, Van Overwalle, Vanhoomissen, Labiouse, & French, 2003). Results consistently show that participants associate predictors with criteria that are non-contingent but jointly frequent and rare. Crucially, these illusory contingency judgments are shown to persist (a) in attitude ratings after extended observational learning and (b) at asymptote in operant learning. In sum, the results are evidence for the impact of rule-based PCs under conditions that call for associative learning. In a next step, rational arguments (Anderson, 1990) are used to set the AR apart from other rule-based models with similar empirical predictions. Results of two simulations reveal that the AR performs remarkably well under real-life constraints. Under clearly definable conditions, like strongly skewed base rates and small observational samples, the AR performs even better than other models, like ΔP (Allan, 1993) or the Sum-of-Diagonals (SoD, Inhelder & Piaget, 1958). Finally, the AR is claimed to be a natural by-product of the learning history with strong contingencies. Suggestive evidence from a simulation is provided that shows an increased likelihood of jointly skewed base rates, the precondition for ARs, in the presence of strong contingencies. Thus, ARs might develop from a confusion of the learned above chance probability *p* (joint-skew | strong-contingency) with an above chance probability *p* (strong-contingency | joint –skew) that justifies an AR inference. Possible future research on how joint observations and base-rates interact to influence contingency judgments is outlined.

1. Introduction

Imagine you want to decrease the rate at which you get a cold. But you have doubts about whether the common-place explanations apply to your life in particular. Your general world knowledge tells you a myriad of possible causes that might increase the risk, like stress, or decrease the risk, like sports, or might do either but you do not have an expectation in which direction, like the food you eat. To find out about what has a causal impact you are willing to change things. So the question is where to start? Probably, most people would agree that you should identify what is statistically related to catching a cold and in what direction. Exactly these judgments about the direction of contingencies, whether zero, positive or negative, are the subject of the present work.

The example of causal induction illustrates how judgments about contingencies might help us understanding and improving our daily life. In fact, knowledge about contingencies has not only been a key in causal induction (e.g., Cheng, 1997) but, more generally, has been considered the key to "explain the past, control the present and predict the future" (Crocker, 1981, p. 272). Given the prominent status of contingency judgments, a large amount of research has addressed the question of how contingencies are judged. Most models involve the contingency judgments between two binary variables. For example, a predictor with the values P1 and P2 is presented and followed by a criterion with the values C1 and C2 (c.f. Table 1).

		Criterion		
		C1	C2	
Predictor	P1	A	B	BR P1
	P2	C	D	BR P2
		BR C1	BR C2	

Table 1. Frequency table illustrating different sources of information for contingency judgment models.

One of the most prominent models of contingencies between two binary variables is ΔP (Ward & Jenkins, 1965):

$$\Delta P = \frac{A}{A+B} - \frac{C}{C+D} \quad (1)$$

ΔP compares the conditional frequencies of C1 being present given P1 to the conditional frequency of C1 being present given P2. If the conditional frequencies are different, ΔP indicates a contingency different from zero, with the direction according to which conditional frequency is larger. As can be seen, ΔP relies on the cell frequencies (A, B, C and D). For the organism trying to use ΔP these cell frequencies correspond to joint observations of predictor and criterion. These observations do not have to be observed at the same time. But the organism has to be able to coordinate the instances at least in memory. ΔP is usually accepted as a normative standard (Allan, 1980). Additionally, it has repeatedly been found to covary with human contingency judgments (e.g., Allan, 1980; Alloy & Abramson, 1979; Shanks, 1985; Wasserman, Dorner, & Kao, 1990) and thus served as a starting point for psychological models for contingency judgments.

A comprehensive review of these psychological models for contingency judgment is beyond the scope of the present work. But it is intriguing to note that judging contingencies between events that are either very frequent or very rare, i.e. characterized by skewed base rates (BR 1 ≠ BR 2 in Table 1), has repeatedly played a major role in the development of models for contingency judgments. For example, when P1 and C1 denote the presence of predictor and criterion, contingency judgments increase with an increasing frequency of C1 independent of ΔP, the so called outcome density effect (Allan, 1980; Allan & Jenkins, 1983; Alloy & Abramson, 1979). Similar effects have been found for increasing frequencies of the predictor (Allan & Jenkins, 1983, Exp. 3) and, in a somewhat unrelated body of research on stereotype formation, for jointly skewed base rates of predictor and criterion (Hamilton & Gifford, 1976; Mullen & Johnson, 1990). These influences have sparked interest in models that are sensitive to skewed base rates (Cheng & Novick, 1990; Wasserman et al., 1990). For example, an easy model initially proposed by Inhelder and Piaget (Inhelder & Piaget, 1958) compares the frequencies of predictor-criterion combinations that support a (positive) contingency (cells A and D) to those combinations that disconfirm it (cells B and C). This sum-of-diagonals (SoD) model increases with an increasing skew of predictor and criterion base rates for the same value of ΔP. A different set of models has accommodated the influence of skewed base rates within associative learning processes (Van Rooy et al., 2003). As we will see below, these models explain base rate influences as a result of pre-asymptotic learning with unequal learning opportunities. In sum, the non-normative influence of skewed base rates has been a touchstone of models of contingency judgments and will form the heart of the present work.

Until recently it has been shared consensus that skewed base rates influence contingency judgments by changing the frequencies of joint observations. For example, all of the 41 models competitively reviewed by Hattori and Oaksford (2007) converge in the assumption that the cell entries A, B, C and D of a contingency table are processed to arrive at a contingency estimate. This focus on cell entries has been challenged by the development of the pseudocontingency framework (PC, Fiedler & Freytag, 2004; Fiedler, Freytag, & Meiser, 2009) that subsumes a broader array of base rate influences.

The most radical novelty in this PC framework assumes that jointly skewed base rates are directly integrated in contingency judgments. If two variables both have a frequent and an infrequent level, PCs predict a positive contingency between the frequent levels and between the infrequent levels of the variables. Thus, in the PC framework to contingency judgments it is possible that the alignment of jointly skewed base rates is uniquely responsible for contingency judgments.

In essence, the present work seeks to gain insight into how the direct influence of jointly skewed base rates interacts with the indirect influence mediated by the cell frequencies. In three steps, I will seek to establish a rule-based model, the alignment rule (AR), to capture the direct influence. First, I will provide empirical evidence that is easily explained by rule-based models like the AR, but not by competing associative models. Then, I will use rational arguments to show that the AR is superior to other to cell-entry based models, like SoD, given clearly definable circumstances. Finally, I will illustrate possible mechanisms how the AR is acquired. I will conclude that jointly skewed base rates are likely to exert a direct influence on contingency judgments even if joint observations are available. I start by illustrating the PC framework and by reviewing empirical evidence, before defining the AR.

2. Pseudocontingencies: contingency judgments under direct base-rate influence

To define PCs, consider the simplest case of a 2 by 2 frequency table (cf. Table 1). The direct base-rate influence of PCs states that judgments about the direction of a contingency (negative, positive or zero) are directly based on base-rate information (c.f. BR P1, BR P2, BR C1, BR C2). A PC different from zero is implied if the attributes' base rates are jointly skewed, e.g. BR P1> BR P2 and BR C1 > BR C2. A PC inference then associates the frequent level of one attribute with frequent level of the other attribute, and analogously, the infrequent level with the other infrequent level. Formally, a PC can be defined as (see also, Freytag, Kutzner, Vogel, & Fiedler, 2009):

$$
\begin{aligned}
&if\!:\left.\begin{matrix} BR\ P1 > BR\ P2\ \cap BR\ C1 > BR\ C2 \\ BR\ P1 < BR\ P2\ \cap BR\ C1 < BR\ C2 \end{matrix}\right\} PC := positive \\
&if\!:\left.\begin{matrix} BR\ P1 > BR\ P2\ \cap BR\ C1 < BR\ C2 \\ BR\ P1 < BR\ P2\ \cap BR\ C1 > BR\ C2 \end{matrix}\right\} PC := negative \\
&if\!: BR\ P1 = BR\ P2\ \cup BR\ C1 = BR\ C2 \} PC := zero
\end{aligned}
\tag{2}
$$

Thus, we speak of a PC if the alignment of predictor and criterion base rates produces a contingency judgment of the form:*"If observations have frequent and rare attributes, then frequent attributes coincide and rare attributes coincide".* This different way to contingency assessment is called "pseudo," because it ignores the cell entries necessary for normative contingency assessment. In sum, PCs represent a new model of contingency judgments because it allows jointly skewed base rates, in contrast to joint observations, to cause a judgment about the direction of a contingency.

By now, there is a considerable body of evidence for this new way to contingency judgments (for a review, Fiedler, Freytag, & Meiser, 2009). The most straight forward demonstration of the base-rate-only influence arises when contingency judgments follow the aligned base rates in the absence of joint observations. For example, in a modification of an illusory correlation paradigm, McGarty and colleagues showed that a majority group was more strongly associated with a frequent valence. Crucially, participants had just been informed about the majority and the minority status of the groups and, independently, observed a stream of mainly positive behaviors (McGarty, Haslam, Turner, & Oakes, 1993). Similarly, Fiedler and Freytag (2004, Exp. 1) found contingency judgments between the diet of patients and the strength of their symptoms without being able to coordinate the both pieces of information for each patient. This evidence illustrates how jointly skewed base rates alone are sufficient to trigger contingency judgments. However, it does not provide evidence for the impact of jointly skewed base rates when joint observations are available.

Evidence that PCs intrude the realm of cell-entry based models for contingency judgments stems from research on tasks providing joint observations. Previous work has controlled the normative model ΔP while manipulating the joint skew of the base rates. An example of these conditions is the seminal demonstration of frequency based illusory correlations by Hamilton and Gifford (1976). In this paradigm, the base rates of groups and valence of behaviors are both skewed at a ratio of around 2:1. This creates the conditions for a PC resulting in a positive contingency between group A and positive behaviors. In contrast, the conditional frequencies of positive and negative behaviors are equal in both groups, thus that the ΔP model would predict a zero contingency (c.f. Table 2).

	Positive behavior	Negative behavior	
Group A	18	8	26
Group B	9	4	13
	27	12	39

Table 2. Frequency table indicating the stimulus distribution used by Hamilton and Gifford (1976, Exp. 1).

A large body of evidence indicates that judgments are in line with PCs under these illusory correlation conditions (for a review see, Mullen & Johnson, 1990) or even when the ΔP model indicates a contingency of opposing sign (Fiedler, 2009; Fiedler & Freytag, 2004). PC-inferences despite diverging

ΔPs have also been shown to generalize to different domains such as personality assessment in a clinical setting (Freytag, Vogel, Kutzner, & Fiedler, 2009), over different procedures like the IAT (Blümke & Fiedler, 2009) and affective priming (Fiedler, Freytag, & Blümke, 2009) and to situations involving multiple variables (Fiedler, 2009). In sum, there is ample evidence that contingency judgments between variables with jointly skewed base rates result in the trade-off between PCs and ΔP. However, as anticipated, different mechanisms might be involved in producing these PC effects, some rule based some associative.

The present work aims at gaining insight into how different mechanisms interact to produce these PCs that are a trade-off with the cell-entry based ΔP-model. To do so, I will study contingency judgments in the simplest case of two dichotomous variables in a single context and make joint observations available that constitute a zero contingency for the ΔP-model. Creating conditions for asymptotic learning, I will provide empirical evidence easily explained by rule-based but not associative models for PCs. I follow Shanks (Shanks, 2007, p. 297) in distinguishing the two classes of models by assuming that associative models involve the processing of the “raw” events; whereas rule-based models transform the stream of observations into concepts, such as joint frequencies or base rates. Several labels have been used to describe these rule-based approaches to contingency judgments, such as intuitive judgments strategies (McKenzie, 1994), inferential judgments (Shanks, 2007) or rules-based judgments (Allan, 1993). In the current work I will use the term rule based to avoid the ambiguity involved in the term intuition and because the models discussed are less elaborate than the inferential models that have been proposed (see 5.3.).

Subsequently, I will provide arguments for the superiority of a particular set of rules, the AR, which illustrates the nature of PCs in its most radical form. Finally, I will address the question how the AR is acquired to claim that it is likely to exert an influence on every day contingency judgments with jointly skewed base rates. Before summarizing the empirical evidence provided in the empirical articles, the next two sections are meant to illustrate how different rule-based and associative models can result in PC effects.

3. Rule-based models behind PCs

As noted by Fiedler and colleagues (2009, p. 199), jointly skewed base rates can result in PC effects via multiple mechanism, like propositional reasoning, contingency models like SoD (White, 2000) or reliability differences (Dougherty, Gettys, & Ogden, 1999). I will discuss the first two under the notion of rule-based accounts and will illustrate the latter one using the example of connectionist models of associative learning.

3.1. The Alignment Rule (AR)

The AR is meant to be an illustrative case of a rule-based PC, exclusively relying on the alignment of skewed base rates. As noted by Fiedler and colleagues (2009, p. 197), PC based reasoning is reminiscent of the atmospheric effect in syllogistic reasoning (Klauer, Musch, & Naumer, 2000). Logic syllogisms have two premises and a conclusion, which leaves room for many different forms of common relational-reasoning that bear similarities to PCs. For conditions with different ecologies "Z", a "PC-syllogism" of the following form might be considered valid: Mostly X in Z and mostly Y in Z therefore most X are Y (adapted from, Fiedler, Freytag, & Meiser, 2009, p. 197). For the present work that will deal with a single ecology, a PC-syllogism can be phrased as:

- More X than Y.
- More M than N.
- Therefore most X are M and most Y are N.

When applied to the case of illusory correlations this could read like: "More people are from Group A than from Group B and more behaviors are positive than negative; therefore most people from group A are positive and most people from Group B are negative." Of course, this PC-syllogism does not qualify as a logical syllogism as it does not have a valid figure and does not use conventional quantifiers (Johnson-Laird & Byrne, 1991, p. 106). However, it serves well to illustrate what the difference is between normative logic and a PC. Even with adequate quantifiers, logic could not relate the first and the second premise, as it requires one term to be present in either premise, in other words a joint observation. Exactly this is what a PC does not need to infer a relation.

Consequently, and central to the present work, I define the "alignment rule" (AR) as any kind of rule that relates two variables by the alignment in the skew of their base rates without making use of joint observations. The PC-syllogism above is but one form. In a more simple form, the AR can also be formulated similar to fast and frugal heuristics (Gigerenzer & Todd, 1999): *"If two variables have a frequent and an infrequent level, then the frequent levels occur together and the infrequent levels occur together"*. Even more basic, ARs could simply match observations by similarity, specifically by similarity of skew. As Goodie and Fantino (1996) showed, observations are readily matched by perceptual or intensional (Fiedler, 1985, p. 48) features, creating responses mirroring contingency judgments. In analogy, the statistical feature "skew" might be used to match observations in the same way.

Depending on the task, it is possible that even more basic processes contribute to PC effects. In evaluative priming or the IAT jointly skewed base rates might simply lead to an alignment of response tendencies, i.e. most of the time using one key to respond, which can explain PC effects (Blümke & Fiedler, 2009; Fiedler, Freytag, & Blümke, 2009). The present work however, does not create task conditions that allow for simply aligning response tendencies. Consequently, I will focus on the slightly more demanding rule-based mechanisms.

Before discussing the relation between the AR and other rule-based models behind PC effects, I will introduce a possible quantification. I do so to further specify an assumption. Specifically, I assume that using the AR becomes more likely given a stronger skew in the base rates. This assumption is driven by the fact that, to my knowledge, there is no evidence for PC effects for a joint skew lower than 2:1. I quantify the AR by the product of the log-transformed base-rate ratios:

$$\text{logAR} := \log_{10} \frac{\text{BR P1}}{\text{BR P2}} \times \log_{10} \frac{\text{BR C1}}{\text{BR C2}} \tag{3}$$

This definition captures the crucial properties of the AR: exclusive focus on base-rate information, the direction implied by a PC (i.e. positive if skewed in the "same" direction, negative otherwise, and zero if at least one base rate is not skewed) and stronger predictions given a larger skew of the base rates. Additionally, it is symmetric with regard to the base-rate ratios.

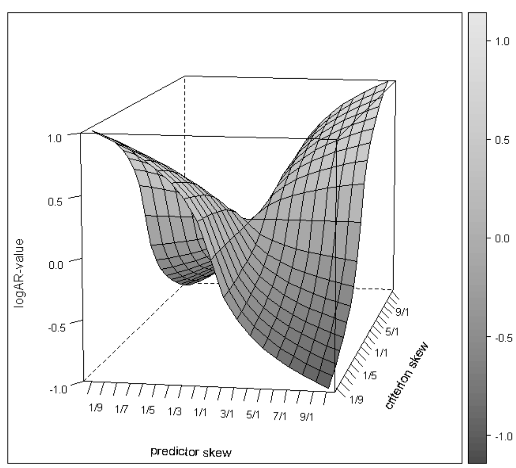

Figure 1. Illustrates the predictions of the AR model as quantified by logAR (see Equation 3) as a function of predictor and criterion base rates varying from 1/10 to 10/1.

Figure 1 graphically illustrates these properties by plotting the size of logAR as a function of predictor and criterion skews. In sum, a mechanism involves an AR if it exclusively makes use of base-rate information and predicts a contingency along the alignment of jointly skewed base rates.

3.2. Other rule based models behind PCs

This section serves to illustrate how the AR coincides with other rule-based models of contingency judgments that rely on cell frequencies. This is important for the empirical part of the present research because some cell-entry based models arrive at the same predictions as AR for the specific tasks and stimulus distribution of the present experiments (see Table 3).

	C1	C2	
P1	180	60	240
P2	60	20	80
	240	80	

Table 3. Example of the frequency tables used for the present experiments.

Specifically, the following rules indicate the same contingency judgment, a positive relation between P1 and C1, as the AR: relying on cell A having most entries (A-Cell model); relying on the confirmatory diagonal (A+D) having more entries than the disconfirmatory diagonal (B+C, Sum-of-Diagonals model, SoD); relying on positive hits, i.e. cell A entries, being more frequent than false positives (cell B, Hits-minus-False-Positives model, HFP). All of these models (see also, Freytag, Kutzner et al., 2009) have received at least some empirical support with a subset of participants using them (Arkes & Harkness, 1983; Shaklee & Mims, 1981; Shaklee & Tucker, 1980; Wasserman et al., 1990). Thus, it is possible that participants' contingency judgments in the present work reflect either of these models or a combination thereof. Because of this possible confound in the empirical demonstrations of the AR, I will use rational arguments for setting AR apart from the above mentioned other strategies concluding that it is remarkably accurate under ecologically plausible conditions (see, Freytag, Kutzner et al., 2009).

However, it is also of theoretical interest to note that different models designed to capture deviations from the normative ΔP-model arrive at similar predictions. Under which conditions the predictions of cell-entry based models and the base-rate based AR coincide might be a first hint how both sources of information interact to create contingency judgments. To assess the relation between the predictions of the AR, logAR and the other models, I extended the simulations reported by Freytag and colleagues (2009). For all possible 2-by-2 frequency tables that result from varying the cell frequencies independently from 1 to 20[1] (160.000 tables) the models are calculated and subsequently correlated. Since the AR is assumed to gain impact with increasing joint skew of the base rates, the

[1] The range of this simulation is restricted in comparison to the remaining simulations (1 to 50) due to computing limitations.

results are not only aggregated over the whole range of skews but are also split by the joint skew of the base rates. I created four exclusive and exhaustive categories of joint skews (see Table 4) ranging from no-skew (1:1) to a rather strong skew (4:1). For example, contingency tables were included in the 3:1 category if predictor and criterion base rates were skewed at a ratio of at least 3:1, but not both stronger than 4:1.

Table 4. Correlations between the results of different models for contingency judgments for two dichotomous variables as a function of the joint skew in the base rates.

strategy	AR	logAR	A-Cell	SoD	HFP	ΔP
AR	1	.59	.00	.00	.00	-.09
logAR	.97/.98/.94/.70	1	.13	.21	.14	.06
A-Cell	.41/.43/.40/-.02	.45/.44/.39/.09	1	.71	.86	.70
SoD	1.00/1.00/.78/-.06	.97/.98/.81/.09	.41/.43/.33/.74	1	.58	.67
HFP	.67/.62/.45/-.04	.66/.61/.46/.05	.92/.92/.90/.86	.67/.62/.50/.59	1	.67
ΔP	.87/.71/.36/-.12	.86/.70/.39/-.06	.58/.56/.48/.72	.87/.71/.57/.80	.62/.54/.47/.68	1

Note: Values in the upper half correspond to correlations for all tables (N=160.000). Values in the lower half correspond to all tables of skews 4:1 (N=524), 3:1 (N=1.216), 2:1 (N=8.040), 1:1 (N=150.220), respectively.

The results can be summarized as follows. Over the whole range of skews (upper part of the matrix), the AR is largely independent of all the other models. However, given the preconditions of the AR become prominent, i.e. a joint skew of at least 2 to 1, the AR correlates to a considerable degree with SoD, HFP and A-Cell (minimal r=.40). This implies that, given jointly skewed base rates, cell-entry based and base-rate based models will often yield the same judgment. From an ecological perspective, the AR and other models based on joint observations do not contradict each other. To the contrary, they might often be integrated without confusion in the presence of joint observations. In the absence of joint observations, the AR might smoothly compensate the missing information of the other models.

For decisive evidence isolating AR-based mechanisms, future research could follow the rule-analytic approach of Shaklee and Mims (1982), selectively using frequency distributions where the strategies yield different results. In the empirical part of the present work however, all these models are possible rule-based mechanisms behind PCs that are pitted against another class of models that can explain PC effects by differential sample size. The next section will illustrate the basic idea behind connectionist models as an example of this set of models.

4. Associative models and sample size behind PCs

Models of associative learning propose that jointly skewed base rates bias contingency judgments because of the size of the psychological sample on which the judgments are based. They build on the fact that skewed base rates necessarily leave one predictor value to be considered more often than the other in combination with the criterion. These models, as the rule-based models except the AR, rely on joint observations of predictor and criterion. I will now highlight the basic mechanism that elicits this sensitivity to sample size in connectionist models for contingency judgments, competition for associative strength (e.g., Van Rooy et al., 2003). It is crucial for the present work that connectionist models imply a decreasing impact of jointly skewed base rates for an increased amount of overall learning.

Competition for associative strength is at the heart of models for associative learning (e.g., Rescorla & Wagner, 1972) that have been used to describe animal and human contingency learning (Allan, 1993) and have been integrated into more complex models for social judgments (Van Rooy et al., 2003). The critical assumptions are that associative strength between concepts, like predictors and criteria, increases with pairings and is limited to a maximum to which learning converges. The delta-learning rule in the contiguity theory by Rescorla and Wagner (1972) describes how an association gradually increases with every pairing of predictor and criterion:

$$\Delta V_n = \alpha\beta\ (\lambda - \Sigma V_{n-1}) \tag{4}$$

where ΔV_n is the change in associative strength for the predictor when it is paired for the n^{th} time with the criterion and α and β are learning rate parameters that depend on the salience of the predictor and the strength of the criterion. Crucially, the decrease in the amount of association change results from the difference between the maximal associative strength (λ), the limit of associative strength supported by the criterion, and the previously acquired associative strength (ΣV_{n-1}). Thus, the stronger the previously acquired associative strength of a predictor with a criterion, the less another pairing will change the association. Finally, when the change reaches zero and the learning process has reached its asymptote.

Applied to the simplest case of contingency learning between two dichotomous variables, the delta-learning rule implies that both predictor levels are competing for associative strength with both criterion levels. If there is a contingency, for example p(C1|P1)=.75 and p(C1|P2)=.25, it can be shown that the difference in associative strengths between P1 and C1 and between P2 and C1 will equal the normative contingency index ΔP at asymptote (Chapman & Robbins, 1990). This is also true for the case of PCs as they are studied in the current work. Since ΔP is zero, i.e. the conditional probability of C1 given both predictor values is .75, the difference in associative strength should be zero at asymptote. However, since P1 is presented more frequently, associative strength will increase faster for this predictor value. This implies that preasymptotic differences will mimic a contingency between P1 and C1.

In sum, the very nature of PCs implies psychological samples of different sizes. Some predictor values are observed more frequently than others and some criterion values are observed more frequently than others. If, as in the present research, joint observations of predictor and criterion are possible and the contingency between predictor and criterion is zero, more frequent predictor values will be associated faster with more frequent criterion values than less frequent predictor values. Thus, preasymptotic learning results in a PC. Crucially for disentangling these sample size mechanisms from rule based accounts, asymptotic learning should not result in a PC. This idea underlies the following two empirical demonstrations of rule-based mechanisms behind PCs.

5. Empirical evidence: rule based and associative

The goal of the following experiments is to disentangle possible mechanisms behind the influence of jointly skewed base rates on contingency judgments, i.e. for PCs. Competition for associative strength in connectionist models explains PCs for preasymptotic learning in short learning sequences. In contrast, the AR predicts PCs for preasymptotic and asymptotic learning, because the extraction of jointly skewed base rates can be assumed to happen rapidly and remain largely invariant with extended experience (for the robustness of base-rate extraction see, Reips & Waldmann, 2008).

Consequently, the experiments create conditions where both mechanisms can be expected to operate. For the connectionist models this implies learning proper. That is predictors have to be followed by criteria to create joint observations. Additionally, associative learning seems more appropriate when the criterion is associated with reinforcement. Thus, I created operant conditions, i.e. presented valenced stimuli as the criterion and/or reinforced correct predictions of the criterion. In contrast, creating conditions where the AR can be expected to operate implies using stimuli with markedly skewed base rates. I followed previous experiments which have reliably produced PC-effects with a base-rate skew of 3:1. Over all experiments, two crucial characteristic of the stimulus distribution were maintained to pit the mechanisms against each other (a) learning was extended as compared to previous studies on PCs (e.g. 96 observations in Fiedler & Freytag, 2004, Exp. 1) and (b) the contingency, i.e. ΔP, between predictors and criteria was zero (c.f. Table 3).

Based on both models, connectionist and AR, I hypothesized that participants' judgments would reflect a contingency between frequent predictors and frequent criteria early on in the learning sequence. Based on the AR, I additionally anticipated the same contingency judgment after extended experience.

5.1. Extended operant learning (Kutzner, Freytag, Vogel, & Fiedler, 2008)

This series of three experiments extends PCs to operant learning. On every trial of a learning sequence, either 160 or 320 trials long, participants hear one of two acoustical tones and have to press one of two keys. Irrespective of the tone, one of the keys was defined to be the "correct" response in 75% percent of trials, the other in the remaining 25% of trials. Correct responses were reinforced with 3 cent and incorrect responses punished with 3 cent. As dependent measure, I assessed the rate with which participants chose the more frequently reinforced response given the two tones. If the response rates differ conditional on the tone, this can be seen as evidence for a contingency judgment (c.f., Allan, 1993).

Results clearly indicate that response rates after 160 or 320 learn trials differ. The rate for choosing the more frequently reinforced response is higher after the more frequently presented predictor-tone. Additional results (not reported in Kutzner et al., 2008) indicate that the same is true early on in the learning sequence, with rates differing already for the first 32 trials ($T(20)=2.07$, $p=.052$). It also seems that associative learning had reached an asymptote since the difference in choice proportions did not change from the first to the last half of trials in Experiment 3. Thus it is unlikely, though possible, that further training would have reduced or eliminated the difference in response rates.

These results contribute to our understanding of PCs in at least three ways. For the first time PCs have been shown in an operant learning setting. This speaks to the impact and robustness of the phenomenon. PCs are shown in that the more frequently reinforced response is chosen at a lower rate after the infrequent predictor value. Thus they are shown even though this implies foregoing the chance to maximize reinforcement. Second, manipulating only the criterion base rate was not sufficient to produce a PC effects. Though predicted by associative models and the AR, usually both base rates have been manipulated at the same time. Thus, the results are direct evidence that it is the joint skew, or alignment of base rates that causes PCs. Third, and crucial for the current argument, the results are not easily explained by the current models of associative learning. A difference in choice proportions remains at asymptote in spite of an experienced zero ΔP between the cues and correct choices. The next series of experiments will extend these insights to a social context.

5.2. Extended learning in stereotype formation (Kutzner, Vogel, Freytag, & Fiedler, 2009)

Building on the first series of experiments, I studied PCs in extended learning in a social domain by adopting a standard paradigm, illusory correlations (ICs, Hamilton & Gifford, 1976). On every trial, participants first saw a group label, either referring to an "orange" or a "purple" group and then a valenced behavioral description. Again, the behavior was positive or negative in 75% percent of the time, irrespective of the group (c.f. Table 3). Participants either passively observed the information like in standard IC paradigms or had respond to the group labels, predicting the upcoming valence or making approach-avoidance decisions towards the groups. Subsequently, the groups were evaluated.

Again, results provide clear evidence for PCs after extended experience with information about the social groups. Under all conditions, the more frequent group, the majority, was evaluated more positively if the behaviors were frequently positive and the minority was evaluated more positively if the frequent behaviors were negative. Additionally, analogous to the previous experiments we found evidence for the usage of this illusory stereotype, which constitutes a first demonstration of IC-based discrimination. When behaviors were frequently positive, majority members were predicted to show positive behaviors more frequently than minority members and were also approached more frequently.

In sum, the results of the two series of experiments provide clear evidence for mechanisms that go beyond associative learning. Still, when learning had reached an asymptote, statistically non-predictive cues influenced online responses and end-of-learning judgments as predicted by PCs. One possible mechanism behind this influence is the AR, a rule based mechanism that is based on predictors and criteria sharing levels that are frequent and rare.

In the current experiments, other rule based models that rely on joint observations instead of base rates, specifically SoD, A-Cell and HFP, arrive at the same predictions as the AR. Thus, it is possible that participants did rely, at least in part, on other rules as well[2]. However, providing unequivocal evidence

[2] Note that it is impossible (SoD) or arbitrary (A-Cell, HFP) to create stimulus distributions that diverge in predictions for AR and the other models, given a ΔP of zero.

for but one model, the AR, was not the aim of the empirical work as this has been done elsewhere (Fiedler & Freytag, 2004; see also, McGarty et al., 1993). Instead, it served to establish that rule-based mechanisms are at work even under conditions that clearly call for associative learning.

5.3. Discussion

The conclusion that rule-based mechanisms were at work in the present studies can be challenged either by suggesting that other associative models or a different class of models, neither associative nor rule based, predicts the same pattern of results.

PCs, a new form of blocking? I have applied the associative models for the current experiments by assuming that both criterion levels compete for associative strength with each predictor independently. For example, it is assumed that each group will be associated with each valence according to the relative frequencies with which that group is paired with the two valences. However, one might suspect that the faster increase in association between the frequent predictor and the frequent criterion level "blocks" or hinders the formation of the association between the infrequent predictor and the frequent criterion level. Indeed, the obtained pattern could be the result of a blocking experiment (e.g., Dickinson, Shanks, & Evenden, 1984). For example, in a forward blocking experiment, first the majority alone and then majority and minority together would be paired with positive and negative behaviors. Under these conditions one could expect that (a) the majority is more strongly associated with the frequent valence than the minority and that (b) both groups are to the same degree weakly associated with the infrequent valence. This would result in a larger difference between frequent and infrequent valence for the majority. These predictions follow from the assumption that cues, i.e. the predictor values, compete for a limited amount of associative strength.

Though plausible, there is a crucial difference between the current empirical demonstration of PCs and blocking experiments: majority and minority have never been paired. To my knowledge, no associative accounts exist that would predict this forward blocking without pairings of the competing cues. Thus, the current versions of associative models do not provide an explanation of the present results.

However, one could frame PCs as a new form of blocking within inferential accounts to forward blocking. These accounts assume inferences like *"If cue A on its own causes the outcome to occur with a certain intensity and probability, and if cue A and T together cause the outcome to occur with the same intensity and probability, this implies that cue T is not a cause of the outcome "* (De Houwer & Beckers, 2003, p. 346). A PC account would simply remove the necessity for observing both causes together and imply an inference like *"If cue A on its own causes the outcome to occur with a certain intensity and probability, and if ~~cue A and~~ T ~~together~~ causes the outcome to occur with the same intensity and probability, this implies that cue T is not a cause of the outcome. "* This can be taken as another illustration of how PCs work: they remove the need for joint observations.

However, to render this PC-blocking an alternative explanation of the present results an additional assumption is needed to explain that the frequent predictor, i.e. cue A, is the focus of attention that blocks the infrequent predictor. Recently, Sherman and colleagues (Sherman et al., 2009) tried to argue that attention is allocated to the majority first exactly because it is more frequent. Even though plausible, this argument seems circular. Additionally, it is based on attention theory (Kruschke, 2001) which has been developed using compound stimuli, i.e. cues A and T presented together. Thus, it remains unclear whether PCs can fruitfully be considered a new form of blocking. Crucially for the current results, this model could also be considered rule based, though the rule seems less parsimonious than the AR. Thus, associative accounts for blocking do not threaten the validity of the current conclusions. Before discussion how associative models could be revised to accommodate the present results, I will discuss another class of models sharing the prediction that pre - but not asymptotic learning will reveal PC effects.

Memory trace models and noise. Exemplar based models have also been used to explain the influence of jointly skewed base rates on contingency judgments, most prominently the account by Smith (1991) based on Hintzman's (1986) memory model and the BIAS model by Fiedler (1996). For example, Smith's model assumes that every observation of a predictor and a criterion leaves a trace in long term memory that has different sections to code several aspects of the observations, such as group membership and valence. If a judgment about the valence of one group is required, these models prompt the memory with an ideal probe of the group-section and receive an echo with an aggregate over all relevant valence sections. Both models implement the final valence judgment as the correlation between the aggregated memory echo for the valence section and the ideal valence section. Crucially for sample-size effects, the models also assume random noise at different stages of the process, e.g. forgetting or less than perfect learning. Consequently, more frequent exposure to one groups results in a more reliable valence aggregate for this group. Thus, the predictions regarding an extended array of observations are similar to those of the connectionist models. The reliability advantage for the more frequently observed group or predictor should decrease with increasing observations. This is due to the fact that, as the number of observations increases, the correlation of the memory output and the ideal prompt will approach its maximum of 1 for the frequent and the infrequent predictor value.

However, these models could be modified to accommodate the results. One possibility would be to render forgetting for a target trace a function of how many traces have been stored since this target trace. This would effectively create a subset of traces that only consists of recent observations. Psychologically, this could be interpreted as retroactive inference. If forgetting increases fast enough with new observations, this would mimic the limited set of observations for which the models explain PC effects. The same is in principle possible for connectionist models even though they do not usually implement forgetting.

In sum, the evidence is in line with the AR as one possible mechanism behind PCs. In the next two sections I will argue that the AR is also a likely mechanism behind PCs. First, I will broaden the scope from the mechanisms that drive PC effects, to the mechanisms that should drive them. I do so to argue that ARs are an adaptive rule for detecting the sign of a contingency and are well suited to work under real life constraints.

6. Adaptive value of the AR (Freytag, Kutzner, Vogel, & Fiedler, 2009)

The analysis of the adaptive value of the AR has two parts, one that could be termed “possibilistic”, the other rational, as it follows the steps 1 to 4 of a rational analysis (Anderson, 1990): define the goals of the cognitive system, develop a formal model of the environment to which the system is adapted, make assumptions about processing limitations and derive the optimal model.

The first part follows the seminal work of McKenzie (1994) who showed that seemingly naïve strategies, like HFP, are quite congruent with normative indices when considering every possible distribution of a 2 by 2 frequency table in a range from 1 to 50. When including the AR (c.f. Equation 2) into the simulations, it indicates the same sign of the contingency as normative indices in 73% of the cases when the skew is at least 3:1. When considering only contingencies of moderate to strong size, larger than .2, the accuracy increases to 95% (see, Freytag, Kutzner et al., 2009, Simulation 1). Thus, the AR is valid for detecting the sign of a normative contingency when its preconditions are met, jointly skewed base rates. Additionally, it performs at a similar level as the other models. However, the AR’s validity in such a possibilistic set of environments is but a first step to determine its adaptive value for a decision maker in real life.

In a second step, the performance of the AR is evaluated after specifying its exact task, taking into account real life constraints and specifying environments in which it possibly has to work. Four assumptions are made. To begin with, the critical task is defined as inferring the sign of a contingency in a space of possible observations based on a sample. This is different from the previous analysis where the task was to correctly infer the sign of a contingency in the sample, i.e. the specific 2 by 2 distribution. This inference task might be closer to how contingency judgments are used in real life, to generalize beyond the observed sample. Coming back to the opening example, trying to determine whether catching a cold in our life is contingent on stress implies being interested in the future. The past observations only serve as a tool for inferring this future contingency. Consequently, we changed the simulation by creating spaces of possible observations in a first step, and by repeatedly drawing samples from these spaces in a second step. Note that this change in the task opens the possibility that the contingencies that hold in the spaces of possible observations are not the same ones as in the samples due to sampling error. Consequently, the standard for determining accuracy is now the contingency in the space, not the sample. This renders normative indices that work on the sample just another fallible strategy.

Further, the assumption is made that information might be unavailable in two ways. First, the amount of observations is restricted for a given contingency-inference task. This might be due to memory loss in the cognitive system when events are spaced in time. It might also be due to the fact that making observations comes at a cost. Thus the simulation implements samples of limited size, i.e. 14 observations. Second, not all of these observations are joint observations of predictor and criterion. For example, when confronted with the task of determining the stress-cold contingency for you, joint observations should be available. However, when trying to do so for your partner or child, you might only know about a cold but not about stressful events, leaving univariate observations. This constraint is implemented by varying the likelihood of making joint observations from 10% to 100%.

Finally, it is assumed that the contingency to be detected is of moderate size, i.e. .3. Though arbitrary, it seems plausible that many important contingency inferences do neither involve close to deterministic relations nor extremely weak relations[3]. Because, our aim is to evaluate the performance of the AR, we also vary the joint skew of predictors and criteria. This creates conditions where the AR has been found to be influential, i.e. when there is a joint skew of at least 2 to 1.

---

[3] Changing the size of the contingency to .7 does not change the qualitative conclusions drawn from the results (see also, Freytag, Kutzner et al., 2009).

The results clearly indicate when AR inferences are valid and even outperform other rule-based models (Freytag, Kutzner et al., 2009, Simulations 2 and 3). When the joint skew in the space of observations reaches a level of 2 to 1, AR correctly predicts the sign of the contingency in about 75% of the cases. Because the base-rate estimates can make use of univariate observations, AR's performance is only slightly impaired when the likelihood of joint observations decreases. This is not the case for its closest competitor, SoD. Although performing well on average, the SoD model is sensitive to the proportion of joint observations. This property results in a relative advantage of the AR over the SoD when joint observations are unlikely. Additionally to being sensitive to the likelihood of joint observations, the ΔP model shows a decrease in performance when the joint skew of the base rates increases. This leads to a relative advantage for the AR when the joint skew is high and joint observations are unlikely.

In sum, the AR can be deemed a rational model under certain constraints. Specifically, the AR is the optimal rule for inferring the sign of a (moderate) contingency when predictors and criteria are jointly frequent or rare, at a ratio of at least 2 to 1, and when the likelihood of observing predictors and criteria together is no more than 30%. If the latter condition is violated, the AR is at least as good as the other rule-based models.

7. The origins of the AR

So far I argued that the AR is a possible mechanism behind PCs and, under real-life constraints, it is rational to use it. However, how should organisms develop exactly the AR as a strategy among the myriads of other possible rules? Answering this question is crucial because the demonstration of usage or adaptive value alone runs the risk of fitting data in a post hoc fashion by yet another rule-based model. Therefore, in the next section I will provide reasons why organisms are prone to acquire this specific rule. I will do so for operant and stimulus-stimulus learning conditions. The basic argument builds on the fact that, as I will show, experienced base rates will be jointly skewed given strong experienced contingencies in the environment.

Operant conditions. I will start discussing operant conditions that assume a causal relation between the behavior shown by the organisms and reinforcement. For example, imagine you finally found out that those weeks of stress are almost always followed by catching a cold, i.e. stress and health are strongly correlated in your life. Most likely you will try to avoid stressful weeks. This example depicts an operant learning situation because your behavior is reinforced and punished along a systematic schedule defined by the causal relation in the environment. The example also illustrates the dominant reaction to such scenarios, probability matching (e.g., Shanks, 1990). Over a very wide range of conditions, humans as well as animals choose behaviors at the rate at which they are reinforced. Given a strong contingency between two behaviors and reinforcement, for example *p* (reinforcement | behavior A) = .8 and *p* (reinforcement | behavior B) = .2, resulting in a ΔP of .6, organisms will show the behavior A 80% of the time.

Crucially, this probability-matching behavior creates jointly skewed base rates. To illustrate this consider Table 5. In the left part the reinforcement schedule described above is applied to an organism that guesses, i.e. behaves at random. This might be an illustration of an early stage in a learning process. In the right part, the same reinforcement schedule is applied to an organism showing probability matching, i.e. choosing the behaviors at the rate at which they are reinforced. The causal relation between behaviors and reinforcement results in a skewed base rates of the reinforcement as well, in this example a logAR of .2 or an average skew of 3 to 1.

Guessing	Reinforcement	Punishment		Probability matching	Reinforcement	Punishment	
Behavior A	40	10	50	Behavior A	64	16	80
Behavior B	10	40	50	Behavior B	4	16	20
	50	50	100		68	32	100

Table 5. Expected frequencies for the same operant reinforcement schedule for guessing (left part) and probability matching (right part).

Thus, given differential reinforcement for behaviors we can control, the normal way of responding will create jointly skewed base rates of behaviors and reinforcement in our experience. Before discussing how this might lead to rely on ARs, consider situations in which we do not have control over what is happening and simply observe.

Stimulus-stimulus learning. A second way how the presence of a strong contingency in the environment might result in doubly skewed base rates in our experience is through sampling error. To illustrate this consider Table 6. The upper line depicts a perfect contingency between predictor and criterion. The lower line is meant to illustrate three random samples drawn from the distribution in the upper line.

Environment	C1	C2	
P1	50	0	50
P2	0	50	50
	50	50	100

Sample 1	C1	C2	
P1	2	0	2
P2	0	8	8
	2	8	10

Sample 2	C1	C2	
P1	5	0	5
P2	0	5	5
	5	5	10

Sample 3	C1	C2	
P1	6	0	6
P2	0	4	4
	6	4	10

Table 6. Frequencies tables illustrating an environment with a perfect contingency between predictor (P1 and P2) and criterion (C1 and C2) and three illustrative cases of randomly drawn samples.

In the presence of sampling error (sample 1 or 3) the base rates will necessarily be jointly skewed as no observations are possible for cells B and C. Of course, this is only true for a perfect contingency. Thus, the following simulation assesses how strong this joint skew is on average when there are non-perfect contingencies in different possible environments. In the simulation I created non-skewed environments with varying ΔP's, ranging from 0.0 to 1.0 (for the algorithm behind the environments, see, Freytag, Kutzner et al., 2009). From these environments samples of varying sizes (2 to 14) were drawn for 10.000 times and the joint skew in the samples was determined using logAR (c.f. Equation 3).

Figure 2 depicts the average logAR value in the samples, i.e. logAR's sampling distribution. Most prominently, the joint skew in the samples is inflated in contrast to the joint skew in the environment (i.e. 1:1) with increasing contingencies in the environment. This effect is strongest for small samples of around 5 observations. At maximum logAR reaches levels close to .15 that correspond to an average skew of 2.5 to 1.

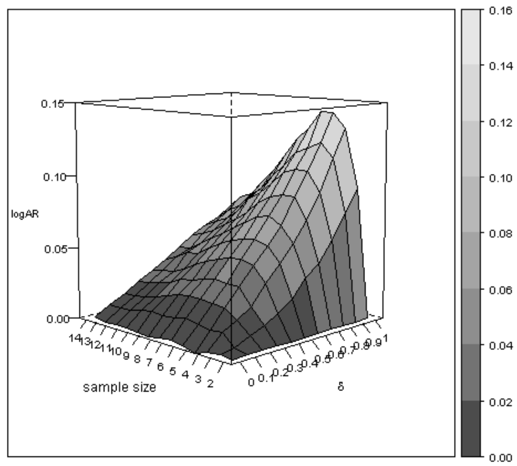

Figure 2. Average joint base-rate skew (logAR) in samples of different sizes drawn from an environment with varying ΔP's and no skew in the base rates.

Even though not illustrated, similar inflations of the sampling distributions can be found for environments that are skewed. These results coincide nicely with the inflation of the normative contingency in small samples (Kareev, 2000). Note additionally, that at under similar conditions the normative predictions are not only inflated, but also correctly extract the sign of the contingency in the environment most of the time[4]. Thus, even if predictors and criteria in our environment are not jointly skewed, a strong contingency causes sample base rates to exhibit a joint skews on average.

One of two starting points. I have just shown that given the experience of strong contingencies in the environment it is likely to experience jointly skewed base rates as well. Note, that this is the complement to what has been shown demonstrating the adaptive value of the AR, namely that there is an above chance probability of experiencing a contingency given jointly skewed base rates. Thus, an organism has two possible starting points to learn that jointly skewed base rates and contingencies are related: situations with strong contingencies and situations with strongly skewed base rates.

To provide a starting point for the development of the AR, the only assumption that has to be made is that strong contingency situations make up an easily accessible category of situations. I argue that this is plausible given the importance of strong contingencies for controlling and predicting the environment. However, the more this is the case, the higher should be the chance that organisms use their learnt above chance probability of skew given contingency, to infer that there is a contingency

[4] Supplementary simulations to those provided in Freytag, Kutzner and colleagues (2009) yield ΔP-accuracies of 83% correct sign detections for an average sample size 2.5 and an environment-ΔP of .7.

given joint skew (Gavanski & Hui, 1992). Once this confusion of the inference direction has elicited the AR, it might self-perpetuate due to its validity for inferring contingencies.

In sum, the fact that organisms inevitably learn that strong contingencies are indicative of jointly skewed based rates is probably conducive to the reverse inference that jointly skewed base rates are indicative of strong contingencies. Of course, this argument is based on the assumption that organisms are sensitive to the contingencies entailed by the joint observations of predictors and criteria in the first place (e.g., White, 2000). Thus, reversing the inference direction is yet another instance of joint observations and base-rates interacting to produce contingency judgments.

8. Concluding remarks

The present work puts forward a three-step argument for a rule-based model for contingency judgments that exclusively relies on jointly skewed base-rates, the alignment rule (AR). First, a rule-based model is shown to be necessary to explain evidence from a trial-by-trial learning task. Then the specific set of rules, termed ARs, is shown to perform well in ecologically plausible conditions when the base-rates of predictor and criterion are jointly skewed at a rate of at least 2 to 1. Finally, the claim that the AR is a psychologically plausible mechanism is substantiated by showing how it naturally follows from an organism's learning history with strong contingencies.

However, the present work not only argues for the direct influence of jointly skewed base rates on contingency judgments. It is also a step towards integrating the direct impact of jointly skewed base rates and the impact of joint observations on contingency judgments. Joint observations are provided in the learning experiments, the rational analysis presupposes that cell-entry based and base-rate driven models compete and the origin of the AR is traced back to instances where joint observations reveal strong contingencies. Additionally, it is shown that the direct base-rate influence as modeled by the AR and the predictions of cell-entry based models often coincide and might thus supplement each other (see 4.2.). Research I currently follow, seeks to elaborate on this interaction between skewed base rates and contingency models based joint observations. For example, the predictions of cell-entry based models and the AR are independently manipulated (see also, Fiedler, 2009) in paradigms similar to the illusory correlation research. It seems reasonable to assume an integration of base-rate information and joint observations into contingency judgments given the empirical evidence on humans' sensitivity to the ΔP model in combination with the skewed criterion base rate (e.g., Allan & Jenkins, 1983, Exp. 3).

However, there might be many ways how jointly skewed base rates and joint observations interact to influence contingency judgments. For example, within a signal detection framework I currently follow the hypotheses that jointly skewed base rates might lower the threshold to decide in favor of a contingency. Another possible interplay could involve different stages of testing a contingency hypothesis. When confronted with multiple important outcomes and multiple possible causes, a starting hypothesis is needed. Pairs of outcomes and causes with jointly skewed base rates might create the expectation of being related. Thus subsequent search for joint observations might favor these pairs. Taken together, I consider the present work part of the long tradition of research on contingency judgments that is stimulated by skewed base rates.

References

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society, 15*(3), 147-149.

Allan, L. G. (1993). Human contingency judgments: Rule based or associative? *Psychological Bulletin, 114*(3), 435-448.

Allan, L. G., & Jenkins, H. M. (1983). The effect of representations of binary variables on judgment of influence. *Learning and Motivation, 14*(4), 381-405.

Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General, 108*(4), 441-485.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.

Arkes, H. R., & Harkness, A. R. (1983). Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General, 112*(1), 117-135.

Blümke, M., & Fiedler, K. (2009). Base rate influence: response criterion shift determine IAT and ST-IAT effects. (submitted).

Chapman, G. B., & Robbins, S. J. (1990). Cue interaction in human contingency judgment. *Memory & Cognition, 18*(5), 537-545.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*(2), 367-405.

Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology, 58*(4), 545-567.

Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin, 90*(2), 272-292.

De Houwer, J., & Beckers, T. (2003). Secondary task difficulty modulates forward blocking in human contingency learning. *The Quarterly Journal of Experimental Psychology B: Comparative and Physiological Psychology*(4), 345-357.

Dickinson, A., Shanks, D., & Evenden, J. (1984). Judgement of act-outcome contingency: The role of selective attribution. *The Quarterly Journal of Experimental Psychology Section A, 36*(1), 29-50.

Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review, 106*(1), 180-209.

Fiedler, K. (1985). *Kognitive Strukturierung der sozialen Umwelt*. Göttingen, Toronto, Zürich: Verlag für Psychologie.

Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environments. *Psychological Review, 103*(1), 193-214.

Fiedler, K. (2009). Pseudocontingencies override genuine contingencies between multiple cues. *submitted*.

Fiedler, K., & Freytag, P. (2004). Pseudocontingencies. *Journal of Personality and Social Psychology, 87*(4), 453-467.

Fiedler, K., Freytag, P., & Blümke, M. (2009). Base rate effects in evaluative priming. *in preparation*.

Fiedler, K., Freytag, P., & Meiser, T. (2009). Pseudocontingencies: An integrative account of an intriguing cognitive illusion. *Psychological Review, 116*(1), 187-206.

Freytag, P., Kutzner, F., Vogel, T., & Fiedler, K. (2009). On the adaptive rationality of Pseudocontingencies. (under review).

Freytag, P., Vogel, T., Kutzner, F., & Fiedler, K. (2009). Pseudocontignencies by base-rate reproduction. *(submitted)*.

Gavanski, I., & Hui, C. (1992). Natural Sample Spaces and Uncertain Belief. *Journal of Personality and Social Psychology, 63*(5), 766-780.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. New York, NY, US: Oxford University Press.

Goodie, A. S., & Fantino, E. (1996). Learning to commit or avoid the base-rate error. *Nature, 380*(6571), 247-249.

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology, 12*(4), 392-407.

Hattori, M., & Oaksford, M. (2007). Adaptive non-interventional heuristics for covariation detection in causal induction: Model comparison and rational analysis. *Cognitive Science: A Multidisciplinary Journal, 31*(5), 765-814.

Hintzman, D. L. (1986). 'Schema abstraction' in a multiple-trace memory model. *Psychological Review, 93*(4), 411-428.

Inhelder, B., & Piaget, J. (1958). Probable dispersions and correlations. In *The growth of logical thinking: From childhood to adolescence.* (pp. 224-242). New York, NY US: Basic Books.

Johnson-Laird, P. N., & Byrne, R. M. (1991). *Deduction*. Hove [u.a.]: Lawrence Erlbaum.

Kareev, Y. (2000). Seven (indeed, plus or minus two) and the detection of correlations. *Psychological Review, 107*(2), 397-403.

Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review, 107*(4), 852-884.

Kruschke, J. K. (2001). The inverse base-rate effect is not explained by eliminative inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(6), 1385-1400.

Kutzner, F., Freytag, P., Vogel, T., & Fiedler, K. (2008). Base-rate neglect as a function of base rates in probabilistic contingency learning. *Journal of the Experimental Analysis of Behavior, 90*(1), 23-32.

Kutzner, F., Vogel, T., Freytag, P., & Fiedler, K. (2009). Pseudocontingencies in stereotype formation: Extending the scope of illusory correlations. (under review).

McGarty, C., Haslam, S. A., Turner, J. C., & Oakes, P. J. (1993). Illusory correlation as accentuation of actual intercategory difference: Evidence for the effect with minimal stimulus information. *European Journal of Social Psychology, 23*(4), 391-410.

McKenzie, C. R. M. (1994). The accuracy of intuitive judgment strategies: Covariation assessment and Bayesian inference. *Cognitive Psychology, 26*(3), 209-239.

Mullen, B., & Johnson, C. (1990). Distinctiveness-based illusory correlations and stereotyping: A meta-analytic integration. *British Journal of Social Psychology, 29*(1), 11-27.

Reips, U.-D., & Waldmann, M. R. (2008). When learning order affects sensitivity to base rates: Challenges for theories of causal learning. *Experimental Psychology, 55*(1), 9-22.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.

Shaklee, H., & Mims, M. (1981). Development of rule use in judgments of covariation between events. *Child Development, 52*(1), 317-325.

Shaklee, H., & Mims, M. (1982). Sources of error in judging event covariations: Effects of memory demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*(3), 208-224.

Shaklee, H., & Tucker, D. (1980). A rule analysis of judgments of covariation between events. *Memory & Cognition, 8*(5), 459-467.

Shanks, D. R. (1985). Continuous monitoring of human contingency judgment across trials. *Memory & Cognition, 13*(2), 158-167.

Shanks, D. R. (1990). Connectionism and the learning of probabilistic concepts. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 42*(2), 209-237.

Shanks, D. R. (2007). Associationism and cognition: Human contingency learning at 25. *The Quarterly Journal of Experimental Psychology, 60*(3), 291-309.

Sherman, J. W., Kruschke, J. K., Sherman, S. J., Percy, E. J., Petrocelli, J. V., & Conrey, F. R. (2009). Attentional processes in stereotype formation: A common model for category accentuation and illusory correlation. *Journal of Personality and Social Psychology, 96*(2), 305-323.

Smith, E. R. (1991). Illusory correlation in a simulated exemplar-based memory. *Journal of Experimental Social Psychology, 27*(2), 107-123.

Van Rooy, D., Van Overwalle, F., Vanhoomissen, T., Labiouse, C., & French, R. (2003). A recurrent connectionist model of group biases. *Psychological Review, 110*(3), 536-563.

Ward, W. C., & Jenkins, H. M. (1965). The display of information and the judgment of contingency. *Canadian Journal of Psychology/Revue canadienne de psychologie, 19*(3), 231-241.

Wasserman, E. A., Dorner, W. W., & Kao, S. F. (1990). Contributions of specific cell information to judgments of interevent contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 509-521.

White, P. A. (2000). Causal judgment from contingency information: Relation between subjective reports and individual tendencies in judgment. *Memory & Cognition, 28*(3), 415-426.

Erklärung

Ich erkläre hiermit, dass ich die vorliegende Dissertation selbstständig angefertigt und keine anderen als die angegebenen Quellen oder Hilfsmittel verwendet habe. Weder die vorliegende Arbeit noch Teile davon sind oder waren Grundlage einer anderen akademischen Prüfung.

Heidelberg, Juli 2009

Florian Kutzner