Doctoral thesis submitted to
the Faculty of Behavioural and Cultural Studies
Heidelberg University
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy (Dr. phil.)
in Psychology

Title of the thesis

Presence and Absence: Reference Sets and Contingency Learning

presented by
Max Ihmels

year of submission
2016

Dean: Prof. Dr. Birgit Spinath
Advisors: Dr. Florian Kutzner & Prof. Dr. Klaus Fiedler

Acknowledgements

The years in the CRISP department leading to this dissertation were wonderful. I could not have been happier. In times when I was not sure my body would allow me to continue my journey, I thought a lot about my life choices and not a single second did I regret making the decision to work in this department. I want to thank all the people who were part of this experience.

I want to thank Florian for teaching me so much while giving me so much space at the same time. Thank you for being such an amazing mentor and friend.

I want to thank Klaus for giving me the opportunity to work and learn in this incredible environment. You are a true inspiration and a role model for scientists everywhere.

I want to thank Mandy for believing in me and for playing a big part in my decision to stay in academia.

I want to thank Ute for repeatedly saving me from overwhelming amounts of paperwork and for keeping the entire department together. While science is not too difficult, bureaucracy is – and you have mastered it.

I want to thank the entire CRISP team for being such great colleagues and friends. It really was an honor and so much fun to work in the same department as all of you.

I want to thank the HiWis for making work so easy around here.

I want to thank my parents for caring so much about what I do and for always supporting me – from the first day of kindergarten to the last day of my PhD (and beyond?).

I want to thank Elisabeth for filling my life with so much joy and for always being there – in good times as in bad times.

I will miss you - see you soon,

Max

Table of contents

Presence and Absence: Reference Sets and Contingency Learning

Human behavior is best described as a function of the individual and the environment surrounding the individual. Different individuals behave differently, different contexts lead to different behavior and importantly, different individuals react differently to specific contexts. Simon emphasized this interplay of individuals and context already in 1956. To date, however, most research approaches still focus on explaining psychological phenomena by only investigating the individual, while the role of the environment is often neglected (Le Mens & Denrell, 2011). At the same time, focusing on environmental aspects can lead to very elegant alternative or additional explanations of well-known phenomena, such as intergroup bias or mere exposure (Denrell, 2005).

As already indicated above, it is not sufficient to look at environmental structures in isolation, but it is important to look at how people interact with the environment and to look at their representation and interpretation of the environment. It is important to differentiate between information which is objectively available and the way available information is actually represented. Information is relative: The representation of objects is not independent of other objects and not independent of the context and the way objects are presented.

The present work is concerned with the importance of reference sets for evaluations, mental representations and contingency judgments. Objects are not processed in a vacuum, but rather relative to other available information. Importantly, this does not only include objectively available information, but also inferred and construed information about the stimulus world. Attributes can occur in various forms. The present work investigates differences between attributes which are either always present on some level, or either present or absent. We are interested in the differences in reference sets and comparison processes evoked by the different types of attributes. The first part of the present volume investigates mental representations and evaluations of attributes which are presented in different ways and have to be processed either in isolation or in combination with other attributes. In the second part of the present work, we investigate downstream consequences of these differences in reference sets in contingency learning. Judging contingencies is a relative process. To optimally judge contingencies, frequencies of observations have to be compared to one another and have to be integrated correctly. Optimally, two conditional probabilities are compared to form a contingency judgment. If the reference sets for such an inference are not the correct ones, judgments are going to be biased. We show that attribute presentation format (present vs. present or present vs. absent) can influence reference sets in contingency learning and as a consequence, bias judgments.

1. Introduction.

An object or situation is not the same as the resulting representation. The representation and interpretation of the same situation can drastically vary across different people and, within the same people, across different contexts. Observing a person who helps an elderly lady to cross the street might lead to the inference that the helping person is a nice person. But if we learn that the person was paid to help the lady, the inference might change. In this case, context knowledge would lead to a different attribution of the behavior (Kelley, 1967) and would influence the interpretation of the situation. Knowing whether 10000 words in a dictionary is a good amount or not is hard to evaluate without knowing anything about how many words are included in other dictionaries (Hsee, 1996). Values can often only be interpreted relative to other information and change their meaning once additional information is introduced. Further, information is not only interpreted relative to other information, but this relativity can also be moderated by context. For example, comparing oneself to another person can either result in assimilation or contrast, dependent on whether one tests the hypothesis that one is similar to the standard or the hypothesis that one is different from the standard (Mussweiler, 2001). This idea of context dependency and variability of possible mental representations is at the heart of social cognition research. Individuals do not always accurately represent the objective environment, but construct their own social reality based on the perception of the input (Bless, Fiedler, & Strack, 2004). The way people process given information can be influenced by presentation format. In that way, mental representations are flexible. In those cases where different presentation formats of the same problem lead to differences in processing or behavior, we speak of framing effects.

1.1 Framing.

We call something a framing effect when *equivalent* presentations of the same problem lead to differences in impressions or choices (Sher & McKenzie, 2006). The term *equivalence* will be discussed in a bit. The most prominent example of framing probably is the *Asian Disease Problem* (Tversky & Kahneman, 1981), the classic of risky choice framing (Levin, Schneider, & Gaeth, 1998). In risky choice framing, participants are presented with a set of alternatives which differ in the amount of risk, but (often) have the same expectancy value. Additionally, problems are presented either in terms of gains or in terms of losses. In this specific scenario, participants are confronted with the problem of deciding which medical program to engage in, trying to combat a disease which is expected to kill 600 people. In both framings, there are two options. A risk free one and a risky one. In the gain framing, participants have to decide whether they want to choose the option where 200 people will survive for sure, or the option where they have a 1/3 chance to save 600 people and a 2/3 chance to save 0 people. In the equivalent loss framing, participants have to choose between an option where 400 people die for sure and the option where they have a 1/3 chance that 0 people die and a 2/3 chance that 600 will die. In both framings, the expectancy value is the same for both alternatives (200 will live, 400 will die). Mathematically, these two frames are therefore equivalent. Still, participants show a preference reversal across the different frames. In the gain framing, the majority of participants chooses the risk free option, while in the loss framing, the majority of participants chooses the risky option. This effect is most commonly explained by referring to the famous prospect theory which introduces different value functions for gains and losses (Kahneman & Tversky, 1979). Findings like this challenge one of the central aspects of expected utility theory: the invariance axiom of rational choice (Bernoulli, 1954 [1738]; Von Neumann & Morgenstern, 1944). If the goal is to maximize expected value or utility, choices should not vary across different presentation modes. While the present volume is concerned with framing effects, the goal is not to make a point for irrationality of human behavior. In fact, as discussed later in this volume, many framing effects might very well be rational if the concept of rationality is extended by ecological aspects and aspects of conversational logic.

There are a lot of phenomena which have been summarized under the label of framing. Manipulating whether a scenario is about gains vs. losses when it comes to risky choices such as in the *Asian Disease problem* is not the only way the idea of framing can be applied. For example, framing does not always have to be about creating preference reversals between independent choice options. It can also simply be about changing the evaluation of an object, by reframing an attribute used to describe the object. Either the attribute is presented in a positive, or in a negative way. For example, one could present either success or failure rates of a medical treatment (with p(Success) = 1 - p(Failure)), leading to differences in the evaluation of the treatment. Or one could describe beef as either 75% lean or 25% fat (Levin & Gaeth, 1988). Here, 75% lean meat is perceived to be better tasting and less greasy than the 25% fat meat. In such a case, the variable of interest is the basic evaluation process (Levin et al., 1998) and it is especially interesting, because it allows for a very basic examination of framing effects.

While there is a plethora of documented instances of framing, the present work is not the place to discuss all of them (for a review, see Krüger, Vogel, & Wänke, in press). This brief description of framing effects should rather serve as a way of illustrating the context dependency of mental representations. The notion and the definition of framing however, remain relevant for the present work.

1.2 Information equivalence.

To recall, we call something a framing effect when two *equivalent* presentations of the same problem lead to differences in judgments or choices (Sher & McKenzie, 2006). Unfortunately, for the most part, researchers are pretty vague when it comes to defining what they actually mean when talking about equivalence. One type of equivalence, however, is well defined: Logical or mathematical equivalence (see e.g. Johnson-Laird & Shafir, 1993; Rubinstein, 1998; Shafir, 1993). Two statements are logically equivalent if they have the same logical content and therefore necessarily entail the other. For example, a winning probability of 60% and a losing probability of 40% are logically equivalent, given there are only the two outcomes winning and losing. Sher and McKenzie (2006) however argue that logical equivalence is not always enough to guarantee or even predict choice invariance. For (rational) choice invariance, there must also be *information equivalence*. Per definition, two statements A and B are information equivalent when there is no choice-relevant background condition C about whose probability a listener can draw inferences from the speakers choice between frames. More formally, A and B are information equivalent if p(C| speaker chooses "A") = p(C| speaker chooses "B"). If this is not the case, there has been information leakage by the speaker's choice of frame. Understanding that the speaker chose the frame A instead of B can lead to inferences which can change a decision. For example, presenting something in a positive (e.g. 60% win rate) vs. a negative (e.g. 40% loss rate) frame might be interpreted as a persuasion attempt or a recommendation. Using this approach, Sher and McKenzie (2006) shed new light on a classic example. They asked participants to sit down at a table with two cups on it, one empty, one full. They then told participants: "*Just to get things started, could you pour water from one cup to the other and set a half-full cup at the edge of the desk*". In the other condition, the "half-full" part was replaced by "half-empty". Confirming the idea, when the "half-empty" cup was requested, participants chose the initially full cup more often than when the "half-full" cup was requested. The findings are interpreted as such that the different frames communicate different reference points (full vs. empty). These reference points are picked up by the receivers and then used for their decision.

This experiment also illustrates that framing is about communication. The chosen way of communicating something can influence judgments and decisions. Framing does not only happen in the laboratory, but also in everyday communication. Some frames might be consciously chosen, others might just be the result of randomness, habits or situational aspects. The same holds for the receiver side of the communication act. The chosen frame (including its intended purpose) might be actively recognized or not. Further, this example illustrates the importance of reference points. Objects are not evaluated, represented or interpreted in a vacuum, but relatively to (communicated) reference points or other relevant objects.

1.3 Conversational implicatures.

The logic behind possible inferences which are a consequence of non-equivalent information is similar to the idea of conversational implicatures (Grice, 1975). These implicatures refer to what is suggested by an utterance, going beyond its literal meaning. In the example above, the literal meaning of half-full and half-empty might be the same, they do however suggest different actions. These implicatures go hand in hand with the cooperative principle, which tries to describe how people communicate and interact with one another:

Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged (1975; p. 45).

This cooperative principle summarizes several sub-principles, most often referred to as maxims: Maxim of Quantity, Maxim of Quality, Maxim of Relation, and Maxim of Manner. Communicated content is expected to be informative, truthful, relevant and clear. This is not only concerned with what information is communicated, but also with how the information is communicated. Information is expected to be presented in the most informative and meaningful way.

Conversation does not only happen in private social situations. It is important to remember that every experiment or survey also is some sort of communicative process. In every experiment or survey situation, respondents need to know what exactly the interviewer wants to know. Scales in a survey will be interpreted as a function of survey context. For example, Wänke (2002) used scales including vague quantifiers such as “very rarely” and “very often” to access how often participants visit the cinema. When the study was presented as a study for a student population, lower frequencies were reported compared to when the study was presented as a study for the general population. Reference points were again inferred as a consequence of the chosen communication context.

These concepts highlight the relative nature of information. Information is evaluated relative to reference points or reference sets and importantly, these reference sets can vary as a function of how information is communicated.

1.4 Objects as complex and rich entities.

Research on similarity judgments (Tversky & Gati, 1978) offers a framework to think about why varying representations of a given object are even possible. To allow varying representation of objects, we need to make some assumptions. Let us assume that each object or entity of interest can be described using a variety of attributes. Further, assume that our internal data bases of these entities are rich and complex. They contain information about a large variety of attributes, such as appearance, function, relation to other objects, etc.. Generally speaking, these data bases contain everything we can deduct from our world knowledge and the observations we make. These can be concrete things such as color or length, but also more abstract things such as quality or value. Note that while Tversky and Gati (1978) use the terms *feature* and *feature sets* to describe object properties, the term *attribute* will be used here to avoid confusion with terms which will be introduced later (features and dimensions in the sense of Garner, 1978). The idea is that when people judge similarity, objects are judged to be more similar the more common attributes they have and judged to be less similar the more distinctive attributes they have. Further, and more importantly for the present work, not all of the attributes inherent in an object are always prominent, salient or relevant. Instead, the representation of an object can vary. For example, in this model, similarity judgments are not symmetrical. The similarity of A to B is not necessarily the same as the similarity of B to A. When judging the similarity of A to B, A is the subject of comparison and therefore more salient than the referent B. The attributes of A then get a higher weight than the attributes of B. This results in similarity being more reduced by the distinctive features of the subject than by the distinctive features of the referent (Gati & Tversky, 1982). Extending this idea also leads to the prediction that the variant is more similar to the prototype than vice versa: North-Korea is judged to be more similar to China than China to North-Korea and an ellipse is more similar to a circle than a circle to an ellipse. Dependent on the presentation format, attributes differ in their relevance for the comparison. The representation of the objects of interest can vary as a function of context. Attributes can vary in their relevance for a comparison.

Further evidence for this idea of variability of attribute importance comes from work on the Cancellation and Focus Model (Dhar & Sherman, 1996; Houston & Sherman, 1995; Sherman, Houston, & Eddy, 1999). Here, the idea is that when comparing choice alternatives, shared and unique attributes differ in their salience. The prediction is that for a decision, shared attributes get cancelled out, whereas unique attributes are salient and focused on. Recent working using eye tracking provided further support for this idea (Sütterlin, Brunner, & Opwis, 2008).

Representations cannot only vary as a function of how the objects themselves are presented, but also as a function of other available objects. There are many settings in which the introduction of other available objects changes evaluations and preferences. Research on attraction effects and evaluability theory offer insights into how attributes are processed in multi-option multi-attribute environments.

1.5 Attraction effects.

Attraction effects can occur in situations where people have to choose between different alternatives. Imagine two objects A and B which vary on two attributes which are both of importance for a decision. You are interested in buying an electronic product and the important attributes are quality and price. The alternatives are equally attractive, one being more expensive but of higher quality, the other being cheaper but of lower quality. From a basic normative or utility focused point of view, preference between the two options A and B should be independent of other available alternatives. It should just be a function of their quality and price. Still, research suggests that introducing a third alternative C can indeed influence preferences between A and B. In the classical paradigm, a third alternative C is introduced, which is dominated by one of the alternatives, but not by the other, meaning that one alternative (say A) is better than C on both attributes but the other alternative (B in this case) is only better on one of the attributes. Then, the preference for A goes up (Huber, Payne, & Puto, 1982; Ratneshwar, Shocker, & Stewart, 1987). This violates the concept of regularity, which is inherent in most choice models (Luce, 1977): the percentage of choosing a specific option should not go up after adding yet another choice option. There are a variety of explanations for this effect, focussing on differences in the perception of the attributes and alternatives, differences in evoked decision strategies, familiarity and meaningfulness, etc. (for a review, see Ratneshwar et al., 1987). Important for the present work is once again the idea that a decision is not made in a vacuum, but in a context. Including additional objects changes the reference set. Not only does the objectively available information change, but also how attributes are understood.

1.6 Evaluability.

More support for the idea of relativity of information comes from research on evaluability (Hsee, 1996; Hsee, Lowenstein, Blount, & Bazerman, 1999; Hsee & Zhang

2010). The idea is to differentiate between attributes which are easy to evaluate independently and attributes which are hard to evaluate independently. Hard to evaluate means that an evaluator does not know how good a certain level of an attribute is without comparisons, while easy to evaluate attributes are easily judged without direct comparison. For example, it is really hard to tell whether 10.000 words in a dictionary is a good amount or not. On the other side, it is really easy to tell that defects in the form of dog-ears and a scratched cover are a bad thing in a dictionary. A mixture of easy and hard to evaluate attributes can lead to preference reversals when two objects are judged in isolation (separate evaluation) compared to when they are judged at the same time (joint evaluation). Imagine there a two dictionaries: Dictionary A has 10.000 words and no defects, dictionary B has 20.000 words and defects. If the dictionaries are now judged in isolation (without knowledge of the other alternative), evaluation will mostly be driven by the defect attribute, because it is difficult to judge the quality of the number of words attribute. Therefore, dictionary A would be preferred. In contrast, when A and B are presented at the same time, the number of words becomes more easily evaluable, because the two alternatives are now comparable. A reference point is created. Then, dictionary B might be preferred.

Sher and McKenzie (2014) highlight that preference reversals of that kind must not necessarily be irrational. To be exact, they argue that when closely investigated, these effects are not even true preference reversals. The idea is that when knowledge of the stimulus environment is limited, the presentation of options alters the information state of the decision maker. Each option which is presented provides information. Before any information is presented, there is an initial belief, a prior. This prior is updated when any information (in the form of another option) is provided. This updated model includes a preference order for the available options. Importantly, these posterior models can differ as a function of the provided sample. Then, rational evaluations will be coherent within contexts but not across contexts.

Both the evaluability account as well as the normative analysis presented above show once again that presentation and representation of the stimulus world, including assumed distributions of attributes, can influence judgments and decisions. The present work is concerned with different attribution presentation formats. Similar to the idea of evaluability, we aim to show that different attribute presentation formats allow different inferences about the world and are more or less easy to deal with, dependent on what other information is available.

1.7 Mindsets.

Representations and decisions cannot only vary as a function of how a stimulus or a problem is presented, but also as a function of states activated within the individual. Representations can differ dependent on the mindset a person is in. This mindset can be temporarily activated or chronically inherent in the person and influence the subjective construal of the objects or problems at hand. Some prominent examples of this mindset dependency are construal level or psychological distance (Trope & Liberman, 2011), influences of mood (Forgas & Bower, 1987), regulatory focus (Crowe & Higgins, 1997) or assimilation vs. accommodation (Fiedler, 2001). While research on mindsets can offer a variety of interesting insights, the present work will rather focus on effects of stimulus and context properties than on effects of different mindsets.

1.8 Mental representations and reference sets – Interim summary.

The examples so far presented highlight the relativity of mental representations. Information is not evaluated in isolation, but evaluated relative to reference sets. How options are evaluated and how they are represented can vary as a function of presentation format and context. Importantly, there are two parts to this. Firstly, how the options with their attributes themselves are presented has an impact on evaluations and representations. Secondly, other available objects also have an impact on how an option with its attributes is evaluated and represented. Attributes are understood relative to comparison sets and reference points which are inferred from the stimulus environment and experimental context. Building on the concepts presented so far, the next part of the present volume aims to extend the notion of attribute framing by differentiating between attributes which are always present on some levels and attributes which have one present and one absent level. It will then be discussed how these attributes differ in the inferences they allow about the stimulus world and how they differ in the reference sets they evoke.

1.9 Extending attribute framing: Presence vs. absence.

As evident from multiple examples of the introduction, attributes can be framed in different ways. In the classical sense, we call something attribute framing when one attribute is presented either in a positive or a negative way, such as success rates vs. failure rates (Levin & Gaeth, 1988). A teacher could either tell his students that normally, about 80% of people pass the exam or he could say that normally, around 20% fail the exam and evoke different reactions in his students.

Importantly, attributes come in different forms and have more properties than just valence. In experimental psychology, especially in judgment and decision making, a lot of variables are presented in a dichotomous fashion. Across a number of trials, participants see two levels of a variable. These two levels of a variable can now be presented in different ways. In the causal learning literature for example, cues as well as outcomes are often presented in a present vs. absent fashion: they have one present level and one absent level. For example, in a scenario where participants have to learn about the effect of fertilizer on plant growth, a plant might receive the fertilizer (present) or not (absent) and develop blossoms (present) or not (absent). In other scenarios, cues or outcomes might appear in a present vs. present fashion. A plant might receive fertilizer A or fertilizer B and then develop green or blue blossoms. Here, both cue levels are positively defined. When the task is about learning a relationship between two dichotomous variables, from a statistical point of view, it does not make a difference whether attributes are presented in a present vs. absent or a present vs. present fashion. It is still just a 2x2 table based on which indices of relation can be calculated. Goal of the present work is to show that it indeed does make a difference for subjective representations and subsequent judgments and decisions whether attributes are presented in a present vs. present or a present vs. absent fashion. The different attribute presentation formats should highlight different reference sets and hence influence judgments and decisions. The goal is to show that framing cannot only influence the encoding and interpretation of a single object, but also how relationships between different objects are learned.

In the broader sense, one might classify the difference between present vs. absent and present vs. present attributes as a variant of attribute framing. While there is not necessarily a manipulation of valence involved, the way attribute levels are presented is still manipulated within a single cue. Based on the distinction of features and dimensions by Garner (1978), a theoretical framework around this instance of attribute framing is built and differences in mental representations of reference sets as well as learning processes in contingency judgment tasks are investigated.

1.10 Features and dimensions.

Sometimes we think about attributes as being present or absent, sometimes we think about attributes as always being present on some level. For example, one might talk about whether a person has children (present) or not (absent). At the same time, when talking about the biological sex of that person, the question is not about presence or absence, but rather about the level of the attribute. Is the person male or female? We argue that this often overlooked distinction of attribute presentation format influences the way we think about attributes and reference sets and is of crucial importance in the domain of judgments about relation.

Garner (1978) introduced a taxonomy relevant for this distinction. In his work on "aspects of a stimulus" (cf. figure 1), he differentiates between *features* and *dimensions* which are nested within the class of attributes. Attributes should here be distinguished from wholistic properties (simple wholes, templates and configurations). Simple wholes refer to all aspects being processed in parallel; templates represent properties that are more than the sum of the stimulus aspects, like prototypicallity. Configurations describe emergent properties of a stimulus, like symmetry. While every stimulus has wholistic as well as attribute or component properties, the present work will focus on the attribute aspect of a stimulus. More specifically, the difference between features and dimensions is investigated. For now, we will compare features which are always dichotomous and dichotomous dimensions. Extensions to continuous or categorial dimensions with more than two levels will be discussed at a later point of the present volume.

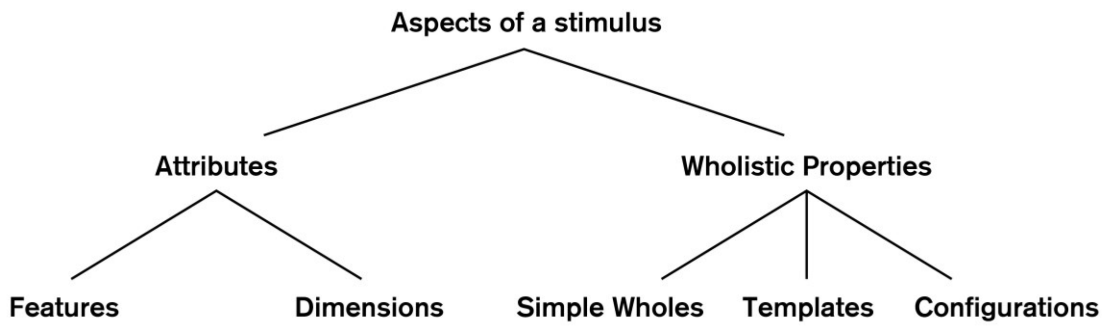

Figure 1. Aspects of a stimulus building on Garner (1978). Features and dimensions are nested within the class of attributes.

A *feature* is defined as an attribute with two asymmetric and qualitatively different levels: present and absent. Hence, the question “Is X present?” is perceived to be adequate. For example, after a therapy, side effects might be present or not.

A *dimension* can take on at least two different positively defined levels. For example, the side effects just used as an example for features can be reframed to a dimension by simply changing the labels to numerous vs. few or strong vs. weak side effects. Here, the question “What level does X have?” is adequate.

Features refer to attributes of a present vs. absent nature and dimensions refer to attributes of a present vs. present nature.

Building on Garner (1978) we assume that different stimuli elicit different sets of stimuli, the *stimulus space*, against which the stimulus at hand is spontaneously compared. This stimulus space refers to a mental representation of the stimulus world and it includes all other alternative stimuli, representing all other attribute constellations that might have been encountered. Every encountered stimulus is embedded in this stimulus space and the central idea is that the exact structure of the stimulus space differs as a function of whether presented attributes are features or dimensions. Features and dimensions allow different inferences about the stimulus world and therefore lead to stimulus spaces differing in elaboration and explicitness, highlighting different aspects of the stimulus world. The core idea is that this mental representation of the stimulus world serves as a reference set relative to which information is processed and evaluated.

1.11 Inference asymmetries of features and dimensions.

Features and dimensions differ in the inferences they allow about the stimulus world when they are presented in isolation. Let us consider dimensions first. In the present work, a dimension is defined as an attribute which is always present on some level. After a therapy, a patient might develop weak or strong side effects. By presenting any one of the attribute levels (say weak side effects), the nature of the attribute is immediately clear ("this is about side effects"). It is easy to think about other stimuli (other people) with other levels of the same attribute. Each cue level reveals the nature of the underlying dimension and allows a straight forward representation in the stimulus space. Comparisons to other levels within the same attribute are well defined, easily accessible and should be highlighted in the stimulus space for dimensions.

For features, this is different. Other than in the case of a dimension, the levels are qualitatively different, producing an asymmetry. One level is present, while the other is absent. For example, after a therapy, a patient might develop fever and stomach pain. The presence of these attributes allow an inference about the nature of the attributes. Given the fever is there, it is clear that it could be present or absent in other people. The absence of a feature on the other hand does not reveal the underlying attribute - It does not reveal what is missing. For every stimulus (or in this example for every patient), there is an endless list of absent features (there could be many more side effects) and this endless list can never be assessed exhaustively. This asymmetry also appears on a linguistic level. The name of the attribute typically is the same as the positive value (e.g. fever).

Importantly, it is not the case that absence never allows any inferences. Given context knowledge and enough processing resources, absence can be informative and allow important inferences about the situation. Not seeing the wedding ring on your friend's hand, where it has been for the last 10 years, is alarming. Still, absence renders the stimulus space much more vague and less elaborate. Reference sets and categories should therefore not be as well defined and there should be more uncertainty about how attributes are structured. As a result, the presence of attributes should be highlighted in a stimulus space construed on the basis of features. As a consequence, in contrast to dimensions, other levels of the same attribute should not constitute the main reference set. Instead of comparing presence to absence within the same attribute, presence should be compared across different objects.

While dealing with presence is straight forward, dealing with absence imposes some problems. Counting the presence of events seems to happen without much cognitive effort (Zacks & Hasher, 2002), while the counting of absence does not: Sequencing or segmenting absence is ambivalent (Gallistel, 2007). It is easy to say how often something came up in a conversation, but counting the absence of a topic is difficult. Should it be counted once per word, per sentence or per conversation? It is assumed that continuous time is decomposable into trials, but life is not composed of trials. Another problem comes into play for events which are hard to detect. In such a situation, the absence of an attribute can be attributed to its true absence or a failure to detect it (Blaisdell, Leising, Stahlman, & Waldmann, 2009). Processing information often is a relative task and levels of a cue might often be compared to other levels of the same cue because they serve as a natural reference and comparison set. The idea is that when one of the cue levels is hard to grasp (absence), the natural reference and comparison set should not be highlighted as much and other reference sets (e.g. other present cues) should get more relevant.

At this point, it is important to note that absence and missing information are not equivalent. The term absence as we use it here is a clearly defined cue level which is contrasted to a positively defined present level. Missing information on the other hand is no clear indicator of a cue level. Not knowing whether a person has children or not does not mean that the person does not have any children. Absence and missing information might be the same visually, but the true value behind an instance of missing information might as well be presence. Problematically, in some situations and communication contexts it might just not be clear if absence of information is an actual indicator of absence or missing information about presence, leaving behind ambiguity.

1.12 Feature-dimension framing.

It might be the case that we are more likely to think about some attributes as dimensions instead of features and vice versa. Still, as the therapy side effects example already illustrated, attributes do not strictly occur as either features or dimensions but can be framed in different ways. In the sense of decision framing (Tversky & Kahneman, 1981), it is possible to communicate two levels of a cue in a present vs. absent fashion or in a way that implies that the attribute is always present on some level (present vs. present). The wording "strong side effects" implies that side effects can be present on different levels, while the wording "side effects" implies that there are either side effects or not. Again, the idea is that this can create differences in the reference set or stimulus space the stimulus is compared against. Using the list context of an experiment, it is even possible to create pseudo-features and pseudo-dimensions to maximize parallelism between experimental conditions. For example, when communicating cultural interests of a person, the dimension framing can be "Music" vs. "Art" (present vs. present) while the feature framing can be "Music" vs. *nothing* (present vs. absent). This does not rely on linguistic details but uses the list context to define features and dimensions. Some of the following experiments will integrate this type of pseudo-features and pseudo-dimensions to parallelize experimental conditions as much as possible.

1.13 Information equivalence of features and dimensions.

We speak of a framing effect when two equivalent presentations of the same problem lead to differences in judgments or decisions. As discussed before, it is important to distinguish between different types of equivalence. Presenting a dichotomous variable as a feature compared to a dimension might not make a difference in terms of logical equivalence (still just two levels of a dichotomous variable), but it does make a difference in terms of inferences they allow. Information equivalence is only given when two frames do not differ in the inferences they allow about some background condition. Clearly, as depicted above, possible inferences differ between present vs. present and present vs. absent attributes. The present level of a cue reveals its nature while the absent level does not. In case of a dimension, both levels reveal the nature of the attribute. Further, in the sense of conversational implicatures (Grice, 1975), a communicative intent might be inferred when something is presented in a present vs. absent fashion (feature). The explicit mention of the present level of a cue highlights its importance. It might be inferred that this level is the important one in the current situation, for example because it is rare (McKenzie, Ferreira, Mikkelsen, McDermott, & Skrable, 2001). Communicating evidence as dimensions implies that multiple cue levels are of importance. Communicating evidence as features creates uncertainty about whether and how the absence of features is informative and important.

1.14 Prior research on features and dimensions.

The feature-dimension distinction or, more precisely, the difference between present vs. absent and present vs. present stimuli, has only received sporadic attention in diverse domains. For example, Gati and Tversky (1982) investigated differences between qualitative and quantitative attributes and showed that the they can differ in their effects on similarity judgments. Typically, the similarity between objects described with features would be a function of feature overlap (Tversky, 1977). In contrast, similarity between objects described with quantitative variables (or dimensions) is typically described as a function of distance in a multi-dimensional space defined by the different quantitative variables (Gati & Tversky, 1982). Similar to the presented ideas about differences in evoked reference sets, these frameworks also speak towards within attribute comparisons for dimensions and across object and attribute comparisons for features.

Bröder, Newell and Platzer (2010) showed that cue (re-)presentation format can influence multiple cue decisions. They confronted participants with information about symptoms of patients and participants had to infer sickness. The cues were either presented as alternative symptoms (e.g. “fever” vs. “hypothermia”) or as presence vs. not present (“fever” vs. “no fever”). Results showed rule-dominance for present vs. absent cues and higher percentages of exemplar based strategies with alternative cues (present vs. present). Note that this differs from the operationalizations used in the present work. Here, presence was contrasted to the negation of presence rather than absence. Also, in this scenario, attribute presentation format is here confounded with the implied direction of relationship. “Fever” and “hypothermia” are both bad while “no fever” is not.

Other research domains such as classification learning (Nosofsky, Kruschke, & McKinley, 1992) or selective attention (Pansky & Algom, 1999) mention that it might be important to differentiate between different types of attributes, but do not engage in a systematic testing of the differences. The most systematic, though still rudimentary, treatise of the issue can be found in research on contingency judgments (Allan & Jenkins, 1983; Beyth-Marom, 1982; Crocker, 1982). A detailed analysis of features and dimensions in contingency learning is discussed later in the present volume. The present work provides an analysis of the role features and dimensions play for mental representations of the stimulus world, as well as an analysis of the role they play in contingency learning. Differences in evoked reference sets can affect how relationships between cues are learned.

1.15 Features and dimensions – Interim summary.

Features and dimensions differ in the inferences they allow about the stimulus world. As a consequence, the mental representation of the stimulus world, the stimulus space, should differ as a function of this type of attribute framing. In a stimulus space initiated by dimensions, the different levels of attributes should be equally represented. This should highlight comparisons within attributes. Hence, the natural reference set for a dimension value are the other values of the same dimension. In contrast, in a stimulus space instigated by features, the presence of attributes should be highlighted, while the absence should be neglected. This should hinder comparisons within the same attribute and rather shift the reference set to present occurrences of other attributes in other objects.

There are multiple things influencing the final mental representation of the stimulus world. Firstly, it is important what objects and attributes are presented, as for example experiments on evaluability or attraction effects show. Secondly, it is important how these objects and attributes are presented, as for example the typical attribute framing experiments show (e.g. 75% lean vs. 25% fat). The mental representation of the stimulus world does not simply reflect the objectively available information, but is rather the result of a constructivist process (e.g. Bless et al., 2004).

1.16 Overview of the empirical work.

There are two main parts to the empirical contribution of the present volume. The shorter first part is concerned with the effects of presence and absence (attribute framing) on stimulus set representations and utilized reference sets. The second part is concerned with attribute framing and reference sets in contingency learning. The second part also includes a computer simulation based analysis of different indices of contingency. Here is a very short overview of the documented experiments and simulations.

- Experiment 1: *Stimulus Space*; Effects of attribute framing on construed stimulus spaces
- Experiment 2-4: *Presenters Paradox Revisited*; Highlighting presence and absence in joint and separate evaluation
- Computer simulations: R framework for understanding contingency indices and creating stimulus distributions in 2x2 tables
- Experiment 5: *All Things Being Equal*; First exploration of feature-dimension framing in contingency learning

- Experiment 6: *Skewed Base Rates*; Attribute framing and base rates in contingency learning
- Experiment 7-9: *Diamond I, II, III*; Systematic variation of attribute framing, ΔP, A-cell and PC in multi cue environments

2. Experiment 1: Stimulus space.

This first experiment is an attempt to access differences in the stimulus spaces construed by participants when confronted with attributes which are either framed as features or dimensions. In line with the reasoning by Garner (1978) we argue that an attribute presented in the form of a feature will prompt thinking about the presence of (other) features. The reference set for a present feature should consist of other present features. Comparisons with other levels of the same attribute should be less prominent. An attribute presented in the form of a dimension should prompt a more diverse stimulus space representation. Especially, other levels of the same cue should be more prominent. Comparison within attributes are expected for dimensions and comparisons across different objects and attributes are expected for features.

2.1 Method.

To test for differences in constructed stimulus spaces, we confronted participants with different consumer products (including specific attributes) and asked what other attributes products of the same type might have.[1]

Participants. 115 students (88 female, $M_{age} = 23.31$, $SD_{age} = 8.99$) participated in our experiment which took roughly 10 minutes. The experiment was distributed via social networks. Participants received course credit for their participation. The experiment was created with SoSci Survey (Leiner, 2014).

Materials and Design. We presented very short descriptions of consumer products to our participants. In each trial, we informed them what product the trial was about, followed by a very brief description of a label written on the product. The labels were either presented in a feature or a dimension framing, which was manipulated between participants. For example, in

[1] I want to thank Malin Hildebrandt, Julia Karl and Kira Weinberg for their support in running this experiment.

one trial, the product of interest was “Cookies”. The label read “with Nuts” in the feature condition and “15% Nuts” in the dimension condition. All of the labels were generated in a similar fashion, with the feature condition stating the presence and the dimension condition stating the contained amount of the attribute. This was done to get as close as possible to the original definition of features and dimensions by Garner (1978). For features, the question “Is X present?” should be more adequate. For dimensions the question “What level does X have?” should be more adequate. Participants were then asked to generate attribute labels which might be written on similar products offered by the same or other competing companies. It was up to the participants to decide how many attributes they wanted to write down, but there was a visual cue of a list, suggesting around 5 answers per product. There were 12 products in total, resulting in 12 trials. Framing was manipulated between participants. Participants either saw all attributes presented as features or all attributes presented as dimensions.

Dependent variable - Generated attributes. The responses given by participants were categorized into four different categories resulting from a 2x2 logic: Attribute (other vs. same) x Framing (feature vs. dimension). For example, when “cookies” was presented in combination with the label “15% Nuts”, a response might have been “20g chocolate”. That would be a different attribute (chocolate instead of nuts) for which the question “What level does X have?” would be adequate. It would therefore be classified as “other attribute” & “dimension”. “With nuts” would be classified as “same attribute” & “feature”. The attributes were coded by two independent raters who achieved a satisfying inter rater reliability, $\kappa = .81$. We were mainly interested in differences on the “same attribute vs. other attribute” factor, but also exploratively looked at how participants framed their attributes (“feature vs. dimension”) to learn more about how exactly participants construed their stimulus spaces.

Data Preparation. To reduce inter trial variance, after coding participants' responses, the proportion of category responses was calculated for each trial. This was done by simply dividing the number of responses for a given category by the total number of responses for that product (for each trial and participant). Afterwards, these values were averaged within each participant, resulting in four proportion scores. For the main analysis, we only differentiated by the attribute (same vs. other) factor. For the explorative analysis, we considered both the attribute (same vs. other) as well as the framing (feature vs. dimension) aspect of responses. Still, we were mainly interested in differences in whether participants generated instances of other attributes or the same attribute as a function of attribute framing.

2.2 Results.

A t-test on the types of attributes generated revealed a difference between the framings, $t(109) = 6.20$, $p < .001$, $d = 1.18$. As expected, participants who were confronted with features were more likely to generate other attributes (91.77%) than participants who were confronted with dimensions (63.62%). Figure 2 shows the explorative analysis including not only a differentiation by attribute, but also by framing (2x2). It is evident that in the feature condition, participants mainly generated other attributes framed as features. So for example, after being confronted with "with nuts", they generated labels such as "with chocolate", "with fruit", etc.. In the dimension condition, participants generated all kinds of labels. Importantly, as shown by the above analysis, the amount of labels referring to the same attribute is higher than in the feature condition. So for example, after being confronted with "15% nuts", participants generated labels such as "10% nuts".

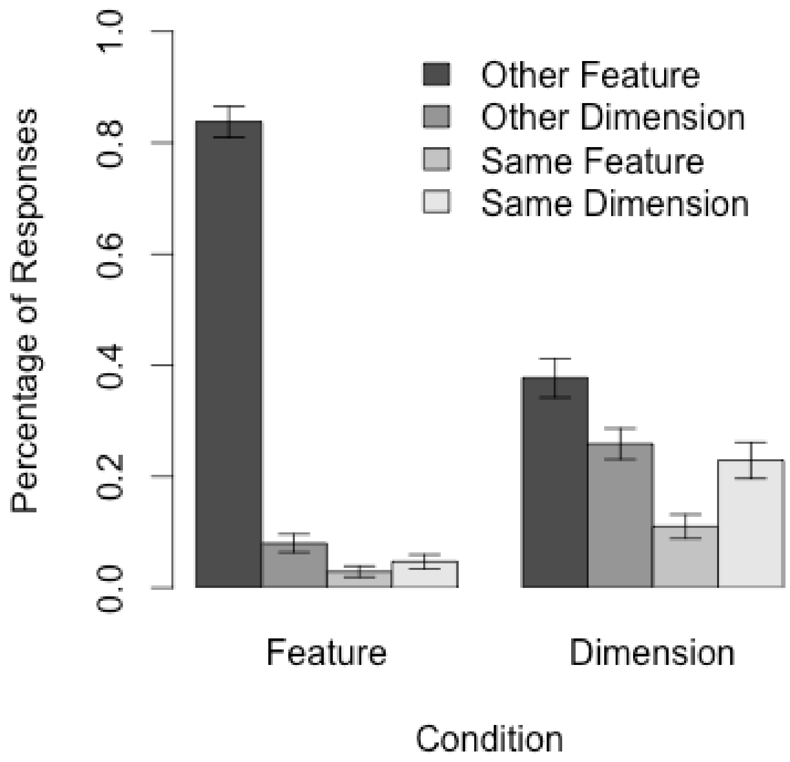

Figure 2. Mean percentages of different types of generated attributes as a function of framing. Attributes are classified into the four categories defined by other vs. same attribute and by the framing participants chose: feature vs. dimension. Error bars represent standard errors (± 1 SE).

2.3 Discussion.

Participants were confronted with a variety of products which had differently framed attributes written on them. It was manipulated between subjects whether these attributes were presented in a feature or a dimension framing and participants were asked to generate attributes which might be written on similar products. When attributes were presented in a feature framing (implying a present vs. absent nature of the attribute), the vast majority of responses fell into the category of *other attribute & feature.* Therefore, when the present level of one attribute was stated (implying the other level is absent), participants mainly thought about the present levels of other attributes. When the attributes were presented as dimensions, responses were more divers. As expected, there were more responses falling into the category of same attribute & dimension than in the feature condition. Compared to the feature condition, responses were more equally distributed across the four categories. Features resulted in a stimulus space full of other present features. Dimensions resulted in a more balanced and diverse representation, including multiple levels within the same attribute.

One drawback of the chosen paradigm is that it might not have been able to grasp when participants actively thought about the absence of an attribute feature. Participants could have responded with something like “without X”, but they rarely ever did. They could have also submitted empty text boxes to respond with absence, but this response is really unlikely in the communication setting of an experiment. Actively telling participants that entering absence is also a possibility might have biased the results as well. Still, participants could have produced other present levels of the same attribute (e.g. “10% nuts” after seeing “with nuts”). The experiment illustrates that peoples’ representation of the stimulus environment is not only a function of what objects are presented, but also a function of how these attributes are framed. When attributes are presented in a feature framing, the presence of other attributes in other objects is relevant. When attributes are presented in a dimension framing, amongst other things, different levels of the same attribute are relevant. Framing effects cannot only have an influence on the representation of the object at hand, but also on the representation of other possible objects and subsequent comparison processes.

3. Experiment 2-4: The presenters paradox revisited: An evaluation mode account (Krüger, Mata, & Ihmels, 2014).

Preference reversals in joint vs. separate evaluation are often explained by general evaluability theory or the evaluability hypothesis (Hsee, 1996; Hsee & Zhang, 2010). Further, as discussed above, Sher & McKenzie (2014) highlight the importance of reference points in joint and separate evaluation. In our work (Krüger, Mata, & Ihmels, 2015) we suggest yet another account why some alternatives are preferred in joint evaluation while other alternatives are preferred in single evaluation. The idea is that evaluation mode influences how much attention is paid to individual attributes.

Revisiting work on the presenters paradox by (Weaver, Garcia, & Schwarz, 2012), we investigated how people evaluate single products and product bundles. Bundles were defined as a combination of a single product plus an extra add-on. For example, participants were asked to evaluate a coffee maker (single product) and a coffee maker including an additional milk frother (single product + add-on, product bundle). Importantly, we also manipulated whether the single product and the product bundle were evaluated separately (by different participants) or jointly (by the same participant). Based on the cancellation and focus model (Sherman et al., 1999) we now expected participants to value the bundle more in joint evaluation than in separate evaluation, because the added value of the bundle (i.e. the add-on) should become more salient when it is contrasted to just the single product. Again, the idea is that the evaluation of an object is context dependent. What attributes people focus on varies as a function of the (constructed) stimulus environment.

We ran three experiments, providing support for our theory. Dependent on evaluation mode (joint vs. separate), objects were evaluated differently. In joint evaluation, the bundle was preferred over the single option. In separate evaluation, consumers were largely indifferent. Also, the relative preference for the bundle was more pronounced in joint evaluation than in separate evaluation. Using a causal chain design (see Spencer, Zanna, & Fong, 2005), we first showed that evaluation mode affects how much attention is paid to the different components of a product bundle. In joint evaluation, the contrast of bundle and single highlights the difference (the add-on) between the two. In a second step, we manipulated add-on salience and measured its impact on preferences. And indeed, participants indicated higher preference ratings for product bundles when the add-on was made salient. The results show that evaluation mode matters for the assessment of product bundles.

The framework presented in the present volume allows a reinterpretation of these findings. An add-on can be understood as a feature. While it is present in some options (product bundle), it is absent in other options (single object). When asked to evaluate a product, this evaluation cannot happen without a reference set. One needs to start thinking about other possible products, their possible attributes and their possible values. In our terminology, a stimulus space is elicited against which the stimulus at hand is compared. To realize the added value of the add-on (present feature), it needs to be contrasted to the possibility of it being absent in other products. If absent is not a prominent possibility, the add-on does not provide any additional value. At the same time, when only being confronted with the single option (absence of the add-on), the absence does not reveal what is missing. As a consequence, people are indifferent in separate evaluation. Presenting both options at the same time changes things. The presence of the add-on in the bundle highlights the absence in the single option and reveals the nature of the attribute. The explicit absence of the add-on in the single option highlights the presence of the add-on in the product bundle and provokes within attribute comparisons.

The experiments showed several things. In line with the presented framework and in line with experiment 1, within attribute comparisons for features do not seem to happen spontaneously when a feature is presented in isolation. When something is present, the possibility of absence is not prominent and when something is absent, one does not know what is missing. The experiments also show that explicitly contrasting presence to absence can help to restructure the mental representation of the world and promote within attribute comparisons, even for features.

4. Conclusion reference sets.

Information is not evaluated and processed in a vacuum, but rather in context, relative to other available information. Importantly, information is not only processed relative to objectively available information, but also relative to subjectively constructed information. Building on Garner (1978), our assumption is that presented stimuli elicit a mental representation of reference objects, the stimulus space, against which the stimuli at hand are compared. As outlined theoretically and as illustrated in the first experiment and the work on the “presenters paradox revisited”, this stimulus space varies as a function of what attributes and objects are presented. Importantly, it also varies as a function of how the attributes are presented. Presence and absence are evaluated differently, dependent on whether other levels of the same attribute are explicitly shown or not.

The different levels of a dimensions are qualitatively the same. Both are positively defined and hence there is no asymmetry. The natural reference set for a value on a dimension are other levels of the same dimension.

The levels of a feature are qualitatively different from one another. While one is present, the other is absent. Absence comes with ambiguity. As discussed, there are problems of segmentation (Gallistel, 2007) as well as problems of detectability (Blaisdell et al., 2009). Additionally, absence does not allow the same inferences about the stimulus world as presence does. Hence, the natural reference set for features does not consist of other levels of the same attribute, but rather of other present attributes in other objects.

With these insights about the relativity of mental representations of multi object or multi cue environments as a foundation, the rest of the empirical work of the present volume investigates the role of feature-dimension framing in contingency learning. The core idea is that dependent on how attributes are framed in a multi cue scenario, different aspects of the stimulus environment become more relevant for comparisons and invite different contingency learning strategies. Framing is not only about how single objects are evaluated, but also about how relationships between attributes are learned. An introduction to judgments of relation and an analysis of density biases and pseudocontingencies are presented before feature-dimension framing and general ideas of reference sets are introduced into the logic of contingency learning.

5. Judgments of relation.

Judgments about relationships of variables are of enormous importance. They help us understand the past, influence the here and now and predict the future (Crocker, 1981). Inferences about correlation and causality are at the heart of scientific progress and to accumulate valid knowledge, it is important that these inferences are correct. If the question is whether a certain treatment can cure depression or not, today, everyone with sufficient scientific training can easily come up with an experimental design to find an answer. One group gets the treatment while the other group does not and then one measures if depression gets better or not. It is crucial to collect all four possible combinations of treatment and outcome of this design and to weight all of them equally: Treatment (Yes vs. No) x Outcome (Good vs. Bad). If we miss one or multiple cells of this design, we cannot make valid inferences. If, for example, we just take a group of depressive people and apply our treatment to all of them, we can only measure how many of these people feel better or worse after the treatment. Importantly, in this case, we cannot infer anything substantial about our treatment. Even if everyone feels better afterwards, it might just have been an effect of time.

The idea of a control group is easily accessible today, but nevertheless a relatively new concept. Only in the beginning of the 20th century did the understanding of experimental control grow enough to slowly establish control groups in clinical and medical as well as psychological research (Boring, 1954). People show a positive testing strategy when testing hypotheses. Rather than looking for confirming as well as disconfirming information, people prefer to mainly look for confirming evidence (Snyder & Swann, 1978). And even when disconfirming evidence is present, the evidence is often not interpreted adequately (e.g. Nickerson, 1998). There are many more examples like this, all illustrating that some information is more readily used than other information. Problematically, this can lead to biased judgments about how variables relate. The task of drawing correct inferences about relationships of variables is not an easy one. It can be difficult to be aware of all the important information and to adequately use it. People seem to focus more on some types of observations while they neglect other observations. Goal of the present work is to show that one factor influencing what types of observations people focus on is the way in which attributes are presented: Either as being present vs. absent or as always being present on some level.

5.1 Normative contingency indices.

Most research on correlation and causation is based on variations of normative indices (for an overview, see Hattori & Oaksford, 2007). In the simple case of two dichotomous variables (the setting of interest in this work), one prominent and intuitive normative index is ΔP. It expresses a difference in conditional probabilities. Figure 3 shows the standard labeling of a 2x2 frequency table. From left to right and top to bottom, the four cells are labeled A, B, C, and D. Which cue is presented row-wise and which cue is presented column-wise as well as the order of the cue levels are arbitrary. However, when attributes are presented in a present vs. absent fashion, the A-cell typically refers to the present-present combination and the D-cell to the absent-absent combination (right part of Figure 3).

Using these labels, ΔP can be defined as:

$$\Delta \mathrm{P} = \frac{A}{A+B} - \frac{C}{C+D}$$

If this index is not 0, there is a relationship between the cues. This means that relatively speaking, one cue level of Cue 2 appears more often with a given level of Cue 1 than it is the case for the other cue level of Cue 1. The word relative is of importance here, because absolute frequencies can differ without the conditional probabilities being different. The crucial part is that this requires all cue combinations to be weighted equally (according to their frequency). Over- or underweighting of cue combinations leads to distortions in the contingency estimation. If, for example, people give a higher impact to the A cell than to the other cells, contingency judgments will be inflated.

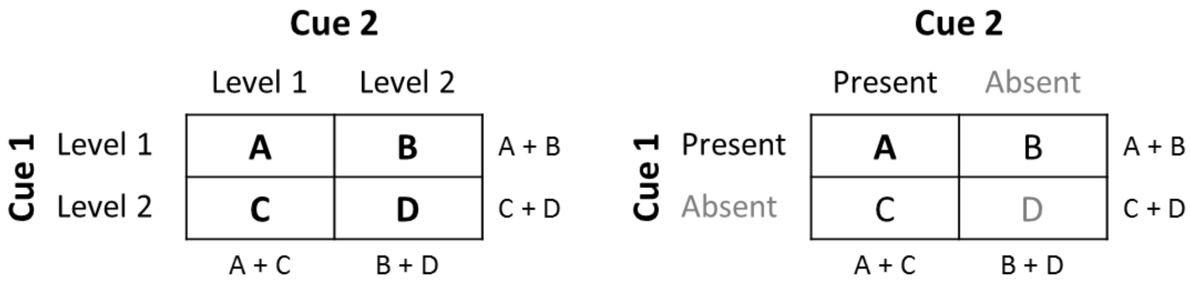

Figure 3. Illustration of standard cell labeling in a 2x2 contingency table. The two cues have two levels each and the different cue combinations (cells) and their frequencies are labeled A,B,C, and D. Cue base rates are the unconditional probabilities of the cues and can be calculated as indicated by the small formulas next to the cells (e.g. A+B). In the case of two cues which are presented as being either present or absent, the A-cell is defined as the present-present combination and the D-Cell is defined as the absent-absent combination (right table).

ΔP is not the only "correct" way to calculate a contingency, but there are other prominent normative indices, such as the phi-coefficient (ϕ, e.g. Hattori & Oaksford, 2007) The main difference being that the phi-coefficient is symmetric. For ΔP, it can make a difference whether you look at the relationship A $\rightarrow$ B or the relationship B $\rightarrow$ A. The distributions used in this experiment are always symmetrical in terms of row- vs. column wise ΔP. We use ΔP because it is easily accessible by asking for two conditional probabilities and because it is the most intuitive index when making predictions from one variable onto the other one. While people have shown to be generally sensitive to varying correlations (varying ΔP; e.g. Hattori & Oaksford, 2007; Mata, in press), people seem to also be influenced by other factors, such as stimulus base rates and cue labels.

5.2 Density biases.

When variables are presented as either being present or absent, contingency inferences usually increase with the probability of the attributes being present (density). Formally speaking, the more skewed a base rate is towards the present level of the cue, the higher the contingency judgments are on average. For example, in one early experiment, participants were asked to relate the (observed) movement of a joystick, left movement vs. absence of movement, to the movement of a dot, left movement vs. absence of movement (Allan & Jenkins, 1983). Holding the actual correlation constant at zero, correlation judgments increased with the probabilities or densities of the joystick and the dot moving (see also Hannah & Benetau, 2009). Since these effects are often examined in research on causality, researchers differentiate between cue (here joystick movement) and outcome (here dot movement) density. The underlying idea, however, is the same: The more present instances, the higher the contingency judgment. Ample evidence supports both the idea of cue-density bias (e.g. Matute, Yarritu, & Vadillo, 2011; Wasserman, Kao, van Hammer, Katagiri, & Young, 1996) and the idea of outcome-density bias (e.g. Allan, Siegel, & Tangen, 2005; Perales & Shanks, 2007; Wasserman et al. 1996).

As variations on Δp, weighted averaging models have been proposed to capture these so called density biases. In these models, the four different cells receive different weights with which they contribute to a contingency judgment, usually in the order $A > B > C > D$ (using the notation above with A being the present-present cell; Wasserman, Dorner, & Kao, 1990). In some cases, the best fitting weights include a zero weight for cell D (absent-absent cell; Hattori & Oaksford, 2007) or even a zero weight for all except the A cell (present-present; White, 2009). Thus, in a heuristic fashion, some of the evidence relevant for assessing the contingency is neglected while other evidence is overweighted. While the exact proposed weights for the different cells vary, they share a common aspect: When attributes are presented in a present vs. absent fashion, contingency judgments are heavily influenced by the A-cell, i.e. the frequency of present-present observations.

5.3 Pseudocontingencies.

For a proper contingency judgment, information about joint occurrences of attributes is necessary. Only when there is information about how often the attributes occur together, a normatively correct judgment is possible. Unfortunately, this information is not always available and even when it is actually available, there might not be enough resources to use all of that information. In those cases, people seem to approximate the relationship by considering category-level data (Fiedler & Freytag, 2004). If the base rate of a certain event (e.g. high number of criminals) is jointly skewed with the base rate (BR) of another event (e.g. high number of foreigners) in a certain category (a city district), people seem to infer a relationship between the two on the exemplar level. Even though the category data is not the appropriate data level to make inferences about exemplars, people seem to use that information. When a judgment of relation is made about two cues, it has been shown that this judgment varies as a function of skewness of both base rates. If both cues have a frequent and a rare level, frequent and frequent get associated and rare and rare get associated. These effects have been explained as the results of a heuristic alignment rule named pseudocontingency (PC, Fiedler & Freytag, 2004). Akin an aggregation bias (Hammond, 1973) and unlike a variation on ΔP, judgments are no longer based on the cell frequencies A to D, hence the term "pseudo", but on the aggregated base rates or unconditional probabilities of the variables. As soon as both variables have frequent and less frequent levels, following the PC strategy leads to the belief that frequent levels belong together and rare levels belong together. With reference to the notation in Figure 3, PCs can be formalized in the PC-index as:

$$\mathrm{PC} = \log_{10}\left(\frac{A+B}{C+D}\right) * \log_{10}\left(\frac{A+C}{B+D}\right)$$

The formula is taken from Kutzner (2009). Any deviation from a 50/50 distribution in the same (different) direction yields a positive (negative) PC-index. When one of the base-rates is not skewed (BR = 0.5), the PC-index is always 0. Note that even though PC-indices can be computed using the cell frequencies, they are defined without any knowledge of joint occurrence of the attributes and can therefore lead to wrong conclusions. The computer simulations presented in chapter 6 of the present volume show further properties of the PC-index. Quantifying the joint skewness of base rates of a distribution in an index allows for very detailed testing of predictions. Rather than being able to only test joint skewness vs. no joint skewness, different levels of joint skewness can be compared. Also, subjective base rates and subjective cue frequencies can be used to estimate a PC-index for every participant and every judgment. This PC-index can then be used to predict contingency judgments. If participants do follow such a PC strategy, their contingency judgments should be higher the higher the PC-index is.

Ample evidence from social and cognitive psychology speaks to pseudocontingency inferences (for reviews see Fiedler, Freytag, & Meiser, 2009; Fiedler, Kutzner, & Vogel, 2013). Interestingly and importantly, most research on pseudocontingencies has been done using attributes with two present levels. For example, in one experiment (McGarty, Haslam, Turner, & Oakes, 1993), participants were informed that they were to learn about members of two social groups with one of the groups forming the majority (being more frequent). When participants observed a stream of many positive and few negative behavior descriptions *without* reference to a group membership, the estimated probability of observing a positive behavior for a majority group member was higher than the same likelihood for a minority group member. Thus, without any joint observations, contingency inference linked frequent attributes and rare attributes (see also Eder, Fiedler, & Hamm-Eder, 2011). The present work will not only investigate pseudocontingencies in an environment where all attributes have two present levels, but also in environments where attributes have one present and one absent level. To anticipate, we expect pseudocontingency inferences to be reduced for present vs. absent attributes because the "necessary" category level data should not be as prominent if the comparisons within attributes are more ambiguous.

Importantly, pseudocontingencies inferences are not limited to situations in which no joint observations are available. Also in situations in which all information about cue combinations is given, jointly skewed base rates lead to an overestimation of contingency. For example, Fiedler (2010) showed that pseudocontingencies can override genuine contingencies between multiple cues, even when all necessary information is presented simultaneously. Participants were presented with information about 4 cues describing different people (gender, place, subject matter, and hobby) and later had to judge the relationship between the different cues, which all had skewed base rates. It was manipulated within participants whether the actual contingency was zero, consistent or inconsistent with the pseudocontingency. Even though participants were presented with the complete joint information about all attributes at the same time, their contingency judgments were solely driven by the skewed base rates.

5.4 Density biases vs pseudocontingencies.

The described phenomena seem similar at first glance. For the density bias / A-cell based strategy, as well as for pseudocontingencies, base rates for both cues are of importance. Both phenomena are prominent in the literature, but there are no attempts of integrating the two. A careful analysis shows that there are important differences between the two. For the PC-index, cue level labels do not matter - it is just about the alignment of skewed base rates. Given a double skewed stimulus distribution such as in Figure 4, the PC prediction does not vary as a function of cue labels. Both base rates are skewed and therefore a high contingency judgment is predicted in all cases. In contrast, the prediction based on an A-cell strategy does vary. Remember that in our terminology the A-cell is always defined as the present-present cell. Contingency judgments should be the highest when the present-present cell is the most frequent cell and the lowest (or even negative) when the present-present cell is the rarest cell. When investigating these two phenomena at the same time, it is therefore important to create constellations where PCs and an A-cell based strategy differ in their prediction. Based on the following rational, we argue that for present vs. present attributes, people rely on base-rate driven judgments (PCs) while for present vs. absent attributes, joint present observations (A-cell frequency) drive contingency judgments.

	Present	Absent	
Present	9	3	12
Absent	3	1	4
	12	4	

	Absent	Present	
Present	9	3	12
Absent	3	1	4
	12	4	

	Present	Absent	
Absent	9	3	12
Present	3	1	4
	12	4	

	Absent	Present	
Absent	9	3	12
Present	3	1	4
	12	4	

Figure 4. All four tables show the same frequency distribution. The actual ΔP is zero in all cases, but both base rates are skewed. A pseudocontingency strategy will lead to the same contingency judgment in all cases (both cues are equally strongly skewed, leading to a high contingency judgment). A strategy based on the A-cell (present-present) leads to different contingency judgments depending on which cell is the present-present one. Contingency judgments should then be highest in the upper left scenario and lowest in the lower right scenario.

The example in Figure 4 illustrates that the PC-index and the A-cell (present-present) can be disentangled to some extent. Still, at this point we do not fully understand how exactly they are related. To fully understand how the PC-index reacts to variations in single cell sizes and base rates and how it relates to ΔP, a more thorough analysis is needed. A computer analysis presented on the next pages will give insight into how different indices of contingency tables relate to one another. It will show how the different properties of contingency are bound together and to what extent they can still vary independently.

6. Understanding contingency indices: An R-framework for stimulus distributions.

When trying to disentangle A-cell and PC-index, it is important to properly understand how the two are related across the world of possible stimulus distributions. The computer simulations presented in this work provide the desired analysis. They allow an understanding of how different indices of contingency relate to each other in randomly generated 2x2 contingency tables. Simple correlations as well as more elaborate plots between different indices can be generated. To anticipate and to illustrate, it can for example be visualized how the PC-index and A-cell size relate to each other. That way, we can understand if they are completely independent or if they put specific boundaries on each other. This goes beyond simply looking at correlations. Whenever the relationship is not linear, simply calculating a correlation might give the wrong impression. Also, the framework can be used to randomly generate a large number of 2x2 distributions which all follow specific demands and is free to use for everyone. For example, 10000 random distributions with skewed base rates between 0.75 and 0.80, a sample size between 40 and 45, and a correlation between 0.5 and 0.55 can be generated with just a single line of code and a minimum of time. This can be really helpful when trying to integrate a representative design (Brunswick, 1955; Dhami, Hertwig, & Hoffrage, 2004; Hammond, 1954) where not only participants, but also stimuli are sampled. The importance of stimulus sampling has been highlighted in the recent past (Judd, Westfall, & Kenny, 2012) and is important for power concerns as well as generalizability of results. It is important to note that randomly sampling stimulus distributions based on a computer simulation does not guarantee representativeness. The idea behind a representative design is that stimuli are randomly sampled from the organism's natural environment (Dhami et al., 2004) to which the experimenter wishes to generalize. The presented simulations cannot give insight into the frequencies with which certain constellations appear in our environment, but they can give insight into what constellations are possible.

The fully documented R Code (R Development Core Team, 2008) for the framework can be found in the Appendix 1. You can also contact the author (Max Ihmels) for the most recent version.

6.1 General properties of the simulations.

To understand the implications of the simulations, it is important to understand how the distributions are generated. The mechanism is very simple: The user specifies as many parameters as required to have specific values and then distributions are randomly generated until all the conditions are met. To be exact, the size of each cell is randomly drawn from a uniform distribution between the specified cell minimum and cell maximum. For example, each cell would have a random frequency between 1 and 20 (default setting). Therefore, all combinations of cell frequencies are equally likely, but are not necessarily represented in a simulated data set.

When thinking about the relationship of two correlation indices, there are two main questions one can ask. First: What combinations are possible? Second: What combinations occur in our environment? The presented simulations will mainly address the first question. Still, illustrations will give some insight towards likelihood of specific combinations, given something is generated randomly the way it is done here. A slightly different approach to similar questions can be found in Kutzner (2009).

In the present work, all simulations are created randomly with random cell sizes between 1 and 15 for each cell. No other parameters are specified. Each illustration is based on 30000 simulated distributions.

As a quick reminder, these are the formulas used to calculate ΔP and the PC-index.

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

$$(2)\quad PC = \log_{10}\left(\frac{A+B}{C+D}\right) * \log_{10}\left(\frac{A+C}{B+D}\right)$$

6.2 Understanding the PC-index.

For all simulations, the PC-index is calculated as specified in formula 2. Figure 5 illustrates the properties of the PC-index and its dependency on the base rates. If one of the two base rates is BR = 0.5, the PC-index is always 0. Only when both base rates are skewed, the index starts to become bigger (in absolute terms). When both cues are skewed into the same direction, the index becomes positive. If they are skewed in different directions, it becomes negative. Note that calling a PC positive or negative is as arbitrary as calling a contingency positive or negative. It fully depends on how the data is presented. Just by swapping rows or columns in a 2x2 table, a positive PC (or contingency) can be reversed (and vice versa). However, once a specific labeling of a 2x2 table is determined, it is helpful to talk about positive or negative PCs and contingencies, because then one can derive which cue levels are associated. We always define the PC-index in a way that when the PC-index is positive, it suggests associating the same cue levels as when the contingency is positive (and vice versa).

The graphs also highlight a specific property of the PC-index. Given one base rate is fixed on any value but BR = .5 (say at BR = .75), the PC-index is bigger the more extreme the other base rate is. Another possible implementation would have been to include some sort of matching advantage. In such a case, the index would be biggest if both base rates are skewed exactly the same amount and get smaller once the base rates are less similar, even if one of the base rates gets more extreme. It is yet to determine which of the two modeling approaches is more accurate, but we chose to implement the non-matching index for simplicity. The more extreme the base rates, the more extreme the PC-index.

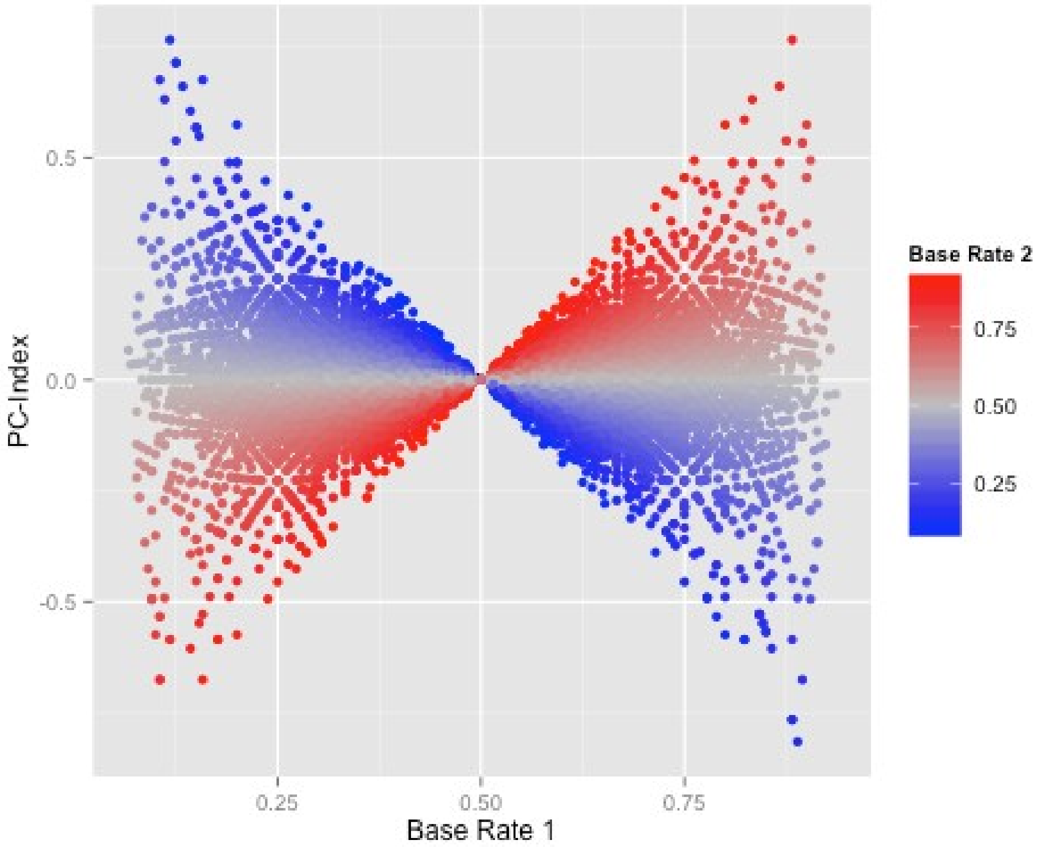

Figure 5. Relationship between cue base rates and the PC-index. If at least one of the base rates is BR = .5, the index is always zero. In all other cases, the index becomes more extreme as base rates get more skewed.

6.3 PCs and ΔP.

The PC-index and ΔP are not completely independent. While the overall correlation in this random sample is close to zero with r = 0.06, the two indices put limiting boundaries on each other. The higher ΔP is in a distribution, the higher the lowest possible value of the PC-index. The lower ΔP is in a distribution, the lower the highest possible value of the PC-index. Naturally, this relationship can also be described the other way around. With a PC-index close to zero, almost every ΔP is possible. The higher the PC-index, the higher the minimum possible ΔP. The lower the PC-index, the lower the maximum possible ΔP. Similar to the idea behind the method of bounds (Duncan & Davis, 1953) one cannot infer the exact value of one of the indices based on the other, but one can infer the possible minimum and maximum. Figure 6 illustrates the relationship.

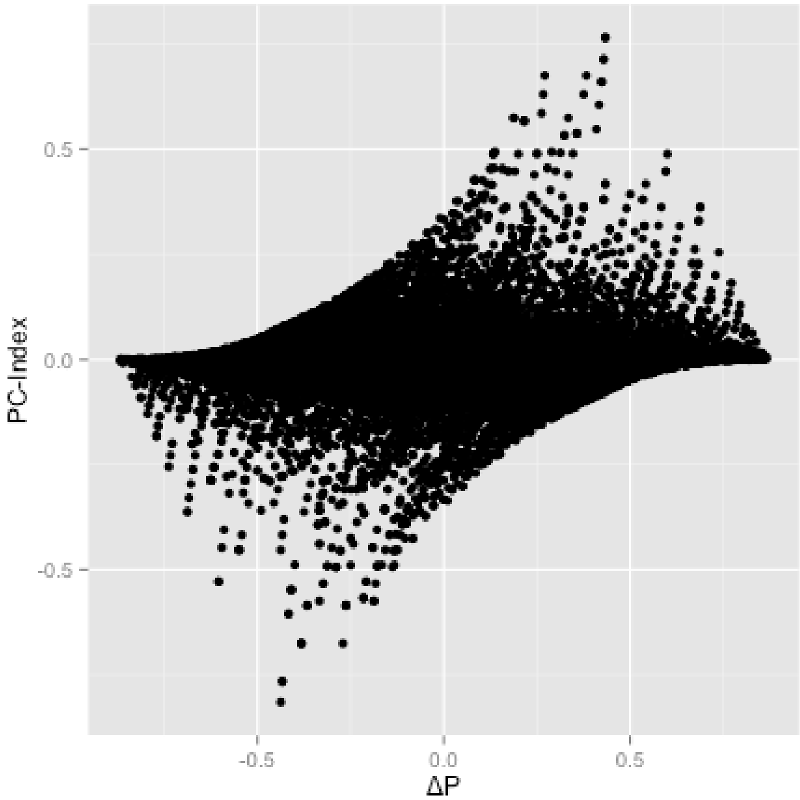

Figure 6. The relationship between ΔP and the PC-index. A fixed value on one of the indices puts boundaries on the possible maximum and minimum of the other index.

The low correlation of PC and ΔP in this sample should not be seen as an indicator for independence in the environment. It should rather be understood as an artifact of the simulation procedure. Different analyses have shown that following a PC strategy can often be adaptive (e.g. Fiedler et al., 2013). Overall, the results show that while there is dependency between the two indices, it is definitely possible to manipulate them independently in an experiment.

6.4 Base rates and ΔP.

The preceding paragraph already gave some information about the relationship of base rates and ΔP. However, base rates were only investigated in the combined form of the PC-index. First of all, the base rates are naturally completely independent from one another. When looking at the individual base rates and ΔP in figure 7, it becomes evident that there is no correlation between base rates and ΔP, r = .00. Still, they are not completely independent. Similar to PC and ΔP, base rates and ΔP put limiting boundaries on each other.[2]

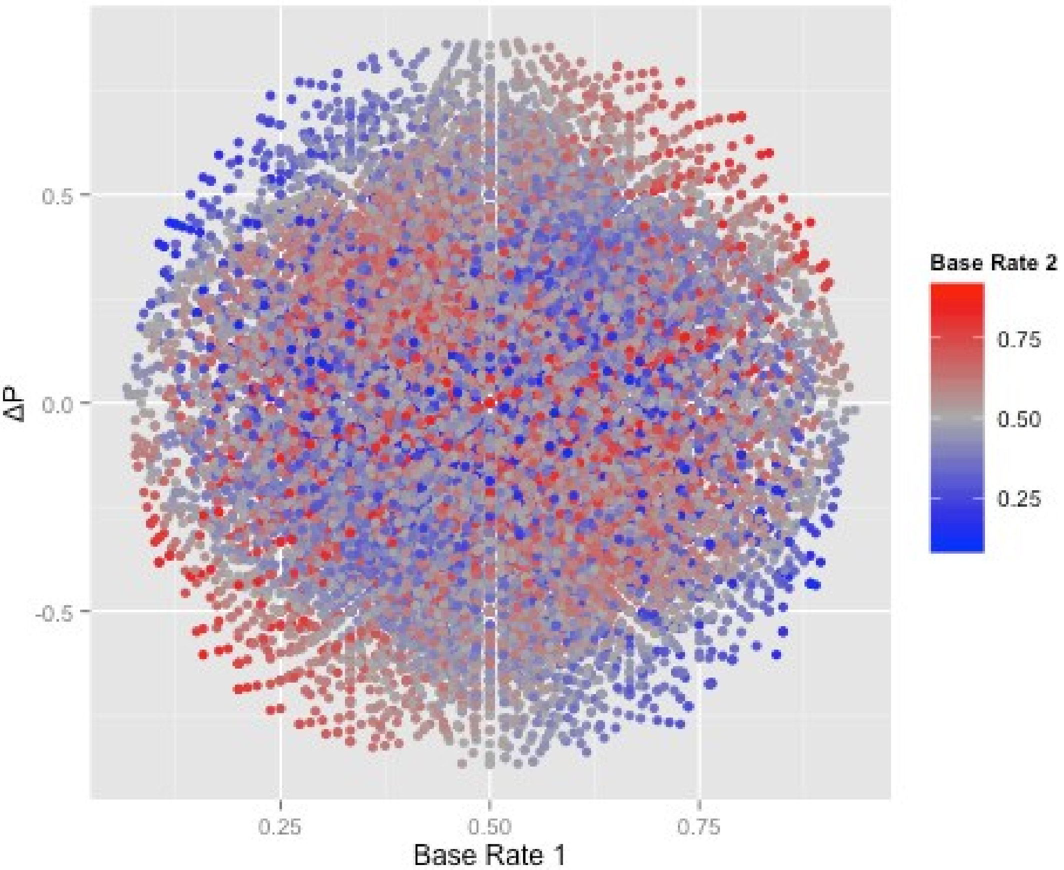

Figure 7. Relationship between cue base rates and ΔP. While the two base rates are naturally independent from one another, the base rates determine what range of possible ΔP values are possible.

---

[2] The phi coefficient has also been shown to depend on the base rates. See Zysno (1997) for an overview and a modified version which depends less on base rates.

With a base rate of 0.5, all values of ΔP are possible. The more extreme the base rate gets, the less extreme ΔP are possible. Again, this can also be viewed the other way around. With a ΔP close to 0, all base rates are possible. The more extreme ΔP gets, the less extreme the base rate can be. Also, the logic of comparing just one base rate to the actual correlation is equivalent to the idea of a density bias. Therefore, these analyses can also be understood as an investigation of the relationship between cue density, outcome density and ΔP. The three-way relationship of both base rates and ΔP is quite complex, but one notable aspect is that once one of the base rates is skewed and ΔP is desired to be high, the other base rate has to be skewed as well, as visible by the only red and only blue areas around the outer areas of the circle.

6.5 Single cells (A-cell) and ΔP.

Single-cell size (relative to the total N of a distribution, e.g. A / N) and ΔP are clearly related. Dependent on which cell we look at, this relationship is either positive (A & D) or negative (B & C), $M_{abs(cor)} = .55$. Still, there is room for ΔP variation, given a specific Cell size. Figure 8 shows the relationship between ΔP and the single cells. Regardless of which cell is manipulated in size, the impact is always the same. Note that for the Figure, the labels A,B,C, and D were given based on how they are used in the ΔP formula. When we talk about the A-cell in an experiment however, it always refers to the present-present cell of the 2x2 table when both cues are presented in a present vs. absent fashion. These examples once again illustrate that given a constant ΔP, A-cell size (present-present cell) can be varied just by changing the labels of the contingency table.

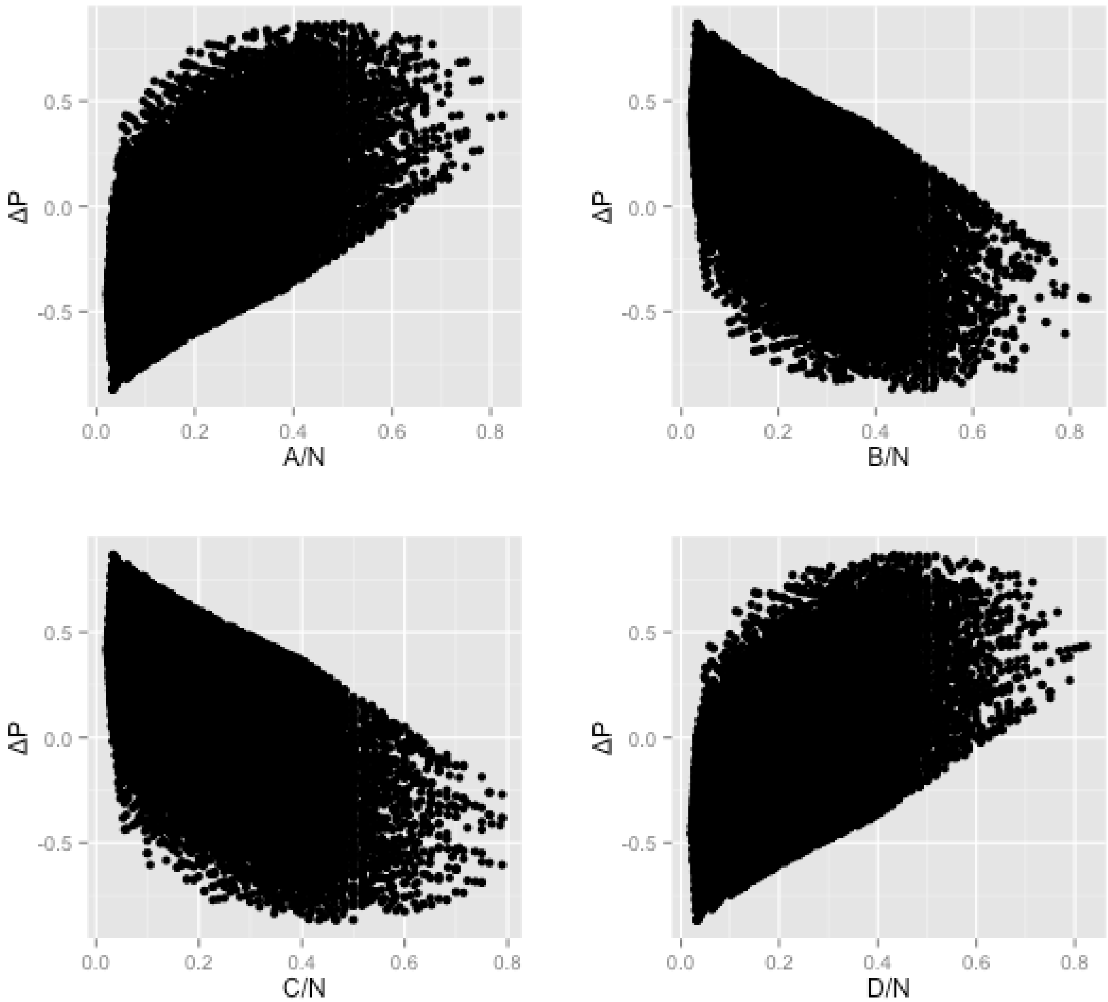

Figure 8. Relationship between single cell sizes and ΔP. The cell sizes are normalized on the sample size of the distribution.

6.6 A-cell & PC.

The relationship between a single cell size and the PC-index is not as straight forward. While there is an overall correlation, $M_{abs(cor)} = .18$, the relationship differs across different levels of single cell size. When the cell of interest only constitutes a small or chance level fraction of the entire set of observations, the PC-index can take any value. With growing size of the cell of interest, the possible variation gets smaller (quite symmetrically) but with even higher cell sizes, variability decreases further while the absolute value goes up.

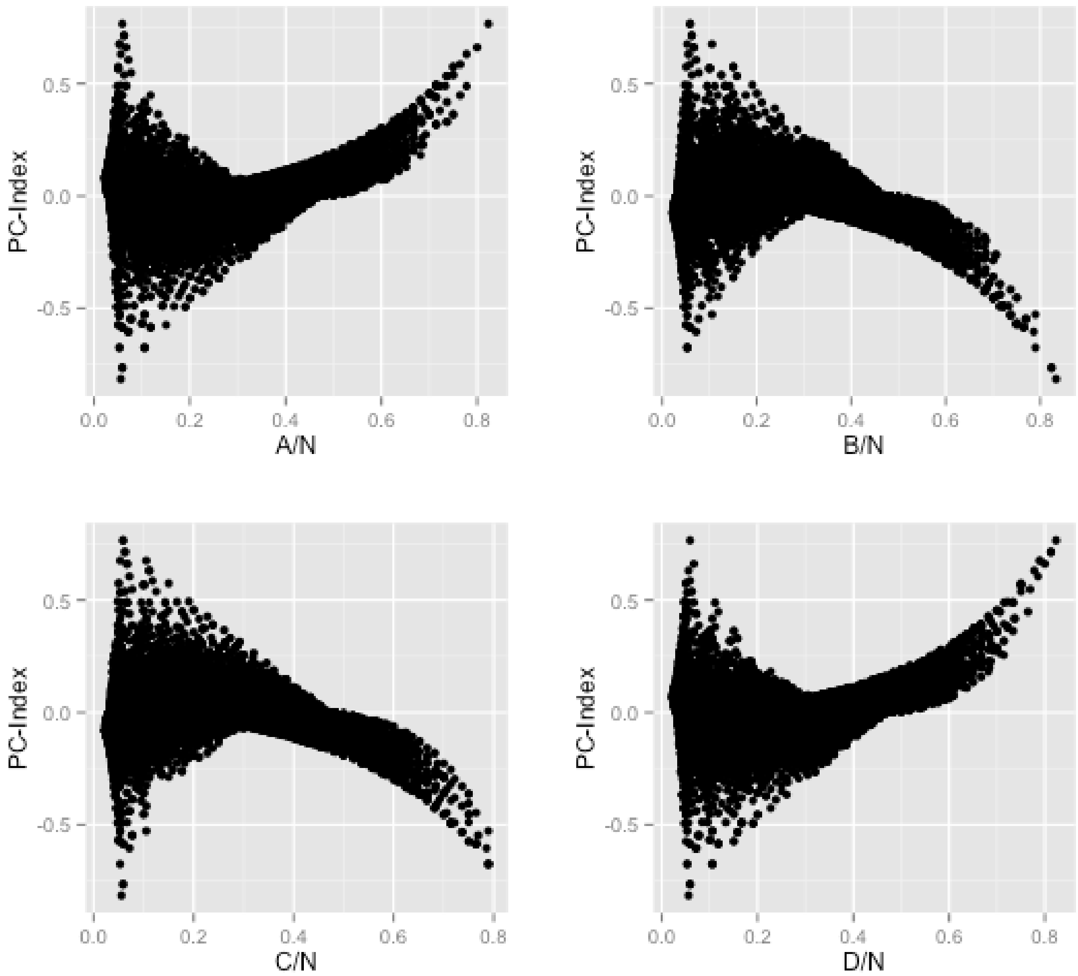

Figure 9. Relationship between single cell sizes and PC-index. The cell sizes are normalized on the sample size of the distribution. Lower cell sizes allow a large variation of the PC-index. Higher cell sizes restrict the possible variation of the PC-index.

Put simply, when the A-cell is small, PC can take any value, when the A-cell is large, PC will be large as well. Top left of figure 9 illustrates this relationship. Similar to the analyses of cell size and ΔP, it is again arbitrary which individual cell we look at. The other three graphs of figure 9 illustrate this aspect. Naturally, this relationship can also be viewed from PC to single cell size, instead of the other way around. Given a specific value of ΔP, only a certain range of possible single cell proportions are possible.

6.7 A-cell, PC & ΔP.

After investigating the pairwise relationships between A-cell size, the PC-index and ΔP, the last step is to investigate the three-way relationship. Given the relationship of PC and A-cell size was already quite complex, the three-way relationship is even more complex. Figure 10 illustrates how ΔP is distributed across the different combinations of A-cell size and ΔP. Again, not all combinations are possible. For the lower fifth of A-cell sizes, ΔP is more positive, the higher the PC. For the rest of possible A-cell sizes, ΔP is more negative, the higher the PC. Additionally, the bigger the A-cell, the higher the minimum ΔP. The simulation gives insight into what can actually be manipulated within an experiment. For example, an A-cell size of 0.6 with a PC value of 0 and a ΔP of -0.25 is not possible (in this range of sample sizes).

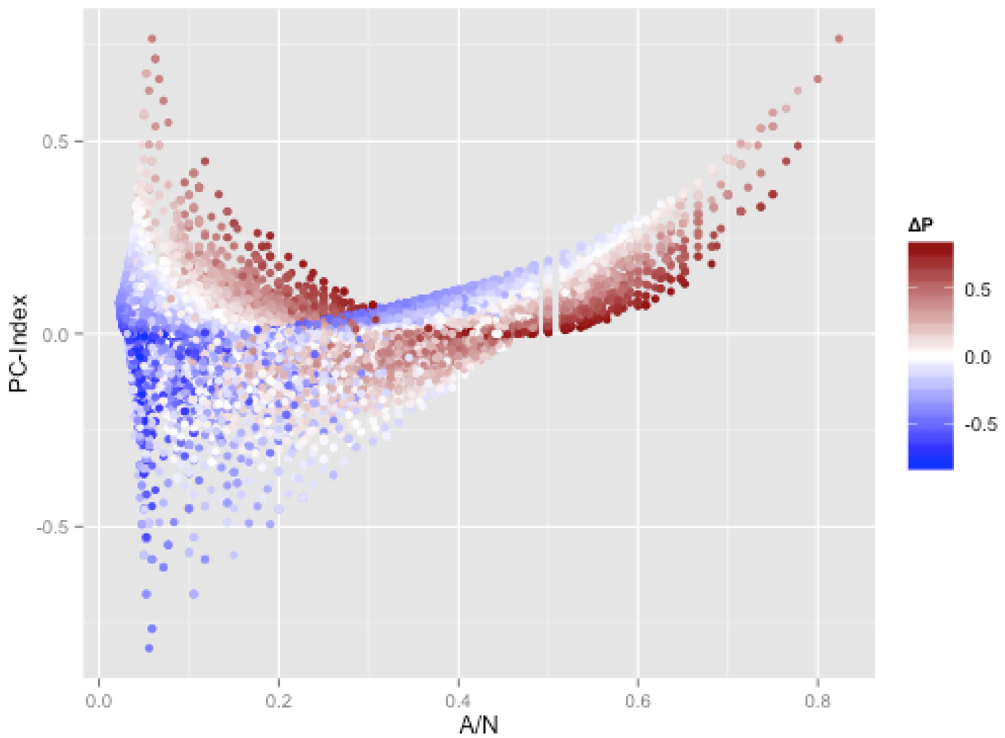

Figure 10. Three way relationship between A-cell size, PC-index and ΔP.

6.8 Simulations – Conclusion.

The presented simulations illustrate that the different indices used in the present work are not completely independent from one another. They put certain boundaries on one another. Given a certain value of one index, only a certain range of another index is possible. Just looking at correlations would not have allowed the same insights, since they lose a lot of information and can therefore be misleading when the relationship is not linear. In that regard, the present analyses go beyond other analyses also concerned with the relationship between contingency indices and strategies which only look at correlations between indices (e.g. McKenzie, 1994). The simulations also illustrate that within certain boundaries, the indices can be manipulated independently in an experiment. Similar to the method of bounds (Duncan & Davis, 1953), one cannot infer the exact value of all indices just by knowing about one, but one can determine a possible minimum and a possible maximum.

Again, purpose of the simulation was not to investigate how adaptive the different indices are (e.g. serve as a proxy for ΔP). The idea was rather to promote the understanding of the different indices and their relationships. Further, the provided R framework can be used to generate countless distributions which follow specific demands with little effort.

7. Features and dimensions in contingency learning.

As illuminated theoretically, as well as analyzed on basis of the simulation results, an A-cell (proportion of present-present occurrences) based strategy and a pseudocontingency strategy do share some similarities, but can at the same time make very different predictions. An A-cell strategy as it is defined here can be summarized as follows: When there are two cues which have one present and one absent level each (i.e. are both features), contingency judgments are higher the more present-present combinations are observed. A pseudocontingency strategy (PC) is about an alignment of skewed base rates. When base rates of two cues are both skewed (i.e. they each have one frequent and one rare level), the frequent levels get associated and the rare levels get associated. The question now is what strategy is used in what kind of environment. The introduced framework of features and dimensions provides clear predictions for differently framed attributes.

The term feature is here used to refer to attributes which have one present and one absent level (notation: present vs. absent). The term dimension is used to refer to attributes which have two present levels (notation: present vs. present). As analyzed and illustrated before, features and dimensions differ in the inferences and comparisons they promote with regards to stimulus environments: they evoke different reference sets. These differences in reference sets inherent to features and dimensions map onto the heuristic processes behind density biases and PC-inferences.

To recall, for features, the stimulus space is expected to mainly be defined by present occurrences. For dimensions, a more elaborate stimulus space is expected in which comparisons between different levels of the same attribute are feasible.

For features, it is easier to give more weight to events that are present and therefore effortlessly stored than to those that are absent and therefore not stored or if so, with uncertainty. As a consequence, in the stimulus space, the mental representation of the stimulus world, present occurrences should be highlighted while absent occurrences should be less prominent. For a contingency judgment to be accurate, all cells have to be weighted equally. If some cells are given a bigger impact than other cells, contingency judgments are going to be biased. We believe that due to the ambiguity of absence, people will put more weight on the present occurrences than on the absent occurrences. Relying on the A-cell (or present occurrences in general) should then lead to density biases. If the frequency of present-present observations goes up, contingency judgments should go up. As illustrated in experiment 1, for features, comparisons within the same attribute are less prominent than comparisons with other present attributes. When something is presented as a feature, the reference set is mainly defined by other present occurrences instead of by other levels of the same attribute. Given that comparisons within the same attribute are not as prevalent, this should also have an impact on base rate accuracy and base rate utilization. Base rate perception should not be as accurate when attribute are presented as features, because comparisons within the same attribute are not as well defined and not highlighted. As a consequence, people should also not use the base rates for their contingency judgments. We therefore do not expect participants to engage in pseudocontingency inferences for features.

For dimensions, the assumed process is different. Here, all levels of an attribute are qualitatively the same. All are positively defined (present). Hence, there is no reason to assume an unequal weighting of the different cue levels or cells. Different levels of a dimension are very well comparable and should invite within attribute comparisons. The reference set for dimensions are other levels of the same attribute. As a consequence, base rates should be more accurately assessed. On the downside, they should also be more prevalent when making contingency judgments, promoting pseudocontingency inferences. We expect better accuracy for base rates, but also more base rate utilization in the form of PC inferences for dimensions. Note that the A-cell as it is defined here (being the only present-present cell) is not defined for dimensions.

Put simply, while features should invite comparisons with other present attributes, dimensions should invite comparisons within the same attribute.

Learning and judging contingencies is a difficult task. Relevant information has to be extracted from complex environments and then has to be correctly integrated. Judging a contingency is a relative task in which different sets and probabilities have to be compared. In the sense of ΔP, the correct or normative way to do this, is by comparing two conditional probabilities. It has been shown again and again that people make mistakes when selecting the information they want to use for (contingency) judgments. For example, when asked for the inference A → B, people use information which was sampled the other way around, B → A (Fiedler, 2008). The introduced work on density biases and pseudocontingencies show the same picture. People use base rates or individual cells to make their contingency judgment, while they should actually weigh all the individual cells equally and use exemplar level data. Goal of the present work is to show that attribute framing in the sense of feature-dimension framing influences the kind of information people rely on and the comparisons they make. Features should lead to people focusing on joint presence while dimensions should highlight base rates.

So far, the contingency learning literature lacks a structured analysis of the feature-dimension distinction, or the present vs. present and present vs. absent distinction in general. However, there is some initial evidence in line with our ideas. Judgments will more likely and more strongly take both levels of an attribute into account when the attributes are framed as dimensions as opposed to features. For example, when attributes are features, the absolute number of joint observations of the features' presence (e.g. therapy and side effect) describes contingency judgments best (e.g. Wasserman et al., 1996). When variables are dimensions (e.g. alternative vs. traditional therapy and severe vs. mild side effects), the relative number of confirming minus disconfirming instances more closely describes judgments (Allan & Jenkins, 1983). Analogously, when instructions explicitly mention just one pole of an attribute, information search focuses on just one attribute level (Crocker, 1982) and the meaning of a contingency is interpreted as the absolute number of positive pieces of evidence (Beyth-Marom, 1982). Thus, evidence from contingency judgments indicates that the framing of stimuli as features can indeed induce selective processing.

Taken together, we assume that features and dimensions trigger different processes and comparison sets for contingency judgments. For features, the important reference set consists of other present occurrences rather than on other levels (absence) of the same attribute. This should lead to growing contingency judgments with growing A-cell sizes and to inaccurate base rate estimates at the same time. For dimensions, the important reference set consists of other levels of the same attribute. Comparisons within the same cue are instigated and should highlight stimulus base rates. This should lead to more accurate base rate perception and also to more base rate utilization in the form of PC inferences.

Taken together, we expect the following:

Hypotheses

H1: Base rate estimates are more accurate in the dimension than in the feature framing (main effect of framing).

H2: A PC-index based on base rates predicts contingency judgments and choices in a dimension framing, but not in a feature framing (interaction of PC-index and framing).

H3: A-cell size predicts contingency judgments and choices in the feature framing (main effect of A-cell size for feature framing).

8. Experiment 5: All things being equal.

One simple type of distribution in a 2x2 contingency table is a flat or equal distribution where all four combinations of cue levels occur with the same frequency, say 15 times (A = B = C = D = 15). In this case, none of the base rates are skewed and the actual correlation in the set is $\Delta P = 0$. The first contingency learning experiment presented in this volume uses such an equal distribution while manipulating framing of two cues in a simple contingency learning task. The experiment should be understood as a first exploration of feature-dimension framing in a very simple scenario.

In such a case where both base rates are symmetrical with $BR_{cue} = BR_{outcome} = 0.5$, a pseudocontingency based strategy leads to a correlation estimate of $\Delta P = 0$. A dimension framing should therefore result in contingency estimates close to zero. For features, there is more information participants might rely on. Cues then have one present and one absent level each. In line with the idea of density biases, participants might focus on the number of joint present occurrences (A-cell size) to come up with their estimate. However, with regards to an A-cell based strategy, it is hard to make point predictions. Assuming the number of joint present occurrences is seen relative to the entire sample size, a chance level of A-cell size (25% of all observations) might also lead to a correlation estimate of $\Delta P = 0$. However, participants may lean towards associating similar cue levels with one another. In a feature setting, cue levels might be seen as similar on a higher level when they are of the same qualitative nature, i.e. when both are present or when both are absent. Based on this physical similarity, people might match those two (Goodie & Fantino, 1996). In line with the idea of natural input-output compatibility (Allan & Jenkins, 1983), participants might associate present levels to other present levels and absent levels to other absent levels, leading to a positively biased contingency judgment.

8.1 Method.

Effects of feature dimension framing on contingency judgments and choices are investigated in a very simple scenario using abstract materials.

Participants. 79 people (M_{age} = 32.11, SD_{age} = 11.13) participated in our experiment which took roughly 10 minutes. Participants were recruited via MTurk (Buhrmester et al, 2011) and were rewarded 5$ / hour plus a variable amount which was dependent on their performance. The experiment was created with SoSci Survey (Leiner, 2014).

Materials and Procedure. We confronted participants with the task of learning about the relationship between lines (cue) on the upper half of cards and colored bars (outcome) on the lower half of those cards. There were always two levels for the cue and two levels for the outcome, resulting in four possible combinations of cue and outcome. All combinations were presented 15 times, resulting in a total of 60 trials and a perfectly equal distribution of the 2x2 table (with $\Delta P = 0$). We manipulated framing of cues between participants. In the dimension condition, participants saw right and left tilted lines in the upper half of the cards and green and red bars in the lower half of the cards. In the feature condition, participants saw one type of tilted line (counter-balanced across participants) or nothing / absence in the upper half of the cards and a red bar or nothing / absence in the lower half of the cards.

Participants received the following instructions in the beginning of the experiment: "You are about to see a number of cards with different symbols on them. At several points in the study, the lower half of the card will show a red bar. In the second part of the study, you will only see the upper half of two cards. Your task is then to pick the card which is more likely to show a red bar in the lower half. For each correct prediction, you will be financially rewarded and get an additional 10 cents."

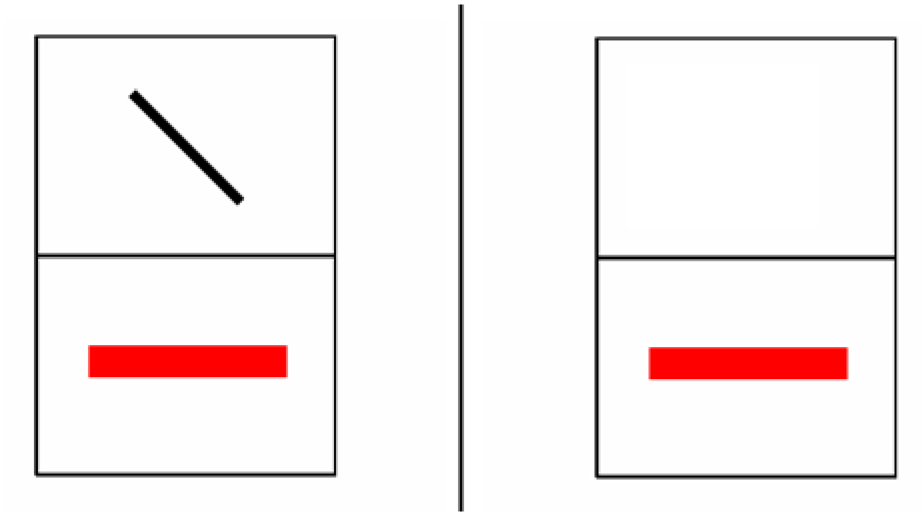

Figure 11. Exemplary trials of experiment 5. In the trial on the left, both cues are present. This could therefore either be a trial from the feature condition where both cues happen to be present, or a trial from the dimension condition where both cues are always present. The right trial shows a trial from the feature condition where the top cue shows the absent level.

Each trial was shown for at least 1.5 seconds. After this period of time, the "next" button appeared, allowing participants to continue to the next trial. The screen always turned white for a short amount of time before presenting the next stimulus to better visually separate the trials. See figure 11 for screenshots of two exemplary trials. Trials were presented in a random order.

Dependent variables - Choices. After the learning phase, participants were presented with two cards showing the two different levels of the cue (lines) in the upper half. The lower half of the cards however, was covered by a large question mark. Their task was to choose the card which has a higher chance of showing a red bar in the bottom half of the card. Participants were also told that they would receive 10 cents per correct decision. Figure 12 shows an example of such a choice trial.

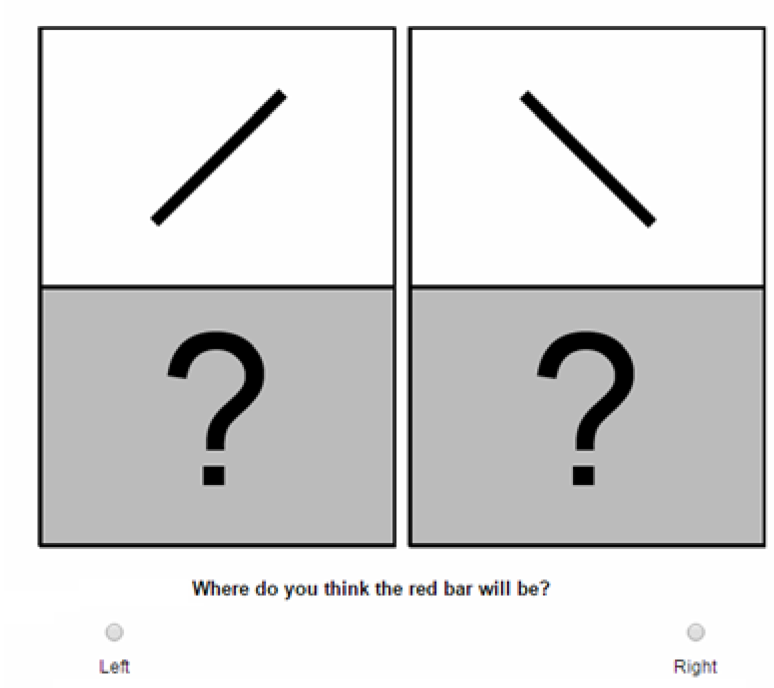

Figure 12. Exemplary choice trial from the dimension condition (both levels of the predictor cue are present). The conditional probability assessment uses a similar presentation format ('?' in the lower half of the card).

Dependent variables - Conditional probabilities. After the choice phase, participants were asked for two conditional probabilities. They were asked to estimate the probability of the card showing a red bar in the lower half of the card, given the cue levels of the upper half. We used a graphical representation similar to the choice task. The upper half of the card showed the cue level and the lower half was covered by a question mark. Based on these two conditional probabilities, a ΔP estimate for every participant was calculated by simply calculating the difference of the two conditional probabilities.

After sociodemographic variables were collected, participants were then thanked, payed accordingly and debriefed carefully. Given the equal distribution in this experiment, there were no correct or incorrect decisions. Both conditional probabilities were $p = 0.5$. Therefore, with regards to the additional payoff, for each choice, it was simply randomly decided whether a choice was "correct" or not.

8.2 Results.

As expected, the different balancing conditions (left tilted line vs. right tilted line in the feature condition) did not differ on either the choice measures or the conditional probability estimates and were therefore combined.

Choice Task. There were 10 choice trials for every participant based on which a simple choice score was calculated. In the feature condition, the score expresses how often participants chose the present level out of these 10 times. In the dimension condition, it was simply counted how often the left tilted line was preferred over the right tilted line. A t-test revealed a difference between the framing conditions, $t(77) = 1.98$, $p = .059$, $d = .45$. As expected, the two cue levels within the dimension framing were chosen nearly equally often (46.7% vs. 53.3%). In the feature condition however, participants chose the present level of the cue more often (60%) than the absent level (40%), indicating a positive perceived contingency between the two cues.

ΔP. The two conditional probability estimates were used to generate a ΔP contingency judgment for every participant by simply calculating the difference between the two values. A t-test again revealed a difference between the framing conditions, $t(77) = 3.14$, $p = .002$, $d = .72$. In the dimension framing, the estimated correlation was close to zero ($M = -.01$, $SD = .22$) while it was positive in the feature condition ($M = .15$, $SD = .24$).

Choices and ΔP. The choice measures and the ΔP based on the conditional probabilities were strongly correlated with $r(77) = .505$, $p < .001$. The more participants chose a given cue level suggesting a correlation, the higher was their contingency judgment (suggesting the same direction of relationship).

8.3 Discussion.

In a very simple environment, we showed that attribute framing in the sense of feature-dimension framing can have an impact on contingency learning processes. When a cue and an outcome were presented as dimensions (had two present levels each), the actual correlation of $\Delta P = .00$ was accurately assessed. When cue and outcome were presented as features (had one present and one absent level each), participants expressed a positive contingency, associating the present levels with one another and the absent levels with one another. This happened for incentivized choices as well as for a contingency judgment in the form of two conditional probability judgments.

To some extent, it is surprising that even in such a simple environment with only one cue and one outcome and a self-paced learning phase, it can make a difference whether attributes are features or dimensions. As discussed, dealing with absence can be an effortful and active process. When nothing is present, it has to be actively construed what could have been present. Further, when not detecting a stimulus, one cannot be sure whether the stimulus was actually absent, or whether one was just not able to detect it. This ambiguity can lead to positive contingency judgments. If absence is not as well encoded as presence, this biases judgments towards a positive relationship between the present cue levels. In other words, emphasis on the A-cell does lead to a positive contingency judgment.

In a simple setting such as in this experiment, people should have the necessary resources to deal with absence. Still, participants ended up associating the present levels of two features. This does not necessarily have to be the result of problems when dealing with absence. It might also be the result of pre-existing beliefs about causal relation. Arguably, there is a natural input-output compatibility of two features (Allan & Jenkins, 1983) and it fits causal schemas that action rather than no action leads to action rather than no action (White, 1995). Also as Goodie and Fantino (1996) showed, people seem to match physically similar items.

One might be tempted to interpret these results simply in the sense of a main effect. Contingency judgments for features are higher than contingency judgments for dimensions, given the same distribution. However, one should be careful with such an inference. In this experiment, only one type of distribution was used and it was a very specific distribution. Namely, a completely symmetrical distribution without skewness of any kind. As derived earlier in the present work, we expect framing to interact with different aspects of the stimulus environment, especially with base rates. To repeat, while we expect people to follow a pseudocontingency strategy for dimensions, we expect people to mainly focus on the number of joint present occurrences for features. We test these predictions in a number of experiments systematically varying base rates and A-cell sizes in complex multi cue environments.

9. Experiment 6: Skewed base rates.

After illustrating that people react differently to the same stimulus distribution dependent on how the information is presented, the next experiment provides a more elaborate test of how base rate skewness drives contingency judgments in different framings. Instead of using abstract material as in experiment 6, this experiment uses meaningful material in a consumer setting. Different consumer products are presented with labels about ingredients. In the feature condition, the labels say "With" or no label is present (absence) while in the dimension condition, the labels always give information about one of two levels. Further, a more complex environment is used where participants have to learn about two products with two attributes each at the same time. It is manipulated within participants whether both cues in a pair are skewed or whether only one is skewed and the other is symmetric, while keeping the actual contingency constant at zero. We do not make predictions about a main effect between the different framings, but rather about the effect of base rate skewness within the framings. For dimensions, joint skewness of two cues should lead to an overestimation of the contingency (PC). If only one cue is skewed, contingency estimates should be close to zero. For features, joint skewness should also lead to higher contingency estimates than single skewness, since in this experiment, joint skewness also leads to a bigger A-cell.

9.1 Method.

This experiment is the first exploration of feature-dimension framing in an environment where cues have skewed base rates. Framing is manipulated between participants while base rate skewness and A-cell size are manipulated within participants.[3]

[3] I want to thank Joana Brokelmann, Engin Devekiran, Lucas Donnerstag and Luisa Voßbeck for their support in running this experiment.

Participants. 169 students (111 female, M_{age} = 25.20, SD_{age} = 8.84) of the University of Heidelberg participated in our experiment which took roughly 10 minutes and was part of an experimental block of about 60 minutes. Participants were financially rewarded (8€ / hour) or received course credit for their participation. The experiment was created with SoSci Survey (Leiner, 2014) and participants were mostly recruited via the online recruiting platform hroot (Bock, Nicklisch, & Baetge, 2012).

Materials and Procedure. In this experiment, participants had to learn about two consumer products with two attributes each. In a trial by trial fashion, we presented them with one product at a time. We later asked questions about the relationship between the attributes within each product category. The different products with their attributes differed in the frequency with which they appeared. For one of the products, both attributes were skewed in their base rate (3:1, BR = .75). For the other product, only one of the base rates was skewed (3:1, BR = .75), with the other base rate being symmetrical (1:1, BR = .50). The actual contingency between the attributes within each product was always held constant at $\Delta P = 0$. See table 1 for the exact distributions. The number of total learning trials for the two products was equal. We counterbalanced which of the two products had two skewed attributes and which one had only one skewed attribute. There were a total number of 32 trials.

In addition to the within subjects manipulated skewness of the attributes, we manipulated framing between participants. In the dimension condition, all attributes were always present, but on two different levels each. In the feature condition, each attribute had one present and one absent level. For this experiment, we chose to always set the present level as the frequent level in the feature condition. Note that this does not allow an independent manipulation of the PC-index and A-cell size. This problem will be tackled in later studies of the present volume.

Table 1

Stimulus distribution for experiment 7.

Product 1 (double skew)	Level 1	Level 2	Sum
Level 1	9	3	(12)
Level 2	3	1	(4)
Sum	(12)	(4)	
Product 2 (single skew)	Level 1	Level 2	Sum
Level 1	6	6	(12)
Level 2	2	2	(4)
Sum	(8)	(8)	

Note. Participants learned about both products at the same time. For one product, base rates of both attributes were skewed while for the other product, only one base rate was skewed. It was counterbalanced across participants which product was used for which stimulus distribution. In the feature condition, the present level was always the frequent one.

We presented images of “smoothies” and “shampoo” as the two products. For the smoothies, the two attributes were “beetroot” and “iron”. In the feature condition, it read “with beetroot” and “with iron” for the present level. The absence level was simply represented by not showing any information about the attribute. For the dimension condition, we used the same attributes but presented labels which indicated different amounts of the attribute: “30% beetroot” vs. “10% beetroot” and “contains 6mg of iron” vs. “contains 1.5mg of iron”. The shampoo attributes were operationalized in a similar fashion. See figure 13 for some examples of the visual presentation. Each image was shown for 3000ms. 200ms later, a “next” button appeared, allowing the participants to get to the next trial. After an inter trial interval of 1000ms, the next image was shown.

Contingency estimation: Cell frequencies. After the learning phase, we asked participants to indicate frequencies for the different products and attribute combinations they saw. For each of the two products, this resulted in 4 judgments (2x2 of the attributes), resulting in 8 cell frequency estimates in total. Each cell frequency was assessed by showing an image of the product (with a given attribute combination) separately and asking for a frequency estimation in an open text box. Based on these cell frequencies, we calculated a ΔP for each product and each participant accordingly to the formula $\Delta P = \frac{A}{A+B} - \frac{C}{C+D}$.

Figure 13. Four exemplary trials from experiment 7. The two pictures on the left each show an exemplary trial from the dimension condition. The labels read "with X% red beet" and "with X mg of iron" and are present in both cases. The two pictures on the right each show an exemplary trial from the feature condition. The labels read "with herbs" and "with argan oil". While both are present in the shampoo on the left, only one of them is present in the shampoo on the right.

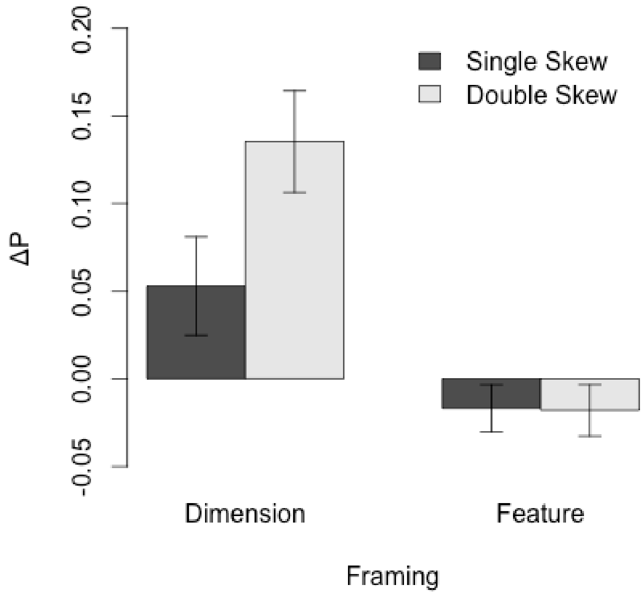

Figure 14. Mean contingency estimates (ΔP) as a function of framing and skewness of the cues. Error bars represent standard errors (± 1 SE).

9.2 Results.

Framing was manipulated between participants while skewness was manipulated within participants. Therefore, a 2 (Framing: feature vs. dimension) x 2 (Skewness: Single vs. Double) mixed ANOVA with subjective ΔP as the dependent variable was conducted. The analysis revealed a main effect of framing, $F(1, 166) = 18.64$, $p < .001$, $\eta_p^2 = .101$. Contingency judgments were generally higher in the dimension than in the feature condition. There also was a main effect of skewness, $F(1, 166) = 3.57$, $p = .061$, $\eta_p^2 = .021$. When both cues were skewed, contingency judgments were higher than when only one cue was skewed. Importantly, both of the main effects were mainly due to the high contingency judgments in the dimension * double skew condition, which is evident in an interaction of framing and skewness, $F(1, 166) = 3.78$, $p = .054$, $\eta_p^2 = .022$. Figure 14 shows the pattern of results.

9.3 Discussion.

The presented experiment delivers a first hint towards the idea that different processes play a role for contingency learning, dependent on whether attributes are either presented in a present vs. present or a present vs. absent fashion. When cues were presented as dimensions, judgments were in line with the pseudocontingency logic. When both base rates of a cue pair were skewed, the contingency was overestimated. When cues were presented as features, there was no such effect. Contingency judgments were always around zero. This was unexpected, because in this experiment, the jointly skewed cues also had the bigger number of joint present occurrences (A-cell). Possibly, the situation was not complex enough for the framing to affect the learning. Maybe participants managed to properly deal with the absence in this setup. It could also be the case that the increase from 6 to 9 pairings was just not big enough to produce the expected effects. The experiment showed that attribute framing is not just about a main effect, but rather about an interaction of attribute framing and stimulus distribution properties. To fully understand the differences in processes, stimulus properties have to be manipulated carefully alongside a manipulation of framing. Interestingly, there also was a difference between the framing conditions when only one of the cues was skewed. In that case, contingency judgments were higher for the dimension than for the feature framing. The last experiment (diamond III) of the present volume will investigate similar setups where only one base rate is skewed, while the other one is symmetrical.

It is important to note that the experimental setup did not include a fully independent manipulation of A-cell size and joint skewness (PC). With the present cell always being the frequent one in the feature condition, an increase in base rate skewness also lead to an increase of A-cell size. The next three experiments in the present volume tackle this problem. Framing will be manipulated between participants while base rates, PC-indices and A-cell sizes will be manipulated within participants at the same time.

10. Experiment 7: Diamond I.

The presented computer simulations helped to illustrate that pseudocontingencies and density biases (or an A-cell based strategy) can be conceptually distinguished. One main goal of the present work is to show that attribute framing in the sense of presenting attributes as either present vs. present or as present vs. absent can make people shift between using pseudocontingency inferences and using an A-cell based strategy. Experiments 5 and 6 showed first evidence for differences in contingency learning as a function of attribute presentation format and stimulus distribution properties. The next three experiments are the core of the present volume and present an extensive test of the moderating impact of attribute framing on contingency learning processes.

For contingency judgments to be accurate, all cells have to be weighted equally. If one of the cells gets assigned a bigger impact than the other cells, judgments are going to be biased. Further, one has to look at the correct aggregation level of data. One needs to look at the data on the exemplar level. Simply looking at category level data (base rates) is not enough to guarantee correct inferences. We believe that feature-dimension framing can lead to shifts in how the individual cells are weighted and how much people use category level data.

A representation of the stimulus world based on a feature space should highlight the presence of attributes. As discussed, absence can lead to ambivalence. Absence is hard to sequence, absence can be difficult to interpret when detectability is not perfect and absence does not allow the same inferences about the stimulus world as presence does. While the present level of an attribute does reveal the nature of the attribute, the absent level does not tell what is missing. This should lead to an overweighting of present occurrences (especially the A-cell) when dealing with features. At the same time, comparisons within a feature (comparing presence to absence) should not be as prominent in the stimulus space, since the reference group is not as clearly defined when one of the levels is absent and therefore ambivalent. On the downside, this should lead to an impaired perception of base rates. On the positive side, base rates should also not be very prominent in the representation of the stimulus world and hence not trigger pseudocontingency inferences.

For dimensions on the other hand, we expect a more balanced representation of the stimulus world. Since all of the cells in a 2x2 table based on dimensions are qualitatively the same, there is no reason to assume any sort of cell weighting inequality. At the same time, comparisons within an attribute are well defined and should therefore be easy. The reference set for both levels of a dimension is clearly defined (i.e. the other present level of the dimension). On the positive side, this should lead to accurate base rate estimation. On the downside, this should also highlight base rates and hence lead to pseudocontingency inferences.

In summary, participants are expected to rely on the number of joint occurrences (A-cell) when dealing with features and are expected to rely on an alignment of skewed base rates (PCs) when dealing with dimensions. Participants should also be better at accessing skewed base rates in a dimension framing. The experiments also serve as a test to whether participants are able to additionally detect differences in actual correlations in a complex multi cue environment. In the three experiments, we built on work by Fiedler (2010) who presented participants with a complex multi cue environment in which base rates of all cues were skewed and the actual correlation between cues varied. In the original work, participants were not able to detect differences in the actual correlation but mainly relied on inferences in line with pseudocontingencies to arrive at their contingency judgments. We hope to show that learning of actual contingencies can happen at the same time as people are influenced by skewed base rates and different A-cell sizes. We tried to improve the paradigm, instructions and questions to make the task a little easier for participants. In the paradigm, there are four cues with two levels each. In the version presented by Fiedler (2010), all cues had two present levels each. In our terminology: All stimuli were presented as dimensions and in line with the theory presented here, these dimensions led to PC inferences.

Experiments 7 to 9 build on this multi cue learning paradigm and provide support for the idea that features promote an A-cell based strategy and dimensions promote PC inferences. The first of the three experiments varies framing (feature vs. dimension) between participants and actual contingency as well as A-cell size within participants while measuring base rate and contingency perception. The already presented hypotheses are related below to facilitate the reading process.

Hypotheses.

H1: Base rate estimates are more accurate in the dimension than in the feature framing (main effect of framing).

H2: A PC-index based on base rates predicts contingency judgments and choices in a dimension framing, but not in a feature framing (interaction of PC-index and framing).

H3: A-cell size predicts contingency judgments and choices in the feature framing (main effect of A-cell size for feature framing).

10.1 Method.

This is the first of three experiments systematically disentangling actual contingencies, the number of joint occurrences and joint skewness between cue pairs while investigating the moderating impact of feature-dimension framing.

Participants. 176 students (114 female, $M_{age} = 23.27$, $SD_{age} = 5.12$) of the University of Heidelberg participated in our experiment which took roughly 15 minutes and was part of an experimental block of about 60 minutes. Participants were financially rewarded (8€ / hour) or received course credit for their participation. The experiment was created with SoSci Survey (Leiner, 2014).

Overview of the experimental procedure. We used an experience based multi cue learning procedure based on a paradigm by Fiedler (2010). In our version of this paradigm, participants were told that they would see a large number of students' profiles containing information about four domains: Sports, culture, city and discipline. They were asked to pay attention to relationships between the different cues to later be able to answer question about these relationships. After the learning phase of 96 trials, participants were asked to give base rate estimates as well as two conditional probability estimates for all possible cue combinations, resulting in six contingency judgments.

Stimulus distribution and visual presentation. All of the four cues had two levels each. In the dimension framing, both levels were positively defined (e.g. for the "Culture" cue: "Music" vs "Art"). In the feature framing, only one level was positively defined (present) while the other level was absent (e.g. for the "Culture" cue: "Music" vs *nothing*). Figure 15 shows an exemplary trial for both framings. The cues are positioned in the form of a diamond. The experiments are therefore referred to as "diamond experiments". Each cue was always represented by a colored box, even when the absent level of a cue in the feature framing was shown. We chose to change some of the original cue labels (Fiedler, 2010) for two reasons: We did not want the cue levels to allow the inference about the other cue levels (for gender: Male implies female and vice versa) and we did not want the cues to create an ingroup vs. outgroup distinction ("Psychology" vs. "Medicine" or "Heidelberg" vs. "Mannheim").

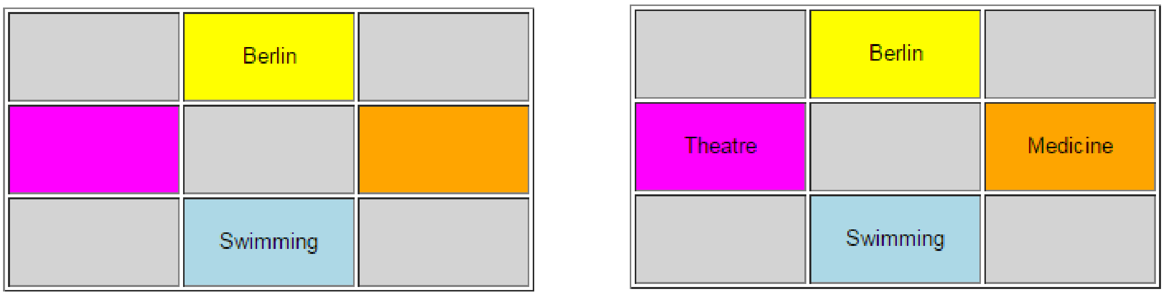

Figure 15. Exemplary trial for feature (left) and dimension (right) framing. All four cues are always represented by a colored box (even when they show the absent level in the feature condition).

The color supported presentation format allowed to later ask questions about the absent level by showing the empty colored box at a certain screen position and facilitated a clear separation of the cues. Each trial lasted at least 3750ms after which a “next”-button appeared which allowed participants to continue to the next trial. After an inter trial interval of 750ms, the next stimulus was shown.

All of the cues had skewed base rates of 3:1, meaning that one cue level was frequent and appeared in 75% of the trials, while the other cue level was rare and appeared in only 25% of the trials. Therefore, following a pseudocontingency based strategy will lead to positive relation judgments between all the cues (frequent levels are associated with frequent levels). Further, four of the six contingencies between the cues were held at $\Delta P = 0$, one was positive with $\Delta P = .33$ and one was negative with $\Delta P = -.33$. Note that a positive contingency is here defined as a contingency pointing into the same direction as the pseudocontingency inference indicated by the skewed base rates. Figure 16 shows a visual representation of the paradigm’s underlying structure.

There were 16 trial types in total resulting from all different combinations of cue levels. Each trial type was presented six times, resulting in 96 trials. The order of the trials was completely randomized. Position of the cues and cue levels were also completely randomized with the exception that in the feature framing, two cues were present frequent (the frequent attribute is the positively defined one, e.g. “Tennis”) and two cues were present rare (the frequent attribute is the absent one, i.e. *nothing*). This ensured that for some cue combinations, a strategy based on pseudocontingencies leads to different predictions than a strategy focusing on the present-present cell (A-cell).

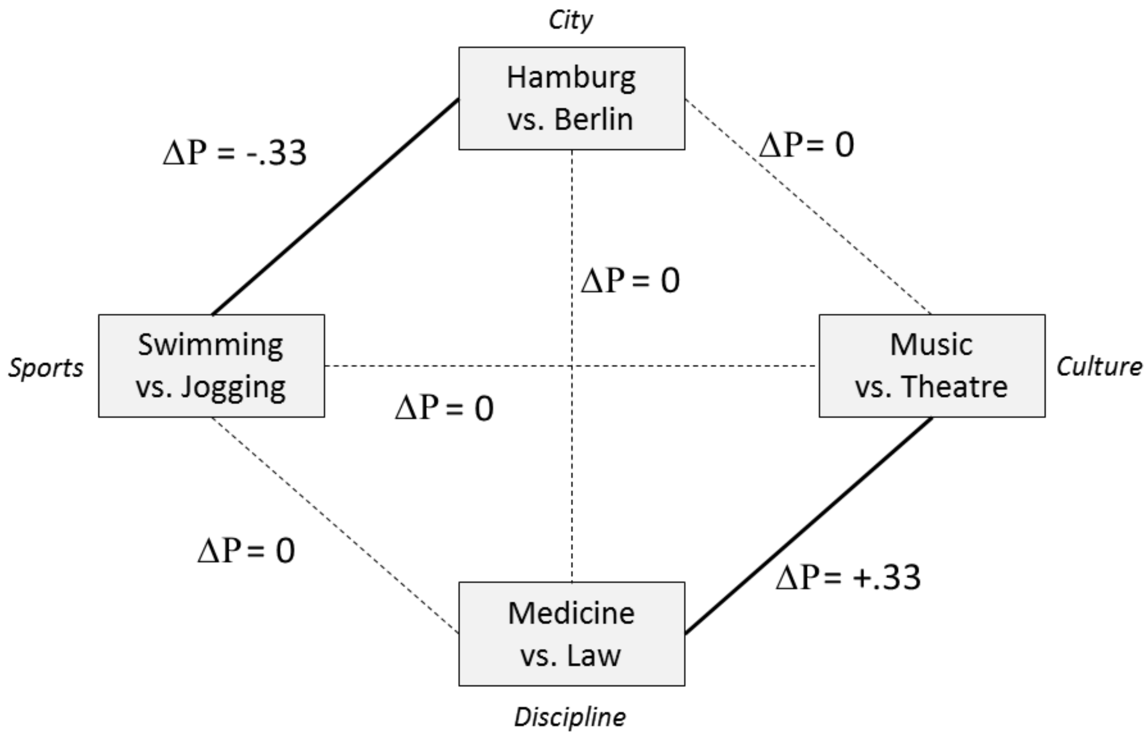

Figure 16. The structure underlying the experiment (with exemplary labels). All base rates are skewed 3:1 (one cue level appears in 75% of the cases). Four of the contingencies are held at $\Delta P = 0$, one is positive with $\Delta P = +.33$ and one is negative with $\Delta P = -.33$. Position of the four cues (City, Culture, Discipline, and Sports) is randomized. The frequent level for each cue is also selected randomly. The picture shows a possible constellation for the dimension condition (all cues have two present labels). In the feature condition, two cues have a present level as their frequent level and two cues have an absent level as their frequent level.

Dependent variable - Base rates. After the learning phase, participants were asked to estimate base rates for the four cues. For each cue either the rare or the frequent level was presented together with the question “In what percentage of the trials did the upper (right /

bottom / left) field look like this?”. Base rates for two cues (at random) were assessed using the frequent level while the base rate for the other two cues were assessed using the rare level. In the feature framing we also fully crossed this with present vs. absent levels.

Dependent variable - Conditional probabilities. There were six relationships to be assessed between the four cues. We assessed each of these relationships using a pair of two conditional probability judgments. Figure 17 shows an example of such a pair of questions. Based on these conditional probabilities, a ΔP (which is per definition the difference in conditional probabilities) was calculated for each pair of cues. Note that for each cue combination, there are four possible ways to assess this relationship. The direction can vary and the target cue level can vary. For each of the six relationships, direction and target attribute were randomized.

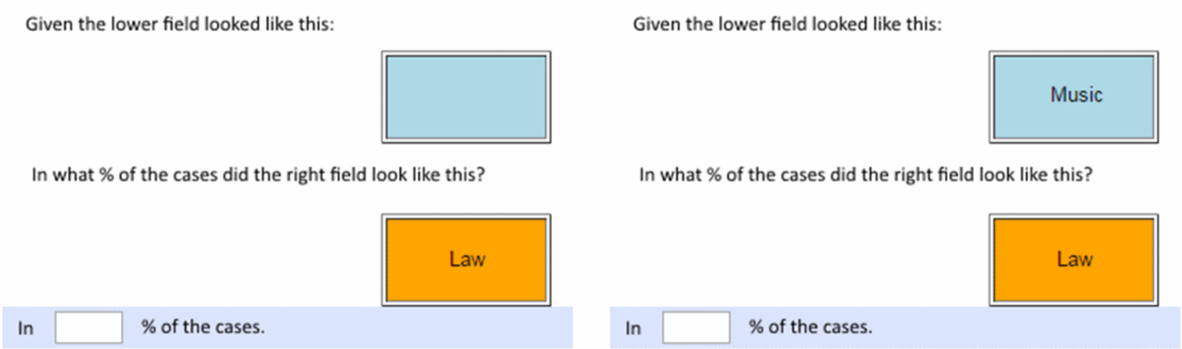

Figure 17. An example of a pair of conditional probability questions (in the feature condition). Based on these conditional probability pairs, a subjective ΔP can be calculated for each cue pair.

10.2 Results.

All data preparation and analysis was conducted in R (R Development Core Team, 2008). Within regression analyses were conducted using the lmer command in the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). Unless stated otherwise, all regression analyses use random uncorrelated slopes and a random intercept for each participant as well as z-standardized predictors.[4]

Data preparation. Goal of the data preparation was to have the following information for every participant: Information about the experimental condition, four base rate estimates, six contingency judgments, a subjective A-cell size for each contingency judgment (only in the feature Condition), and a subjective PC-index for each contingency judgment.

Base rates. We counterbalanced whether we asked for the frequent or the rare level of a cue. When asking for a rare level, we later transformed that value by applying NewEstimate = 100 - Estimate. All statistics and graphs are based on the (transformed) base rates of the frequent levels.

Contingency judgments. We assessed two conditional probabilities for each cue pair. For each pair, we computed a ΔP by taking the difference between the two conditional probability estimates. It was set up in a way that a positive ΔP always indicated a correlation in the same direction as indicated by a pseudocontingency inference.

Subjective A-cell. Knowing what levels were present and having base rate as well as conditional probability estimates, we were able to compute a subjective A-cell size for each cue pair (estimated frequency of present-present observations). Naturally, this was only done for the feature condition.

Subjective PC. Using the same PC-index formula as presented before, we were able to calculate a PC-index for each cue pair based on the base rate estimates of the cues:

$$\mathrm{PC} = \log_{10}\left(\frac{A+B}{C+D}\right) * \log_{10}\left(\frac{A+C}{B+D}\right)$$

[4] Note that degrees of freedom for predictors vary due to differences in number of observations and as a function of the model fitting process (Satterthwaite approximations).

Data analysis. Results of the manipulation checks will be presented first, before several analyses on the contingency judgments are described.

Base Rates. A regression analysis showed the expected effect of framing on estimated base rates, $b = 3.17$, $t(174) = 3.59$, $p < .001$. On average, participants in the dimension condition were more accurate and more extreme in their base rate judgments than participants in the feature condition. Still, in both conditions, judgments were regressive. The left part of figure 18 shows the result.

Subjective PC - Manipulation check. As expected, the difference in base rate perception also translated into a difference in PC-indices. The PC-indices (which are based on base rate estimates) were larger for the dimension than for the feature condition, $b = 0.04$, $t(174) = 2.81$, $p < .01$. See the middle part of Figure 18 for the difference in PC-indices.

A-cell - Manipulation check. In the feature condition, subjective A-cell size varied as a function of objective A-cell size (the number of cues having present as their frequent level in a cue pair), $b = 0.10$, $t(462) = 11.33$, $p < .001$.[5] The higher the actual number of present-present occurrences, the higher participants rated the frequency of present-present occurrences. Note that given our definition of the A-cell, this only includes feature-condition participants. The right part of Figure 18 shows this effect.

[5] The model did not converge assuming random slopes. Therefore, a fixed slope was estimated.

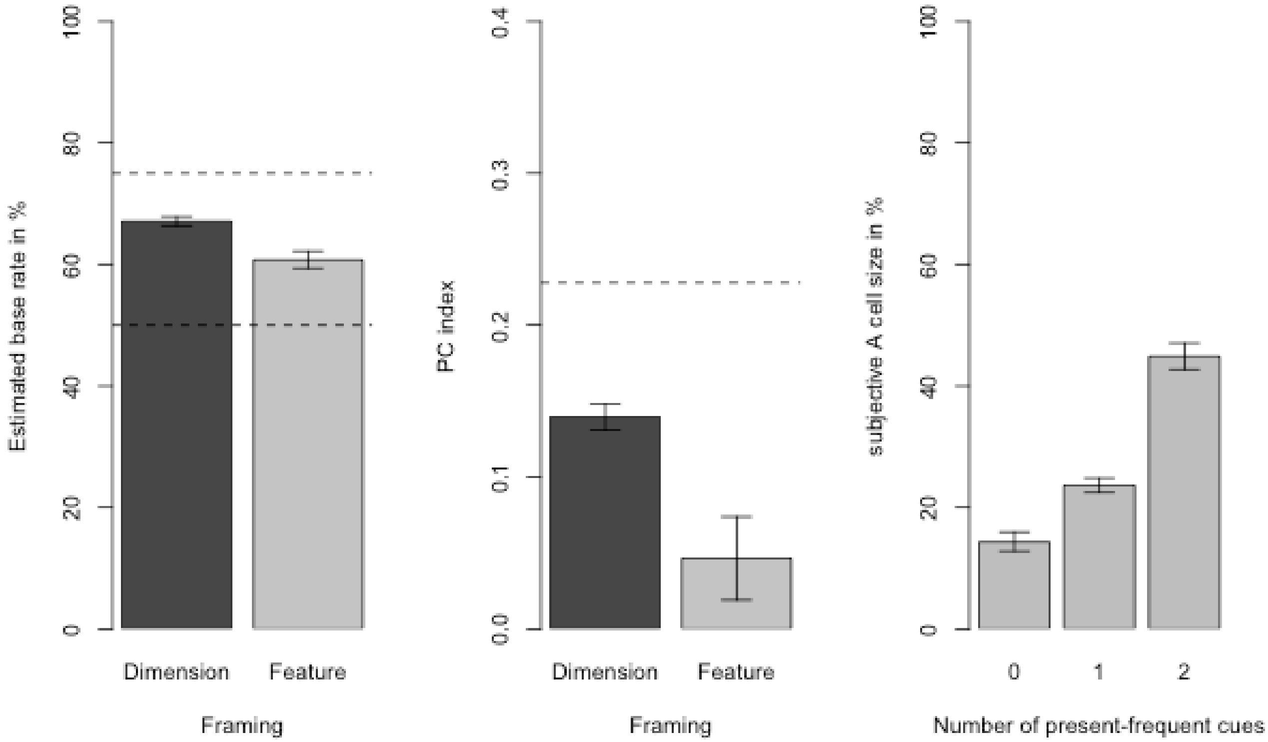

Figure 18. The left part shows average base rate estimates as a function of framing with the dotted lines showing the correct (75%) and the ignorant response (50%). The middle part shows the average PC-indices as a function of framing with the dotted line showing the theoretically accurate PC-index. The right part shows subjective average A-cell size as a function of objective A-cell size (number of cues being present-frequent for a cue pair) for the feature condition. Error bars represent standard errors (± 1 SE).

Contingency judgments analyses – Overview. There are several relevant analyses to be conducted for the contingency judgments. We can differentiate between the analyses of the effects of the actual experimental manipulations and the analyses of the predictive value of the mediators (subjective A-cell size and subjective PC-index). In other words, we can check if objective A-cell size, as well as subjective A-cell size (and PC-index), predict contingency judgments. Since all base rates were equally skewed, there is no objective manipulation of the

PC-index. Still, we can use the subjective PC-index to predict contingency judgments in both framings. Additionally, the available predictors vary between the two framings. For the feature condition, we have the actual contingency (-.33. vs. 0 vs. +.33), the PC-index, as well as the A-cell size to predict contingency judgments. For the dimension condition, the A-cell is not defined and hence only the actual contingency and the PC-index are used to predict judgments. For this reason, we always present a global analysis first and then present separate analyses for the two framing conditions. We will first present the analyses of the experimental factors and then the analyses of the subjective PC-indices and subjective A-cell sizes.

Experimental factor analysis. A within regression analysis with ΔP as criterion and framing and actual contingency as predictors showed a main effect of framing, $b = 4.82$, $t(174.30) = 4.31$, $p < .001$. Participants in the dimension framing judged contingencies to be more positive than participants in the feature framing. There also was a significant effect of the actual contingency, $b = 6.41$, $t(875.30) = 7.31$, $p < .001$, indicating general sensitivity to variations in the correlation: the higher the actual contingency, the higher the contingency judgments. Notably, the intercept positively differed from zero, $b_0 = 7.22$, $t(174.30) = 6.45$, $p < .001$, indicating a bias towards positive contingency judgments. This is in line with the idea of pseudocontingencies biasing judgments towards the direction suggested by joint skewness. There was no interaction of contingency and framing, $b = .62$, $t(857.40) = .71$, $p = .48$.

A separate analysis for the dimension condition also showed the main effect of the actual contingency on the contingency judgments, $b = 5.81$, $t(428.00) = 5.07$, $p < .001$. The intercept did again differ from zero, $b = 11.88$, $t(90) = 6.81$, $p < .001$, suggesting the presence of pseudocontingency inferences. The A-cell is not defined for this condition and was therefore not used as a predictor. The left part of figure 19 shows this pattern.

A separate analysis for the feature condition also showed the main effect of actual contingency, $b = 6.16$, $t(126.09) = 4.58$, $p < .001$. Further, the objective number of joint present occurrences (A-cell) in a cue pair predicted the contingency judgments, $b = 7.64$, $t(71.01) = 4.66$, $p < .001$. As expected, the more cues in a cue pair showed presence as their frequent level, the higher were the contingency judgments. In contrast to the dimension condition, the intercept was smaller and only differed close to marginally significant from zero, $b = 2.29$, $t(95.94) = 1.62$, $p = .108$. This indicates no (or weaker) pseudocontingency inferences.[6] The right half of figure 19 shows this pattern.

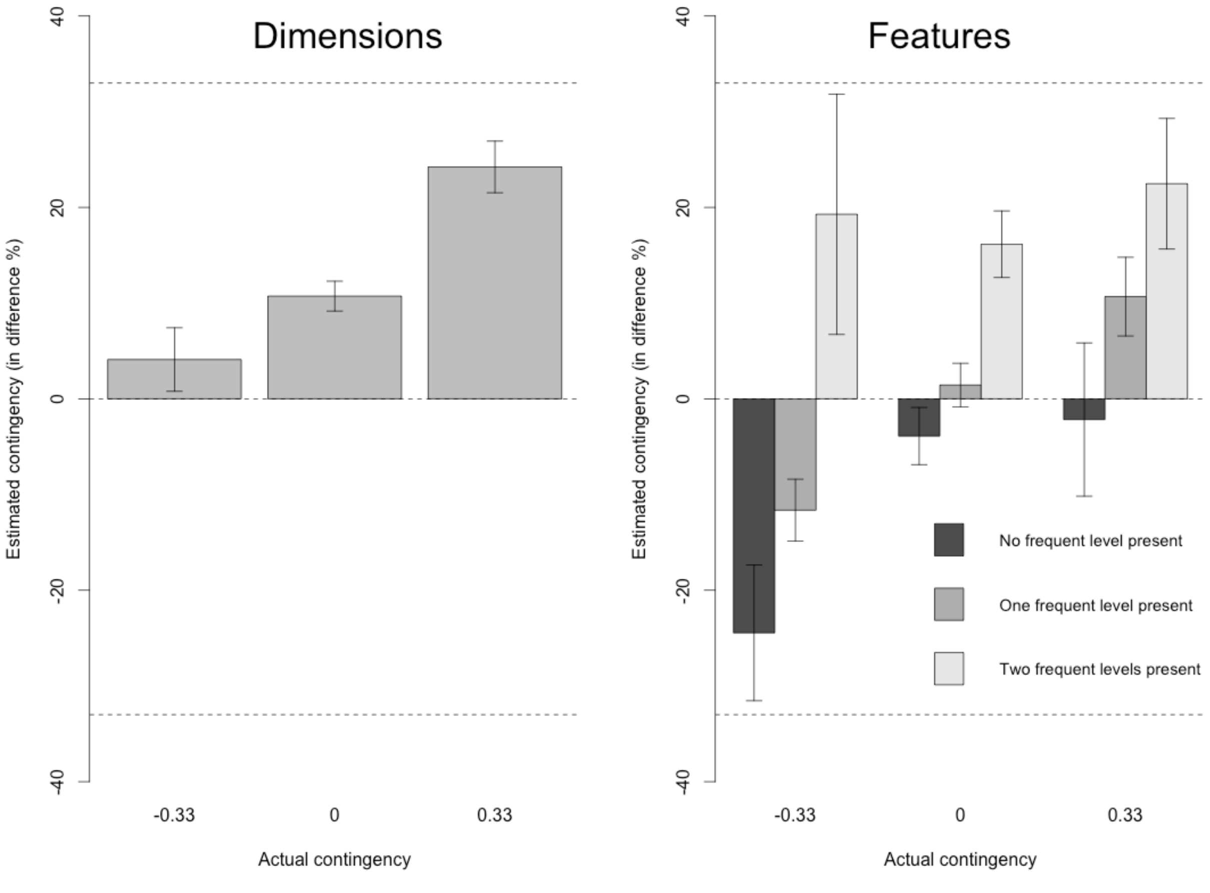

Figure 19. Average contingency judgments (in differences of conditional probabilities) as a function of framing, actual contingency and objective A-cell size as a proxy for subjective A-cell size (feature framing only). Error bars represent standard errors (± 1 SE).

[6] Ignoring the A-cell as a predictor in the feature condition does not change this result. There again was a main effect of contingency, $b = 7.01$, $t(102.85) = 5.09$, $p < .001$, and a close to marginal significant intercept, $b = 2.24$, $t(112.42) = 1.63$, $p = .106$. Hence, differences in results cannot be attributed to differences in the number of predictors between the feature and the dimension condition.

Predictive value of subjective indices. The following analyses try to predict contingency judgments as a function of the subjective perception of joint skewness (PC-index) and subjective number of joint present occurrences (A-cell, feature condition only). Actual contingency will again be used as a predictor to reduce noise and to provide a more conclusive analysis. However, the pattern of results for the subjective variables is completely identical when the actual contingency is not used as a predictor.

A within regression analysis with subjective ΔP as criterion and framing, actual contingency, and PC-indices as predictors on contingency judgments showed a main effect of framing, $b = 3.97$, $t(174.90) = 3.64$, $p < .001$. Participants in the dimension condition judged contingencies to be more positive than participants in the feature condition. Further, there was a main effect of actual contingency in the expected direction, $b = 6.35$, $t(530.20) = 7.20$, $p < .001$. The higher the actual contingency, the higher the contingency judgment. There also was a main effect of PC-index, $b = 9.09$, $t(129.80) = 5.44$, $p < .001$. The higher the PC-index, the higher the contingency judgment. Importantly, there also was an interaction of PC-index and framing, $b = 8.27$, $t(111.50) = 5.08$, $p < .001$. The separate analyses for the feature and the dimension framing give insight into the nature of this interaction.

In the dimension framing, a main effect of contingency, $b = 5.87$, $t(452.90) = 5.22$, $p < .001$, and a main effect of the PC-index, $b = 7.43$, $t(374.50) = 5.43$, $p < .001$, were evident.[7] Participants in this condition were sensitive to the actual contingency but at the same time influenced by an index based on their base-rate estimates. The more the cues of a cue pair were perceived to be skewed in the same direction, the higher participants estimated the contingency to be.

[7] The model did not converge assuming correlated random slopes. Therefore, fixed uncorrelated effects were estimated.

In the feature framing, the main effect of contingency was also prevalent, $b = 6.53$, $t(480) = 4.51$, $p < .001$[8]. Importantly and in contrast to the dimension framing, the PC-index did not predict contingency judgments, $b = 0.08$, $t(480) = 0.05$, $p = .96$. Subjective A-cell size on the other hand did strongly influence contingency judgments, $b = 4.12$, $t(480) = 2.98$, $p < .01$. The higher the perceived frequency of present-present occurrences, the higher the perceived contingency.[9]

Note that the figures only show the aggregation of data across the experimental factor levels, but do not represent the relationship between subjective indices and contingency judgments.

10.3 Discussion.

This experiment shows that when people are learning frequencies and relationships of cues in a complex multi cue environment, the way information is presented is of crucial importance. In an adaptation of an experience based learning paradigm by Fiedler (2010), we show that it does make a difference whether cues in a cue pair have two present levels or whether they are presented in a present vs. absent fashion.

When cues were presented as present vs. present (dimensions as we refer to them in the sense of Garner, 1978) other levels of the same cue seemed to be highlighted. Other levels of the same cue served as the reference set. Participants were accurate in assessing base rates and their contingency judgments were predicted by their own perception of stimulus base rates (in the form of the PC-index). The more participants perceived two cues to be skewed, the stronger was the inferred correlation. While participants picked up on differences in

---

[8] The model did not converge assuming correlated random slopes. Therefore, fixed uncorrelated effects were estimated.

[9] Ignoring the A-cell as a predictor in the feature condition does not change this result. There again was a main effect of contingency, $b = 7.06$, $t(90.72) = 5.03$, $p < .001$, but no effect of the PC-index, $b = .12$, $t(44.69) = .085$, $p = .933$. Hence, differences in results cannot be attributed to differences in the number of predictors between the feature and the dimension condition.

correlation, they were overall biased towards positive contingencies, which are suggested by the pseudocontingency.

When confronted with present-absent cues (features in the sense of Garner), presence of the cue itself and presence of other cues seemed to be highlighted. Other present attributes served as the reference set. When participants had to deal with absence, they were less accurate in their base rate estimations. More importantly, other than in the dimension condition, participants seem to not have used base rates for their contingency judgments - At least not in the same way (no predictive value of the PC-index & small intercept). Objective and subjective frequency of joint presence (A-cell) on the other hand, did predict their contingency judgments. The more often participants perceived the present levels of two cues to occur together and the more often they actually occurred together, the higher their contingency judgments. Put simply, other levels of the same attribute are the relevant reference set for dimensions while present levels of other attributes are the relevant reference set for features.

In both the feature and the dimension condition, participants were sensitive to the actual correlation between the cues. The more positive the actual correlation, the more positive was their contingency judgment. While they were not always correct on an absolute level, being able to detect relative difference is still quite impressive and encouraging, given the complexity of the task. In the original work by Fiedler (2010), participants were not able to detect the actual correlation masked by a PC. This difference might possibly be explained by some details of the experimental procedure. In our work, we used colored boxes to visually separate the cues and to reduce complexity and difficulty of the dependent variables. The visual presentation format might have helped participants to reduce complexity of the task.

Contingency judgments are relative tasks and therefore within subject comparisons give more insight than between subject omparisons. Still, it is interesting to note that there was an overall main effect difference between the feature and the dimension condition. All mean values for contingency judgments in the dimension condition where above zero, while they did appropriately vary around zero in the feature condition (figure 21). This is an indicator for the presence and strength of PC inferences in the dimension condition. In that sense, pseudocontingencies were able to overwrite the actual negative contingency in the dimension condition. Only in the feature condition participants were actually able to detect the negativity of the contingency, given there were only a small or moderate amount of joint present occurrences.

A possible limitation of this study lies in the ordering of the experimental procedure. After the learning phase, participants were first asked to give their base rate estimates and then asked to give their contingency judgments in the form of two conditional probabilities for each cue pair. Asking for the base rates might have highlighted certain aspects of the stimulus world, which might not have been that prominent in participants' representations otherwise. While this cannot explain differences between the different framings, it might still bias results. Simply reversing the order is also not a perfect solution. People might as well infer base rate estimates on the basis of their correlation judgments. Further, differences in contingency judgments do not necessarily mean differences in choices or behavior in general (e.g. Lichtenstein & Slovic, 1971). Learning contingencies in real life is not only about the learning itself, but about empowering people to make smart decisions based on what was learned. It is therefore important to show that differences between feature and dimension framings do not only make a difference for contingency judgments, but also for choices which should be based on estimated contingencies. The next experiment therefore constitutes a replication of experiment 7, replacing the base rate and ΔP assessment with a performance contingent choice task.

11. Experiment 8: Diamond II.

Analogue to the previous experiment, we predict differences in choice strategies dependent on whether attributes are presented as either features or dimensions. In the dimension condition, participants should be better at detecting the actual contingency if it is in line with a PC inference (i.e. when the contingency is positive instead of negative), while this should not matter in the feature condition. In the feature condition, participants should be better at detecting the actual contingency if an A-cell based strategy ("where are the most present-present occurrences?") falls onto the same response compared to when the actual contingency and an A-cell based strategy are in conflict and predict different things. Instead of using subjective base rates and A-cell size, we rely on this logic of redundant and conflicting trials with regards to different strategies and actual contingencies. For choices to be correct, the contingency must be detected correctly.

11.1 Method.

Participants. 95 students (61 female, Mage = 24.40, SDage = 5.70) of the University of Heidelberg participated in our experiment which took roughly 15 minutes and was part of an experimental block of about 60 minutes. Participants were financially rewarded (8€ / hour) or received course credit for their participation. The experiment was created with SoSci Survey (Leiner, 2014). Three participants were removed because due to a technical error, none of their responses were stored.

Procedure. The learning phase of the experiment was the same as in experiment 7. This means that 96 trials of four cues were presented, with all cues having 3:1 skewed base rates. Of the six correlations between the cues, one was positive with $\Delta P = .33$, one was negative with $\Delta P =.-33$ and the other four correlations were $\Delta P = 0$. Note that a positive correlation refers to a correlation pointing into the same direction as a pseudocontingency inference, while a negative correlation refers to a correlation conflicting with a pseudocontingency inference. Other than in experiment 7, we did not assess base rates and conditional probabilities. Instead, participants were asked to engage in a choice task. Importantly, there again was one condition were stimuli were presented in a present vs. absent fashion (feature condition) and a condition were stimuli were presented in a present vs. present fashion (dimension condition).

Choice task. After the learning phase, participants were instructed about a choice task. In this choice task, participants were presented with one target cue level at a time (e.g. “Berlin”). Their goal was to find a person having the specified attribute. To do this, they were able to define one other cue level. They were instructed to choose this other cue level in such a way that the profile has the highest probability of also showing the target cue level. The choice task had two steps: In a first step, participants selected which of the three cues they want to use, and in a second step they decided what level of the chosen cue they would like to use. Figure 22 shows an illustration of this procedure. Note that in our setup, there was always a normatively correct solution for each attribute. Looking back at Figure 19, you can see that for each cue, exactly one of the three correlations with the other cues is not zero. It is therefore correct to choose the cue with the non-zero correlation and then choose the appropriate cue level. If the correlation is positive, one should choose the cue level that has the same frequency as the target cue level (frequent & frequent or rare & rare). If the correlation is negative, one should choose the cue level with the opposite frequency (frequent & rare or rare & frequent). In other words, sometimes there was a match and sometimes there was a conflict between the correct solution and the solution suggested by a PC inference. The same logic holds for the A-cell strategy (in the feature condition). Sometimes the solution suggested by an A-cell based strategy matched the correct solution and sometimes it did not. At the end of the experiment, participants were rewarded 0.25€ for each correct decision.

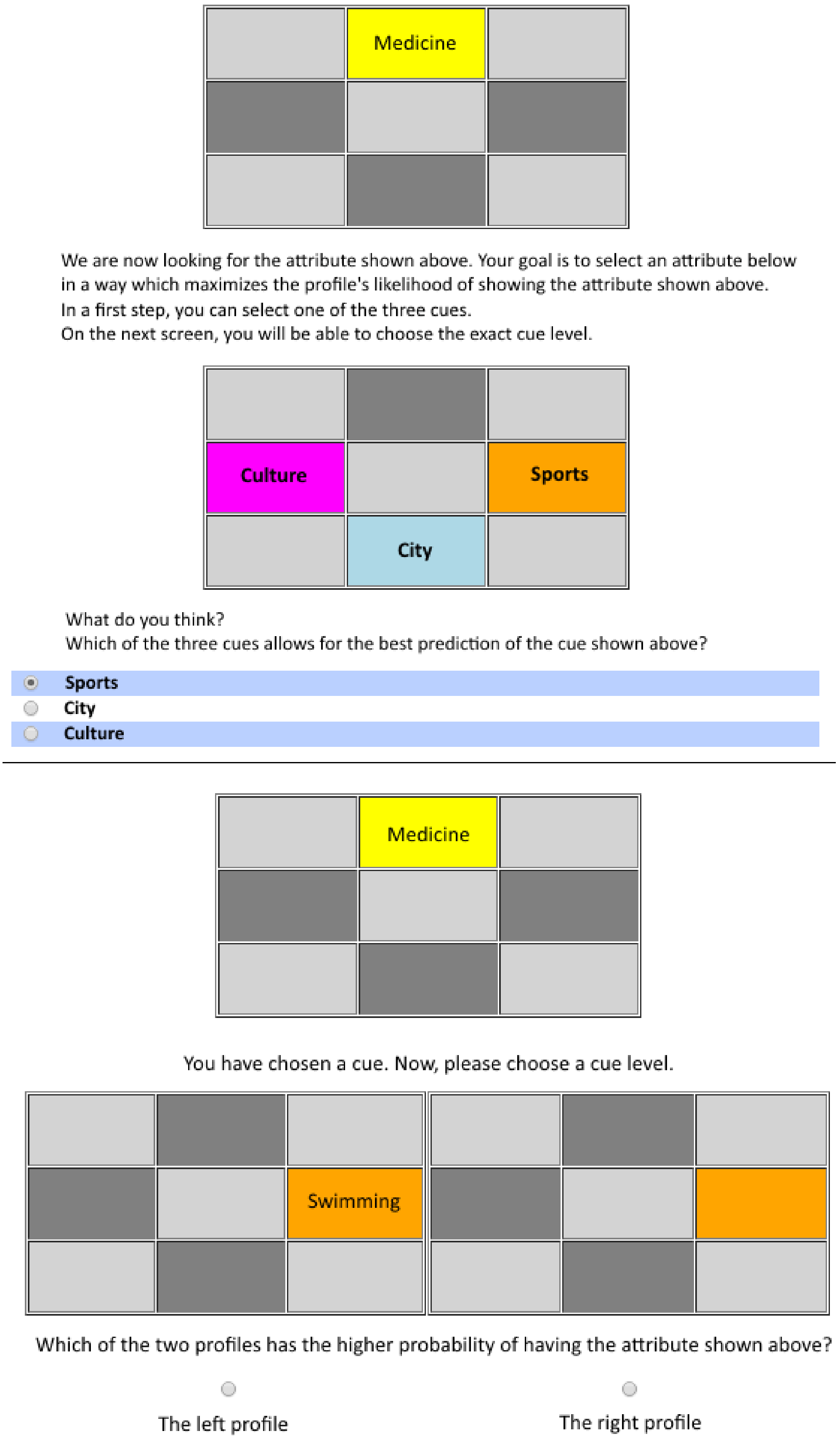

Figure 20. Example of a choice task item in the feature condition. In a first step (top), participants select a cue. In a second step (bottom), they select the cue level of the chosen cue.

11.2 Results.

We analyzed the percentage of correct choices (cue and level) made by the participants. For this, it is important to understand that not all trials were equal. While there always was a correct solution, trials differed in what responses were suggested by the strategies we were interested in: Pseudocontingency inferences and an A-cell based strategy.

PC-Inferences. All of the cues had skewed base-rates of 3:1, suggesting a PC-inference between all of them. This inference suggests associating the frequent with the other frequent and the rare with the other rare level for each cue pair. Given the base-rates are all the same, the PC does not predict the choice of a specific cue, but if a cue is chosen, it predicts what level should be chosen. This inference can now either be in line with the choice suggested by the contingency or not. In case of the positive contingency, the contingency as well as the PC-inference suggest associating frequent with frequent and rare with rare. In case of the negative contingency, however, the contingency and the PC are in conflict – given the correct cue is chosen in the first place. While the contingency suggests associating frequent with rare and rare with frequent, the PC still suggests the association of frequent with frequent and rare with rare. This allows us to compare how often participants choose the correct cue and level if contingency and PC are redundant to how often they choose the correct cue and level when contingency and PC are in conflict. This could be done for the dimension as well as the feature condition. Figure 21 shows the percentage of correct responses split by cases where the actual contingency and the PC are redundant and cases where they are in conflict.

A-cell. Similar to experiment 7, the skewed base rates of 3:1 lead to qualitatively different cues in the feature condition. The present level of a cue could either be the frequent or the rare level. Again, we made sure that two cues were present-frequent while the other two cues were present-rare. Therefore, the size of the A-cell for different cue pairs varied. The A-cell strategy was modeled in a way that it selects the cue with the most present occurrences, leading to the largest A-cell of the possible cue pairs. Then, if the target attribute is present, the present level is selected. If the target attribute is absent, the absent level is selected. Similar to the logic above, we can now compare cases in which A-cell and contingency are redundant in their predictions to cases where they are in conflict. Note that in half of the cases, there were two other cues with an equally high number of present occurrences. In that case, both alternatives are predicted by the A-cell strategy. Again, this was only done for the feature condition, because the A-cell was not defined for the dimension condition. In figure 21, the proportion of correct responses is split by cases where the A-cell prediction and the contingency prediction are redundant additionally to the differentiation by PC-redundancy.

Percentage of correct choices. Available predictors for the choices being correct or not differed between the framing conditions. For both conditions, we were able to differentiate between trials where the PC-inference was the same as the choice suggested by the actual contingency and trials where they were in conflict. Note that this is equivalent to trials where a positive (redundant), compared to a negative (conflicting) contingency needed to be inferred to make the correct choice. For the feature condition, we were able to additionally differentiate between trials where an A-cell based strategy and the choice suggested by the contingency were the same and trials where they were in conflict.

For dimensions, the data structure is pretty straight forward. We can compare how often participants chose the correct cue and level, when the PC and the contingency are redundant compared to when they are in conflict (which is analogue to comparing how often they are correct when the correlation is positive compared to negative). As expected, when PC and contingency were redundant, participants were way more likely to choose to correct cue and level than when they were in conflict, $b = .32$, $t(314) = 7.99$, $p < .001$. The right part of Figure 21 illustrates this effect.

For features, the structure is more complicated. We can also compare cases were PC and contingency are either redundant or in conflict, but we can additionally also compare cases where an A-cell based strategy and the contingency are redundant and where they are in conflict. Other than in the dimension condition, it did not matter whether the PC and the contingency were redundant or not, $b = -.02$, $t(372) = 1.36$, $p = .174$. Importantly and as expected, participants were more likely to choose the correct cue and level when contingency and A-cell pointed towards the same solution than when they differed, $b = .12$, $t(372) = 5.91$, $p < .001$. The interaction was close to being marginally significant, $b = .03$, $t(372) = 1.609$, $p = .108$. The left side of figure 21 gives insight into the nature of this interaction trend. As a reference point, random guessing should lead to 16.67% of choices being correct (6 cue levels to choose from, only 1 correct: 1/6).

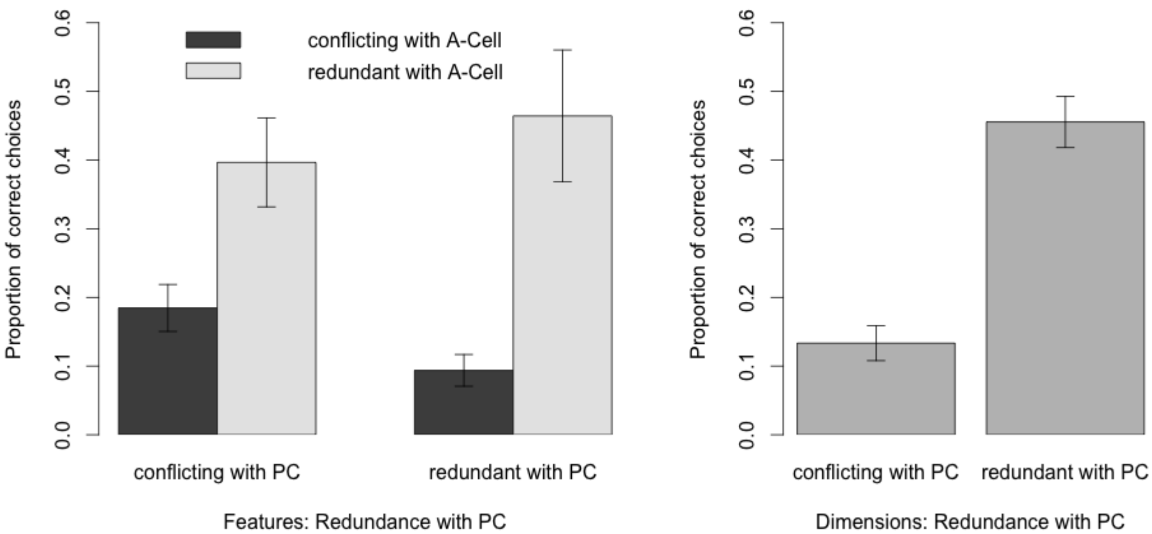

Figure 21. Mean number of correct choices for the different framing conditions and the different trial types. For features (left half of the graph), trials are split for the contingency being redundant or in conflict with the PC and/or the A-cell. For dimensions (right half of the graph), they are only split for the PC being redundant or in conflict with the contingency. Error bars represent standard errors (± 1 SE).

11.3 Discussion.

We confronted participants with the task of selecting and controlling the highest correlated cue in a multi cue environment to be able to maximize the likelihood of producing a target cue level. As expected, participants differed in how well they performed, dependent on what kind of trial type they were presented with. In the dimension condition, participants were way better at finding the correct cue and level when the cue level they had to choose was in line with the cue level suggested by a PC strategy. When the actual contingency and a PC-inference predicted conflicting things (suggested different associations between the cue levels), participants performed way worse. In other words, when the PC-index and the correlation had the same sign (e.g. were both positive), participants were better than when the signs differed. For features, the pattern was different. It did not matter whether the contingency was in line with a PC-inference or not. It did however matter whether the A-cell strategy was in line with the actual contingency or not. Participants chose the correct cue and level more often when the number of joint present occurrences between two cues was high.

In this experiment, participants were not asked to give any base rate, cell size or contingency judgments but were simply asked to make a choice (which should be based on a contingency judgment). That way, the pattern of results cannot be attributed to an artificial highlighting of certain attributes of the stimulus world, such as base rates.

While A-cell size was varied to some degree in Experiment 7 and 8, all base rates were always skewed, resulting in the same PC-index between all cues in both experiments. While it was varied whether this PC-index pointed towards the same relationship as the actual correlation or not, the absolute value of the PC-index always stayed the same. We relied on variance in participants' perception of the stimulus distributions to predict their contingency judgments. The next and final experiment includes a more elaborate manipulation of base rates resulting in multiple base rate combinations (PC-indices) and A-cell sizes. To be able to control how good people are at detecting these differences and to examine how they use these values for their contingency judgments, the last experiment uses the same dependent variables as experiment 7 (Base rates and conditional probabilities, resulting in subjective ΔP, subjective PC-indices and subjective A-cell size).

12. Experiment 9: Diamond III.

Experiment 7 and 8 provide support for the idea that a feature framing promotes contingency inferences based on the number of joint present occurrences between two cues, while a dimension framing promotes pseudocontingency inferences. In those two experiments, three levels of contingencies, three levels of A-cell size but only one level of joint base rate skewness (PC-index) appeared. We relied on variation in participants' base rate perception (and the contrast to zero) to investigate PC effects. In this last experiment, base rates are manipulated within participants, creating more levels of PC-indices and at the same time, more different A-cell sizes.

12.1 Method.

Participants. 117 students (81 female, Mage = 24.22, SDage = 6.67) of the University of Heidelberg participated in our experiment which took roughly 15 minutes and was part of an experimental block of about 60 minutes. Participants were financially rewarded (8€ / hour) or received course credit for their participation. The experiment was created with SoSci Survey (Leiner, 2014).

Structure - Setup. The core setup of this experiment was the same as in experiment 7 and 8. Participants again had to learn the relationships between four cues with two levels each. The surface characteristics of the learning phase as well as the dependent variables were the same as in experiment 7. Framing (feature vs. dimension) was again manipulated between participants. The stimulus distributions however, were different and manipulated within participants. This time, all correlations between all cues were held at $\Delta P = 0$. Further, there now always were two cues with a base rate of $BR = .5$ (equal number of trials showing cue level 1 as cue level 2) and two cues with a skewed base rate of $BR = .75$ (3 times as many trials for cue level 1 compared to cue level 2). This created a within participants variation of base rates, leading to a within manipulation of the PC-index in both conditions, as well as a within manipulation of the A-cell in the feature condition. In the dimension condition, the implementation of that idea is pretty straight forward. Given the cue levels are not qualitatively different, it does not matter which cue level is frequent and which one is rare (in the BR = .75 cues). For the feature condition, there are different ways to implement the concept. For the two cues which have a base rate of BR = .5, it does not matter which level is present and which one is absent. Both occur equally often. For the two cues where the base rate is skewed with BR = .75, there are two possibilities each. The present level can either be the frequent one (present in 75% of the time →absent in 25% of the time) or it can be the rare one (present in 25% of the time → absent in 75% of the time). Therefore, there are 4 possible combinations. Two cues can both be present frequent, both can be present rare or one can be present frequent while the other one is present rare. Note that for the last possibility, both cues can switch roles, but given that they are equivalent, there is no need to differentiate between the two. In summary, this results in 3 interesting feature conditions with regards to the skewed cues: Both present frequent vs. both present rare vs. one present frequent + one present rare. We implemented all of them to get as much variation into A-cell size as possible. The next paragraphs provide more formal details on the three feature conditions and on the one dimension condition.

Cue pair types - base rate combinations & PCs. Given there are 2 levels of base rates, there are 3 unique combinations of them between two cues.

1. $BR_1 = BR_2 = 0.75$: Both base rates are skewed. A PC strategy leads to a positive contingency judgment ($\text{PC-index}_1 = .227$).

2. $BR_1 = .75$, $BR_2 = .5$; Only one base rate is skewed. A PC strategy leads to a zero contingency judgment ($\text{PC-index}_2 = .0$).

3. $BR_1 = .5$, $BR_2 = .5$; No base rate is skewed. A PC strategy leads to a zero contingency judgment ($\text{PC-index}_3 = .0$).

Cue pair types - A cell size. The number of present-present observations for each cue pair directly varies as a function of stimulus base rates, as well as labeling of the cue levels. Across the different feature conditions presented above, there are 6 different A-cell sizes between cue pairs. Due to the zero correlation between all cues, we can simply multiply the probabilities of the attributes being present in each cue to get to the objective A-cell size (in percent).

In case of the symmetrical base rate combination, A-cell size is always the same. Independent of how the cue levels (present vs. absent) are distributed.

1. $BR_1 = BR_2 = .5$; $A_{prop} = .5 * .5 = .25$;

In all other cases, it depends on whether in a cue with a base rate of BR = .75, the present level is the frequent or the rare one. The following labels will be used: “Present frequent” meaning a feature cue where the present level is the frequent one (present appears in 75% of the cases) and “present rare” meaning a feature cue where the present level is the rare one (present appears in 25% of the cases).

2. $BR_1 = BR_2 = .75$; Both present frequent; $A_{prop} = .75*.75 = .5625$
3. $BR_1 = BR_2 = .75$; One present frequent, one present rare; $A_{prop} = .75 * .25 = .1875$
4. $BR_1 = BR_2 = .75$; Both present rare; $A_{prop} = .25*.25 = .0625$
5. $BR_1 = .75$; $BR_2 = .5$; Skewed cue present frequent; $A_{prop} = .75 * .5 = .375$
6. $BR_1 = .75$; $BR_2 = .5$; Skewed cue present rare; $A_{prop} = .25 * .5 = .125$

In case participants follow an A-cell based strategy, their contingency judgments should be higher the higher the proportion of present-present occurrences for a cue pair. Also note that this setup again allows a test of the PC strategy vs. the A-cell strategy in the feature condition. For example, look at the case where both base rates are $BR_1 = BR_2 = .75$. Independent of the cue level distribution (present / absent), the PC always makes the same prediction. The predictions of the A-cell strategy however, depend heavily on the chosen cue levels. In this case, the A-cell can have 3 different sizes: A_{prop} = .5626 vs. A_{prop} = .1875 vs. A_{prop} = .0625. The actual correlation in this case is always the same with $\Delta P = 0$.

Procedure & Data preparation. Instructions and procedural details were exactly the same as in experiment 7. After a learning phase of 96 trials, base rates and conditional probability pairs (ΔP) were assessed. Data was also prepared similar to experiment 7. Goal of the data preparation was again to have the following information for every participant: Information about the experimental condition (framing), four base rate estimates, six contingency judgments (ΔP), a subjective A-cell size for each contingency judgment (only in the feature condition), and a subjective PC-index for each contingency judgment. Further, the data set also contained information about what cue pair type a contingency judgment belonged to (objective base rates and cue labels).

12.2 Results.

All data preparation and analysis was conducted in R (R Development Core Team, 2008). Within regression analyses were conducted using the lmer command in the lme4 and lmertest package (Bates et al., 2015). Unless stated otherwise, all regression analyses use random uncorrelated slopes and a random intercept for each participant as well as z-standardized predictors.[10]

Manipulation check - Base rates. A within subject regression analysis with base rate estimates as criterion and objective base rates (.5 vs. .75) and framing as predictor showed a significant effect of objective base rate, $b = 9.24$, $t(458) = 10.32$, $p < .001$. Participants correctly gave higher base rate estimates when a cue was skewed compared to when it was not skewed. Further, there was an effect of framing, $b = 2.20$, $t(458) = 2.38$, $p = .02$. Participants in the dimension condition gave higher base rate estimates than those in the

[10] Note that degrees of freedom for predictors vary due to differences in number of observations and as a function of the model fitting process (Satterthwaite approximations).

feature condition. The interaction was not significant, $b = 0.94$, $t(458) = 1.06$, $p = .28$.[11] The left half of figure 22 illustrates these results.

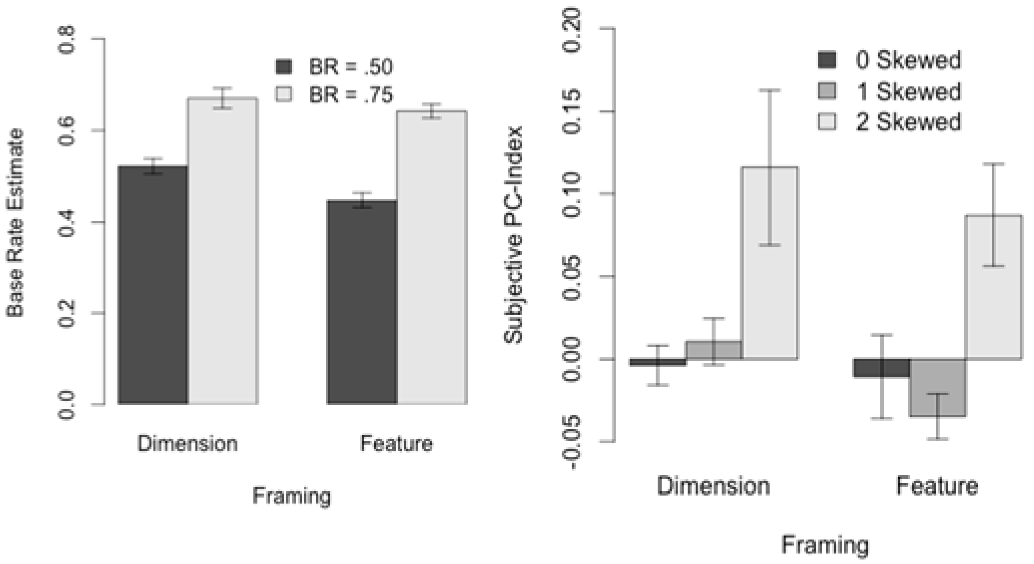

Figure 22. The left half shows mean base rate estimates as a function of framing and objective base rates. The right half shows mean PC-indices based on the base rate estimates as a function of framing and the number of skewed cues in a cue pair. Error bars represent standard errors (± 1 SE)

Manipulation Check - PC-index. A within subject regression analysis with the subjective PC-index as criterion and framing and number of skewed cues in a cue pair as predictors showed a significant effect of the number of skewed cues in a pair, $b = .03$, $t(686) = 3.18$, $p = .002$. When both cues in a pair were skewed, this resulted in a high PC-index. When zero or only one cue were skewed, the PC-index was close to 0. Further, there was a marginal effect of framing, $b = .02$, $t(686) = 1.64$, $p = .10$. PC-indices in the dimension condition were higher than those in the feature condition. There was no interaction, $b = .00$, $t(686) = .30$, $p = .76$. The right half of figure 22 illustrates the results.

[11] The model did not converge assuming a random effect for the objective base rate. Therefore, a fixed effect was estimated.

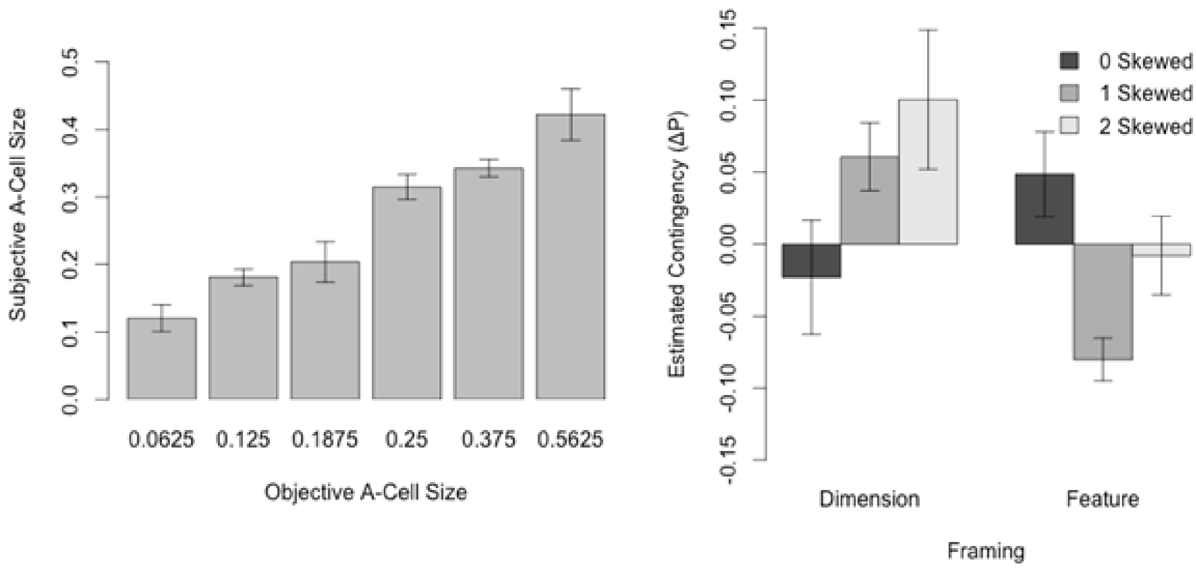

Figure 23. The left half shows mean subjective A-cell size as a function of objective A-cell size in the feature condition. The right half shows contingency judgments (ΔP) as a function of framing and number of cues skewed in a cue pair. Error bars represent standard errors (± 1 SE).

Manipulation check - Feature. In the feature condition, as a result of the base rate and cue label manipulations, six different objective A-cell sizes were possible. A within regression analysis with subjective A-cell size as criterion and objective A-cell size as predictor was conducted to test if participants were sensitive to the A-cell size manipulation. As expected, there was an effect of objective A-cell size, $b = .09$, $t(82.80) = 10.65$, $p < .001$. The bigger the A-cell objectively was, the higher were participants' judgments. The left half of figure 23 illustrates this effect.

Contingency judgments. As in experiment 7 and 8, the number of available predictors differed between the feature and the dimension condition. Both conditions allow the calculation of a PC-index which can be used to predict contingency judgments. The A-cell (the only present-present cell in a cue pair) however, is only defined for features. Similar to the analysis of experiment 7, we will first present an analysis of the objective experimental factors and then in a second step, the analyses using the subjective measures.

While there are three different levels of base rate skew combinations in a cue pair in our experiment, the objective PC-index as it is defined here, only has two levels. It is PC = 0 when none or when only one cue is skewed and PC = .227 when both cues are skewed. Therefore, for the main analysis of the experimental manipulations, we only use these two levels to predict judgments. Additionally, we also present an exploratory analysis of the same experimental factor understood as linear with three levels: None skewed, one skewed, both skewed.

Experimental factors – Analysis. A within regression analysis with framing and objective PC-index as predictors and contingency judgments as criterion revealed a main effect for framing, $b = 3.97$, $t(645.00) = 3.84$, $p < .001$. Contingencies were judged to be more positive in the dimension than in the feature condition. There was no main effect of PC-index, $b = .80$, $t(645.00) = .782$, $p = .43$ and no interaction of framing and PC-index, $b = .56$, $t(645.00) = .60$, $p = .54$.

A separate analysis for the dimension condition showed no effect for the objective PC-index, $b = 2.17$, $t(26.45) = .95$, $p = .35$. The right half of figure 23 shows the result. Note that to express this comparison, the "0 skewed" and "1 skewed" bars have to be mentally averaged.

A separate analysis for the feature condition showed no effect for the objective PC-index, $b = .59$, $t(263.10) = .52$, $p = .60$, but the expected effect of A-cell size, $b = 3.13$, $t(74.20) = 1.96$, $p = .053$. The more joint present occurrences there were objectively, the higher participants judged the contingency. There was no interaction of PC-index and A-cell size, $b = .22$, $t(475.90) = .239$, $p = .81$.[12] The right half of figure 23 shows the pattern of

[12] Ignoring the A-cell as a predictor in the feature condition does not change this result. There again was no main effect of the PC-index, $b = .48$, $t(500) = .40$, $p = 69$. Hence, differences in results cannot be attributed to differences in the number of predictors between the feature and the dimension condition.

results when ignoring A-cell size. Figure 24 shows the more differentiated pattern when also differentiating by A-cell size.

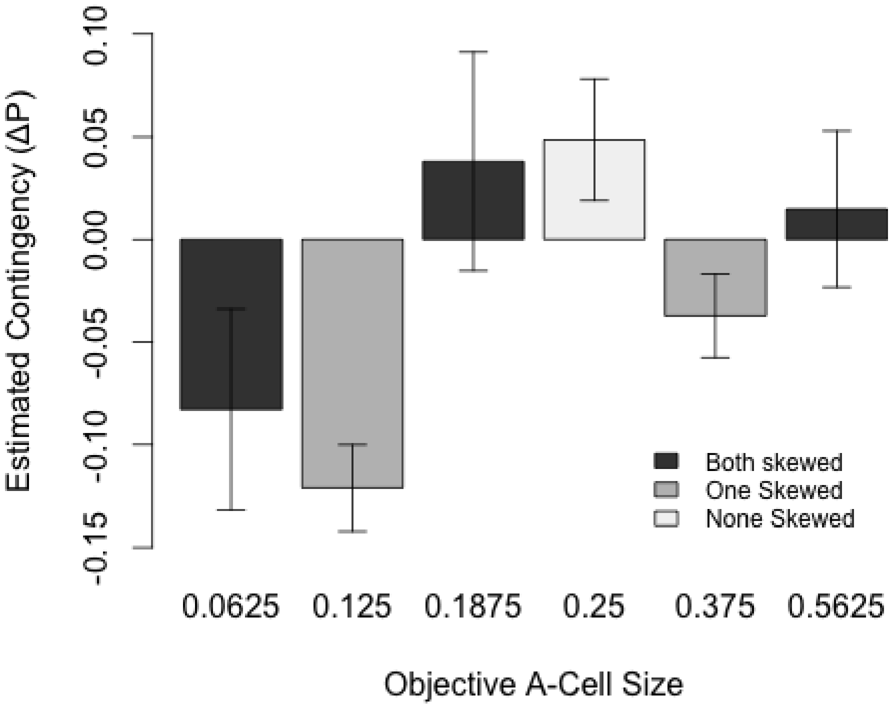

Figure 24. Mean ΔP as a function of objective A-cell size and number of cues skewed in a cue pair for the feature condition. Comparisons within equally colored bars show effects of A-cell size variation given a fixed PC-index. Error bars represent standard errors (± 1 SE).

Explorative analysis: Three levels of skewness. We redid the same analysis as presented above replacing the two levels of the PC-index by the three levels of skewness in a cue pair: none skewed vs. one skewed vs. both skewed.

The overall analysis with framing and skewness as predictors and contingency judgments as criterion revealed a main effect of framing, $b = 4.08$, $t(645) = 3.99$, $p < .001$. Judgments in the dimension condition were higher than in the feature condition. There was no main effect of skewness, $b = .46$, $t(645) = .45$, $p = .65$, but in contrast to the analysis above, there was an interaction of skewness and framing, $b = 2.16$, $t(645.00) = 2.12$, $p = .034$. The separate analyses for dimensions and features give insight into this interaction.

The separate analysis for dimensions revealed a marginally significant effect of skewness, $b = 3.60$, $t(41.83) = 1.84$, $p = .07$. The more cues in a cue pair were skewed, the higher was the contingency judgment.

The analogue analysis for features did not reveal a main effect of skewness, $b = .08$, $t(303.24) = .68$, $p = .49$, but it did again reveal an effect of A-cell size, $b = 3.48$, $t(70.14) = 2.32$, $p = .02$. The bigger the objective A-cell was, the higher were the contingency judgments. There also was an interaction of skewness and A-cell size, $b = 3.14$, $t(96.20) = 2.79$, $p < .01$.[13] Figure 24 gives insight into the complicated nature of this effect.

Predictive value of subjective indices. Similar to the analyses of experiment 7, we will now present the analysis using the subjective measures instead of the experimental factors. PC-index and A-cell size therefore refer to subjective values instead of objective values in this section.

A within regression analysis with framing and PC-index as predictors and ΔP as criterion revealed a main effect of framing, $b = 3.67$, $t(314.63) = 3.59$, $p < .001$. Participants in the dimension condition gave higher contingency estimates than those in the feature condition. Further, there was a marginal effect of the PC-index, $b = 3.10$, $t(78.90) = 1.90$, $p = .06$. The higher this index was, the higher were participants' contingency judgments. The interaction of framing and PC-index did not reach significance, $b = 2.26$, $t(95.98) = 1.16$, $p = .25$.

[13] Ignoring the A-cell as a predictor in the feature condition does not change this result. There again was no main effect of skewness, $b = .16$, t(500.00) = 1.37, p = .17. The model did not converge assuming random slopes and was therefore estimated assuming fixed effects. Differences in results cannot be attributed to differences in the number of predictors between the feature and the dimension condition.

A separate analysis for dimensions with subjective PC-index as predictor and ΔP as criterion revealed a marginally significant effect of the PC-index, $b = 5.77$, $t(6.69) = 1.87$, $p = .10$. The higher the PC-index, the higher were the contingency judgments.

A separate analysis for with subjective PC-index and subjective A-cell size as predictors and ΔP as criterion revealed a main effect of the PC-index, $b = 2.99$, $t(66.88) = 2.03$, $p = .04$. The higher the subjective PC-index, the higher the contingency judgments. Further, there was a main effect of subjective A-cell size, $b = 4.21$, $t(92.13) = 3.01$, $p < .01$.The bigger the A-cell was perceived to be, the higher were the contingency estimates. There was no interaction of PC-index and A-cell size, $b = .75$, $t(23.40) = .63$, $p = .53$.[14]

12.3 Discussion.

We presented participants with a complex multi cue environment in which we manipulated base rates and cue labels. By doing this, we manipulated the PC-index between cue pairs and the A-cell size independently (as far as this is possible) from one another while keeping the actual correlation constant at ΔP = 0 between all cue pairs.

Again, the manipulations were very successful. Participants were really good at picking up differences in base rates and A-cell sizes. This also reflected in accurate PC-indices. As expected and in line with experiment 7, participants were again better at assessing base rates in the dimension condition than in the feature condition and this also translated into more extreme PC-indices in the dimension than in the feature condition. The overall pattern of contingency judgments also provided more support for our hypotheses. A dimension framing led participants to pseudocontingency inferences, while a feature framing led participants to A-cell driven judgments.

[14] Interestingly, this time it did make a difference whether the subjective A-cell was included as a predictor or not. When the analysis was done only using the subjective PC-index, there was no significant effect, $b = 1.12$, $t(629.00) = 1.13$, $p = .26$. The model did not converge assuming a random slope and was therefore done assuming a fixed effect.

The pattern of results was not as clear cut as in experiment 7. The objective PC-index did not predict judgments in the dimension condition, while the number of skewed cues in a cue pair and the subjective PC-index did. Due to variance in base rate perception, the subjective PC-index was not always zero in a cue pair where one of the cues was skewed. The explorative analyses of the different levels of skewness and analyses of the subjective PC show that when taking this into account, there is evidence for PC inferences in the dimension condition. Similar to the single-skew condition in the “skewed base rate” experiment (experiment 6), contingency judgments were slightly positive, even when only one of the base rates was skewed. Unlike in the other experiments, we also find some evidence for PC-inferences in the feature condition. The subjective PC-index did predict judgments, but only when subjective A-cell size was also put into the equation.

This experiment illustrates that in our framework, it is possible to make and test very differentiated predictions of patterns, rather than just predicting some difference in means between two conditions. Participants did remarkably well in assessing differences in base rates and specific cue combinations, but still showed biases in contingency judgments. Judgments based on dimensions were driven by joint skewness (PC-inferences) while judgments based on features were driven by the number of joint present occurrences (A-cell).

13. Summary diamond experiments (7-9).

Across three experiments, we found evidence for differences in contingency learning, dependent on how attributes were presented. People were overall sensitive to variations in actual contingency, but showed different biases dependent on how the attributes were presented. As expected, dimensions lead to pseudocontingency inferences while features lead to A-cell driven judgments.

The different levels of a dimension are qualitatively the same. Hence, there is no reason to expect any cell inequality in terms of weighting. At the same time, comparisons within dimensions are very well defined. Other levels of the same attribute serve as the natural reference set. As a consequence, base rates are picked up well. On the downside, these base rates are then also used in the sense of pseudocontingency inferences to form contingency judgments.

The different levels of a feature on the other hand, are qualitatively different. Absence is more ambivalent to process and as a result, presence becomes more relevant in the mental representation of the stimulus world. The reference set does no longer consist of other levels of the same attribute, but of present levels of other attributes. As a consequence, base rates are not as well picked up and also not used for contingency judgments (no PC-inferences). The number of joint present occurrences (the A-cell) on the other hand, is used to infer contingency judgments.

While experiment 7 and 9 showed the expected pattern for judgments, experiment 8 showed the pattern for choices. Feature-dimension framing cannot only influence judgments, but also influence choices. Experiment 8 also rules out the possibility that the effects in experiment 7 and 9 were only due to highlighting base rates in the experiment by asking for them. Even without asking for base rates, choices were in line with a PC strategy for dimensions and in line with an A-cell based strategy for features.

14. General Discussion.

In this last part of the present volume, a summary of the theoretical framework, a summary of all the experiments and a discussion of conceptual limitations and associated future research are presented.

14.1 Relativity of presence and absence.

The present work makes a strong point for relativity of information. Attributes are not processed in isolation, but rather in comparison to reference sets. Building on a theoretical framework by Garner (1978), we investigated different types of attributes: Features and dimensions. A feature was defined as an attribute with two asymmetric and qualitatively different levels: present vs. absent ("is X present?"). A dimension was defined as an attribute with at least two different positively defined levels: present vs. present ("what level does X have?").

In the first part of the present volume, impact of presence and absence of attributes on mental representations and evaluations was investigated. The core idea was that whenever a stimulus is presented, this stimulus is compared to a stimulus space, a mental representation of the possible stimulus world. The exact architecture of this stimulus space is assumed to differ dependent on what objects are presented and importantly also dependent on how attributes within these objects are presented. A stimulus space elicited by features is assumed to mainly highlight the presence of attributes, because dealing with absence comes with several problems. Absence is ambiguous because it is hard to sequence, produces uncertainty when it comes to events which are hard to detect and because it does not allow the same inferences about the stimulus space as the presence of attributes does. Absence does not tell what is missing. As a consequence, comparisons between attributes become more relevant than comparisons within an attribute. A stimulus space elicited by dimensions on the other hand is assumed to be more elaborate and balanced. There is no inequality between the different cue levels and comparisons within attributes are well defined.

14.2 Mental representations - Summary of experiments 1 to 4.

In the "stimulus space" experiment (experiment 1), we tried to change people's representation of the stimulus world not by showing additional options, but by framing single attributes within objects differently. To some participants, we presented consumer products with labels suggesting a feature nature (present vs. absent) of the attributes, while we presented labels suggesting a dimension nature (different levels of presence) to other participants. We then asked participants to generate other possible product labels and assessed what kind of labels they produced. As expected, a feature framing of attributes triggered thinking about other attributes, while a dimension framing triggered more diverse thinking, also including other levels within the same attribute. We understood this as an indicator of differently construed stimulus spaces as a consequence of attribute framing.

In the "presenters paradox revisited" work (experiment 2-4, Krüger, Mata, & Ihmels, 2014), we showed that attribute importance for an option can vary as a function of other available alternatives. In separate evaluation, product bundles, consisting of a main as well as an add-on product, were evaluated as attractive as only the main product. In joint evaluation however, bundles were preferred. The absence of the add-on in the "main product only" option highlighted the presence of the add-on in the product bundle and the absence in the main product only option and led to higher differences in attractiveness. This goes beyond the idea of evaluability (Hsee, 1996) because it is not only about creating a reference point for a value on a scale, but also about highlighting the presence or the absence of an attribute. The results are in line with the idea that for features, comparisons within the same attribute are not readily available. Presence is not spontaneously compared to absence and absence does not tell what is missing. Unless absence was made explicit by contrasting the bundle to the single option, participants did not seem to consider the possibility of absence of the add-on in other products. Showing the explicit absence of a feature helped structuring the stimulus space and therefore highlighted the relevance of the present level of the feature.

14.3 Contingency learning.

After investigating differences in present vs. absent levels of attributes in mental representations and evaluations, we were interested in downstream consequences of this type of attribute framing in contingency learning. Contingency learning generally is about how people learn relationships between variables - in the simplest form, the relationship between two dichotomous variables. While people seem to generally be sensitive to varying normative indices of correlation, they are also prone to different biases. Building on the framework of features and dimensions, we worked towards a better understanding of biases in contingency learning. To be more precise, we tried to integrate two similar but hitherto largely unconnected phenomena: Density biases and pseudocontingencies (PCs). Both phenomena are very present in the literature, but there were no attempts trying to integrate the two. Both are driven by skewed base rates and might therefore be confused. Thinking about different ways of labeling cues, however, shows that there is a gap between the two and we need a framework to explain when and why which strategy is used, because they can make very different predictions. We show that thinking about attribute framing helps understanding contingency learning processes.

We speak of density biases in settings where cues have one present and one absent level. The typical finding is that contingency judgments go up the more present occurrences there are, irrespective of the actual correlation. The cell of the 2x2 contingency table where both cues are present is commonly referred to as the A-cell. Hence, we talk about an A-cell based strategy when contingency judgments vary as a function of joint present occurrences.

Pseudocontingencies on the other hand occur in situations where both cues have skewed base rates. People then tend to associate the frequent level of one cue to the frequent level of the other cue, and the rare level of one cue to the rare level of the other cue, irrespective of the actual correlation. We also presented and evaluated a formula for quantifying the strength of the PC in a stimulus distribution: the PC-index (Kutzner, 2009). While the two phenomena both depend on cue base rates and therefore seem quite similar, there are important differences. An A-cell based strategy is all about cue labels. It is important whether the present or the absent levels of the cues are frequent or rare. The PC strategy does not react to changes in cue labels. It is only about skewness of base rates.

14.4 Contingency Learning – Summary of simulations.

A computer simulation framework was created to facilitate a better understanding of how different indices of correlations relate to one another in 2x2 tables. It is very important to understand how certain indices and strategies of contingency statistically relate to one another when trying to disentangle them theoretically as well as empirically. The framework also provides a very easy way to create a large number of stimulus distributions all fulfilling desired constraints. The framework can therefore be used to easily implement stimulus sampling in contingency learning experiments. As recently highlighted (e.g. Judd et al., 2012), it is not enough to think about getting larger samples to increase power and validity of experiments, but one also has to think about sampling stimuli to allow for generalization. Additionally, the simulations help to understand the way we quantified pseudocontingencies in the form of the PC-index.

The simulations showed that the relationship between single cell sizes (A-cell), ΔP and the PC-index is quite complex. Simply correlating these indices with one another can be deceiving because correlations only tell us about linear relationships. It is more helpful to think about the different indices putting boundaries on one another. Similar to the idea of the method of bounds (Duncan & Davis, 1953), one cannot infer the exact value of one of the indices based on the other, but one can infer what the possible minimum and maximum are. For example, single cell size (A-cell) and PC are related the following way. Given the number of A-cell observations is really small compared to the entire sample size, the PC-index can take almost all values. When the number of A-cell observation increases, the possible variation in the PC-index decreases and it gets bigger the bigger the A-cell is.

The simulations also highlighted a special property of the PC-index as it was implemented here. Unless one base rate is $BR = .5$, the index gets bigger the more skewed the other base rate is. An alternative would have been to implement a matching advantage into the index. This index would be biggest if both base rates are skewed exactly the same amount and get smaller, once the base rates are less similar, even if one of the base rates gets more extreme. Investigating which of these process assumptions is more accurate might improve our understanding of PC inferences in the future.

14.5 Contingency learning - Summary of experiments 5 to 9.

The “All things being equal” experiment (experiment 5) served as a first exploration of the effects of feature-dimension framing in contingency learning. In a very simple contingency learning setting, we manipulated whether attributes were presented as being present vs. absent or present vs. present. Even in the very simple setting where all four cells (A,B,C,D) appeared equally often, we found differences between the two framings. In the dimension framing, participants accurately assessed the zero correlation. In the feature framing, participants associated the two present and the two absent levels, suggesting a positive correlation between the two cues. However, one should not infer that there simply is a main effect of attribute framing on contingency judgments. As shown in the rest of the experiments, the effects of framing are moderated by stimulus distribution properties such as base rates.

In the “skewed base rates” experiment (experiment 6), we investigated effects of attribute framing in a more complex environment. Participants had to learn about the relationship of two cues within two consumer products each. For one of the products, both base rates were skewed, while for the other product, only one base rate was skewed. The experiment did not include a full crossing of cue labels and base rates and did therefore not allow for a perfect disentangling of different processes or strategies. Still, the different framing conditions did produce different contingency estimates. In the feature condition, it did not matter whether both or only one base rate was skewed. The contingency was always estimated to be around 0. For the dimension condition, participants indicated a positive contingency for the double skewed distribution and a 0 contingency for the single skewed distribution. This is in line with the idea of a pseudocontingency strategy. When both base rates of a distribution are skewed, the frequent levels get associated with one another and the rare levels get associated with one another. The experiment did not vary the PC-index and A-cell size independently. The last three experiments are at the heart of the present volume and present a more elaborate and systematic variation of framing and frequencies in complex multi cue environments.

In the “diamond I” experiment (experiment 7), we confronted participants with a multi cue environment based on work by Fiedler (2010). All cues had skewed base rates of 3:1. We manipulated framing between participants (feature vs. dimension) and actual contingency within participants (positive vs. zero. vs. negative). Additionally, in the feature condition, we manipulated whether the frequent or the rare level was the present or the absent one. This allowed for varying the A-cell size (number of joint present occurrences) independently of cue base rates and therefore independently of the PC-index. As predicted, we found that in a dimension framing, contingency judgments were accurately predicted using the PC-index which was based on participants’ base rate estimates. Judgments were generally biased towards the direction suggested by the pseudocontingency. In the feature framing, participants' contingency judgments were not predicted by the PC-index, but by the number of joint present occurrences (A-cell). On top of that, we found an overall sensitivity for the actual varying correlation. Supporting the idea of differently structured stimulus spaces dependent on attribute framing, we also found effects of framing on base rate estimates. In the dimension condition, base rates were more accurately assessed than in the feature condition.

The "diamond II" experiment (experiment 8) was very similar to the "diamond I" experiment. This time however, we did not use contingency judgments as a dependent variable, but we were interested in choices. Further supporting our ideas, participants in the feature condition were better in finding the correct solution when the correct solution was also suggested by an A-cell based strategy. In the dimension condition, participants were better in finding the correct solution, when the solution was also suggested by a PC-based strategy. In this experiment, we also avoided asking for base rates or conditional probabilities, but rather directly asked for choices. This way, we can rule out that asking for base rates might highlight several aspects of the stimulus world which would normally not be used for judgments or decisions.

The "diamond III" experiment (experiment 9) was again very similar to the "diamond I experiment. Rather than only relying on variability in base rates due to variability in participants' judgments, we manipulated base rates within participants. We kept correlations between all cues constant at $\Delta P = 0$, but varied base rates (and cue labels) within participants ($BR = .75/.25$ vs. $BR = 5$). This setup allowed for an even more exhaustive differentiation between different strategies by disentangling PC-index and A-cell size even more thoroughly. Participants were again very accurate at assessing the base rates and A-cell sizes. Overall, the results were again in line with our ideas, showing evidence for the use of a PC-strategy in the dimension condition and for the use of an A-cell strategy in the feature condition. Also, participants were again better at assessing base rates when attributes were presented as dimensions compared to when they were presented as features.

14.6 Attribute framing and contingency learning – Summary.

When learning frequencies and relationships of cues, the way information is presented is of crucial importance. We show that when a cue has two present levels, other processes are involved than when a cue is presented in a present vs. absent fashion. When confronted with present vs. present cues (dimensions as we refer to them in the sense of Garner, 1978), other levels of the same cue seem to be highlighted. When confronted with present vs. absent cues (features in the sense of Garner), presence of the cue itself and presence of other cues seem to be highlighted. This can influence base rate perception and also judgments of contingency. We showed that for dimensions, base rates were estimated more accurately than for features. Participants perceived the actually skewed distributions to be more skewed when all of the cue-levels were positively defined as compared to when they had to deal with the absence of stimuli. While participants were sensitive to differences in actual contingencies, contingency judgments and influencing factors differed between the framings. For dimensions, actual contingency as well as base rate alignment (PC-index) predicted contingency judgments. For features, actual contingency as well as subjective A-cell size predicted contingency judgments while the PC-index did not. Put simply, while the reference set for dimensions seems to consist of other levels of the same attribute, the reference set for features seems to consist of other present attributes in other objects.

At first glance, density biases (or an underlying A-cell focused strategy) and pseudocontingencies seem similar. In both cases, contingency judgments are influenced by heightened frequencies of certain cells. And indeed, in some cases both findings allow similar predictions. However, a closer look clearly shows that in certain environments, predictions can be independent. In a pseudocontingency framework, cue labeling does not matter. It is only about the alignment of skewed base rates. When two cues have frequent and rare levels, frequent levels get associated and rare levels get associated. Which of these cue combinations is present-present or absent-absent is irrelevant for the prediction. For a strategy based on present-present observations, cue labels are of obvious importance. We show that when cue levels are presented in a way that all levels are positively defined (dimensions), pseudocontingency inferences are promoted by the heightened importance and accuracy of base rates. When absence is involved (features), people seem to put less weight on base-rate assessment and rather focus on the presence of other cues, producing density biases.

A thorough look into the literature allows an interesting observation: While a lot of causation research uses stimuli that are presented in a present vs. absent fashion, a lot of contingency assessment research uses stimuli that are presented in a present-present fashion (Kutzner & Ihmels, 2016). Systematically controlling for this difference might help to understand some of the differences in findings and theories between those two well established research areas.

14.7 Conceptual limitations.

Building on a feature-dimension framework by Garner (1978), we extended the notion of attribute framing by investigating differences between attributes which were either present vs. absent or present vs. present. While we did engage in different operationalizations of our ideas, the present work is by no means an exhaustive investigation of all related aspects. The next paragraphs discuss some conceptual limitations and offer starting points for future related research.

All examples and operationalizations of variables presented in this volume were always dichotomous. There were always two levels of every variables or every cue. In the case of features, one present and one absent level and in the case of dimensions, two present levels. Obviously, in our environment, a lot of variables have more than two levels and some are even continuous with an infinite number of levels. The way a feature is defined here ("is X present?") strictly implies a dichotomous fashion: Present and absent. For dimensions, this is different. There can easily be more than two answers to the question "What level does X have?". So at first glance, there seems to be an asymmetry between features and dimensions. While features are always dichotomous, dimensions can have more than two and sometimes even an infinite number of levels. Importantly, as discussed in the introduction, it is not so much about whether attributes are features or dimensions but more about how we think about attributes. One could think about whether a person wears glasses or not or one could think about what kind of glasses someone is wearing. A hierarchical structure of first thinking about the presence or absence and then, given the attribute is present, thinking about what level the attribute has seems adaptive in many situations. The question what kind of glasses someone is wearing only makes sense once it is known that the person is wearing glasses. Hierarchical structures as well as variables with more than two levels might be considered in future research.

In the present work, dimensions were introduced as variables where both cue levels are of similar (symmetric) quality, while the two levels of a feature were introduced as qualitatively different. While all of the dimensions used in the experiments were symmetric, one can also think of asymmetric dimensions. By this we mean attributes for which one level (or one end) of a dimension is more relevant than the other - for whatever reasons. Think about smog in a big city such as Los Angeles. To simplify, let us think about days of harmless smog and days where the smog gets so strong that it gets dangerous. While the smog is always present on some level in this case, one level (dangerous smog) might be seen as more relevant than the other. In this case, because it implies special actions and precautions. This is also related to the distinction of marked and unmarked adjectives (Bierwisch, 1967). An unmarked adjective can be used in a neutral way and refers to both an area on the scale, as well as the scale itself (Lee, 2014). For example, one could ask how *tall* a man is (tall = unmarked). The marked adjective in this case would be *short*. Asking how *short* a man is would suggest that the man actually is short (Clark & Card, 1969). In a way, markedness is also related to typicality (Fraenkel & Schul, 2008). Labels implying a certain frequency should also bias contingency judgments. Using asymmetric dimensions (marked vs. unmarked) as illustrated here might trigger processes similar to those expected for features. Extending this thought, it might be helpful to think about features and dimensions not as two exclusive categories, but as endpoints of a continuous scale.

In the present work, features and dimensions were presented as two levels of a dichotomous variable. Either something was clearly framed as a feature or clearly framed as a dimension. Dimensions were supposed to allow easy structuring of the stimulus space while features were supposed to result in a more vague representation due to absence being hard to deal with. However, information can come in a lot of different forms and some cases might lead to something in between a clear feature and a clear dimension, or rather in between a very well structured and an impoverished stimulus space. For example, as discussed earlier, a situation with marked adjectives might be such a setup. Framing as well as context can probably help structuring feature based stimulus spaces or lead to less structured dimension based stimulus spaces. Therefore, one might also think about this type of attribute framing in a continuous rather than a dichotomous fashion. We are currently working on a meta-analysis on the effects of attribute / feature-dimension framing in contingency learning (Kutzner & Ihmels, 2016). In this meta-analysis, we adopt this thought of a continuous conceptualization and test existing experiments / instructions for how they are understood by participants.

Features were always presented as being present or absent. Another way to implement the definition would have been to contrast present vs. not present (or also absent vs. not absent). While we chose to go with *absence* instead of “not present” in a first step, we do also plan to investigate how people react to explicit absence. Perception wise and conceptually, these things are slightly different. If something is completely absent, one has to actively construct the possibility of the presence, based on prior experience in the same context. Visually, absence is qualitatively different from presence. Stating "not present" takes this difference away and allows differentiation of absence and lack of information (Blaisdell et al., 2009). Complete absence leaves some room for insecurity. It can raise questions such as: "Was it really absent or was the information just missing? Did I just not detect it?", etc.. For the explicit absence, this is different. Here it is clear that the attribute really was not present and at the same time, the wording reveals the underlying attribute. Further, if framing and experiments in general are understood as a communicative act, other communicator or experimenter intentions might be inferred when absence is stated explicitly compared to when something is just absent.

14.8 Process details.

We did our best to investigate the role of attribute framing in contingency learning and so far, the experiments have provided what we think are very interesting insights. Nevertheless, there is a lot to be done and we hope to inspire other research with our theoretical as well as empirical work. Regarding our own work, there is a lot that can be followed up on. For example, one might argue that the presented level of analysis was not fine grained enough. In our "diamond" experiments, we chose to balance everything. For example, when asking about base rates, we randomly asked for the frequent or the rare level and thereby also chose randomly if we asked for the present or absent level in the feature condition. Later, we simply looked at base rate estimates after the appropriate transformations, ignoring this variation of properties. The most detailed level of analysis would have been a 2x2 analysis of the base rate estimates: "Did we ask for the frequent or the rare level?" and "Did we ask for the present or the absent level?". The same logic can be applied to how we asked about conditional probabilities, etc.. To strengthen our understanding of the ongoing processes when people deal with present vs. absent attributes, it might provide interesting insights to further illuminate these fine differences in how questions are asked about present vs. absent attributes. Also in general (also for dimensions), it might make a difference whether one asks about the frequent or the rare level, given people do not seem to have a perfectly linear understanding of numerical quantities (Juslin, Nilsson, Winman, & Lindskog, 2011). While the present work provided good first insights into what is happening, there is still much to be explored in terms how exactly and at what stages of the process things happen. The task of learning contingencies is a difficult one and includes multiple aspects. Thinking about what steps are necessary to properly assess and use a contingency can be helpful to think about the impact attribute framing might have on contingency learning. Goal of the present work was to show the existence of attribute framing effects in contingency learning, rather than exactly investigating at what step of the learning and judgment process the effects play a role. Crocker (1981) identified six steps that are involved in the contingency learning process, from deciding what data to look at to decisions based on contingency judgments. According to the model, the steps are (1) deciding what data is relevant for the judgment, (2) sampling cases, (3) classifying instances (interpreting and coding the data), (4) recalling evidence and estimating frequencies of confirming and disconfirming cases, (5) integrating the evidence, (6) using the covariation estimate. Future work might investigate details of the effects of attribute framing on the different steps. For an overview of how accurate people are at the different steps of the contingency learning process, see Mata (in press) or Crocker (1981).

14.9 Extensive model fitting.

While the present work did contain a lot of practices similar to how model fitting contests usually work, the idea was not to find the perfect way of integrating subjective cell sizes to best predict participants contingency judgments. There most probably are better ways (in the sense that they predict more accurately) than the ones we used. For example, instead of just using A-cell size to predict contingency judgments (weights of 1, 0, 0, 0 for A, B, C, D), we could have used a less radical and more complex index such as the dual factor heuristic presented by Hattori and Oaksford (2007). The PC-index also could be built in different ways producing different slopes in the PC-index as a function of joint skewness. It was not our first priority to find the exact index predicting participants judgments, but rather to integrate our theoretical ideas in a simple and straight forward way. The more joint occurrences, the bigger the A-cell index and the more both cues are jointly skewed, the bigger the PC-index.

It might nevertheless be fruitful to look at other contingency indices and to think about how their predictive value might change as a function of how attributes are presented. Take the proportion of confirming instances (PCI: White, 2003) for example. To be able to talk about confirming or disconfirming instances, it is necessary to have some sort of prior expectation about the relationship. Different attribute framings might shift that prior. For example, as already indicated, features might lead to the expectation that both present levels are associated (Goodie & Fantino, 1996).

14.10 Potential applications.

While most of the presented research did not have the primary goal of working towards knowledge which can be applied in a consumer or marketer setting, some of the generated insights might still be fruitful to think about. For example, the experiment on the "stimulus space" (experiment 5) offers a pretty straight forward implication for marketers. The goal for marketers is that their product is preferred over other products. Therefore, being able to control comparison processes can be very useful. When an attribute in a product is of relatively low quality, one should not present it as a dimension, since this would trigger and allow comparisons to the quality of this attribute in other objects. Presenting it as a feature makes the comparison harder and should therefore be preferred. Research shows that in the communication setting of advertising, communicated attributes are expected to be positive. Even if consumers do not know what the attribute actually means, since marketers chose to communicate it, “it must be something good” (Wänke & Reutner, 2010). Presenting the attribute as a dimension would potentially undermine this effect, since attribute levels across products might then be compared. For example, when a snack bar contains only a small amount of the up and coming “chia seeds” compared to other healthy snack bars, one should rather go for “with chia seeds” than for “2g chia seeds” when the other snack bars contain more (say “5g chia seeds”).

As discussed in the introduction, single levels of features and dimensions differ in the inferences they allow about the world. They are not information equivalent. Deciding how to frame individual attributes can impact mental representations and subsequent decisions. Framing also has a communicative aspect. A feature might imply that the presence is more important than the absence while for a dimension, this is not the case. This might be used to strategically guide people towards a certain processing of information, without objectively changing the given information. A recommendation might be inferred based on the chosen frame.

15. Conclusion - Presence vs. absence: Reference sets and contingency judgments.

Presence and absence differ in the inferences they allow about the stimulus world. They differ in how easy they are to deal with and in the comparisons they suggest. It is important to think about the way information is presented and to pay attention to the often overlooked distinction of present vs. present and present vs. absent attributes. Framing is not only about how single objects are evaluated, but importantly also about how relationships between objects are perceived.

Introducing the theoretical distinction of features and dimensions (Garner, 1978), we were able to integrate two prominent, but hitherto unconnected findings: Pseudocontingencies and density biases. Our framework allows clear predictions to when people are susceptible to base rate driven inferences and when they are likely to be mainly influenced by one specific type of cue level combination, namely the present-present cell. Dimensions promote within attribute comparisons and pseudocontingency inferences, while features promote comparisons with other present attributes.

Judgments of relation are relative and sometimes difficult judgments. Observations have to be compared to one another and while there are objectively correct ways to do this, some information tends to be get more weight while other information tends to be neglected. Which information is used in what way varies as a function of how attributes are presented: As present vs. absent or as always being present on some level.

References

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.

Allan, L. G., Siegel, S., & Tangen, J. M. (2005). A signal detection analysis of contingency data. *Learning & Behavior, 33*, 250-263.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-8 [Computer software]. Retrieved from http://CRAN.R-project.org/package=lme4

Bernoulli, D. (1954 [1738]). Exposition of a new theory on the measurement of risk. [Translation of "Specimen theoriae novae de mensura sortis"]. *Econometrica, 22*, 23-36.

Beyth-Marom, R. (1982). Perception of correlation reexamined. *Memory & Cognition, 10*(6), 511-519.

Bierwisch, M. (1967). Some semantic universals of German adjectivals. *Foundations of Language*, 1-36.

Blaisdell, A. P., Leising, K. J., Stahlman, W. D., & Waldmann, M. R. (2009). Rats distinguish between absence of events and lack of information in sensory preconditioning. *International Journal of Comparative Psychology*, *22*(1), 1-18.

Bless, H., Fiedler, K., & Strack, F. (2004). Social cognition: How individuals construct social reality. Sussex, UK: Psychology Press.

Bock, O., Nicklisch, A., & Baetge, I. (2012). hroot: Hamburg registration and organization online tool. *WiSo-HH Working Paper Series No. 01*, 1-14.

Boring, E. G. (1954). The nature and history of experimental control. *The American Journal of Psychology, 67*(4), 573-589.

Bröder, A., Newell, B. R., & Platzer, C. (2010). Cue integration vs. exemplar-based reasoning in multi-attribute decisions from memory: A matter of cue representation. *Judgment and Decision Making, 5*(5), 326-338.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(3), 193-217.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3-5.

Clark, H. H., & Card, S. K. (1969). Role of semantics in remembering comparative sentences. *Journal of Experimental Psychology*, *82*(3), 545-553.

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

Crocker, J. (1982). Biased questions in judgment of covariation studies. *Personality and Social Psychology Bulletin, 8*(2), 214-220.

Crowe, E., & Higgins, E. T. (1997). Regulatory focus and strategic inclinations: Promotion and prevention in decision-making. *Organizational Behavior and Human Decision Processes, 69*(2), 117-132.

Denrell, J. (2005). Why Most People Disapprove of Me: Experience sampling in smpression formation. *Psychological Review, 112*(4), 951-978.

Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin,130*(6), 959-988.

Dhar, R., & Sherman , S. J. (1996). The Effect of Common and Unique Features in Consumer Choice. *Journal of Consumer Research, 23*(3), 193-203.

Duncan, O. D., & Davis, B. (1953). An alternative to ecological correlation. *American Sociological Review*, *18*(6), 665-666.

Eder, A. B., Fiedler, K., & Hamm-Eder, S. (2011). Illusory correlations revisited: The role of pseudocontingencies and working-memory capacity. *The Quarterly Journal of Experimental Psychology, 64*(3), 517-532.

Fiedler, K. (2001). Affective states trigger processes of assimilation and accommodation. In L. L. Martin & G. L. Clore (Eds.), *Theories of mood and cognition: A user's guidebook* (pp. 85-98). Mahwa, NJ, US: Lawrence Erlbaum Associates Publishers.

Fiedler, K. (2008). The ultimate sampling dilemma in experience-based decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(1), 186-203.

Fiedler, K. (2010). Pseudocontingencies can override genuine contingencies between multiple cues. *Psychonomic Bulletin & Review, 17*(4), 504-509.

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

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

Fiedler, K., Kutzner, F., & Vogel, T. (2013). Pseudocontingencies: Logically Unwarranted but Smart Inferences. *Current Directions in Psychological Science, 22*(4), 324-329.

Forgas, J. P., & Bower, G. H. (1987). Mood Effects on Person-Perception Judgments. *Journal of Personality and Social Psychology, 53*(1), 53-60.

Fraenkel, T., & Schul, Y. (2008). The meaning of negated adjectives. *Intercultural Pragmatics*, *5*(4), 517-540.

Gallistel, C. R. (2007). Flawed foundations of associationism? Comment on Machado and Silva (2007). *American Psychologist, 62*(7), 682-685.

Garner, W. R. (1978). Aspects of a Stimulus: Features, Dimensions, and Configurations. In E. Rosch & B. Lloyd (Eds.), *Cognition and Categorization* (pp. 99-121). Oxford England: Lawrence Erlbaum.

Gati, I., & Tversky, A. (1982). Representations of qualitative and quantitative dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 8*(2), 325-340.

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

Grice, P. (1975). Logic and Conversation. In P. Cole, J. Morgan (Eds.), *Syntax and Semantics. 3: Speech Acts* (pp. 41-58). New York, NY: Academic Press.

Hammond, K. R. (1954). Representative vs. systematic design in clinical psychology. *Psychological Bulletin*, *51*(2), 150-159.

Hammond, J. L. (1973). Two sources of error in ecological correlations. *American Sociological Review, 38*(6), 764-777.

Hannah, S. D., & Beneteau, J. L. (2009). Just tell me what to do: Bringing back experimenter control in active contingency tasks with the command-performance procedure and finding cue density effects along the way. *Canadian Journal of Experimental Psychology*, *63*(1), 59-73.

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

Houston, D. A., & Sherman, S. J. (1995). Cancellation and Focus: The Role of Shared and Unique Features in the Choice Process. *Journal of Experimental Social Psychology, 31*(4), 357-378.

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes, 67*(3), 247-257.

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin, 125*(5), 576-590.

Hsee, C. K., & Zhang, J. (2010). General evaluability theory. *Perspectives on Psychological Science, 5*(4), 343-355.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis. *Journal of Consumer Research*, *9*(1), 90-98.

Johnson-Laird, P. N., & Shafir, E. (1993). The interaction between reasoning and decision making: An introduction. *Cognition, 49*(2), 1-9.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: a new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*(1), 54-69.

Juslin, P., Nilsson, H., Winman, A., & Lindskog, M. (2011). Reducing cognitive biases in probabilistic reasoning by the use of logarithm formats. *Cognition*,*120*(2), 248-267.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society, 47*(2), 263-292.

Kelley, H. H. (1967). Attribution theory in social psychology. *Nebraska Symposium on Motivation, 15,* 192-238.

Kutzner, F. (2009). *Pseudocontingencies - rule based and associative* (Doctoral dissertation, University of Heidelberg, Heidelberg, Germany). Retrieved from http://archiv.ub.uni-heidelberg.de/volltextserver/9912/1/Dissertation_Kutzner_UB.pdf

Kutzner, F., & Ihmels, M. (2016). *Attribute framing in contingency learning – A meta analysis.* Manuscript in preparation.

Krüger, T., Mata, A., & Ihmels, M. (2014). The Presenter's Paradox Revisited: An Evaluation Mode Account. *Journal of Consumer Research, 41*(4), 1127-1136.

Krüger, T., Vogel, T., & Wänke, M. (In Press). The Framing of Consumer Decision Making. In C. Jansson-Boyd & M. Z. Zawisza (Eds.), *International Handbook of Consumer Psychology*. Sussex, UK: Taylor & Francis.

Le Mens, G., & Denrell, J. (2011). Rational learning and information sampling: On the "naivety" assumption in sampling explanations of judgment biases. *Psychological Review, 118*(2), 379-392.

Lee, C. (2014). *Consumer Linguistics: A Markedness Approach to Numerical Perceptions.* (Doctoral dissertation, University of Orgeon, Oregon, United States). Retrieved from https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/18352/Lee_oregon_0171A_10957.pdf?sequence=1

Leiner, D. J. (2014). SoSci Survey (Version 2.5.00-i) [Computer software]. Retrieved from http://www.soscisurvey.com

Levin, I. P., & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research, 15*(3), 374-378.

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes, 76*(2), 149-188.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of experimental psychology*, *89*(1), 46-55.

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology, 15*(3), 215-233.

Mata, A. (in press). Judgment of covariation: A review. *Psicologia*.

Matute, H., Yarritu, I., & Vadillo, M. A. (2011). Illusions of causality at the heart of pseudoscience. *British Journal of Psychology, 102*(3), 392-405.

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

McKenzie, C. R., Ferreira, V. S., Mikkelsen, L. A., McDermott, K. J., & Skrable, R. P. (2001). Do conditional hypotheses target rare events? *Organizational Behavior and Human Decision Processes*, *85*(2), 291-309.

Mussweiler, T. (2001). 'Seek and ye shall find': antecedents of assimilation and contrast in social comparison. *European Journal of Social Psychology, 31*(5), 499-509.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175-220.

Nosofsky, R. M., Kruschke, J. K., & McKinley, S. C. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 211-233.

Pansky, A., & Algom, D. (1999). Stroop and Garner effects in comparative judgment of numerals: The role of attention. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 39-58.

Perales, J. C., & Shanks, D. R. (2007). Models of covariation-based causal judgment: A review and synthesis. *Psychonomic Bulletin & Review, 14*(4), 577–596.

R Development Core Team (2008). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing,Vienna, Austria. ISBN 3-900051-07-0. Retrieved from http://www.R-project.org.

Ratneshwar, S., Shocker, A. D., & Stewart, D. W. (1987). Toward understanding the attraction effect: The implications of product stimulus meaningfulness and familiarity. *Journal of Consumer Research, 13*(4), 520-533.

Rubinstein, A. (1998). *Modeling bounded rationality*. Cambridge, MA: MIT Press.

Shafir, E. (1993). Choosing versus rejecting: Why some options are both better and worse than others. *Memory & Cognition, 21*(4), 546-556.

Sher, S., & McKenzie, C. R. (2006). Information leakage from logically equivalent frames. *Cognition, 101*(3), 467-494.

Sher, S., & McKenzie, C. R. (2014). Options as information: Rational reversals of evaluation and preference. *Journal of Experimental Psychology: General, 143*(3), 1127-1143.

Sherman, S. J., Houston, D. A., & Eddy, D. (1999). Cancellation and Focus: A Feature-Matching Model of Choice. *European Review of Social Psychology, 10*(1), 169-197.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*(2), 129-138.

Snyder, M., & Swann, W. B. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology*, *36*(11), 1202-1212.

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology, 89*(6), 845-851.

Sütterlin, B., Brunner, T. A., & Opwis, K. (2008). Eye-Tracking the Cancellation and Focus Model for Preference Judgments. *Journal of Experimental Social Psychology, 44*(3), 904-911.

Trope, Y., & Liberman, N. (2011). Construal Level Theory. In P. A. M. Van Lange, A. W. Kruglanski & E. T. Higgins (Eds.), *Handbook of Theories of Social Psychology: Volume 1* (pp. 118-134). Thousand Oaks, CA: Sage.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327-352.

Tversky, A., & Gati, I. (1978). Studies of similarity. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and Categorization* (pp. 79-98). Hillsdale, NJ: Lawrence Earlbaum Associates.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453-458.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior (1st ed.)*. Princeton, NJ: Princeton University Press.

Wänke, M. (2002). Conversational norms and the interpretation of vague quantifiers. *Applied Cognitive Psychology, 16*(3), 301-307.

Wänke, M. & Reutner, L. (2010). Pragmatic Persuasion or the Persuasion Paradox. In J. Forgas, W. Crano & J. Cooper (Eds.), *The Psychology of Attitudes & Attitude Change* (pp. 183-198). New York, NY: Psychology Press.

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

Wasserman, E. A., Kao, S. F., Van Hamme, L. J., Katagiri, M., & Young, M. E. (1996). Causation and association. *Psychology of Learning and Motivation, 34*, 207-264.

Weaver, K., Garcia, S. M., & Schwarz, N. (2012). The Presenter's Paradox. *Journal of Consumer Research*, *39*(3), 445-460.

White, P. A. (1995). *The understanding of causation and the production of action: From infancy to adulthood*. London, UK: Psychology Press.

White, P. A. (2003). Making causal judgments from the proportion of confirming instances: The pCI rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 710-727.

White, P. A. (2009). Accounting for occurrences: An explanation for some novel tendencies in causal judgment from contingency information. *Memory & Cognition, 37*(4), 500-513.

Zacks, R. T., & Hasher, L. (2002). Frequency processing: A twenty-five year perspective. In P. Sedlmeier & T. Betsch (Eds.), *ETC. Frequency Processing and Cognition* (pp. 21-36). New York, NY: Oxford University Press.

Appendix 1 - Contingency Framework

Note that all lines beginning with '#' are comments and don't have any impact on the actual functions. Just copy the code into R (R-studio recommended) and install the packages mentioned at the beginning of the code. In case of any questions, feel free to contact the author (Max Ihmels).

```
#### Framework for Handeling Contingency Learning - Max Ihmels ####

#### Packages you should install:
# ggplot2
# gridExtra
# wesanderson
# scales

#### General Setup and Labeling ####
#        Outcome A    Outcome B
# Pred A      A           B
# Pred B      C           D
#### General Setup and Labeling for data handling ####
#        Outcome A    Outcome B
# Pred A    x[1]        x[2]
# Pred B    x[3]        x[4]

#### Clear memory and load libraries - START ####
rm(list=ls())
library(ggplot2)
library(gridExtra)
library(wesanderson)
library(scales)
#### Clear memory and load libraries - END ####

#### Base Rate Functions - START ####
# Sample Size
getSampleSize <- function(x) {
  return(x[1] + x[2] + x[3] + x[4])
}
# Simply Inverse Base Rate
getInverseBR <- function(BR) {
  return(100 - BR)
}
# Get BR in percent (0-1) from two values within one row
getBRperc <- function(X, Y) {
  BRperc <- X / (X + Y)
  return(BRperc)
}
## Get Base in percent for upper row and left column
getBR <- function(x) {
  BRtop   <- (x[1] + x[2]) / sum(x)
  BRleft  <- (x[1] + x[3]) / sum(x)
  return(c(BRtop, BRleft))
}
#### Base Rate Functions - END ####
```

```
#### Contingency Functions - START ####
# Get Delta P - Calculated on Basis of the cell frequencies
getDeltaP <- function(x) {
  deltaP <- (x[1] / (x[1] + x[2])) - (x[3] / (x[3] + x[4]))
  return(deltaP)
}
# Get Phi - Calculated on the Basis of the cell frequencies - FORMULA
getPhi <- function(x) {
  phi <- x[1] + x[2] + x[3] + x[4]
  return(phi)
}
# Get H
getH <- function(x) {
  H <- x[1] / ((x[1] + x[2]) * (x[1] + x[3]))^0.5
  return(H)
}
# Get Sum of Diagonals (SOD)
getSOD <- function(x) {
  SOD <-  x[1] + x[4] - x[2] - x[3]
  return(SOD)
}
# Get PC Coefficient from Cell Frequencies - Uses getBRperc to get relative Base Rates
getPC <- function(x) {
  PC <- log10((x[1] + x[2]) / (x[3] + x[4])) * log10((x[1] + x[3]) / (x[2] + x[4]))
  return(PC)
}
# Get proportion of confirming instances (pCI)
getpCI <- function(x) {
  pCI <- (x[1] + x[4] - x[2] - x[3]) / (x[1]+x[2]+x[3]+x[4])
  return(pCI)
}
#### Contingency Functions - END ####

#### Overview functions - START ####
## Contingency Summary: A,B,C,D, SampleSize, BRtop, BRleft, Delta P, PHI, H, SOD, PC
contSummary <- function(x) {
  # Get Base Rates
  A       <- x[1]
  B       <- x[2]
  C       <- x[3]
  D       <- x[4]
  N       <- getSampleSize(x)
  BRtop   <- getBR(x)[1]
  BRleft  <- getBR(x)[2]
  # Get most common Coefficients
  DeltaP  <- getDeltaP(x)
  PHI     <- getPhi(x)
  H       <- getH(x)
  SOD     <- getSOD(x)
  PC      <- getPC(x)
  pCI     <- getpCI(x)
  return(cbind(A, B, C, D, N, BRtop, BRleft, DeltaP, PHI, H, SOD, PC, pCI))
}
#### Overview functions - END ####
```

```
#### Generation of Stimulus Distributions - START ####

## Simply turn 4 values (ABCD) into one object
setDistr <- function(A,B,C,D){
  return(c(A,B,C,D))
}

## Create Distribution with given demands
# nDistr = Number of distributions to be created
# CellMin = Minimum size of a single cell
# CellMax = Maximium size of a single cell
# N       = Desired SampleSize
# Nerr    = SampleSize Error (How far is the Sample Size allowed to deviate from N)
# targetBRtop = Desired BR of predictor (top row)
# BRtoperr = BRtop error
# targetBRleft = Desired BR of outcome (left column)
# BRlefterr = BRleft error
# targetDP = Desired DeltaP
# DPerr = DeltaP error
# targetH = Desired H
# Herr = H error
# targetPC = Desired PC
# PCerr = PC error
# targetPhi = Desired Phi
# Phierr = Phi error
# targetSOD = Desired SOD
# SODerr = SOD error
createDistribution <- function(nDistr = 1,
                               CellMin  = 1, CellMax = 20,
                               targetN    = 40, Nerr = 50,
                               targetBRtop = FALSE, BRtoperr = 0.05,
                               targetBRleft = FALSE, BRlefterr = 0.05,
                               targetDP   = FALSE, DPerr  = 0.03,
                               targetH    = FALSE, Herr   = 0.03,
                               targetPC   = FALSE, PCerr  = 0.03,
                               targetPhi  = FALSE, Phierr = 0.03,
                               targetSOD  = FALSE, SODerr = 10,
                               targetpCI  = FALSE, pCIerr = 0.03) {
  # Randomly determine cell size for the four cells (With regards to CellMax)
  sampling <- function() {
    cellsABCD <- sample(CellMin:CellMax,4, replace = TRUE)
    return(cellsABCD)
  }
  # Create Matrices to store "correct" distributions (Using more than 1 to increase speed)
  resDistr <- NULL
  resDistr2 <- NULL
  resDistr3 <- NULL
  # Create nDistr distrubtions
  for(i in 1:nDistr) {
    keepGoing = TRUE
    while(keepGoing == TRUE) {
      samDis <- sampling()
      if(
        ## Sample Size - Target Sampling Size <= Sample size error?
        abs(getSampleSize(samDis) - targetN) <= Nerr &&
```

```
      ## targetBRtop
      (targetBRtop == FALSE || abs(getBR(samDis)[1] - targetBRtop) <= BRtoperr) &&
      ## targetBRleft
      (targetBRleft == FALSE || abs(getBR(samDis)[2] - targetBRleft) <= BRlefterr) &&
      ## DeltaP - Target Delta P <= Delta P error? (If a targetDP is specified)
      (targetDP == FALSE || abs(getDeltaP(samDis) - targetDP) <= DPerr) &&
      ## H - Target H <=  H error (If a targetH is specified)
      (targetH == FALSE || abs(getH(samDis) - targetH) <= Herr) &&
      ## PC - Target PC <= PC error? (If a target PC is specified)
      (targetPC == FALSE || abs(getPC(samDis) - targetPC) <= PCerr) &&
      ## PHI
      (targetPhi == FALSE || abs(getPhi(samDis) - targetPhi) <= Phierr) &&
      ## SOD
      (targetSOD == FALSE || abs(getSOD(samDis) - targetSOD) <= SODerr) &&
      ## pCI
      (targetpCI == FALSE || abs(getpCI(samDis) - targetpCI) <= pCIerr)
      ) {
     keepGoing = FALSE
     if(i <= nDistr / 3){
      resDistr <- rbind(resDistr, contSummary(samDis))
     } else if(i > nDistr / 3 && i < nDistr / 3) {
      resDistr2 <- rbind(resDistr2, contSummary(samDis))
     } else {
      resDistr3 <- rbind(resDistr3, contSummary(samDis))
     }
     }
   }
  }
  finalDistr <- rbind(resDistr, resDistr2, resDistr3)
  return(as.data.frame(finalDistr))
}
#### Generation of stimulus Distributions - END ####


#### List of Data Handling Functions - START ####
## Base Rates:
# getSampleSize(x)  - Returns N
# getInverseBR(A)   - Inverses Base Rate (100 - A)
# getBRperc(A,B)    - Relative frequency of A in A + B
# getBR(x)          - Returns base rate in % for top row and left column
## Contingencies:
# getDeltaP(x)      - Returns Delta P
# getPhi(x)         - Returns Phi
# getH(x)           - Returns H
# getSOD(x)         - Returns SOD
# getPC(x)          - Returns PC on the Basis of Cell frequencies
# getpCI(x)         - Returns pCI on the Basis of Cell frequencies
#### List of Data Handling Functions - END ####

#### List of Overview functions - START ####
# contSummary(x)  - Returns A, B, C, D, N, BRtop, BRbot, BRleft, BRright, DeltaP, Phi, H, SOD, PC
#### List of Overview functions - END ####
```

```
#### List of Generation functions - START ####

# setDistr(A,B,C,D) - Simply turns A B C D into one object that can then be used for other functions

# Create distributions with given demands
# createDistribution(
#         nDistr = 1
#         CellMin  = 1, CellMax = 20,
#         targetN    = 40, Nerr = 10,
#         targetBRtop = Desired BR of predictor (top row)
#         BRtoperr = BRtop error
#         targetBRleft = Desired BR of outcome (left column)
#         BRlefterr = BRleft error
#         targetDP   = FALSE, DPerr  = 0.03,
#         targetH    = FALSE, Herr   = 0.03,
#         targetPC   = FALSE, PCerr  = 0.03,
#         targetPhi  = FALSE, Phierr = 0.03.
#         targetSOD  = FALSE, SODerr = 10)
# Creates a Distribution with the desired features:
# nDistr  = Number of distributions to be created
# CellMin = Minimum size of a single cell
# CellMax = Maximium size of a single cell
# N       = Desired SampleSize
# Nerr    = SampleSize Error (How far is the Sample Size allowed to deviate from N)
# targetDP = Desired DeltaP
# DPerr = DeltaP error
# targetH = Desired H
# Herr = H error
# targetPC = Desired PC
# PCerr = PC error
# targetPhi = Desired Phi
# Phierr = Phi error
# targetSOD = Desired SOD
# SODerr = SOD error
#### List of Generation functions - END ####
```

```
#### Exemplary Use - START ####
## Example of randomly creating distributions without any restrictions
# Default values for Sample Size / Cell Sizes are used
results <- createDistribution(nDistr = 30000, CellMax = 15)

## Example of plotting two values against each other (plus colour coding by a third variable)
## and arranging them in the next step
p1 <- qplot(DeltaP, H, colour = PC, data = results)
p2 <- qplot(PC, BRtop, colour=BRleft, data = results)
p3 <- qplot(BRtop, BRleft, colour=DeltaP, data = results)
p4 <- qplot(DeltaP, PC, colour = N, data = results)
grid.arrange(p1,p2,p3,p4)

## Example of calculating a correlation of two indices
cor(results$DeltaP, results$PC)

## Example of a more advanced plot including labels and custom colour gradients
# Relationship between base rates and the PC-index
qplot(BRtop, PC, color = BRleft, data = results, xlab="Base Rate 1", ylab="PC-Index") +
scale_colour_gradient2(midpoint=0.5, low="blue", mid="gray", high="red", name ="Base Rate 2")

## Set working directory to desired location and uncomment to write the distributions to a csv file
#write.csv(results,file="Distributions.csv")
#### Exemplary Use - END ####
```

List of figures

List of tables

List of abbreviations

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

**FAKULTÄT FÜR VERHALTENS-
UND EMPIRISCHE KULTURWISSENSCHAFTEN**

**Promotionsausschuss der Fakultät für Verhaltens- und Empirische Kulturwissenschaften
der Ruprecht-Karls-Universität Heidelberg**
Doctoral Committee of the Faculty of Behavioural and Cultural Studies, of Heidelberg University

Erklärung gemäß § 8 (1) c) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften
Declaration in accordance to § 8 (1) c) of the doctoral degree regulation of Heidelberg University, Faculty of Behavioural and Cultural Studies

Ich erkläre, dass ich die vorgelegte Dissertation selbstständig angefertigt, nur die angegebenen Hilfsmittel benutzt und die Zitate gekennzeichnet habe.
I declare that I have made the submitted dissertation independently, using only the specified tools and have correctly marked all quotations.

Erklärung gemäß § 8 (1) d) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften
Declaration in accordance to § 8 (1) d) of the doctoral degree regulation of Heidelberg University, Faculty of Behavioural and Cultural Studies

Ich erkläre, dass ich die vorgelegte Dissertation in dieser oder einer anderen Form nicht anderweitig als Prüfungsarbeit verwendet oder einer anderen Fakultät als Dissertation vorgelegt habe.
I declare that I did not use the submitted dissertation in this or any other form as an examination paper until now and that I did not submit it in another faculty.

Vorname Nachname
First name Family name ______________________________

Datum, Unterschrift
Date, Signature ______________________________