New Insights Into Cognitive Processes in Pseudocontingency Inference by Means of Experimental Methods and Statistical Modeling

Franziska M. Bott

Inaugural Dissertation
Submitted in partial fulfillment of the requirements for the degree of Doctor of Social Sciences in the Research Training Group "Statistical Modeling in Psychology" at the University of Mannheim

Supervisor:
Prof. Dr. Thorsten Meiser

Dean of the School of Social Sciences:
Prof. Dr. Michael Diehl

Thesis Reviewers:
Prof. Dr. Arndt Bröder
Prof. Dr. Benjamin Hilbig

Defense Committee:
Prof. Dr. Arndt Bröder
Prof. Dr. Benjamin Hilbig
Prof. Dr. Thorsten Meiser

Thesis Defense:
September 24, 2020

For my family

Contents

Summary

When estimating the contingency between two variables, individuals often show biases in the association they infer: Sometimes they infer an association where there is none, other times they infer an association that is opposite to the one that is actually observed. These biases have been shown to occur when individuals attempt to infer a contingency based on skewed samples of the variables: frequent categories are assumed to be associated with each other as well as infrequent categories; a phenomenon called *pseudocontingency* (e.g., Fiedler et al., 2009). The present thesis aimed to deepen the understanding of pseudocontingencies through the combination of experimental methods and statistical modeling.

Empirical research demonstrates that pseudocontingencies are relied on for probability judgments and choices (e.g., Meiser et al., 2018). They are also reflected in reconstructive guessing processes, when memory for individual observations fails (e.g., Klauer & Meiser, 2000). In the present thesis, I corroborate this result by analyzing guessing processes based on pseudocontingencies using hierarchical multinomial processing tree models and by validating the model parameters' substantive interpretation (see Manuscript I, Bott et al., in press). In this context, analyses additionally revealed that interindividual differences in cognitive performance as measured by the INSBAT test battery (Arendasy et al., 2009) do not seem to predict interindividual differences in relying on pseudocontingencies (Bott et al., in press).

While early proposals assume differential processing of frequent versus infrequent events (e.g., Hamilton & Gifford, 1976), more recent research suggests that pseudocontingencies are the result of utilizing the marginal frequencies of variables, instead of their joint frequencies, when inferring a contingency (e.g., Fiedler et al., 2009). In line with this notion, in this thesis, I discuss a computational model, the Bayesian Marginal Model (see Manuscript II, Bott et al., 2020; Klauer, 2015), as a normative reconstruction of the pseudocontingency heuristic. The model succeeds in capturing effects found in the literature on pseudocontingency inference by assuming that beliefs about joint frequencies and thus contingencies are updated by observed marginal frequencies.

Most research as well as the Bayesian Marginal Model put individuals in the role of a passive observer of predetermined information. Thus the present thesis furthermore extends empirical research by investigating the role of self-determined information sampling in pseudocontingency inference (see Manuscript III, Bott & Meiser, 2020). The results indicate that pseudocontingencies may result in wrong judgments and sub-optimal choices. However, the probability of misguidance is largely reduced when individuals actively search for information themselves.

Taken together, the experiments and statistical analyses discussed in the present thesis corroborate pseudocontingencies as inferences based on marginal frequencies. Moreover, I provide original evidence that pseudocontingency effects can be captured by a normative model following the norms of Bayesian belief updating, thereby rendering pseudocontingencies as *not* "illogical" as described by some of its proponents. The results additionally highlight that pseudocontingencies are not necessarily wrong contingency inferences, especially when individuals actively search for information.

Articles

This cumulative thesis is based on three manuscripts, one of which has been published, one has been accepted, and one has been submitted for publication. The manuscripts are discussed and appended to this work in the order in which they are listed below. In general, note that in summarizing the articles throughout this thesis, I refrain from reiterating specific details of the respective manuscript which can be found in the manuscripts appended to this work.

MANUSCRIPT I

Bott, F., M., Heck, D. W., & Meiser, T. (in press). Parameter validation in hierarchical MPT models by functional dissociation with continuous covariates: An application to contingency inference. *Journal of Mathematical Psychology*.

MANUSCRIPT II

Bott, F., M., Kellen, D., & Klauer, K. C. (2020). *Normative accounts of illusory correlations.* Invited revision submitted to *Psychological Review*.

MANUSCRIPT III

Bott, F., M., & Meiser, T. (2020). Pseudocontingency inference and choice: The role of information sampling. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Advance online publication. https://doi.org/10.1037/xlm0000840

The present thesis focuses on *pseudocontingency* effects in judgments and choices. In the main text, I clarify when pseudocontingencies might occur and how they can be measured in memory-based judgments, I discuss mathematical models that claim to be able to capture the effect, and provide original evidence for the role of information sampling.

In the first manuscript (Bott et al., in press), we use a hierarchical multinomial processing tree (MPT) model to analyze biased guessing based on inferred pseudocontingencies. Furthermore, we validate the substantive interpretation of the model's guessing parameters. For this purpose, we exemplify a new approach to validating MPT model parameters by testing their convergent validity and discriminant validity in a nomological network. We thereby highlight how cognitive modeling in experimental psychology can profit from its combination with psychometric assessment of relevant constructs (cf. *cognitive psychometrics*, Batchelder, 1998, 2010; Riefer et al., 2002).

In the second manuscript (Bott et al., 2020), we discuss empirical and theoretical limitations of the *Rule of Succession* as normative account of pseudocontingencies. As an alternative, we demonstrate that the *Bayesian Marginal Model* successfully captures pseudocontingency effects found in the literature. Moreover, we illustrate that the model can also readily make new predictions yet to be tested empirically.

While pseudocontingencies are claimed to be highly successful in real-world environments, empirical research has mainly focused on attributing a passive observer role to individuals. In order to investigate the phenomenon in a more naturalistic setting, the third manuscript (Bott & Meiser, 2020) examines the role of free information sampling in pseudocontingency effects and thereby sheds light on differences between passive information intake and active information search.

All in all, these three manuscripts increase insights into the cognitive processes involved in (pseudo-)contingency inference. Moreover, they illustrate that the combination of experimental psychology and statistical modeling is fruitful for gaining a better empirical *and* theoretical understanding of cognitive processes involved in psychological phenomena.

1 Perceived Associations Between Events as Basis for Judgments and Decision Making

Are all BMW drivers road hogs? Does your favorite sports team always win when you are wearing the team's jersey? Or are you always late when taking the train? Indeed, individuals often perceive associations between people, events, or behaviors based on past observations or experiences; for example, between taking the train versus the bike and being late versus on time. As a result, individuals make later judgments and decisions accordingly. For instance, you will avoid taking the train, but instead go by bike to be on time, you will try to wear your team's jersey during all their games as a good-luck charm, or the police carrying out traffic checks will not randomly pick cars to check, but BMW vehicles will be checked particularly often. Such perceived associations may be correct and may lead to informed decisions, however, they may as well be incorrect. In this thesis, I will focus on one specific class of effects in contingency inference, which is referred to as *pseudocontingencies*.

These pseudocontingencies are inferences of a contingency between two variables from the variables' marginal frequencies (e.g., Fiedler et al., 2009). Consider the subtables in Table 1 which lists joint frequencies and marginal frequencies of two binary variables X and Y each taking the values '0' and '1'. The marginal frequencies show that '0' is the most frequent value in both variables. Building on the example from above, this could translate into a person who takes the train more often than the bike (options $X = 0$ and $X = 1$)) and who is late more often than on time (outcomes $Y = 0$ and $Y = 1$). Based on a pseudocontingency, being late would thus be expected to be more likely when taking the train as compared with taking the bike. On an abstract level, empirical evidence suggests that individuals tend to infer an association between the variables X and Y, such that the frequent value of Y is judged to be more likely given the frequent value of X than given the infrequent value of X.

Even if the train is taken most often and being late is more frequent than being on time, however, the genuine contingency between means of transportation and punctuality may indeed be positive (i.e., being late is more likely when taking the train), but could instead also be zero or negative (i.e., being late is actually equally likely or more likely when taking the bike). In order to mathematically quantify an association between two binary variables, their joint frequencies (k_{ij} for $Y = i$ and $X = j$) have to be

Table 1: Example Contingency Tables Depicting Joint Frequencies and Marginal Frequencies of the Binary Variables X and Y

	$\phi_{XY} = 1.00$			$\phi_{XY} = .00$			$\phi_{XY} = -.46$		
	$X = 0$	$X = 1$		$X = 0$	$X = 1$		$X = 0$	$X = 1$	
$Y = 0$	24	0	24	16	8	24	12	12	24
$Y = 1$	0	12	12	8	4	12	12	0	12
	24	12		24	12		24	12	

taken into account. One way of quantifying a contingency is through the Phi coefficient ϕ with $-1 < \phi < 1$:

$$\phi = \frac{k_{00}k_{11} - k_{01}k_{10}}{\sqrt{(k_{00}+k_{01})(k_{10}+k_{11})(k_{00}+k_{10})(k_{01}+k_{11})}} \tag{1}$$

Another way of quantifying the association is by comparing the probability of one possible value of Y conditional on the two possible values of X as in the Δp rule:

$$\Delta p = P(Y = 0|X = 0) - P(Y = 0|X = 1) \tag{2}$$

In both cases, the value $\phi = 0$ or $\Delta p = 0$ indicates that there is no contingency between X and Y (see second subtable in Table 1). ϕ and Δp will be positive/negative when there is a positive/negative association between X and Y (see first and third subtable in Table 1). So why do individuals perceive an association between the variables anyway and expect the most frequent events to disproportionately co-occur as well as the infrequent events?

In the following chapters, I will discuss opposing accounts regarding the emergence of pseudocontingencies. Furthermore, I will present new empirical evidence on the difference between passive information intake and active information search for pseudocontingency inference. Whereas early proposals of pseudocontingencies assumed a memory advantage for infrequent events as compared with frequent events, I show that there is no such difference in recognition performance to account for pseudocontingencies. Using hierarchical Bayesian multinomial processing tree models I further establish that pseudocontingencies manifest in reconstructive guessing processes (Manuscript I, Chapter 2.2.1). Moreover, I will explain that pseudocontingencies are inferred on the basis of statistically inappropriate information (i.e., marginal frequencies), wherefore they are often described as illogical. Yet, I provide evidence that pseudocontingencies can be captured by a computational model that reconstructs contingencies based on marginal frequencies following the norms of Bayesian updating (Manuscript II, Chapter 2.3.2). In Chapter 3, I will shed light on the role of active information sampling in relying on marginal frequencies to infer a (pseudo-)contingency (Manuscript III).

2 Opposing Accounts of Pseudocontingencies

2.1 Memory for Distinct Events

Early research focused on the effect of inferring an association between X and Y when the actual contingency is zero, which is today still known as *illusory correlation*. In their seminal work, Hamilton and Gifford (1976) investigated erroneous impression formation about majority groups and minority groups based on skewed samples. Participants read statements about individuals that were members of either Group A or Group B and who either showed desirable behavior or undesirable behavior. Similar to the stimulus distribution in the second subtable of Table 1, Group A was the majority group and most individuals were described as showing desirable behavior (Hamilton & Gifford, 1976, Experiment 1). Importantly, group membership and desirability of the behavior were uncorrelated ($P(desirable|A) = P(desirable|B) = .69$, with $\Delta p = .00$). Yet, participants estimated the probability of desirable behavior in Group A to be higher than in Group B ($\hat{p}_A = .66$ and $\hat{p}_B = .56$, with $\Delta\hat{p} = .10$).

The traditional explanation of such effects proposes differential processing of frequent and infrequent events (Hamilton & Gifford, 1976). It is assumed that infrequent events (e.g., Group B and undesirable behavior) are more distinct than frequent events (e.g., Group A and desirable behavior), especially so when two rare events are paired. Due to this paired distinctiveness, rare events are more salient and are paid more attention to during encoding. Thus, they are processed more deeply and should be more easily available from memory during retrieval in a judgment phase. It is further assumed that as a consequence, the co-occurrence of the two rare events is overestimated, resulting in the perception of a contingency where there is none (e.g., a contingency between group membership and desirability).

2.1.1 Multinomial Processing Tree Models Testing Increased Memory for Distinct Events

Initial evidence in favor of the distinctiveness-based account was given by Hamilton et al. (1985) showing that in a free recall, more items were recalled which represented the co-occurrence of infrequent events (e.g., undesirable behaviors of Group B) as compared with the other item categories (e.g., desirable behaviors of Group A, undesirable behaviors of Group A, and desirable behaviors of Group B). Analyses of response times further corroborated the distinctiveness-based account: Correct assignments of infrequent events to each other were faster than correct assignments of other combinations (Johnson & Mullen, 1994; McConnell et al., 1994). However, instead of representing memory processes these results of enhanced free recall and faster assignments might as well reflect decision processes driven by expectancies that are in turn based on the inferred contingencies.

In order to explicitly test the memory advantage for doubly infrequent events as proposed by the distinctiveness-based account, more refined experimental techniques and modeling techniques have been proposed. In so-called *source-monitoring* experiments, participants have to discriminate between previously presented target items and new distractor items. Additionally, they have to remember the source of the items (e.g., whether a presented behavioral statement described a member of Group A or a member of Group B). Data from such recognition tests allow for model-based analyses using *multinomial processing tree* (MPT) models to disentangle memory processes from guessing processes (Batchelder & Riefer, 1990; Bayen et al., 1996). MPT models are statistical models that account for observed response frequencies (e.g., frequency of Group A responses) by estimating the probabilities of latent cognitive processes jointly contributing to that frequency data. In the case of source-monitoring experiments, MPT models are used to estimate three types of processes involved: item memory, source memory, and guessing (Bayen et al., 1996; Bröder & Meiser, 2007). The example MPT model depicted in Figure 1 assumes that the individual responses (e.g., Group A responses, Group B responses, and New responses) are each the result of qualitatively distinct processes. Given the example of group membership and type of behavior, the model assumes that participants recognize a target item as old and a distractor item as new with probability D. Alternatively, an item is not recognized with the complementary probability $1 - D$. If item memory fails, participants guess that an item was previously presented with probability b or that it is a new item with probability $1 - b$. The parameter d^{group} denotes the probability of correctly remembering the group as source attribute of a remembered target item. If group membership is not remembered with probability $1 - d^{group}$, the model assumes that a participant guesses Group A with the probability g^{group} and Group B with the complementary probability of $1 - g^{group}$. Thus, a correct response, for instance Group A, may be the result of correctly remembering an item and its source, of only remembering an item but guessing its source, or may even be the result of merely guessing.

According to the distinctiveness-based account (wrongly) inferring a contingency be-

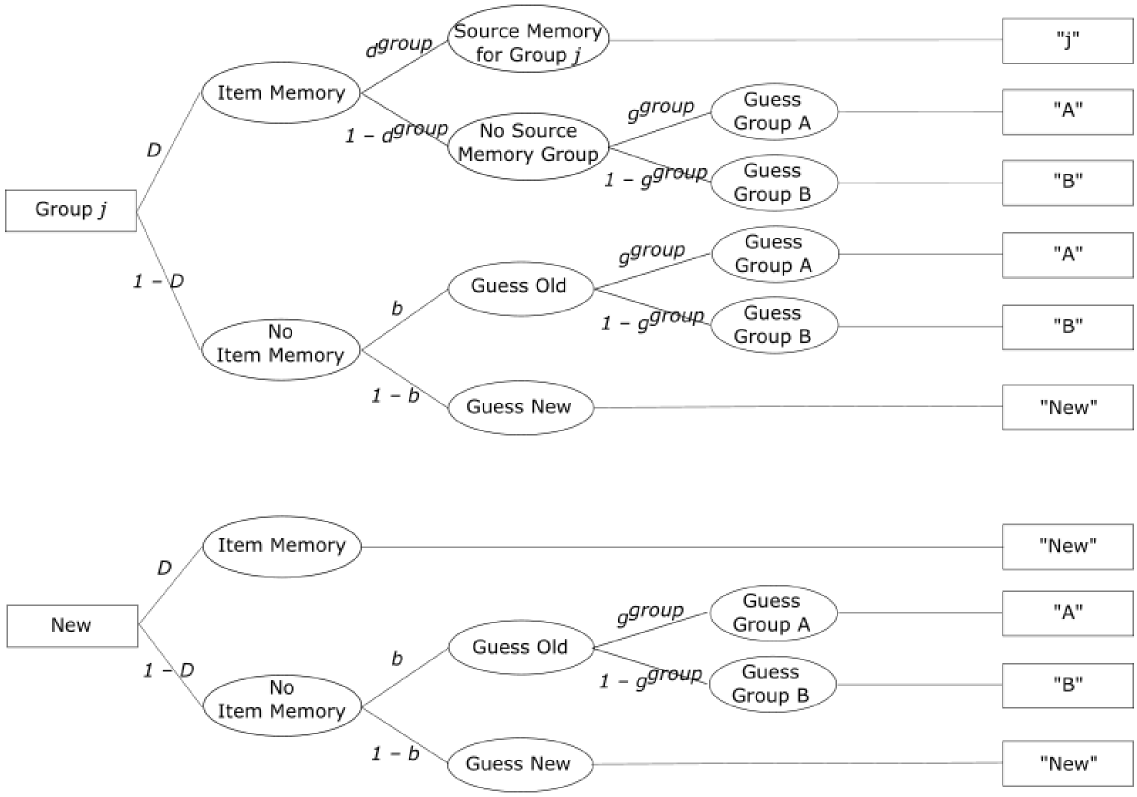

FIGURE 1: Multinomial processing tree model of source monitoring for two sources: group membership $j \in A, B$. Distractor items are referred to as *New*. D = probability to recognize an item; d^{group} = probability to remember Group j; b = probability to guess that an item is old; g^{group} = probability to guess Group A.

tween group membership and type of behavior is based on an enhanced memory for doubly infrequent events, in the example, thus the result of enhanced memory for undesirable behavior performed by a member of Group B. Therefore, the MPT model is fit specifying separate sets of parameters for desirable and undesirable behaviors. Furthermore, studies tested whether recognition performance and source memory differed between Group A and Group B. Yet, contrary to the predictions based on paired distinctiveness, source-monitoring experiments using MPT models showed that there was no difference in recognition performance given behaviors describing members of Group A versus Group B and no differential source memory, but only a general difference in memory for desirable versus undesirable behaviors (e.g., Klauer & Meiser, 2000; Meiser, 2003; Meiser & Hewstone, 2001).

2.2 Pseudocontingency Heuristic

An alternative explanation known as *pseudocontingency heuristic* proposes that individuals infer a contingency based on the observed *marginal frequencies* instead of (specific) joint frequencies. Broadly speaking, given that there is one frequent value per variable, individuals are assumed to heuristically infer that the frequent observations per variable are associated with each other (e.g., Group A and desirable behavior) and that the rare observations per variable disproportionately co-occur (e.g., Group B and undesirable behavior). The pseudocontingency heuristic can thus account not only for illusory correlations, but for a much broader scope of contingency inferences: Pseudocontingencies may occur independent of the joint frequencies and thus independent of the true contingency or even when joint frequencies are not presented. For instance, Fiedler and Freytag (2004) found that inferred associations reflected the skewness found in the marginal frequencies when only they were presented. Furthermore, pseudocontingencies are inferred when the actual contingency is zero (illusory correlation; e.g., Eder et al., 2011; Fiedler et al., 1993; Meiser & Hewstone, 2006) and when the actual contingency is consistent with the inference based on marginal frequencies (e.g., first subtable in Table 1; Fiedler, 2010). Researchers even demonstrated that pseudocontingency inferences may override genuine contingencies (e.g., third subtable in Table 1; Fleig et al., 2017; Meiser et al., 2018; Vogel et al., 2013). Moreover, the pseudocontingency heuristic is further supported by results demonstrating that the combination of the two infrequent values per variable (e.g., the co-occurrence of undesirable behavior and Group B) does not even need to be observed for a pseudocontingency to be inferred (Fleig et al., 2017; Meiser & Hewstone, 2004; Meiser et al., 2018). These results especially are incompatible with the distinctiveness-based account.

Furthermore, pseudocontingencies have been shown to increase in strength with decreasing working-memory capacity (Eder et al., 2011), increasing salience of the marginal frequencies as compared with joint frequencies (Meiser et al., 2018), and the more attention is paid to the marginal frequencies (Fleig et al., 2017). The less likely joint frequencies are used to infer a contingency, the more likely pseudocontingencies are inferred on the basis of marginal frequencies. As a consequence, many experimental studies use more complex scenarios including a third context variable (e.g., desirable versus undesirable behaviors describing members of Group A and Group B who live in Town X and Town Y). Taking the example from the beginning, this would translate into considering the frequencies of being late versus on time when taking the train versus the bike, for instance, not only in Germany, but additionally considering those frequencies when in the Netherlands. Table 2 illustrates one such stimulus distribution. Typically, two contrasting contexts are used: the two focal variables' marginal frequencies are skewed within one context and co-vary across contexts. By implication, they are each associated with the context variable. In the example, in Germany, individuals might be late three times as frequently as being on time and they might take the train more frequently than going by bike, while in the Netherlands, both might be the other way around. Using trivariate stimulus distributions is assumed to increase cognitive load and to increase salience of skewed marginal frequencies, thus increasing the strength of pseudocontingency effects (Fiedler & Freytag, 2004). According to the pseudocontingency heuristic, in the example, the frequent means of transportation and the frequent outcome are associated with each other as well as the infrequent means of transportation and the infrequent outcome, for Germany and the Netherlands. Thus, the same contingency is predicted resulting in a preference for taking the bike in Germany as well as in the Netherlands, when trying to be on time.

Taken together, the pseudocontingency heuristic assumes that contingencies are inferred on the basis of marginal frequencies: joint frequencies are estimated based on the marginal frequencies. Put differently, it is assumed, in principle, that individuals reconstruct joint frequencies by marginal frequencies. Accordingly, illusory correlations, as a special case of pseudocontingencies, do not result from a memory bias, but are instead assumed to be inferred on the basis of perceived statistical regularities (i.e., marginal frequencies).

Coming back to source-monitoring experiments, pseudocontingencies may nevertheless play a role in source-*guessing* processes, whenever source memory for an item fails: instead of randomly guessing, individuals may match their response probabilities to perceived item-source contingencies (e.g., Bayen & Kuhlmann, 2011; Klauer & Wegener, 1998). This perceived contingency does not necessarily correspond to the genuine item-source contingency in the stimulus distribution, but may reflect a pseudocontingency. When participants, for example, do not remember whether a desirable behavior was shown by a member of Group A or by a member of Group B, they will guess group membership based on the inferred (pseudo-)contingency between group membership and type of behavior. In line, Klauer and Meiser (2000) and Meiser and Hewstone (2004) showed that the probability to guess Group A was higher given a desirable behavior than given an undesirable behavior, when the pseudocontingency based on observed marginal frequencies favored Group A over Group B. They interpreted the source-guessing parameters as reflecting evaluative biases based on inferred pseudocontingencies. Taking biased guessing parameters as evidence of inferred pseudocontingencies, however, renders the interpretation of the MPT model's guessing parameters theoretically important. Thus, one goal of the first manuscript (Bott et al., in press) was to investigate the construct validity of guessing parameters in a source-monitoring MPT model.

2.2.1 Parameter Validation in Hierarchical Multinomial Processing Tree Models

Bott, F., M., Heck, D. W., & Meiser, T. (in press). Parameter validation in hierarchical MPT models by functional dissociation with continuous covariates: An application to contingency inference. *Journal of Mathematical Psychology*.

Traditionally, MPT models are fitted to frequency data aggregated across participants and are validated by testing the selective influence of specific experimental manipulations on model parameters (Erdfelder et al., 2009; Hütter & Klauer, 2016). Discrete realizations of factors that are assumed to influence specific cognitive processes involved in the given task (e.g., item memory, source memory, or guessing) are realized as experimental factors in order to test their selective effects on specific parameters. However, using aggregated frequency data to fit an MPT model implies that there are no substantial differences between items or participants (parameter homogeneity). Yet, in case of parameter heterogeneity, parameter estimates may be biased (Klauer, 2006; Smith & Batchelder, 2010). Thus, more recently, hierarchical extensions of MPT models have been developed to explicitly account for parameter heterogeneity between participants (Klauer, 2010; Matzke et al., 2015; Smith & Batchelder, 2010). Besides addressing concerns regarding biased parameter estimates, hierarchical MPT models additionally offer a new possibility of parameter validation. Instead of using discrete variables, selective covariations of interindividual differences in model parameters with continuous variables can be tested. With regard to the source-monitoring experiments in pseudocontingency research, for example, direct measures of group evaluations can be used to test selective covariations with guessing parameters; other model parameters should not be related.

In our manuscript (Bott et al., in press), we thus demonstrated this novel validation technique and validated the guessing parameters of an MPT model as indicators of perceived (pseudo-)contingencies. Using data from an experiment on pseudocontingency inference in a trivariate scenario, we tested for convergent parameter validity and discriminant parameter validity by means of selective covariations with direct evaluative judgments. As in previous studies, participants read statements about desirable versus undesirable behaviors shown by members of Group A versus Group B living in Town X versus Town Y. Using the event distribution as depicted in Table 2, Group A and desirable behavior were the frequent events given Town X, but the infrequent events given Town Y, and vice versa for Group B and undesirable behavior. In a learning phase, the total of 48 behavioral statements was presented in random order. The subsequent test phase included direct evaluative judgments regarding the groups within towns as well as a source-memory test. To assess the participants' evaluations of Group A

TABLE 2: Example 2 x 2 x 2 Contingency Table Depicting Joint Frequencies and Marginal Frequencies of the Binary Variables *Group Membership* and *Type of Behavior* Within Town X and Town Y

	Town X			Town Y		
Behavior	Group A	Group B		Group A	Group B	
Desirable	12	6	18	0	6	6
Undesirable	6	0	6	6	12	18
	18	6		6	18	

and Group B in Town X and Town Y, participants rated the groups on trait attributes and estimated the groups' frequencies of undesirable behavior per town. In the source-memory test, participants decided for each of the 48 target items and additional 48 distractor items whether it was OLD or NEW. If judged as OLD, they were additionally asked whether it occurred in Town X or Town Y and whether it described a member of Group A or Group B (for more details see Bott et al., in press). As we used a trivariate stimulus distribution, the classical source-monitoring MPT model depicted in Figure 1 cannot be used. Instead, an extension of MPT models to account for crossed source dimensions (e.g., town of residence and group membership; Meiser & Bröder, 2002) was applied.

Figure 2 illustrates this source-monitoring MPT model for crossed source dimensions which separately estimates source memory for the source dimension town of residence and source memory for the source dimension group membership (Meiser & Bröder, 2002). As in the classical source-monitoring MPT model, the parameter D denotes the probability that an item (here, behavioral statement) is correctly recognized as presented during a learning phase (OLD) or as a new item (NEW). Item memory fails with the complementary probability $1 - D$. In that case, participants guess OLD with probability b or guess NEW with probability $1 - b$. The parameters d again measure the probability of remembering the source of an item. In the case of crossed source dimensions, however, the MPT model differentiates between source memory for town of residence (d^{town}) and source memory for group membership (d^{group} and e^{group}). If the town of residence is not remembered with probability $1 - d^{town}$, participants guess TOWN X with probability g^{town} and TOWN Y with probability $1 - g^{town}$. The parameter d^{group} describes the probability of remembering group membership, if the town of residence is remembered. In case source memory for town of residence fails, group membership is remembered with probability e^{group}. Thus, the model reflects the assumption that source memory for one source dimension may differ depending on source memory for the other dimension. Differences between d^{group} and e^{group} thereby correspond to stochastic dependency in multidimensional source memory. This assumption can be tested via model comparison by investigating whether model fit decreases when constraining d^{group} and e^{group} to be equal. The order of the source-memory parameters reflect the sequence of retrieval processes in the memory test as they are specified in the same order as the source dimensions are prompted. If the source dimension group membership is not remembered with probability $1 - d^{group}$ or $1 - e^{group}$, a behavioral statement assigned to Town X is attributed to Group A with probability $g^{group}_{|X}$ or to Group B with probability $1 - g^{group}_{|X}$. Likewise, the parameter $g^{group}_{|Y}$ denotes the probability of guessing Group A when an item was attributed to Town Y. The conditional specification of the guessing parameters for group membership again mirrors the order of responses in the source-memory test and allows for differential guessing tendencies that may reflect inferred contingencies (Meiser, 2003; Meiser & Hewstone, 2004).

According to the pseudocontingency heuristic, given the stimulus distribution in Table 2, participants are expected to perceive Group A more favorably than Group B within both towns. The overall evaluation of Town X should be more positive than the evaluation of Town Y in line with the genuine contingency between town of residence and type of behavior apparent in the marginal frequencies of desirable and undesirable behaviors within each town. Thus, participants are expected to estimate the probability of undesirable behavior to be higher given Town Y as compared with Town X and to be higher given Group B than given Group A within each town. If guessing parameters do indeed reflect evaluative judgments based on inferred (pseudo-)contingencies, they should at least differ depending on the statement's desirability. Therefore, separate sets of parameters were estimated for desirable behaviors and for undesirable behaviors. Moreover, as we aimed to validate g^{town}, $g^{group}_{|X}$, and $g^{group}_{|Y}$ as indicators of perceived (pseudo-)contingencies, we tested the parameters' covariations with those evaluations of towns and groups within towns (i.e., the direct measures of source evaluation: frequency estimates of undesirable behavior and trait ratings). Evaluations of Town X and Town Y should selectively relate to the guessing probability g^{town}, while group evaluations should selectively relate to g^{group}. More precisely, evaluations of Group A versus Group B within Town X should selectively co-vary with $g^{group}_{|X}$ and evaluations of Group A versus Group B given Town Y should only co-vary with $g^{group}_{|Y}$. By contrast, parameters D, d, and e measure genuine memory performance and should not be influenced by the evaluations of towns and groups.

Following the hierarchical latent-trait approach (Klauer, 2010), all MPT parameters were modeled as random effects to account for differences between individuals. The hierarchical MPT models were estimated using Bayesian methods as implemented in TreeBUGS (Heck et al., 2018). In order to establish the guessing parameters' validity via functional dissociation, we included continuous variables (i.e., direct measures of source evaluation) as predictors of model parameters. For this purpose, we z-standardized

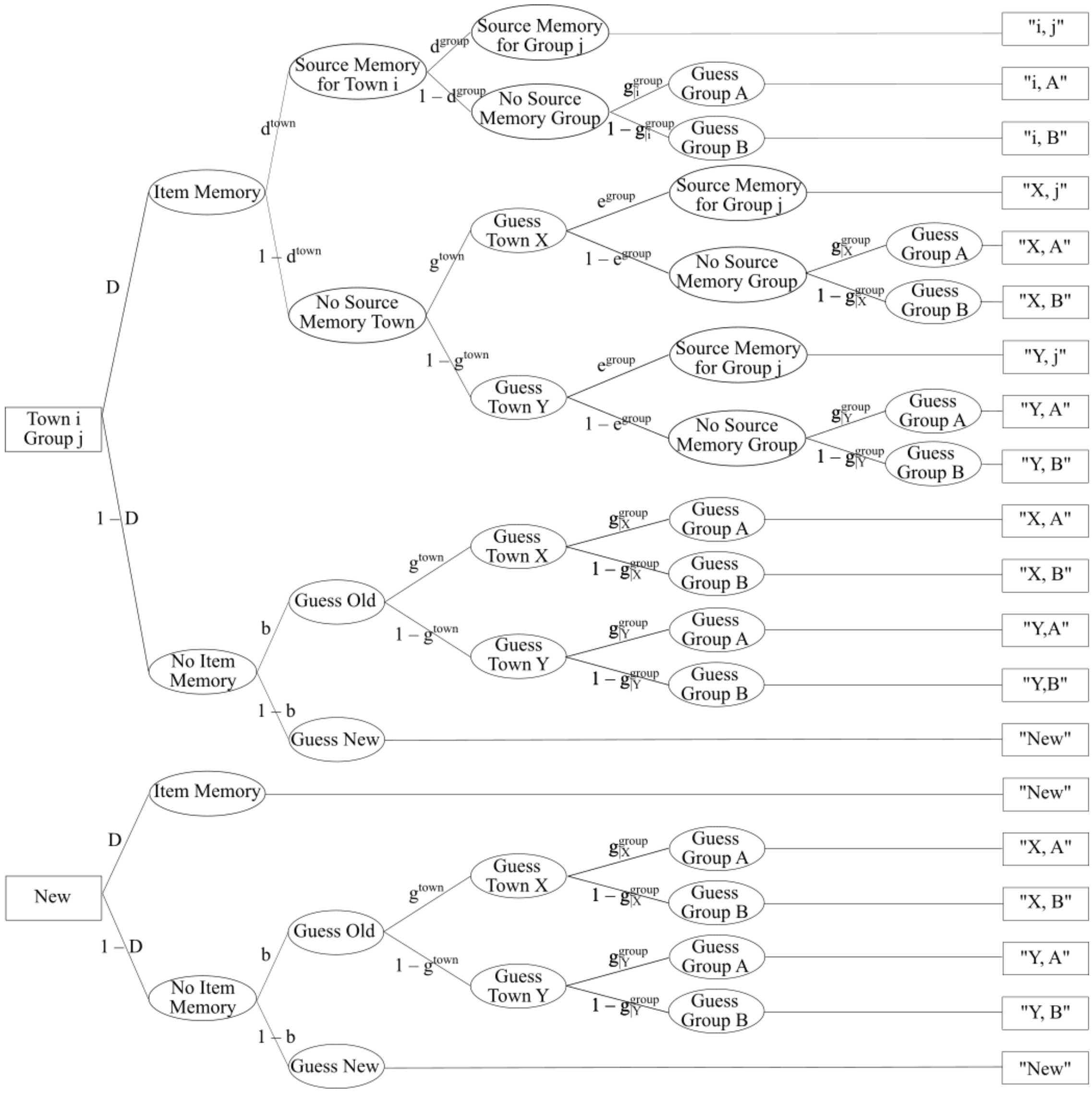

FIGURE 2: Multinomial processing-tree model of source monitoring for 2 × 2 crossed sources: town of residence $i \in X, Y$ and group membership $j \in A, B$. Distractor items are referred to as *New*. D = probability to recognize an item; d^{town} = probability to remember Town i; d^{group} = probability to remember Group j given Town i was recollected; e^{group} = probability to remember Group j given Town i was not remembered; b = probability to guess that an item is old; g^{town} = probability to guess Town X; $g^{group}_{|X}$ = probability of guessing Group A given assignment of item to Town X; $g^{group}_{|Y}$ = probability of guessing Group A given assignment to Town Y.

the predictors and assumed a normal distribution $\mathcal{N}(\mu = 0, \sigma = \sqrt{2}/2)$ as a prior distribution of each slope parameter. Figure 3 illustrates the implied prior predictive distributions when regressing an MPT parameter on a covariate and testing a directional hypothesis ($\mathcal{H}_1 : \beta > 0$ or $\mathcal{H}_1 : \beta < 0$) or a non-directional hypothesis ($\mathcal{H}_1 : \beta \neq 0$).

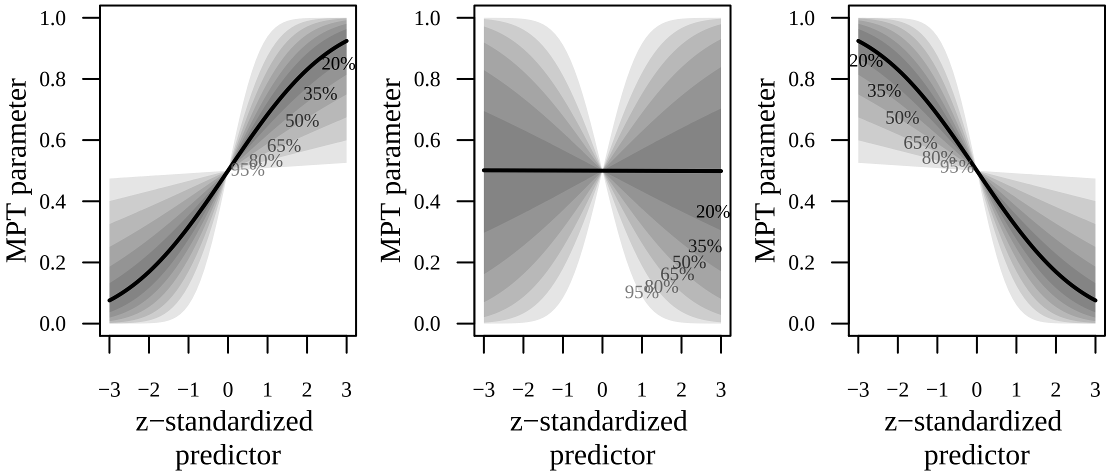

FIGURE 3: Prior predictive distribution for a probit regression of an MPT parameter on a z-standardized covariate implied by the prior distribution of the slope parameter, $\beta \sim \mathcal{N}(\mu = 0, \sigma = \sqrt{2}/2)$. Solid black lines depict the median, while the areas shaded in gray depict prediction intervals. The left and right panels show the prior predictive distribution given a directional hypothesis ($\mathcal{H}_1 : \beta > 0$ and $\mathcal{H}_1 : \beta > 0$, respectively). The middle panel shows the prior predictive distribution given a non-directional hypothesis ($\mathcal{H}_1 : \beta \neq 0$).

Replicating previous MPT analyses of aggregate data (Meiser & Hewstone, 2004, 2006), the probability of guessing Town X when source memory failed was higher for desirable behaviors than for undesirable behaviors. Likewise, the probability of guessing Group A was higher for desirable behaviors than for undesirable behaviors for those statements that were assigned to Town X as well as for those assigned to Town Y. Furthermore, these differences in guessing probabilities paralleled differences in the direct measures of town evaluation and group evaluation (for details see Bott et al., in press).

To validate the guessing parameters, we first analyzed the relations of interindividual differences in source-guessing parameters with interindividual differences in direct source evaluations (e.g., differences in evaluations between Town X and Town Y, differences in evaluations between Group A and Group B given Town X, and differences in evaluations between Group A and Group B given Town Y; Bott et al., in press). Indeed, differences in evaluations of Town X and Town Y selectively predicted the guessing probability g^{town} for desirable behaviors as well as for undesirable behaviors and did not predict any other source-guessing parameter. Evaluative differences between Group A and Group B in Town X specifically related to the guessing parameter $g^{group}_{|X}$ for desirable and undesirable behaviors, while differences between groups in Town Y specifically predicted $g^{group}_{|Y}$ given desirable behaviors and given undesirable behaviors. No other source-guessing parameters were related, each. Additionally, there were no

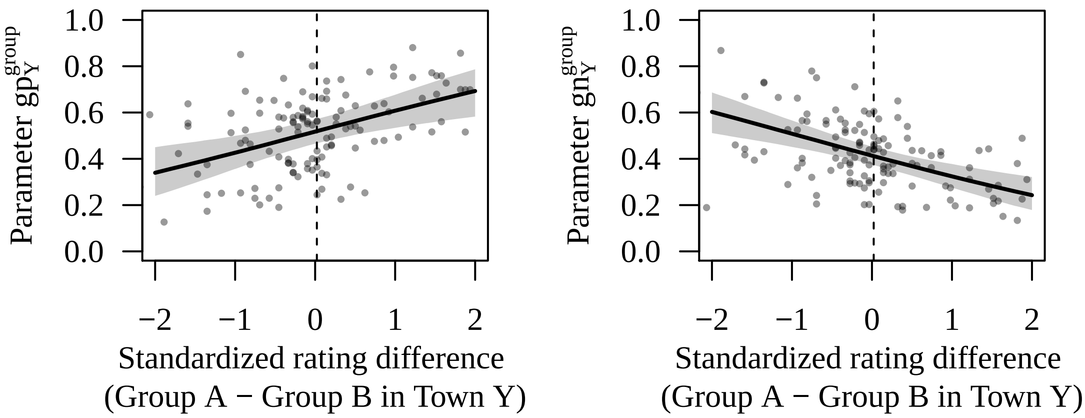

FIGURE 4: Regression of the probability to guess Group A given an item assigned to Town Y ($g_{|Y}^{group}$) on rating differences between Group A and Group B in Town Y. The left panel shows the covariation of $gp_{|Y}^{group}$ for desirable behaviors, while the covariation of $gn_{|Y}^{group}$ is shown in the right panel. The solid lines depict the posterior median of the prediction function, while the 95% Bayesian credibility intervals are shown in gray. Individual-level predictions are depicted by the gray points. Vertical dashed lines show the group mean of the covariate.

covariations with item-memory parameters, source-memory parameters, or the probability to guess OLD. Representative of these results, Figure 4 displays the regression of the $g_{|Y}^{group}$ parameters for desirable behaviors and undesirable behaviors ($gp_{|Y}^{group}$ and $gn_{|Y}^{group}$, respectively) on the difference in evaluative trait ratings between Group A and Group B in Town Y. As can be seen, the more favorably Group A was evaluated, the more likely desirable behaviors were assigned to Group A and the less undesirable behaviors were assigned to Group A. Taken together, the results corroborate the interpretation of source-guessing parameters as reflecting perceived item-source contingencies. In the example, the source-guessing parameters can indeed be interpreted as mirroring evaluative biases based on inferred contingencies and pseudocontingencies, in the case of town evaluations and group evaluations, respectively.

Going beyond, the framework of hierarchical MPT models allows for parameter validation in an even wider context of continuous measures of psychological constructs. Thereby, they offer means of validation and theory testing in a broader nomological network. As an illustration, we investigated relations of the MPT parameters with cognitive performance measures from a standardized cognitive assessment battery (Bott et al., in press). Other researchers relied on this or a similar approach, for instance, to estimate associations between personality traits and dishonest behavior (Heck et al., 2018), moral judgments (Kroneisen & Heck, 2019), or environmental preferences (Klein et al., 2017), or to estimate associations of memory with fluid and crystallized intelligence (Michalkiewicz et al., 2018). To further establish the construct validity of MPT model parameters in pseudocontingency research, we analyzed the specificity of covariations between MPT parameters and cognitive performance measures that may be relevant for source monitoring and contingency inference: Pseudocontingencies are inferences drawn from mathematically inappropriate information, yet they are based on statistical regularities (i.e., marginal frequencies). Thus, interindividual differences in the ability to apply mathematical skills (i.e., quantitative thinking as measured by the INSBAT, Arendasy et al., 2009) as well as interindividual differences in the ability to combine information and to recognize regularities (i.e., fluid intelligence) might explain interindividual differences in inferring pseudocontingencies and thereby in source-guessing parameters. By contrast, memory parameters in the MPT model should be related to the cognitive ability to remember and recognize information. Nonetheless, in exploratory analyses, we did not find evidence for these cognitive performance measures as distinct predictors of parameters of the MPT model for crossed source dimension (for more details see Bott et al., in press).

To recap, the inference of pseudocontingencies has been shown to influence (evaluative) judgments and subsequent decisions. Using hierarchical Bayesian MPT modeling, we demonstrated that inferred pseudocontingencies are related to specific guessing processes that are indeed reflective of non-randomly attributing items to sources (Bott et al., in press). Moreover, there is more corroborative evidence in favor of the inference of contingencies based on marginal frequencies as proposed by the pseudocontingency heuristic than for the distinctiveness-based account: Contingencies mirroring the skewness in marginal frequencies are inferred independent of whether the combination of infrequent events is observed (e.g., Meiser et al., 2018). In addition, MPT analyses did not show enhanced memory for doubly infrequent events as predicted by the distinctiveness-based account (Klauer & Meiser, 2000; Meiser, 2003; Meiser & Hewstone, 2004).

Although pseudocontingencies can be deceptive and research is largely focused on decision problems that intend to induce wrong contingency inferences based on pseudocontingencies, they do have considerable adaptive value: Indeed, marginal frequencies do not determine joint frequencies. Yet, marginal frequencies do restrict the range of possible values, especially so when they are skewed (Duncan & Davis, 1953). Accordingly, Kutzner et al. (2011b) could demonstrate that pseudocontingencies often succeed in capturing the sign of the actual contingency. Therefore, they are sometimes described as "logically unwarranted but smart and useful" (Fiedler et al., 2013, p. 328) which rests on the idea that pseudocontingencies are highly successful in real-world environments as compared with psychological experiments (cf. also *ecological rationality*, Gigerenzer, 2019; Goldstein & Gigerenzer, 2002). Do pseudocontingencies thus reflect normatively unjustified information processing?

2.3 Normative Accounts

2.3.1 Regression-To-The-Mean Accounts

Fiedler (1996) and Dougherty et al. (1999) proposed exemplar-based memory models, the BIAS model and the MINERVA-DM model, to account for pseudocontingencies. Both assume that memory consists of a database of events experienced in the past. Each event or item is stored in memory as a copy of it, called memory trace. However, the memory traces are decayed to varying degrees and do not match the original event exactly. To access memory and form judgments or make decisions, the memory traces will be compared with the original event of interest (memory probe) which additionally requires aggregation across the memory traces (e.g., summation or averaging). In general, aggregation cancels out error variance. Thus, the aggregation across memory traces will be more similar to the original event the more traces are aggregated, The models propose that judgments and decisions are proportional to that similarity. Accordingly, given the stimulus distribution of Town X in Table 2, for example, they predict that the similarity with desirable behavior will be higher given memory traces of Group A than for memory traces of Group B, due to the fact that there are more observations of Group A. The models would thus predict pseudocontingencies. However, they are only able to do so, when asked for an estimation regarding the frequent value of the variable of interest (e.g., desirable behavior). Following the same mechanism, both models would also predict higher probability estimates of undesirable behavior for Group A as compared with Group B, again due to the fact that there are more memory traces for Group A. The latter prediction, however, is neither in line with predictions based on the pseudocontingency heuristic, nor with experimental results (e.g., Eder et al., 2011; Hamilton & Gifford, 1976; Klauer & Meiser, 2000; Meiser, 2003).

While exemplar-based memory models are not successful in accounting for pseudocontingencies, Costello and Watts (2019) recently suggested that probability judgments are the result of updating prior information as per the *Rule of Succession* (Laplace, 1820/1951). Given a sample of N observations and given that one out of two events occurred k times in that sample, the Rule of Succession states that the normative estimation for the probability of that event is

$$\hat{p} = \frac{k+1}{N+2} = \frac{2 \cdot .50 + N \cdot \frac{k}{N}}{2+N} \tag{3}$$

The re-parameterization of the Rule of Succession as weighted average of the probability

.50 and the observed probability of the event in the sample ($\frac{k}{N}$), highlights the assumption that a prior probability of .50 (e.g., that an event occurs at chance level) is updated based on the observed probability. The value 2 in the equation can be interpreted as one pseudocount per possible event, as if each event has been observed once prior the actual sample, wherefore the prior probability of $\frac{1}{2}$ will be updated. The larger the observed sample, the stronger the weight that is given to the observed probability as compared to the prior probability. Thus, with an increasing number of observations N, the inferred probability will be less regressive towards $\frac{1}{2}$. Costello and Watts (2019) demonstrated that wrongly inferring a contingency between group membership and type of behavior is the byproduct of applying the Rule of Succession to a sample in which marginal frequencies are skewed, but the actual contingency is zero: When applying the Rule of Succession to each group separately, the estimate for the majority group will deviate less from the observed probability $\frac{k}{N}$, while the estimate for the minority group will be more regressive towards $\frac{1}{2}$. Thus, the estimates, for instance for the majority group A and the minority group B, will conform to the inequality $\hat{p}_A > \hat{p}_B > .50$ when $p > .50$, or $\hat{p}_A < \hat{p}_B < .50$ when $p < .50$ as predicted by the pseudocontingency account. The estimates will imply a contingency when none is present.

However, applying the Rule of Succession to the example of Town X in Table 2, already discloses that the Rule of Succession has trouble capturing effects found in data on pseudocontingency inference at large. In Town X, $\hat{p}_A = \frac{12+1}{18+2} = .65$ and $\hat{p}_B = \frac{6+1}{6+2} = .875$ (with $\Delta\hat{p} = -.225$) result as probability estimates for desirable behavior. According to the Rule of Succession, the estimate $\hat{p}$ is expected to be higher for the *minority* group as compared with the majority group in line with the genuine group-behavior contingency (i.e., $\Delta p = -.33$). However, this prediction contradicts the inference based on marginal frequencies (i.e., a positive pseudocontingency). Figure 5 displays the probability predicted by the Rule of Succession as a function of the number of observations and different genuine event probabilities. As can be seen, with an increasing number of observations N, the predicted probability approaches the genuine event probability p. If event probabilities for the majority group A and the minority group B are equal and larger .50 (e.g., $p_A = p_B = .67$), the Rule of Succession indeed produces higher probability estimates for the majority group A (e.g., $\hat{p}_A = .653$ for $n_A = 18$) as compared with the minority group B (e.g., $\hat{p}_B = .639$ for $n_B = 9$) and thus pseudocontingencies. Yet, if the actual probability for the minority group is higher than for the majority group (e.g., $p_A = .67$, $p_B = .83$, with $\Delta p = -.16$), predictions derived from the Rule of Succession contradict experimental results of higher probability estimates for Group A than for Group B (e.g., Fleig et al., 2017; Meiser & Hewstone, 2004; Meiser et al., 2018) and thus the predictions by the pseudocontingency heuristic (e.g., $\hat{p}_A = .653$ for $n_A = 18$ and $\hat{p}_B = .770$ for $n_B = 9$).

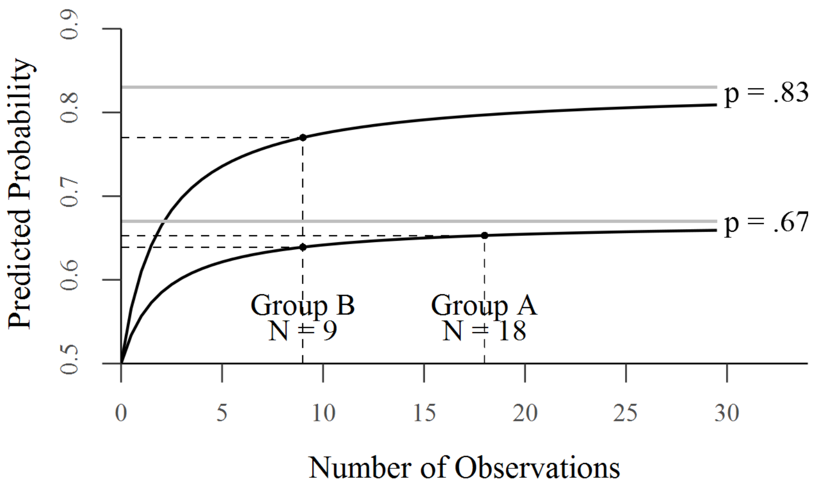

FIGURE 5: Probability of the frequent outcome (e.g., desirable behavior) predicted by the Rule of Succession as a function of number of observations (N) from Group A and Group B.

Another shortcoming of the Rule of Succession, as of any regression-to-the-mean account with a prior probability of $\frac{1}{2}$, is exactly that $\frac{1}{2}$-midpoint which cannot be crossed over: when $p > .50$, the predicted probability estimates will fall into the range $.50 \leq \hat{p} \leq 1$; when $p < .50$, the predictions will be $.50 \geq \hat{p} \geq .00$. However, inspection of participants' probability judgments show that given the infrequent group, the estimated probabilities often do not only reflect an underestimation, but a reversal in which event is frequent (i.e., $\hat{p} < .50$ for the *frequent* outcome; e.g., Bulli & Primi, 2006; Eder et al., 2011; Hamilton & Gifford, 1976; Meiser & Hewstone, 2006; Meiser et al., 2018).

Although neither the exemplar-based accounts nor the Rule of Succession discussed are able to account for pseudocontingencies, in the second manuscript (Bott et al., 2020) we argue that pseudocontingencies nevertheless may reflect an instance of bounded rationality (Simon, 1990). We demonstrate that a computational model, the Bayesian Marginal Model originally proposed by Klauer (2015), predicts pseudocontingencies as the output of updating a prior probability with observed marginal frequencies.

2.3.2 Bayesian Marginal Model

Bott, F., M., Kellen, D., & Klauer, K. C. (2020). *Normative accounts of illusory correlations.* Invited revision submitted to *Psychological Review*.

As previously discussed, pseudocontingencies are inferred on the basis of associating skewed marginal frequencies (e.g., Fiedler et al., 2009). As marginal frequencies do not determine, but only restrict the range of possible cell frequencies and contingencies

(Duncan & Davis, 1953), the Bayesian Marginal Model jointly considers all possible cell distributions and thus contingencies which are in line with the observed marginal frequencies.

Like any Bayesian account, the Bayesian Marginal Model captures prior beliefs about the contingency by a prior distribution. Given the assumption that the data in a contingency table follow a multinomial distribution, the prior beliefs are captured by a Dirichlet distribution $\mathcal{D}ir(\boldsymbol{\alpha})$. The Dirichlet distribution is parameterized by a vector of, in our case, four concentration parameters $\boldsymbol{\alpha} = (\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11})$, that can be interpreted as pseudocounts per cell in the contingency table. Figure 6 illustrates prior beliefs about the contingency ϕ as captured by the Dirichlet distribution. When all α_{ij} are equal, we see that the prior distribution is dispersed over the [-1, 1] range and concentrated around zero. Moreover, the stronger the prior belief that there is no contingency (i.e., the higher values of alpha), the more peaked is the prior distribution around zero. The use of prior beliefs that are unbiased (i.e., setting all α_{ij} to the same value) is reasonable for studies in which little to no prior information is available, for instance, when using abstract or novel variables (e.g., groups labeled 'A' and 'B'). However, when a reasoner does belief a priori that some events occur more frequently than others or that there is a non-zero contingency, this can be captured via unequal values α_{ij}: the respective pseudocounts can be increased in value, for instance, if participants a priori expect that one group is more frequent than the other, that desirable behavior is more common than undesirable behavior, or both (see lower panels in Figure 6 from left to right, respectively). Note that only the last panel of Figure 6 illustrates an example of not expecting a zero contingency a priori.

The prior beliefs are then updated with new observations, however, only with the observed marginal frequencies. In the Bayesian Marginal Model, this implies, that the prior beliefs in terms of the pseudocounts $\boldsymbol{\alpha}$ are updated with each possible cell distribution that is consistent with the observed marginal frequencies. As an example see Table 3 that reports all possible joint frequencies under the marginal frequencies of the example of Town X in Table 2. The contingencies (quantified by ϕ) are positive in five of the seven possible sets of joint frequencies and their range is asymmetric $-.33 \leq \phi \leq 1.00$. In this example, the prior beliefs would be updated with seven possible sets of joint frequencies, each.

The resulting posterior belief is a mixture distribution over the updated sets of joint frequencies, thus a mixture of the posterior Dirichlet distributions with $\boldsymbol{\alpha}^* = \boldsymbol{\alpha} + \boldsymbol{k}$, where $\boldsymbol{k}$ corresponds to the respective joint frequencies consistent with the observed marginal frequencies (e.g., $\boldsymbol{k} = (12, 6, 6, 0)$). In the Bayesian Marginal Model, each posterior Dirichlet distribution is weighted as a function of the marginal frequencies and the prior. The first row of Figure 7 illustrates the weights associated with the possi-

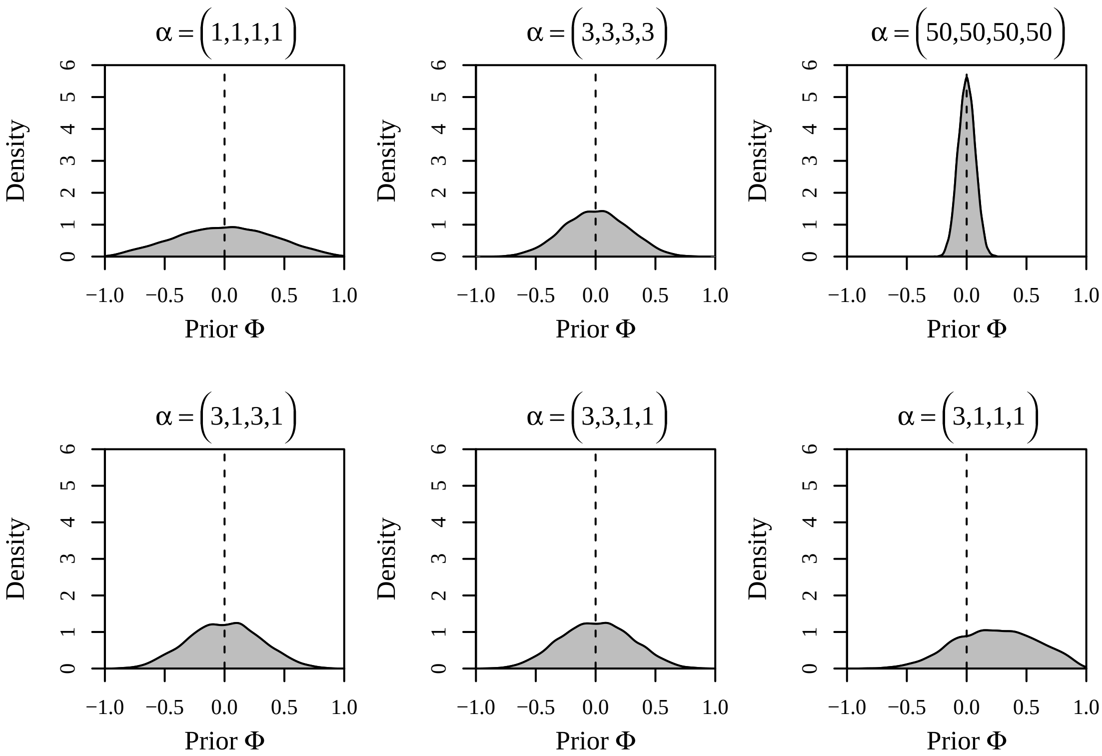

FIGURE 6: Example prior Dirichlet distributions of ϕ.

TABLE 3: Possible Joint Frequencies (k_{ij} for $Behavior = i$ and $Group = j$) and Associations (ϕ) Given the Marginal Frequencies of the Town X Example in Table 2 ($n_+ = 18$, $n_- = 6$, $n_A = 18$, $n_B = 6$)

Index	k_{+A}	k_{+B}	k_{-A}	k_{-B}	ϕ
1	12	6	6	0	-0.33
2	13	5	5	1	-0.11
3	14	4	4	2	0.11
4	15	3	3	3	0.33
5	16	2	2	4	0.56
6	17	1	1	5	0.78
7	18	0	0	6	1.00

Note. True $\phi = -0.33$.

ble joint frequencies given the example marginal frequencies of Town X in Table 2. As can be seen, under the prior of $\boldsymbol{\alpha} = (1, 1, 1, 1)$, each set of joint frequencies is equally weighted. Moving away from this prior would lead to differential weighting: the larger the value of α the more weight will be attributed to the joint frequencies that are similar to the prior beliefs. For instance, assuming $\alpha = (50, 50, 50, 50)$ would result in weighting the moderate joint frequencies most, like Index 2 and 3 in Table 3. Thus, most weight will be attributed to a zero contingency.

The output of the Bayesian Marginal Model is a posterior distribution that will be obtained by taking multiple samples from the Dirichlet distributions. A single sample is taken by first sampling the component Dirichlet distribution (e.g., the Dirichlet distribution corresponding to the joint frequencies identified with Index 2 in Table 3, with $\boldsymbol{\alpha}^* = (\alpha_{00} + 13, \alpha_{01} + 5, \alpha_{10} + 5, \alpha_{11} + 1))$. Secondly, a sample is drawn from that specific Dirichlet distribution. Figure 7 depicts resulting posterior ϕ distributions and corre-

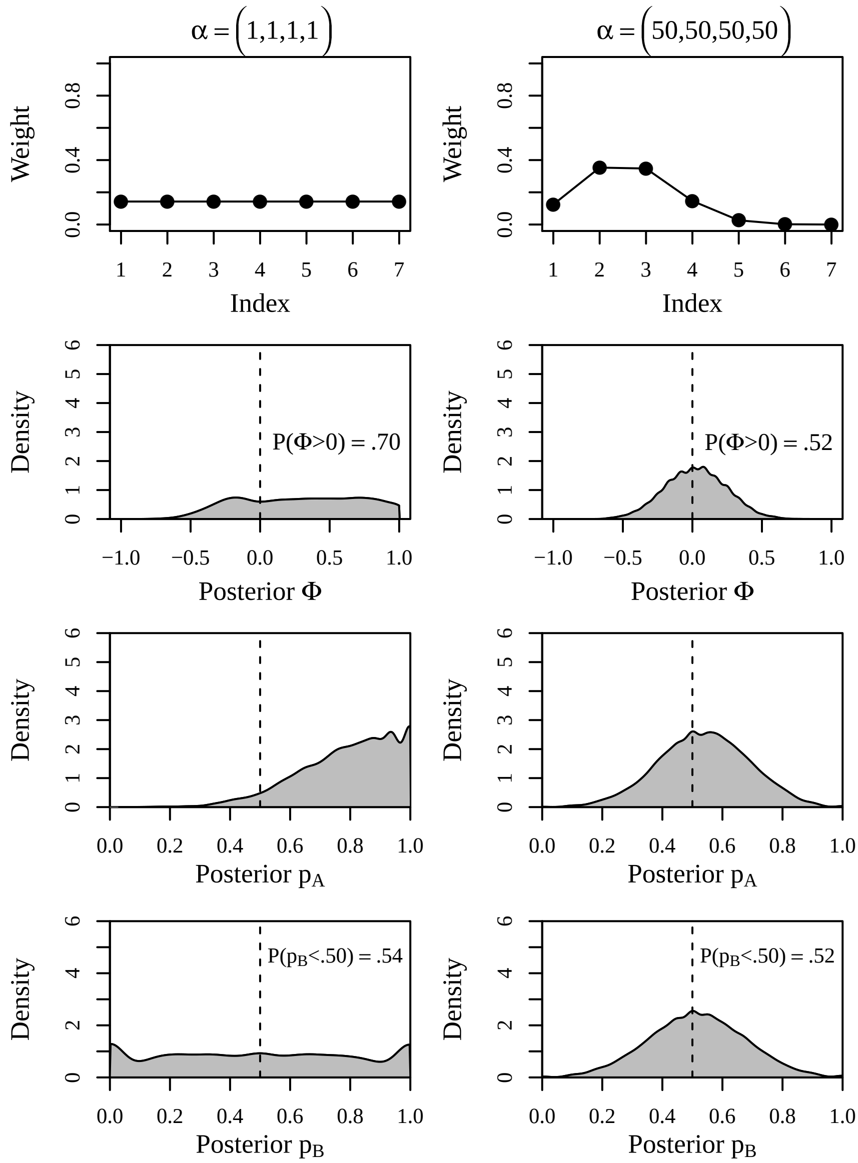

FIGURE 7: Illustration of the Bayesian Marginal Model under different priors given the example of Town X in Table 2.

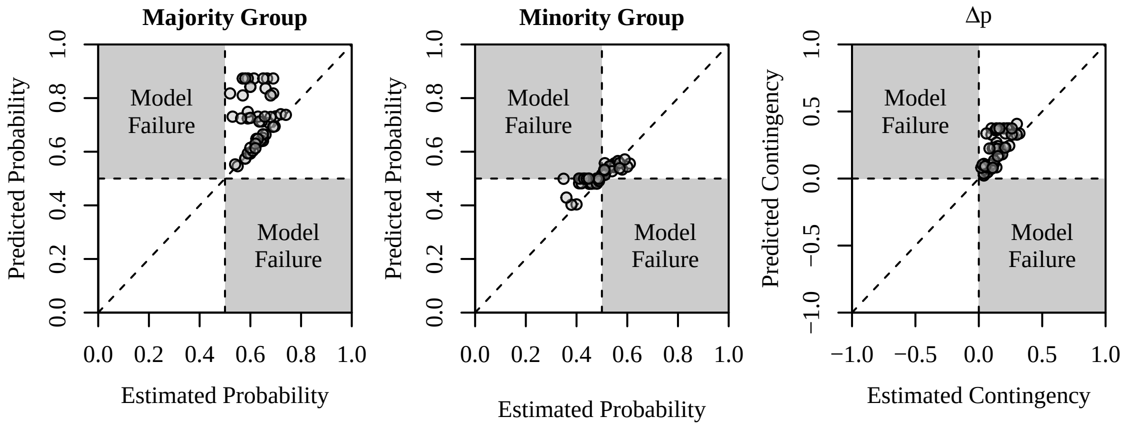

FIGURE 8: Model fit to previously published data. Posterior means are displayed as probabilities predicted by the Bayesian Marginal Model. 'Model failure' corresponds to a qualitative mismatch between participants' mean estimates and the model predictions.

sponding predictions for $\hat{p}_A$ and $\hat{p}_B$. Pseudocontingencies are predicted to be prevalent in the case of $\boldsymbol{\alpha} = (1, 1, 1, 1)$ when the reasoner does not have strong prior beliefs. However, when they do (e.g., $\boldsymbol{\alpha} = (50, 50, 50, 50)$), pseudocontingencies are predicted to be less likely and less strong. Furthermore, note that the model also predicts the occurrence of $\hat{p}_B < .50$; a phenomenon that is problematic for regression-to-the-mean accounts.

To further explore the potential of the Bayesian Marginal Model, we investigated its ability to capture the qualitative patterns of pseudocontingency effects reported in 43 previously published experimental conditions (Bott et al., 2020). As we focused on studies that used artificial groups or options, we used unbiased prior beliefs by restricting α_{ij} to be equal. The results of the model fits are illustrated in Figure 8 and show that the Bayesian Marginal Model can capture the qualitative patterns in the data: The model can not only produce pseudocontingency effects, but can also account for the underestimation of probabilities for the minority group to the point of reversal in which outcome is most frequent (i.e., probability estimates $< .50$ for the minority group).

In the analyses, we provided original evidence that the Bayesian Marginal Model constitutes a normative reconstruction of the pseudocontingency heuristic that successfully predicts pseudocontingencies, including probability reversals, as a result of belief updating using marginal frequencies (Bott et al., 2020). The actual reasoning processes when estimating probabilities or contingencies, however, clearly do not identically correspond to the algorithmic computations of the model, but rather approximate the process of updating and taking samples from a probability distribution. In sum, the pseudocontingency account proposed by Fiedler and colleagues, which assumes that individuals infer contingencies based on marginal frequencies, can thus be corroborated not only empirically, but also normatively.

3 When Information Sampling May or May Not Evoke Biases

Bott, F., M., & Meiser, T. (2020). Pseudocontingency inference and choice: The role of information sampling. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Advance online publication. https://doi.org/10.1037/xlm0000840

In Chapter 2, I discussed pseudocontingencies as an inference from statistical regularities based on predetermined event sequences that are reflected, among other things, in reconstructive guessing processes. On this basis, in the following chapter, I will elaborate on the role of individuals' self-directed information sampling in pseudocontingency inference as compared with information predetermined and presented by the experimenter.

For a long time, research on pseudocontingencies has been mostly limited to studies concerned with impression formation or stereotype formation. Using various scenarios, researchers have demonstrated that associations between, for example, group membership and behavior or between gender and hobbies, abilities, or fields of study are inferred from skewed marginal frequencies and may override genuine contingencies (e.g., Fiedler, 2010; Fiedler et al., 2007; Meiser & Hewstone, 2004). More recently, it has also been shown that pseudocontingencies influence decision making and may lead to sub-optimal choice behavior even when choices are directly relevant for the reasoner and in voluntary choices (Fleig et al., 2017; Kutzner et al., 2011a; Meiser et al., 2018).

In pseudocontingency experiments, scenarios are usually created in which skewed marginal frequencies are observed that should result in the inference of a pseudocontingency contradicting the genuine contingency. Observing one behavior or outcome more frequently than another and additionally encountering one group or option more frequently than another may be the case in some scenarios in everyday life. However, at least equally often, before forming a judgment or making a decision, reasoners have to gather information themselves. Therefore, the question arises whether pseudocontingencies may actually also be crucial, when the available information is not predetermined.

Investigating self-directed information sampling, researchers often realize a *free sampling paradigm*: Participants are free to sample any number of single events by choosing between multiple - typically two - options in any order they desire. None of the draws is incentivized, except for one final choice to be made at the end. Results on free information sampling are inconclusive, though, with regard to whether options are sampled equally often (e.g., Hau et al., 2008; Hertwig et al., 2004) or not (e.g., Lejarraga et al., 2012; Wulff et al., 2018). Moreover, most studies on information sampling use options that only vary in their outcomes' magnitudes and outcomes' probabilities, but not in their mean payoff. Those that do use options with different mean payoffs reported a preference for the option with the higher mean payoff in the final choice (e.g., Hertwig et al., 2006; Hilbig & Glöckner, 2011). In contrast, according to the pseudocontingency heuristic, a preference for the option with the higher mean payoff is only predicted if that option was most frequently observed when positive outcomes are most common or if it was the least frequent option when losses are most common (cf. e.g., Fiedler et al., 2009). Consequently, it is unclear whether skewed marginal frequencies are sampled. By implication, it is an open question whether freely drawn samples give rise to the inference of a pseudocontingency which may or may not contradict the genuine contingency (i.e., does not align with the options' underlying payoff structure). Free information sampling might even foster genuine contingency assessment independent of the skewness in the sampled marginal frequencies as it may draw the focus of attention away from the marginal frequencies towards the joint frequencies (cf. Fleig et al., 2017).

3.1 The Effect of Self-Determined Information Samples on Pseudocontingency Inferences

For this reason, we investigated information sampling behavior and its impact on pseudocontingency inference manifested in probability judgments and choices (Bott & Meiser, 2020). All experiments implemented the following general procedure. The experimental task was to first assess and later trade two shares (option X1 and option X2). They were traded at two times of a day (context C1 and context C2) and yielded either a gain (outcome Y1) or a loss (outcome Y2) with different probabilities also varying over the time of day. The underlying event distribution is displayed in Table 4. After a fixed number of learning trials, participants traded the shares themselves while aiming at maximizing their gains. Finally, they were asked, among other things, to estimate the options' winning probabilities given each context.

Implementing this decision scenario in the traditional experimental paradigm with predetermined learning trials, we presented the individual events summarized in Table 4 in random order during the learning phase: Participants observed option X1 more frequently than option X2 and gains more frequently than losses in context C1 and vice versa given context C2. Replicating previous findings we found pseudocontingency

TABLE 4: 2 x 2 x 2 Contingency Table Representing Joint Frequencies and Marginal Frequencies Underlying the Experiments in Bott and Meiser (2020)

	Context C1			Context C2		
Outcome	Option X1	Option X2		Option X1	Option X2	
Y1 (gain)	15	8	23	2	9	11
Y2 (loss)	9	2	11	8	15	23
	24	10		10	24	

Note. $\phi_{YX|C1} = -.17$; $\phi_{YX|C2} = -.17$;
$\phi_{XC} = .41$; $\phi_{YC} = .35$; $\phi_{YX} = .00$

inferences in self-relevant choice: Participants preferred option X1 in both contexts as indicated by higher choice frequencies and higher winning probability estimates (see Table 5 and Bott & Meiser, 2020, Experiment 1).

In comparison, implementing a free sampling paradigm, when participants sample information themselves during the learning phase, they are free to decide for each learning trial which context and which option to observe. The outcomes were drawn by the experimental program in accordance with the event frequencies in Table 4. We analyzed information sampling behavior by specifying a multivariate generalized linear mixed model to predict information sampling in individual trials as a function of individual previous observations, like the previously sampled option and the previously observed outcome. The analysis revealed that individuals often adopt a sampling strategy of re-sampling an option, especially when it resulted in a positive outcome previously (Bott & Meiser, 2020). Moreover, options more likely resulting in a gain are sampled more frequently than the options with a low winning probability, even though there are no direct consequences of the resulting outcomes during information sampling. Figure 9 summarizes this result by depicting the mean relative frequencies of samples per option

TABLE 5: Relative Frequencies of Choices and Winning Probability Estimates of Options Within Contexts in Experiment 1 in Bott and Meiser (2020)

	Context C1				Context C2			
	Option X1		Option X2		Option X1		Option X2	
	M	SD	M	SD	M	SD	M	SD
Choices	61.33	32.31	38.67	32.31	60.07	31.03	39.93	31.03
Estimates	54.95	20.36	46.95	22.31	42.60	21.32	37.84	19.42

Note. Data in percent conditional on context.

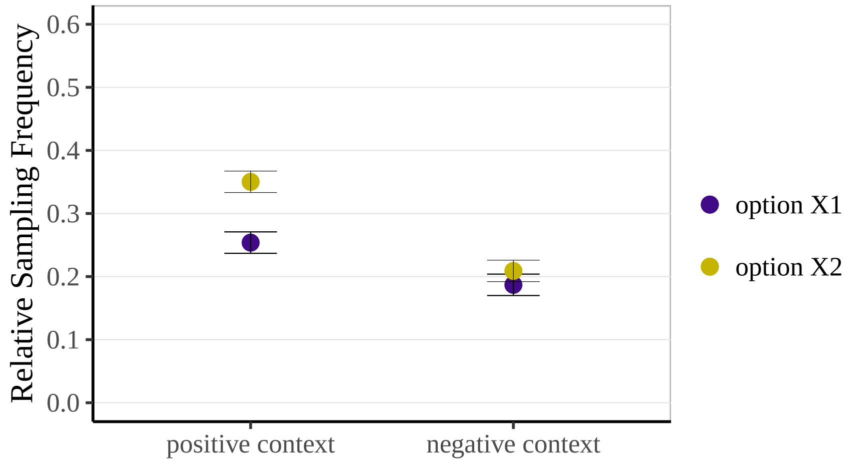

FIGURE 9: Relative frequencies of sampling option X1 versus option X2 within the predominantly positive context C1 versus predominantly negative context C2 in Bott and Meiser (2020, Experiment 4). Note, the options' outcome probabilities corresponded to the event frequencies depicted in Table 4.

within each context: Participants sampled both options within the predominantly positive context C1 more frequently than the options within the predominantly negative context C2. Additionally, option X2 with the higher genuine winning probability was, on average, more frequently sampled within the positive context than option X1; within the negative context both options were sampled roughly equally often.

Furthermore, we analyzed subsequent choices also as a function of the individual samples drawn. Consequential choices following the sampling phase reflected a preference for the most frequently sampled option when gains were the most frequent outcome (i.e., in context C1; Bott & Meiser, 2020). Although this choice behavior seems to be in line with the pseudocontingency heuristic, in most cases, the frequently sampled and subsequently chosen option in the positive context corresponded to the option with the higher underlying winning probability (i.e., option X2). The inference of a pseudocontingency would thus have resulted in the same preference as the inference of the genuine contingency. Moreover, when both options were sampled equally often, the superior option (i.e., option X2) was preferred, as well. For a few cases, nevertheless, self-determined information sampling resulted in sub-optimal choices based on inferred pseudocontingencies. Within the predominantly negative context C2 effects were reduced or even absent, in information sampling as well as later choice. Taken together, the results revealed an asymmetry in choices given contexts of predominantly positive outcomes versus given contexts of predominantly negative outcomes: While there are no differences between options in negative contexts, in positive contexts, choice preferences are already discernible in the sample drawn, independent of whether they reflect genuine contingencies or pseudocontingencies.

Possibly, observing positive outcomes during the learning phase might be reinforcing in itself, even though they have no direct consequences. As a result, individuals may be hedonically motivated to focus on options which they evaluate positively and which they expect to result in positive outcomes. Such effects of evaluation-based preferences in information sampling and choice should be even more discernible in a so called *partial feedback paradigm*. In this paradigm, each sample drawn represents information *and* payoff. Thus, every single draw is incentivized according to the observed outcome. Implementing the experimental design as a partial feedback paradigm, in a more recent experiment, the tendency to re-sample an option that previously resulted in a gain was indeed stronger (Bott, 2019). As a consequence, option X2 was chosen more frequently than option X1 in the predominantly positive context C1. An effect that was additionally stronger during the learning phase as compared with the free sampling paradigm. However, again there was no difference between options given the predominantly negative context C2.

3.2 Asymmetric Pseudocontingency Effects

Similar results regarding the asymmetry between positive contexts and negative contexts have been found in previous research on pseudocontingency effects in impression formation as well as choice behavior (e.g., Fleig et al., 2017; Meiser & Hewstone, 2004; Meiser et al., 2018): Pseudocontingency effects were stronger when mostly positive outcomes were observed, whereas they were reduced or even absent when negative outcomes were most common. Those experiments all used a trivariate scenario, similar to the ones depicted in Tables 2 and 4, in which the overall positivity/negativity (e.g., town or context) were manipulated as within-participant factor. Due to observing contrasting contexts, one possible explanation for the asymmetry may be that the overall positivity/negativity is learned fairly quickly. Assuming that reasoners focus on how to maximize positive experiences (e.g., gains), they might disregard variables within the negative context as less important, as they will generally avoid those anyways. For instance, independent of whether there is a minority group of nice people, towns with a bad reputation are rather avoided as a whole. The results on information sampling discussed before hint at such avoidance: More samples were drawn within the predominantly positive context (Bott, 2019; Bott & Meiser, 2020). Yet, in the given stimulus distribution, this could have also been the natural result of preferring options that result in positive outcomes, independent of the overall positivity/negativity of the context.

To more explicitly test whether attention is allocated to available information depending on the overall positivity/negativity, we examined whether contexts varying in their overall winning probability are sampled to varying degrees (Bott & Meiser, 2020,

Table 6: Relative Frequencies of Choices and Winning Probability Estimates of Options Within Contexts in the Sampling Condition of Experiment 3 in Bott and Meiser (2020)

	Context C1				Context C2			
	Option X1		Option X2		Option X1		Option X2	
	M	SD	M	SD	M	SD	M	SD
Choices	66.54	33.16	33.46	33.16	50.75	32.98	49.25	32.98
Estimates	55.82	21.71	41.56	22.13	39.69	20.05	41.46	21.64

Note. Data in percent conditional on context.

Experiment 3). In this experiment, we again realized context C1 and context C2 as a within-participant factor and participants were only free to decide for each learning trial which context to observe. The options and outcomes were drawn by the experimental program in accordance with the event frequencies summarized in Table 4. We analyzed the probability to sample the predominantly positive context C1 in an individual learning trial by means of a generalized linear mixed model including the previously sampled context and the previously observed outcome as predictors. Overall, we did not find differences between the predominantly positive context C1 and the predominantly negative context C2 with regard to sampling frequency. Even though the order of observations was not random, on average, the information per context did thus not differ from the exact frequencies depicted in Table 4. Accordingly, we replicated results obtained with predetermined learning trials: Pseudocontingencies were inferred and used as basis for judgments and decision making, although leading to sub-optimal choices. Yet again, the pseudocontingency effect was asymmetric with stronger effects given the predominantly positive context. Mean relative choice frequencies and probability estimates are shown in Table 6.

An alternative explanation for the asymmetry also hinges on the notion that the overall positivity/negativity is learned fairly quickly before making inferences about the variables within a context. Prior beliefs about these variables, formed on the basis of the context's positivity/negativity, may thus influence learning. Assuming that reasoners may pay more attention to losses than gains (cf. Yechiam & Hochman, 2013), one could expect stronger effects of prior beliefs in negative scenarios or contexts as compared with positive scenarios. In the Bayesian Marginal Model discussed above, such prior beliefs can be captured via values of $\boldsymbol{\alpha}$ that overweight the frequent outcome relative to other cells (e.g., $\boldsymbol{\alpha} = 2, 2, 1, 1)$). The simulated results in Figure 10 show that expecting the frequent outcome twice as often as the infrequent outcome reduces the expected pseudocontingency effect. Note, that the probability and strength of a pseudocontingency is lower the stronger the prior beliefs. Assuming the prior beliefs are stronger when negative outcomes are more frequent (e.g., context C2), this could produce the asymmetry in the strength of the inferred pseudocontingencies.

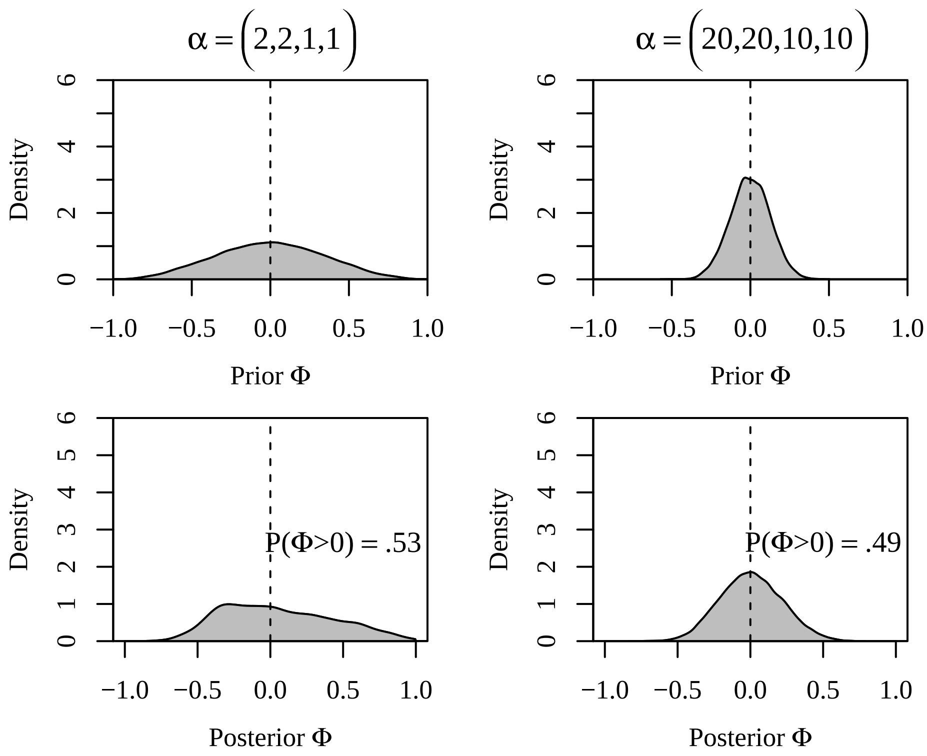

Figure 10: Prior ϕ and posterior ϕ for the example of a total sample of 34 observations comprising 23 occurrences of the frequent outcome and 24 occurrences of the frequent option. The example priors reflect overweighting of the cells corresponding to the frequent outcome relative to the infrequent outcome.

3.3 The Effect of Prior Expectations

Going beyond the analyses and results in Bott and Meiser (2020), beliefs about or evaluations of groups or options probably evolve over the course of time. (Pseudo-) Contingencies inferred based on the current pool of experience may thus already influence further (sampling) behavior. Likewise, it can even be assumed that (pseudo-)contingencies inferred on the basis of passive information, when there are no prior direct experiences with the groups or options, influence future judgments and choices. In order to investigate the effect of prior beliefs experimentally, expectations about the options' winning probabilities, for instance, can be induced prior the learning phase. These expectations may either correspond to the genuine contingency between options and outcomes or correspond to the opposite, incorrect contingency. In order to differentiate between effects on the evaluation of options which can initially be explored without any consequences, and effects on the evaluation of options that have never been directly experienced prior choice, a free sampling paradigm and a partial feedback paradigm can be realized, respectively, as between-participants conditions.

Implementing this design using the stimulus material in Table 4, a new experiment revealed that the options in the predominantly positive context were preferred to the options in the predominantly negative context (Bott, 2019). Figure 11 shows the mean relative choice frequencies in the final eight choices per context in each sampling-expectation condition. When prior expectations corresponded to the true contingency (i.e., higher winning probability of option X2 than option X1), the superior option X2 was chosen more frequently than option X1, during information sampling and in final choices. This difference between options was greater given the predominantly positive context. Even in the case of prior expectations that turn out to be wrong, over the course of trials, participants were able to update evaluations accordingly with the result of at least judging both options within a context as equally good or bad. Interestingly, this pattern of choices was not only revealed when participants were free to explore all choice options without consequences, but also when each choice was incentivized from the outset. The results indicate that individuals can adjust prior and maybe premature evaluations as long as they keep gathering information and experiences.

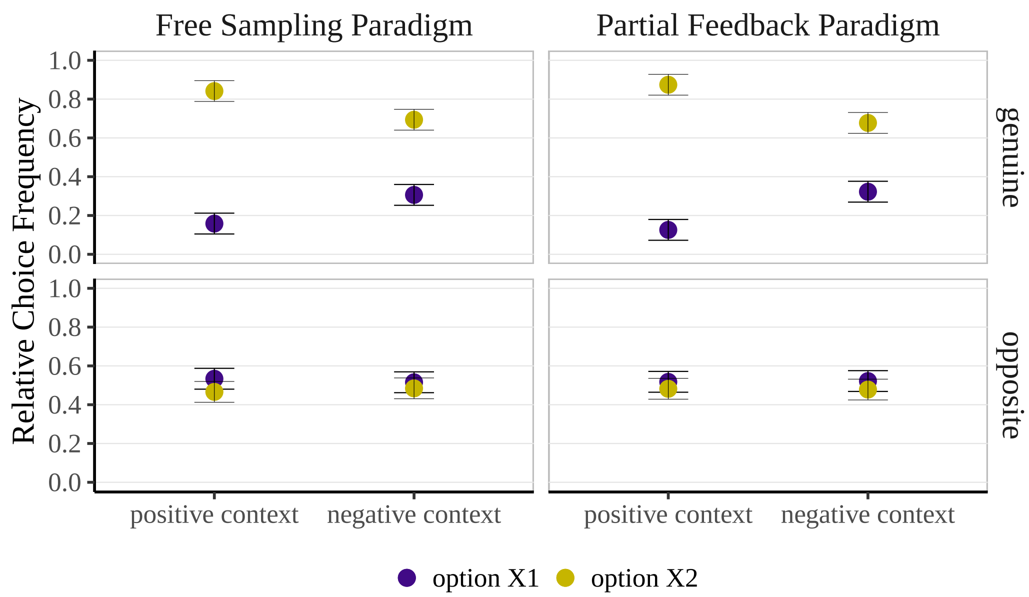

FIGURE 11: Relative choice frequencies for options X1 and X2 per context and per sampling-expectation condition: Participants were randomly assigned to one of two sampling conditions (free vs. partial) and were either told the genuine contingency between options and outcomes or the opposite, incorrect contingency (genuine vs. opposite).

4 Conclusion

Contingency assessment has been widely recognized as essential for everyday life. Thus, not surprisingly, many researchers are interested in the processes involved in contingency learning, may it be in learning of associations between stimuli in conditioning (Rescorla & Wagner, 1972) or between causes and effects in causal reasoning (Cheng & Novick, 1992; Spellman & Mandel, 1999; Waldmann & Holyoak, 1992). Initially, it was assumed that individuals infer contingencies similar to computing Δp by comparing two conditional probabilities. However, various factors may influence which information is used in order to asses contingencies. The environment itself may determine which information is available in the first place and subjective causal models about associations between variables, for instance, may influence which information is considered and how it is used (Waldmann et al., 2006). On the one hand, researchers demonstrated that individuals are able to assess genuine contingencies when considering joint frequencies and are even able to take third variables, like a context variable, into account when being aware of the third variable's moderating role (Cheng & Novick, 1990, 1992; Spellman et al., 2001; Waldmann & Hagmayer, 2001). On the other hand, it has been shown that this is only the case as long as genuine contingency assessment is cognitively manageable. Otherwise, simplifying strategies are employed, like using marginal frequencies instead of pairs of observations to infer a contingency (e.g., Fiedler et al., 2009).

This phenomenon is known as pseudocontingency (e.g., Fiedler et al., 2009). When marginal frequencies are skewed, the pseudocontingency heuristic proposes that individuals associate frequent categories with each other as well as infrequent categories and thereby infer a contingency; when marginal frequencies are uniformly distributed individuals will assume that there is no contingency. Even though pseudocontingencies are inferred on the basis of statistically inappropriate information (i.e., marginal frequencies instead of joint frequencies), they are quite adaptive and useful, for instance, when no co-occurrence information is available. Pseudocontingencies have additionally been shown to facilitate detection of the sign of true contingencies (Kutzner et al., 2011b). Moreover, in the second manuscript (Bott et al., 2020), we presented pseudocontingencies as the expected output of updating prior beliefs about a contingency based on the use of marginal frequencies, while following the norms of Bayesian belief updating. Thus, the pseudocontingency heuristic as well as this Bayesian Marginal Model

(Bott et al., 2020), as normative reconstruction of the pseudocontingency heuristic, assume that contingencies are inferred on the basis of marginal frequencies. This implies that basically joint frequencies are reconstructed by marginal frequencies. While more recent research focuses more on probability judgments and choices, especially research on pseudocontingencies in impression formation asks participants to estimate the frequency of a certain type of behavior (e.g., desirable versus undesirable behavior) given each group to be evaluated. However, in those experiments, participants are given the total number of observations per group (i.e., marginal frequencies of group membership) prior estimation. In order to investigate the reconstruction of joint frequencies (e.g., the frequency of jointly observing Group A and desirable behavior), future research should consider asking for estimates of certain joint frequencies without presenting information on marginal frequencies.

Furthermore, pseudocontingency inferences are traditionally investigated using experimental paradigms that put individuals in the position of a passive observer of information. Over the course of learning trials, predetermined information about the (co-)occurrence of two variables is presented, sometimes in combination with a third, context variable. Using various scenarios, researches demonstrated that when observing skewed marginal frequencies of the two focal variables, individuals heuristically infer that the frequent observations in each variable disproportionately co-occur as well as the infrequent observations per variable (e.g., Fiedler & Freytag, 2004; Fleig et al., 2017; Meiser & Hewstone, 2004; Meiser et al., 2018; Vogel et al., 2013). Even when it comes to assigning events to each other, in case memory fails, individuals have been shown to guess co-occurrences based on inferred pseudocontingency (e.g., Klauer & Meiser, 2000; Meiser & Hewstone, 2004, 2006). In the first manuscript (Bott et al., in press), we corroborate this finding of guessing processes reflecting inferred pseudocontingencies via parameter validation in a hierarchical MPT model disentangling memory processes from guessing processes.

Extending empirical research on pseudocontingency inference, in a series of four experiments, we investigated active information search as compared with passive observation of information and its role in pseudocontingency inference (Bott & Meiser, 2020). We replicated earlier findings that skewed marginal frequencies are indeed used when inferring a contingency. However, the results also revealed that the more freely information are sampled by individuals themselves, the more discernible are later preferences in the sample drawn and the more likely do they align with underlying genuine contingencies between the variables at hand. Moreover, effects in information sampling and choice differ in their strength depending on the observations' predominant valence: While negative experiences are avoided if possible, positive or positively evaluated events are preferred during information sampling and later choices, even when information sampling has no direct consequences, but could be utilized for exploration of all events possible.

The discrepancy in effects depending on the overall positivity/negativity of events may be traced back to different prior expectations or beliefs that additionally may be updated to varying degrees given a positive versus negative context, especially when contrasting contexts are observed. In the Bayesian Marginal Model (Bott et al., 2020; Klauer, 2015), prior beliefs about the presence of a contingency between variables are captured via a prior probability distribution. In its current form, however, it only allows to model a 2 x 2 design (e.g., two options and two outcomes), but a 2 x 2 x 2 design with contrasting contexts only by considering them in parallel: Estimates are computed per context. The integration of two 2 x 2 contingency tables to jointly consider all eight cells, instead of two sets of four cells, poses an important challenge yet to be tackled. Furthermore, to test the hypothesis that the asymmetry between predominantly positive contexts and predominantly negative contexts is due to the valence of events, future research should explore differences in judgments and choices based on inferred (pseudo-)contingencies between valenced stimuli and non-valenced stimuli as well as differences to stimuli to which a valence will be allocated at a later point in time.

To conclude, the present thesis corroborates the notion that pseudocontingencies are inferred on the basis of marginal frequencies and are considered a solid basis for judgments and choices. Demonstrating that the pseudocontingency heuristic can be reconstructed by a model following the norms of Bayesian belief updating and reporting original evidence, I highlight that despite the prefix "pseudo", pseudocontingencies are not necessarily wrong inferences. Instead, they often align with genuine contingencies, especially when reasoners make the effort of actively searching for further information and of gathering new experiences.

5 Bibliography

Arendasy, M., Hornke, L. F., Sommer, M., Häusler, J., Wagner-Menghin, M., Gittler, G., Bognar, B., & Wenzl, M. (2009). *INSBAT – Intelligenz-Struktur-Batterie*. Mödling, Austria, Schuhfried.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, *97*, 548–564. https://doi.org/10.1037/0033-295X.97.4.548

Batchelder, W. H. (1998). Multinomial processing tree models and psychological assessment. *Psychological Assessment*, *10*(4), 331–344. https://doi.org/10.1037/1040-3590.10.4.331

Batchelder, W. H. (2010). Cognitive psychometrics: Using multinomial processing tree models as measurement tools (S. E. Embretson & S. E. Embretson (Ed), Eds.). In S. E. Embretson & S. E. Embretson (Ed) (Eds.), *Measuring psychological constructs: Advances in model-based approaches.* Washington, DC, US, American Psychological Association. https://doi.org/10.1037/12074-004

Bayen, U. J., & Kuhlmann, B. G. (2011). Influences of source-item contingency and schematic knowledge on source monitoring: Tests of the probability-matching account. *Journal of Memory and Language*, *64*(1), 1–17. https://doi.org/{10.1016/j.jml.2010.09.001}

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning Memory and Cognition*, *22*(1), 197–215. https://doi.org/10.1037/0278-7393.22.1.197

Bott, F. M. (2019). [Unpublished data from experiments on information sampling and pseudocontingency inference]. University of Mannheim.

Bott, F. M., Heck, D. W., & Meiser, T. (in press). Parameter validation in hierarchical mpt models by functional dissociation with continuous covariates: An application to contingency inference. *Journal of Mathematical Psychology*.

Bott, F. M., Kellen, D., & Klauer, K. C. (2020). *Normative accounts of illusory correlations*. Manuscript submitted for publication.

Bott, F. M., & Meiser, T. (2020). Pseudocontingency inference and choice: The role of information sampling. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Advance online publication. https://doi.org/10.1037/xlm0000840

Bröder, A., & Meiser, T. (2007). Measuring source memory. *Zeitschrift für Psychologie/Journal of Psychology*, *215*(1), 52–60. https://doi.org/10.1027/0044-3409.215.1.52

Bulli, F., & Primi, C. (2006). Illusory correlation and cognitive processes: A multinomal model of source-monitoring. *Review of Psychology*, *13*(2), 95–102.

Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology*, *58*(4), 545–567. https://doi.org/10.1037/0022-3514.58.4.545

Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, *99*(2), 365–382. https://doi.org/10.1037/0033-295X.99.2.365

Costello, F., & Watts, P. (2019). The rationality of illusory correlation. *Psychological Review*, *126*(3), 437–450. https://doi.org/10.1037/rev0000130

Dougherty, M. R., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review*, *106*(1), 180–209. https://doi.org/10.1037/0033-295X.106.1.180

Duncan, O. D., & Davis, B. (1953). An alternative to ecological correlation. *American Sociological Review*, *18*(6), 665–666. https://doi.org/10.2307/2088122

Eder, A. B., Fiedler, K., & Hamm-Eder, S. (2011). Illusory correlations revisited: the role of pseudocontingencies and working-memory capacity. *Quarterly journal of experimental psychology*, *64*(3), 517–532. https://doi.org/10.1080/17470218.2010.509917

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Journal of Psychology*, *217*(3), 108–124. https://doi.org/10.1027/0044-3409.217.3.108

Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environments. *Psychological Review*, *103*(1), 193–214. https://doi.org/10.1037/0033-295X.103.1.193

Fiedler, K. (2010). Pseudocontingencies can override genuine contingencies between multiple cues. *Psychonomic Bulletin & Review*, *17*(4), 504–509. https://doi.org/10.3758/PBR.17.4.504

Fiedler, K., & Freytag, P. (2004). Pseudocontingencies. *Journal of Personality and Social Psychology*, *87*(4), 453–467. https://doi.org/10.1037/0022-3514.87.4.453

Fiedler, K., Freytag, P., & Meiser, T. (2009). Pseudocontingencies: An integrative account of an intriguing cognitive illusion. *Psychological Review*, *116*(1), 187–206. https://doi.org/10.1037/a0014480

Fiedler, K., Freytag, P., & Unkelbach, C. (2007). Pseudocontingencies in a simulated classroom. *Journal of Personality and Social Psychology*, *92*(4), 665–677. https://doi.org/10.1037/0022-3514.92.4.665

Fiedler, K., Kutzner, F., & Vogel, T. (2013). Pseudocontingencies: Logically unwarranted but smart inferences. *Current Directions in Psychological Science 2013*, *22*(4), 324–329. https://doi.org/10.1177/0963721413480171

Fiedler, K., Russer, S., & Gramm, K. (1993). Illusory correlations and memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 111–136. https://doi.org/10.1006/jesp.1993.1006

Fleig, H., Meiser, T., Ettlin, F., & Rummel, J. (2017). Statistical numeracy as a moderator of (pseudo)contingency effects on decision behavior. *Acta Psychologica*, *174*, 68–79. https://doi.org/10.1016/j.actpsy.2017.01.002

Gigerenzer, G. (2019). Axiomatic rationality and ecological rationality. *Synthese*. https://doi.org/10.1007/s11229-019-02296-5

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*(1), 75–90. https://doi.org/10.1037/0033-295x.109.1.75

Hamilton, D. L., Dugan, P. M., & Trolier, T. K. (1985). The formation of stereotypic beliefs: Further evidence for distinctiveness-based illusory correlations. *Journal of Personality and Social Psychology*, *48*(1), 5–17. https://doi.org/10.1037/0022-3514.48.1.5

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology*, *12*(4), 392–407. https://doi.org/10.1016/S0022-1031(76)80006-6

Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description-experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making*, *21*(5), 493–518. https://doi.org/10.1002/bdm.598

Heck, D. W., Thielmann, I., Moshagen, M., & Hilbig, B. E. (2018). Who lies? A large-scale reanalysis linking basic personality traits to unethical decision making. *Judgment and Decision Making*, *13*(4), 356–371.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*(8), 534–539. https://doi.org/10.1111/j.0956-7976.2004.00715.x

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2006). The role of information sampling in risky choice (K. Fiedler & P. Juslin, Eds.). In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition*. Cambridge, Cambridge University Press.

Hilbig, B. E., & Glöckner, A. (2011). Yes, they can! Appropriate weighting of small probabilities as a function of information acquisition. *Acta Psychologica*, *138*(3), 390–396. https://doi.org/10.1016/j.actpsy.2011.09.005

Hütter, M., & Klauer, K. C. (2016). Applying processing trees in social psychology. *European Review of Social Psychology*, *27*(1), 116–159. https://doi.org/10.1080/10463283.2016.1212966

Johnson, C., & Mullen, B. (1994). Evidence for the accessibility of paired distinctiveness in distinctiveness-based illusory correlation in stereotyping. *Personality and Social Psychology Bulletin*, *20*(1), 65–70. https://doi.org/10.1177/0146167294201006

Klauer, K. C. (2006). Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika*, *71*(1), 7–31. https://doi.org/10.1007/S11336-004-1188-3

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika*, *75*(1), 70–98. https://doi.org/10.1007/s11336-009-9141-0

Klauer, K. C. (2015). Mathematical Modeling (B. Gawronski & G. V. Bodenhausen, Eds.). In B. Gawronski & G. V. Bodenhausen (Eds.), *Theory and explanation in social psychology*. New York, NY, US, Guilford Press.

Klauer, K. C., & Meiser, T. (2000). A source-monitoring analysis of illusory correlations. *Personality and Social Psychology Bulletin*, *26*(9), 1074–1093. https://doi.org/10.1177/01461672002611005

Klauer, K. C., & Wegener, I. (1998). Unraveling social categorization in the "who said what?" paradigm. *Journal of Personality and Social Psychologyocial*, *75*(5), 1155–1178. https://doi.org/10.1037/0022-3514.75.5.1155

Klein, S. A., Hilbig, B. E., & Heck, D. W. (2017). Which is the greater good? A social dilemma paradigm disentangling environmentalism and cooperation. *Journal of Environmental Psychology*, *53*, 40–49. https://doi.org/10.1016/j.jenvp.2017.06.001

Kroneisen, M., & Heck, D. W. (2019). Interindividual differences in the sensitivity for consequences, moral norms, and preferences for inaction: Relating basic personality traits to the CNI model. *Personality and Social Psychology Bulletin.* Advance online publication. https://doi.org/10.1177/0146167219893994

Kutzner, F., Vogel, T., Freytag, P., & Fiedler, K. (2011a). A robust classic: Illusory correlations are maintained under extended operant learning. *Experimental Psychology*, *58*(6), 443–453. https://doi.org/10.1027/1618-3169/a000112

Kutzner, F., Vogel, T., Freytag, P., & Fiedler, K. (2011b). Contingency inferences driven by base rates: Valid by sampling. *Judgment and Decision Making*, *6*(3), 211–221.

Laplace, P. S. (1820/1951). *Philosophical Essays on Probabilities*. New York, Dover.

Lejarraga, T., Hertwig, R., & Gonzalez, C. (2012). How choice ecology influences search in decisions from experience. *Cognition*, *124*(3), 334–342. https://doi.org/10.1016/j.cognition.2012.06.002

Matzke, D., Dolan, C. V., Batchelder, W. H., & Wagenmakers, E.-J. (2015). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika*, *80*(1), 205–235. https://doi.org/10.1007/s11336-013-9374-9

McConnell, A. R., Sherman, S. J., & Hamilton, D. L. (1994). Illusory correlation in the perception of groups: An extension of the distinctiveness-based account. *Journal of Personality and Social Psychology*, *67*(3), 414–429. https://doi.org/10.1037/0022-3514.67.3.414

Meiser, T. (2003). Effects of processing strategy on episodic memory and contingency learning in group stereotype formation. *Social Cognition*, *21*(2), 121–156. https://doi.org/10.1521/soco.21.2.121.21318

Meiser, T., & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning Memory and Cognition*, *28*(1), 116–137. https://doi.org/10.1037/0278-7393.28.1.116

Meiser, T., & Hewstone, M. (2001). Crossed categorization effects on the formation of illusory correlations. *European Journal of Social Psychology*, *31*(4), 443–466. https://doi.org/10.1002/ejsp.55

Meiser, T., & Hewstone, M. (2004). Cognitive processes in stereotype formation: The role of correct contingency learning for biased group judgments. *Journal of Personality and Social Psychology*, *87*(5), 599–614. https://doi.org/10.1037/0022-3514.87.5.599

Meiser, T., & Hewstone, M. (2006). Illusory and spurious correlations: Distinct phenomena or joint outcomes of exemplar-based category learning? *European Journal of Social Psychology*, *36*(3), 315–336. https://doi.org/10.1002/ejsp.304

Meiser, T., Rummel, J., & Fleig, H. (2018). Pseudocontingencies and choice behavior in probabilistic environments with context-dependent outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(1), 50–67. https://doi.org/10.1037/xlm0000432

Michalkiewicz, M., Arden, K., & Erdfelder, E. (2018). Do smarter people employ better decision strategies? the influence of intelligence on adaptive use of the recognition heuristic. *Journal of Behavioral Decision Making*, *31*(1), 3–11. https://doi.org/10.1002/bdm.2040

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

Riefer, D. M., Knapp, B. R., Batchelder, W. H., Bamber, D., & Manifold, V. (2002). Cognitive psychometrics: Assessing storage and retrieval deficits in special populations with multinomial processing tree models. *Psychological Assessment*, *14*, 184–201. https://doi.org/10.1037/1040-3590.14.2.184

Simon, H. A. (1990). Bounded rationality. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *Utility and probability* (pp. 15–18). London, Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-20568-4_5

Smith, J. B., & Batchelder, W. H. (2010). Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology*, *54*(1), 167–183. https://doi.org/10.1016/j.jmp.2009.06.007

Spellman, B. A., & Mandel, D. R. (1999). When possibility informs reality: Counterfactual thinking as a cue to causality. *Current Directions in Psychological Science*, *8*(4), 120–123. https://doi.org/10.1111/1467-8721.00028

Spellman, B. A., Price, C. M., & Logan, J. M. (2001). How two causes are different from one: The use of (un)conditional information in Simpson's paradox. *Memory and Cognition*, *29*(2), 193–208. https://doi.org/10.3758/BF03194913

Vogel, T., Freytag, P., Kutzner, F., & Fiedler, K. (2013). Pseudocontingencies derived from categorically organized memory representations. *Memory & Cognition*, *41*(8), 1185–1199. https://doi.org/10.3758/s13421-013-0331-8

Waldmann, M. R., & Hagmayer, Y. (2001). Estimating causal strength: The role of structural knowledge and processing effort. *Cognition*, *82*(1), 27–58. https://doi.org/10.1016/S0010-0277(01)00141-X

Waldmann, M. R., Hagmayer, Y., & Blaisdell, A. P. (2006). Beyond the information given: Causal models in learning and reasoning. *Current Directions in Psychological Science*, *15*(6), 307–311. https://doi.org/10.1111/j.1467-8721.2006.00458.x

Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General*, *121*(2), 222–236. https://doi.org/10.1037/0096-3445.121.2.222

Wulff, D. U., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin*, *144*(2), 140–176. https://doi.org/10.1037/bul0000115

Yechiam, E., & Hochman, G. (2013). Losses as modulators of attention: Review and analysis of the unique effects of losses over gains. *Psychological Bulletin*, *139*, 497–518. https://doi.org/10.1037/a0029383

A Acknowledgements

"Mischief managed!"

from *Harry Potter and the Prisoner of Azkaban*
by J. K. Rowling

I would like to thank each and every one who supported and guided me during my dissertation. Below, you will find a semi-structured list of my *Purveyors of Aids to Finishing a Dissertation*.

First, I would like to thank Thorsten Meiser for his enthusiasm and for providing advice and support whenever needed, while at the same time giving me the freedom to grow with challenges. I thank Benjamin Hilbig and Tanja Lischetzke who provided helpful insights and advice along the way, and David Kellen for additionally hosting me in Syracuse.

I thank Hansjörg, Simone, and Daniel from whom I learned a lot.

Many thanks to my fellows and friends Malte, Martin, David, Lili, and Nikoletta for their valuable advice and many fun hours in and outside the office. Moreover, I am very grateful to Doro, Ines, Leah, and Nico for being great friends who kept reminding me of my life outside of academia and supported me with laughing, dancing, cooking, vacationing, or simply talking.

Particularly thanks to Mirka and Sophie for additionally being the best office and floor mates I could have wished for; thanks for your invaluable support and wisdom.

Last, but not least, I would like to thank my family, Jürgen, Elke, Julia, Jan, Felix, and Tim who sometimes challenged me, but always believed in me, encouraged me in difficult times and shared my happiness in good times. Thank you.

B Statement of Originality

1. I hereby declare that the presented doctoral dissertation with the title *New Insights Into Cognitive Processes in Pseudocontingency Inference by Means of Experimental Methods and Statistical Modeling* is my own work.
2. I did not seek unauthorized assistance of a third party and I have employed no other sources or means except the ones listed. I clearly marked any quotations derived from the works of others.
3. I did not present this doctoral dissertation or parts of it at any other higher education institution in Germany or abroad.
4. I hereby confirm the accuracy of the declaration above.
5. I am aware of the significance of this declaration and the legal consequences in case of untrue or incomplete statements.

I affirm in lieu of oath that the statements above are to the best of my knowledge true and complete.

Signature:

Date: 06/16/2020

C Co-Authors' Statements

Co-Author: Thorsten Meiser

With this statement, I confirm that the following articles included in the present thesis were primarily conceived and written by Franziska Bott.

Bott, F. M., Heck, D. W., & Meiser, T. (in press). Parameter validation in hierarchical MPT models by functional dissociation with continuous covariates: An application to contingency inference. *Journal of Mathematical Psychology*.

Bott, F., M., & Meiser, T. (2020). Pseudocontingency inference and choice: The role of information sampling. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Advance online publication. https://doi.org/10.1037/xlm0000840

Franziska Bott designed and conducted all experiments reported in Bott and Meiser (2020). She analyzed all data in both manuscripts and wrote most of the articles, including first drafts and revisions of the manuscripts. I contributed the experiment and data for Bott et al. (in press) as well as to developing and refining the research questions in both manuscripts. Furthermore, I provided recommendations for making the messages of the articles clearer.

Prof. Dr. Thorsten Meiser
Mannheim, June 2020

Co-Author: Daniel W. Heck

With this statement, I confirm that the following article included in the present thesis was primarily conceived and written by Franziska Bott.

Bott, F., M., Heck, D. W., & Meiser, T. (in press). Parameter validation in hierarchical MPT models by functional dissociation with continuous covariates: An application to contingency inference. *Journal of Mathematical Psychology*.

Franziska Bott analyzed the data and wrote most of the manuscript, including the first draft and revisions. I contributed to refine the statistical analyses and to clarify their presentation in the manuscript, and gave helpful comments on draft versions of the manuscript.

Prof. Dr. Daniel W. Heck
Marburg, June 2020

Co-Author: David Kellen

With this statement, I affirm that the work in the manuscript

Bott, F., M., Kellen, D., & Klauer, K. C. (2020). *Normative accounts of illusory correlations.* Invited revision submitted to *Psychological Review*.

of which I am a co-author, will be part of the doctoral dissertation by Franziska Bott. I confirm that Franziska Bott wrote the first draft of this manuscript and contributed significantly revising the manuscript in the further writing process. Furthermore, she built the study corpus and performed simulations and analyses. I implemented the Bayesian Marginal Model in R and helped with model fits. Moreover, I provided adjustments to the manuscript and presented ideas to broaden its scope.

Prof. Dr. David Kellen
Syracuse, June 2020

Co-Author: Karl Christoph Klauer

With this statement, I affirm that the work in the manuscript

Bott, F., M., Kellen, D., & Klauer, K. C. (2020). *Normative accounts of illusory correlations.* Invited revision submitted to *Psychological Review*.

of which I am a co-author, will be part of the doctoral dissertation by Franziska Bott. I confirm that Franziska Bott wrote the first draft of this manuscript and contributed significantly revising it. Furthermore, she built the study corpus and preformed simulations and analyses. I provided corrections and recommendations to the manuscript and presented ideas for restructuring the manuscript to broaden its scope.

Prof. Dr. Karl Christoph Klauer
Freiburg, June 2020

D Copies of Articles

Parameter Validation in Hierarchical MPT Models by Functional Dissociation with Continuous Covariates: An Application to Contingency Inference

Franziska M. Bott[1], Daniel W. Heck[2], and Thorsten Meiser[1]

[1]University of Mannheim

[2]Philipps-Universität Marburg

Author Note

This research was funded by the Deutsche Forschungsgemeinschaft (DFG), grants ME 1918/2-3 and GRK 2277 (Research Training Group "Statistical Modeling in Psychology"). Data and analysis scripts will be provided on OSF (`https://osf.io/a6fcz/`).

Correspondence to: Franziska Bott, Department of Psychology, University of Mannheim, L13 15, D-68161 Mannheim, Germany. E-mail: f.bott@uni-mannheim.de

Abstract

In traditional multinomial processing tree (MPT) models for aggregate frequency data, parameters have usually been validated by means of experimental manipulations, thereby testing selective effects of discrete independent variables on specific model parameters. More recently, hierarchical MPT models which account for parameter heterogeneity between participants have been introduced. These models offer a new possibility of parameter validation by analyzing selective covariations of interindividual differences in MPT model parameters with continuous covariates. The new approach enables researchers to test parameter validity in terms of functional dissociations, including convergent validity and discriminant validity in a nomological network. Here, we apply the novel approach to a multidimensional source-monitoring model in the domain of stereotype formation based on pseudocontingency inference. Using hierarchical Bayesian MPT models, we test the validity of source-guessing parameters as indicators of specific source evaluations on the individual level. First, analyzing experimental data on stereotype formation ($N = 130$), we replicated earlier findings of biased source-guessing parameters while taking parameter heterogeneity across participants into account. Second, we investigated the specificity of covariations between conditional guessing parameters and continuous direct measures of source evaluations. Interindividual differences in direct evaluations predicted interindividual differences in specific source-guessing parameters, thus validating their substantive interpretation. Third, in an exploratory analysis, we examined relations of memory parameters and guessing parameters with cognitive performance measures from a standardized cognitive assessment battery.

Keywords: parameter validation; hierarchical MPT; continuous covariates; pseudocontingency

Parameter Validation in Hierarchical MPT Models by Functional Dissociation with Continuous Covariates: An Application to Contingency Inference

Multinomial processing tree (MPT) models are statistical models used to estimate probabilities of latent cognitive processes from observable frequency data (Batchelder & Riefer, 1999; Erdfelder et al., 2009). In order to test assumptions about psychological processes that jointly contribute to behavioral responses, MPT models were traditionally fitted to data aggregated across participants. A good model fit to the data, though, is insufficient to justify an interpretation of the model parameters in terms of the assumed psychological processes. Besides model fit, the substantive interpretation of MPT model parameters has to be validated via selective influence.

In traditional MPT modeling, parameters are validated by testing whether they are selectively influenced by specific experimental manipulations (Erdfelder et al., 2009; Hütter & Klauer, 2016). These experimental manipulations are discrete realizations of factors assumed to enhance or impair specific cognitive processes. In validation studies, the selective influence of an experimental manipulation on one specific model parameter is tested. In contrast, none of the remaining parameters representing other cognitive processes should be affected. Accordingly, using discrete experimental variables, convergent validity and discriminant validity of MPT model parameters are evaluated via model comparisons (e.g., Bayen, Murnane, & Erdfelder, 1996; Erdfelder & Buchner, 1998; Klauer & Wegener, 1998; Meiser & Bröder, 2002; Meissner & Rothermund, 2013).

More recently, hierarchical MPT models have been developed to account for parameter heterogeneity between participants (Klauer, 2010; Matzke, Dolan, Batchelder, & Wagenmakers, 2015; Smith & Batchelder, 2010). Instead of fitting MPT models using aggregate data, model parameters are estimated at the level of individuals and are assumed to follow a continuous distribution at the group level. In such hierarchical models, interindividual differences in model parameters can be tested for covariations with continuous variables. Across various substantive research areas, hierarchical MPT models are increasingly being used to investigate such covariations. For instance, researchers relied on this approach to estimate associations of memory parameters in MPT models with fluid and crystallized intelligence (Michalkiewicz, Arden, & Erdfelder, 2018), associations between guessing parameters and the reliance on both stereotypes and inferred contingencies (e.g., Arnold, Bayen, Kuhlmann, & Vaterrodt, 2013; Ernst, Kuhlmann, & Vogel, 2019; Kuhlmann, Bayen, Meuser, & Kornadt, 2016), and associations between personality traits and dishonest behavior (Heck, Thielmann, Moshagen, & Hilbig, 2018), moral judgments (Kroneisen & Heck, in press), and environmental preferences (Klein, Hilbig, & Heck, 2017). In a similar vein, Anders, Oravecz, and Alario (2017) and Boehm, Steingroever, and Wagenmakers (2018) proposed general regression frameworks for testing covariations between continuous variables and model parameters in Bayesian hierarchical cognitive models.

Besides testing covariations with theoretically relevant model parameters for which a hypothesized effect is of substantive interest, however, hierarchical MPT models present a new approach to parameter validation: instead of showing selective influence via (discrete) experimental manipulations, the construct validity of parameters as valid measures of cognitive processes can be tested by analyzing the specificity of (continuous) covariations. To this end, continuous measures of psychological constructs or traits can be used which are assumed to selectively correspond to the specific cognitive process underlying a given MPT model parameter. Then, one can test for a functional dissociation in terms not only of (a) substantial stochastic associations between interindividual differences in the given MPT parameter and measures of the same construct (i.e., convergent validity), but also in terms of (b) negligible or weak stochastic associations between interindividual differences in the MPT parameter and measures of theoretically distinct constructs (i.e., discriminant validity). By testing convergent and discriminant validity, the approach allows researchers to validate the parameters of an MPT model in a nomological network (Campbell & Fiske, 1959).

MPT models can be flexibly tailored to various experimental paradigms. In the present paper, we use an extended version of the classical source-monitoring MPT model proposed by Batchelder and Riefer (1990) to disentangle memory processes and guessing processes. In source-monitoring experiments, participants have to discriminate between target items presented in a preceding study phase and new distractor items. Additionally, they have to assign items judged to be old to one out of two or more sources that have been learned during the study phase (e.g., whether words judged as old were presented in blue or red). Source-monitoring MPT models allow to disentangle and estimate three types of processes contributing to task performance in those experiments (Bayen et al., 1996; Bröder & Meiser, 2007): item memory, source memory, and guessing. Extending this model to combinations of multiple sources (e.g., by presenting items in red/blue crossed with a left/right position on the screen), multidimensional source-monitoring MPT models account for memory for multiple, crossed dimensions of source information (Meiser & Bröder, 2002).

In research on source memory, the focus is usually on measuring memory for sources as well as memory for items in a process-pure way, that is, independently from guessing processes (e.g., guessing that a word was shown in blue in the absence of source memory). However, in MPT models, guessing parameters may not only represent perfectly random and unbiased guessing, but can instead be interpreted as indicators of expectancy-based guessing. In experimental studies, it has been shown that source-guessing parameters can reflect, for instance, learned item-source contingencies in the experimental stimuli (e.g., Bayen & Kuhlmann, 2011; Klauer & Wegener, 1998) or schemata and stereotypes associated with the items and sources (Bayen, Nakamura, Dupuis, & Yang, 2000; Ehrenberg & Klauer, 2005; Kuhlmann et al., 2016; Wegener & Klauer, 2004). Likewise, Klauer and Meiser (2000) and Meiser and Hewstone (2004) interpreted source-guessing parameters as reflecting evaluative biases in assigning positive versus negative items to sources of different valence in the absence of source memory.

Evaluative biases are assumed to originate in inferred item-source contingencies. Yet, the inferred contingencies do not necessarily correspond to genuine item-source contingencies in the experimental design but may reflect illusory correlations or pseudocontingencies. In general, pseudocontingencies denote contingencies that are inferred on the basis of skewed base rates or joint covariations of two variables with a third one, rather than on the basis of genuine covariation information of the two focal variables (Fiedler & Freytag, 2004; Fiedler, Freytag, & Meiser, 2009; Fiedler, Kutzner, & Vogel, 2013; Meiser, Rummel, & Fleig, 2018). In source-monitoring experiments on pseudocontingencies in stereotype formation (see Meiser & Hewstone, 2010, for an overview), positive behavioral statements and negative behavioral statements (items) are paired with individual members of two groups (sources). Usually, the base rates of (a) positive items versus negative items and (b) Source A versus Source B are skewed in the stimulus distribution (e.g., by presenting a majority of positive items and a majority of items from Source A). A pseudocontingency is observed if a contingency between the item categories and source categories is inferred on the basis of the skewed base rates, such that participants associate the more frequent item category (e.g., positive items) with the more frequent source category (e.g., Source A) and the infrequent item category (e.g., negative items) with the infrequent source category (e.g., Source B; Meiser & Hewstone, 2004, 2006). Because the marginal base rates for item types and sources do not uniquely determine the item-source contingency, however, subjectively inferred pseudocontingencies might not correspond to the genuine item-source contingency (e.g., the proportion of positive versus negative items can be higher for Source B than for Source A).

Coming back to source memory, if the source of an item is not remembered, individuals may guess the source on the basis of the inferred (pseudo-)contingency between items and sources (Klauer & Meiser, 2000; Meiser & Hewstone, 2004). The inferred pseudocontingency, which mirrors a positive covariation between frequent items and frequent sources in the example, should therefore be reflected by participants' guessing parameters in a source-monitoring MPT model. Accordingly, biased guessing parameters are taken as evidence of inferred evaluative (pseudo-)contingencies. At the same time, inferred pseudocontingencies are not expected to be related to any other parameters of the MPT model (especially not to memory parameters). This line of reasoning highlights that, in the domain of source monitoring, the substantive interpretation not only of *memory* parameters but also of *guessing* parameters is of theoretical importance. Hence, the goal of the present research is to analyze the association of guessing parameters with direct measures of source evaluations in a pseudocontingency paradigm. To this end, hierarchical MPT models allow us to investigate the construct validity of parameters by testing covariations of interindividual differences in source-guessing parameters with interindividual differences in direct measures of inferred pseudocontingencies.

In the remainder of the article, we replicate previous results on multinomial modeling of pseudocontingency inference before applying the novel approach of parameter validation with continuous covariates. First, we used the framework of hierarchical Bayesian MPT models (Heck, Arnold, & Arnold, 2018; Klauer, 2010) to analyze biases in source guessing based on pseudocontingency inference while taking parameter heterogeneity into account. Second, we tested the validity of source-guessing parameters by means of their selective covariations with direct, continuous measures of source evaluations. Finally, in an exploratory analysis, we investigated the associations of performance measures from a standardized cognitive assessment battery with memory parameters and guessing parameters of the hierarchical MPT model in order to further exemplify the possibilities of validating MPT model parameters in a nomological network. The joint analysis of MPT parameters and other variables measuring general cognitive capacities also illustrates how cognitive modeling in experimental psychology can profit from combining experimental data with psychometric assessment, an approach labeled *cognitive psychometrics* by William Batchelder and David Riefer (Batchelder, 1998, 2010; Riefer, Knapp, Batchelder, Bamber, & Manifold, 2002).

2 Experiment

In the experiment, we investigated stereotype formation on the basis of pseudocontingency inference. The study addressed the reliance on inferred pseudocontingencies when reconstructing sources of positive and negative behavioral statements in a scenario with two artificial groups and the context factor of town of residence (see Meiser & Hewstone, 2004). A first aim was to replicate the finding that source guessing is biased in a pseudocontingency paradigm with skewed base rates. Previous studies used an MPT model of source memory for two-dimensional source information (Meiser & Bröder, 2002) to estimate (1) item memory for behavioral statements, (2) source memory for town of residence, (3) source memory for group membership, and (4) guessing processes from data aggregated across participants (e.g., Meiser & Hewstone, 2004, 2006). However, since MPT models were fitted to aggregate data, heterogeneity between participants might have resulted in biased parameter estimates and erroneous statistical inference (e.g., Klauer, 2006; Rouder et al., 2007). Therefore, we used a hierarchical Bayesian MPT model to account for interindividual differences in item memory, source memory, and guessing.

2.1 Method

2.1.1 Participants and Design. A sample of $N = 133$ participants was recruited from the University of Mannheim, from universities of applied sciences, and a secondary school in Mannheim. Participants were randomly assigned to one of two experimental conditions with or without base rate information concerning the two groups in the two towns (see below). Participants received course credit in the case of psychology students or a financial compensation of 30 EUR for completing the pseudocontingency experiment and the cognitive assessment.

The data of three participants were discarded from analyses because these participants reported that German was not their first language and either showed scores smaller than 2 SD below the mean in the verbal subscale (see below) or had problems following the instructions. The remaining sample size guaranteed a statistical power of at least $1-\beta = .90$ to detect a difference in the direct evaluative ratings of the two target groups with medium effect size $f = .25$ and $\alpha = .05$ in a conventional frequentist analysis of pseudocontingencies (e.g. Meiser & Hewstone, 2004). Of the $N = 130$ participants, 50% were female and mean age was 23.12 years ($SD = 4.97$). Concerning level of education, 10% of the participants reported that they were attending secondary school or held an exam from a basic or intermediate secondary school, 78% had graduated from high school, and 12% held a university degree. 66 participants were tested in the condition without base rate information, and 64 in the condition with base rate information.

2.1.2 Material and Procedure. Experimental sessions consisted of the pseudocontingency experiment for about 30 min, a brief unrelated questionnaire for about 5 min, and a cognitive assessment for about 2.5 h. Participants were invited to take breaks between the different sections of the session and between different subscales of the cognitive assessment.

The pseudocontingency experiment started with a learning phase of behavioral statements about members of two artificial groups, labeled Group A and Group B, in two artificial towns, labeled Town X and Town Y. For each participant, 24 desirable behaviors and 24 undesirable behaviors were randomly drawn from a larger pool of behaviors. The pool consisted of 76 moderately desirable and 76 moderately undesirable behaviors that showed mean ratings of +1.5 or above for desirable behaviors, and -1.5 or below for undesirable behaviors, on a scale from -3 to 3 in an independent pretest study ($N = 36$; see Meiser, 2003, for details). Examples are "stays friendly and fair in discussions" and "becomes angry when being delayed by an elderly pedestrian". The randomly selected behavioral statements were assigned to the groups and towns according to the stimulus distribution in Table 1.

In the learning phase, each behavioral statement contained a male first name, group membership in Group A or Group B, town of residence Town X or Town Y, and the actual behavior. The behavioral statements were displayed for 6 s with an interstimulus interval of 1.5 s. Behavioral statements describing group members of either group in Town X were presented on the left-hand side of the computer screen and statements describing group members of either group in Town Y on the right-hand side. The order of behavioral statements was randomized per participant. Participants were instructed to read the statements carefully and to form an impression of the two groups in the two towns. The instructions did not ask participants to memorize the individual behaviors or the group and town information for a later memory test.

In the stimulus distribution (see Table 1), Group B shows a larger proportion of desirable behaviors than Group A within either town: In Town X, all members of Group B but only 67% of the members of Group A were described as engaging in desirable behavior.

Table 1

Stimulus Distribution of Desirable and Undesirable Behaviors for Group A and Group B in Town X and Town Y

	Town X		Town Y	
	Group A	Group B	Group A	Group B
Desirable	12	6	0	6
Undesirable	6	0	6	12

In Town Y, 33% of the members of Group B but none of the members of Group A engaged in desirable behavior. The true contingency conditional on the context factor of town of residence should thus give rise to a more positive impression of Group B. Moreover, the higher proportions of Group A and of desirable behaviors in Town X relative to Town Y imply true positive contingencies of town of residence with group membership and desirability. The distributions of Group A and Group B and of desirable and undesirable behaviors are skewed within each town, however, such that a pseudocontingency was expected to be inferred by the participants. The pseudocontingency should associate the frequent categories (i.e., Group A and desirable behaviors in Town X, Group B and undesirable behaviors in Town Y) and the infrequent categories (Group B and undesirable behaviors in Town X, Group A and desirable behaviors in Town Y), leading to a more positive impression of Group A thus contrasting the true contingency in the stimuli. This pseudocontingency inference should be strengthened by the covariation of the skewed distributions of Group A versus Group B and desirable versus undesirable behaviors across the context factor town of residence (Meiser et al., 2018).[1]

[1] Given the fixed base rates of group membership and type of behavior, the stimulus distribution used here (see Table 1) is the most extreme one consistent with those base rates. However, pseudocontingency inferences have also been found with less extreme stimulus distributions that did not contain zero frequencies in the stimulus distribution (cf. e.g., Fiedler & Freytag, 2004; Meiser et al., 2018, Experiment 3).

In order to manipulate attention to the context factor and the varying base rates, explicit information concerning the base rates of the two groups within the two towns was given to half of the participants: in the experimental condition *without base rate information*, participants received no a priori information about the proportions of Group A and Group B for either town. In the experimental condition *with base rate information*, the instructions pointed out that in Town X, 75% of the persons belonged to Group A and 25% to Group B, whereas in Town Y, 25% of the persons belonged to Group A and 75% to Group B. The proportions of Group A and Group B for either town were also displayed throughout the learning phase in the condition with base rate information. We assumed that the explicit information on the skewed base rates of group membership may strengthen pseudocontingency inference, as alternative manipulations using a visualization of base rates via bar charts (Meiser et al., 2018) or using changes of attentional focus (Fleig, Meiser, Ettlin, & Rummel, 2017) have shown more pronounced pseudocontingency effects.

The test phase of the pseudocontingency experiment included a rating task, a source-memory test, and frequency estimates. The order of tasks followed standard practice (e.g. Meiser, 2003; Meiser & Hewstone, 2004) and avoided that numerical information from the frequency estimation task contaminated judgment and guessing processes in the trait ratings or memory test, respectively. To first assess the participants' general impressions about the groups per town, in the rating task, participants rated Group A and Group B in Town X and Town Y on five positive and five negative trait attributes in a fixed randomized order. Trait ratings were assessed on separate 10-point rating scales (0: trait does not apply at all; 9: trait applies completely) for each of the four combinations of group and town. The ratings were averaged across the ten traits after re-coding ratings for negative trait attributes, so that larger trait scores reflect more positive judgments.

In the source-memory test, participants were presented with the 48 target behaviors from the learning phase and 48 new distractor behaviors in random order. The distractor items included 24 desirable and 24 undesirable behaviors that were randomly drawn for each participant from the same pool as the target behaviors. In the memory test, participants had to decide for each behavior whether it was old (i.e., had occurred in the learning phase) or new. If a behavior was judged old, participants had to decide whether it occurred in Town X or Town Y in a second step, and whether it described a member of Group A or Group B in a third step. The individual frequencies of judging desirable and undesirable target and distractor behaviors as old and of assigning them to Town X or Town Y and to Group A or Group B provided the data for the hierarchical MPT analysis.

Frequency estimates formed the final task of the pseudocontingency experiment, as here participants were shown the total number of statements about each combination of group and town in the learning phase (e.g., 18 for Group A in Town X, see Table 1). Next, they had to estimate the number of undesirable behaviors for the given group within the given town. The estimated numbers of undesirable behaviors were transformed to proportions by dividing them by the total numbers of behavioral statements per group-town combination, so that higher values reflect more negative evaluations.

After the pseudocontingency experiment, participants filled in a brief questionnaire that was unrelated to the experiment and were then invited to have a break before the cognitive test was administered. We used the computerized and adaptive INSBAT test battery (Arendasy et al., 2009) to assess cognitive performance in the domains of fluid intelligence, quantitative reasoning, short-term and long-term memory, as well as crystallized intelligence. To assess fluid intelligence, we selected the subscales of figural inductive reasoning and verbal deductive reasoning as part of our cognitive assessment. Fluid intelligence measured by the INSBAT reflects the ability to recognize regularities or to combine information and to draw logical conclusions from them. The phenomenon of pseudocontingency inference might be traceable to this ability: pseudocontingencies are based on inferences drawn from statistical regularities (i.e., skewed base rates). In a similar vein, Fleig et al. (2017) demonstrated that even participants high in statistical numeracy do infer pseudocontingencies when focusing on base rates. Since pseudocontingencies are assumed to form the basis of source evaluations which in turn should be reflected in source-guessing parameters, fluid intelligence and its subscales may potentially relate to the source-guessing parameters. For the domain of quantitative reasoning, we chose the subscales of computational estimation and arithmetic competence. They capture the ability to apply mathematical skills in problem-solving (see Arendasy et al., 2009). Similarly to fluid intelligence, interindividual differences in quantitative thinking might explain interindividual differences in inferring pseudocontingencies and thereby in source-guessing parameters in our experiment. Memory performance was assessed with the INSBAT subscales of verbal short-term and long-term memory. Participants' cognitive ability to remember and recognize information could be related to memory parameters in the MPT model. Finally, subscales of general knowledge and verbal fluency were assessed from the domain of crystallized intelligence. To measure general knowledge, participants had to fill out a cloze test asking to complete word definitions, whereas they had to solve anagrams in the verbal fluency task. Crystallized intelligence and its subscales were thus not included in the following analyses, because these tests appeared less relevant for experimental source memory and pseudocontingency inference. Nevertheless, we did measure those subscales in order to comprehensibly assess participants' cognitive ability and to be able to give feedback to the individual participant on their cognitive performance as measured by the INSBAT. For our analyses we used the specific scores for each subscale as well as the total scores for each factor.

2.2 Results

We first report analyses of the rating scores and frequency estimates as direct measures of pseudocontingency inference. Next, we present the results of the hierarchical MPT analyses of source monitoring. All analyses were conducted in R (R Core Team, 2018).

2.2.1 Direct measures of pseudocontingency inference. In order to analyze mean rating scores and frequency estimates, we conducted Bayesian hypothesis tests for the $2 \times 2 \times 2$ mixed ANOVA design with condition as between-participants factor and town of residence and group membership as within-participant factors (Rouder, Morey, Speckman, & Province, 2012). For this purpose, we used the BayesFactor package (Morey & Rouder, 2018). In a Bayesian ANOVA, the Bayes factor $B_{1,0}$ quantifies the relative evidence for the alternative hypothesis $\mathcal{H}_1$ (e.g., an ANOVA model with two main effects and no interaction) against the null hypothesis $\mathcal{H}_0$ (e.g., the intercept-only model). Similarly, the inverse $B_{0,1} = 1/B_{1,0}$ quantifies the relative support for the null hypothesis. Besides the intercept-only model, we included all models comprising one or more main effects, whereas models with interactions were only included if they comprised all corresponding main effects. We compared all models to the intercept-only model (i.e., by using the option "withMain" in the BayesFactor package). In order to report the results more concisely, we additionally computed posterior probabilities for subsets of models (e.g., all models including the effect of condition) by means of model averaging (Hoeting, Madigan, Raftery, & Volinsky, 1999). When assuming uniform prior probabilities over all models, the posterior probability of a single model can be calculated by dividing its Bayes factor $B_{1,0}$ (i.e., the evidence compared to the intercept-only model) by the sum of Bayes factors $B_{i,0}$ for all models i (including the intercept-only model itself, for which $B_{0,0} = 1$). The posterior probability of a subset of models can then be computed by summing the posterior probabilities of the corresponding models.

The mean rating scores per combination of group and town are displayed in Table 2. The analyses revealed the strongest evidence for the ANOVA model including main effects of town of residence and group membership only, with a corresponding Bayes factor of $B_{(\text{group+town}),\text{null}} = 6.26 \cdot 10^{34}$ compared to the intercept-only model (the Bayes factor comparing this model to the second-best model which additionally included their interaction was $B_{(\text{group+town}),(\text{group+town+group:town})} = 2.44$; Bayes factors for all models are reported in Table A1). Town X was evaluated more positively than Town Y corresponding to the genuine item-source contingency between behavioral statements and town of residence in the stimulus distribution. Moreover, reflecting a pseudocontingency between behavioral statements and group membership, Group A received higher evaluative ratings than Group B. Descriptively, this effect was weaker within Town Y than within Town X. Concerning the experimental manipulation of base rate information, the prior probability of 73.7% for the subset of ANOVA models including experimental condition as an effect (i.e., 14 out of 19 models) decreased to a posterior probability of 12.7%, thus indicating no evidence for an effect of explicit base rate information during the instruction and learning phase.[2]

Table 2

Rating Scores and Estimated Proportions of Undesirable Behaviors for each Combination of Town of Residence and Group Membership

	Town X				Town Y			
	Group A		Group B		Group A		Group B	
	M	SD	M	SD	M	SD	M	SD
Rating score	5.77	1.29	4.95	1.19	4.27	1.04	3.77	1.29
Estimated frequency	0.36	0.17	0.47	0.22	0.53	0.21	0.60	0.19

Note. Rating scores were calculated for each participant by averaging across all trait ratings after ratings on negative traits were re-coded. Higher rating scores indicate more positive evaluations. Estimated frequencies were computed by dividing the estimated number of negative behaviors by the total number of behavioral statements per town-group combination presented during the learning phase. They refer to the estimated probability of negative behavior on a scale from 0 to 1 with higher values indicating more negative evaluations.

The mean frequency estimates are also shown in Table 2. As for the rating scores, the Bayes factors favored the model including main effects for town of residence and group membership only, $B_{(\text{group+town}),\text{null}} = 2.48 \cdot 10^{19}$ (the Bayes factor comparing this model to the second-best model that additionally included their interaction was $B_{(\text{group+town}),(\text{group+town+group:town})} = 3.87$; Bayes factors for all models are reported in Table A2). The estimated frequencies of negative behavior for Town X were lower than for

[2] The absence of an effect of the experimental manipulation in this experiment may be due to the fact that the manipulation highlighted the varying skewed base rate of only one variable (i.e., group membership), whereas other manipulations have focused on the skewed base rates of both proximal variables of a pseudocontingency inference (Fleig et al., 2017; Meiser et al., 2018).

Town Y, and negative behavior was also estimated to be less frequent for members of Group A as compared to Group B. Again, descriptively, the effect of group membership was stronger in Town X than in Town Y. Although the genuine item-source contingency between behavioral statements and group membership favored Group B (see Table 1), the estimated frequencies as well as the rating scores thus reflected an inferred pseudocontingency in favor of Group A. Again, there was no evidence for any effect of experimental condition: the prior probability of the subset of 14 ANOVA models including condition as a factor declined from 73.7% to a posterior probability of 17.9%.

2.2.2 Source memory and source guessing. We used the MPT source-memory model for crossed source dimensions (Meiser & Bröder, 2002) to separately estimate source memory for the source dimension town of residence, source memory for the source dimension group membership, and source guessing. The model specifies the probabilities of source-memory decisions by means of eight parameters (see Figure 1). The parameter D denotes the probability to recognize a target item as old and a distractor item as new. An item is not recognized with the complementary probability $1 - D$. If item memory fails, a participant guesses that an item is old with probability b or that an item is new with the complementary probability $1-b$. The d^{town} parameter measures the probability of remembering the town of residence as one source attribute of the item. If the town of residence is not remembered, with the complementary probability $1 - d^{town}$, participants have to guess. The parameter g^{town} describes the probability of guessing Town X, whereas $1-g^{town}$ is the probability of guessing Town Y. If the town of residence is remembered, group membership is remembered with the probability d^{group}; if source memory for town of residence fails, group membership is remembered with the probability e^{group}. Thus, the model reflects the assumption that the probability of remembering group membership may vary over items for which town of residence is retrieved versus not retrieved. Differences between d^{group} and e^{group} thereby reflect stochastic dependency in multidimensional source memory. The source memory parameters in the MPT model were specified in the order in which the source dimensions were prompted in the memory test (i.e., town followed by group) and thus reflected the sequence of retrieval processes required for the responses. If source memory for group membership fails, with either probability $1 - d^{group}$ or probability $1 - e^{group}$, a behavioral statement assigned to Town X is attributed to Group A with probability $g^{group}_{|X}$ or to Group B with probability $1 - g^{group}_{|X}$. Likewise, a behavioral statement assigned to Town Y is attributed to Group A with probability $g^{group}_{|Y}$ and to Group B with probability $1 - g^{group}_{|Y}$. The specification of guessing parameters for the group assignment conditional on town assignment reflected the order of responses in the source-memory test and allowed for differential guessing tendencies that mirror the actual contingency between groups and towns in the stimuli (Meiser, 2003; Meiser & Hewstone, 2004). While the parameters D, d^{town}, d^{group}, and e^{group} measure genuine memory performance, the source-guessing parameters are assumed to reflect evaluative judgments on the basis of inferred (pseudo-)contingencies between behavioral statements and town of residence or group membership.

As we expected the inference of a pseudocontingency between group membership and type of behavior, the assignment of positive behavioral statements to sources might differ from the assignment of negative statements when guessing is not random but informed. If the pseudocontingency is inferred and reflected in the MPT model's guessing parameters, the probabilities $g^{group}_{|X}$ and $g^{group}_{|Y}$ of choosing Group A will be higher for positive statements than for negative statements. Moreover, the probability g^{town} of guessing Town X may be influenced by the actual contingency between town of residence and desirability, such that positive behaviors are more likely to be assigned to Town X than negative behaviors. Therefore, we estimated separate sets of parameters for positive target and distractor behaviors and for negative target and distractor behaviors, respectively.

Using the hierarchical latent-trait approach (Klauer, 2010), all MPT parameters were modeled using random effects to account for differences between individuals. Bayesian estimation of hierarchical MPT models was conducted with the TreeBUGS package using default priors (Heck, Arnold, & Arnold, 2018). For the slope parameters, we z-standardized the predictors and assumed a normal distribution $\mathcal{N}(\mu = 0, \sigma = \sqrt{2}/2)$ as a prior distribution of each slope parameter on the latent probit scale. Figure 2 displays

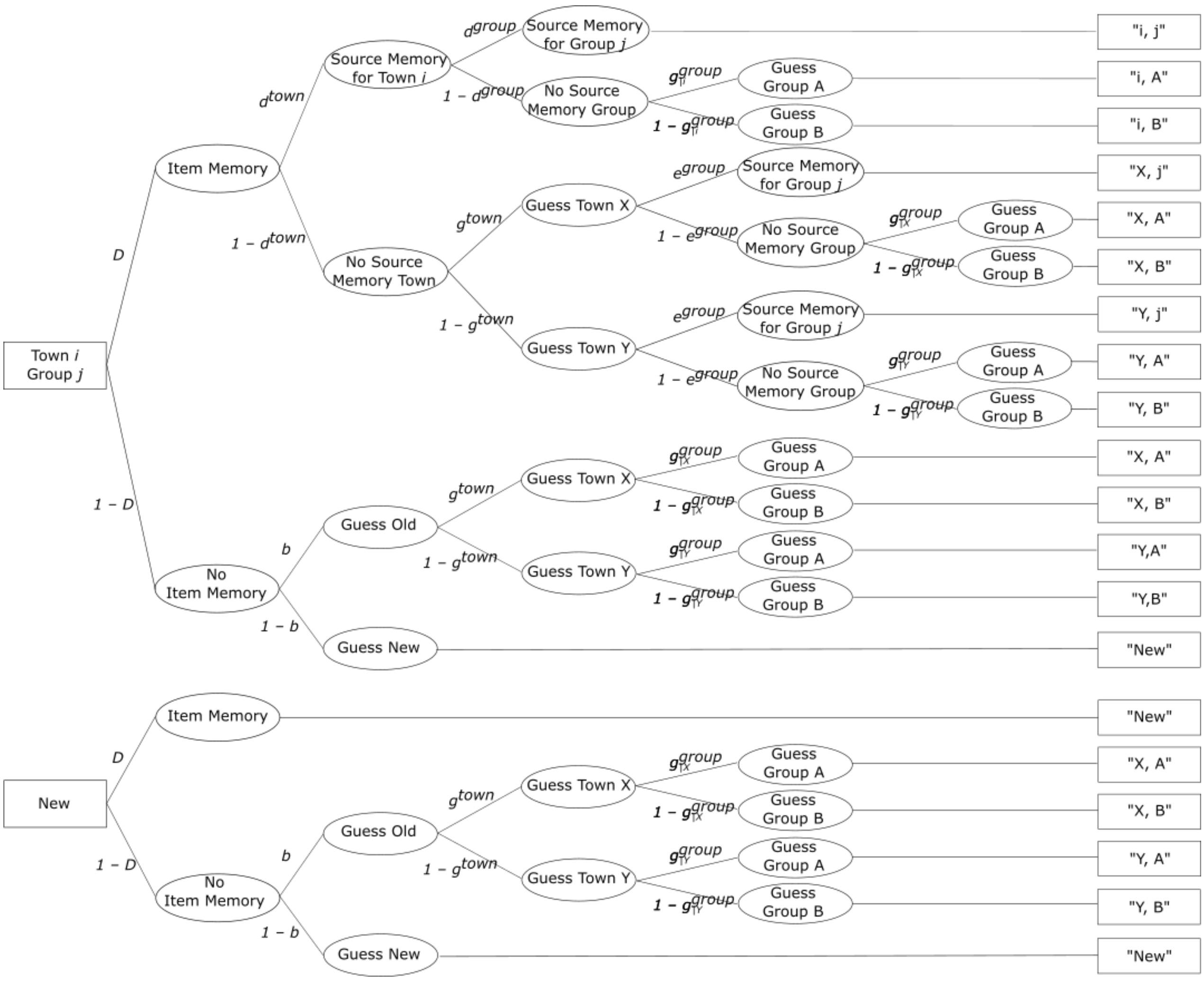

Figure 1. Multinomial processing-tree model of source monitoring for 2×2 crossed sources: town of residence $i \in X, Y$ and group membership $j \in A, B$. Distractor items are referred to as *New*. D = probability to recognize a target item; d^{town} = probability to remember Town i; d^{group} = probability to remember Group j given Town i was recollected; e^{group} = probability to remember Group j given Town i was not remembered; b = probability to guess that an item is old; g^{town} = probability to guess Town X; $g^{group}_{|X}$ = probability of guessing Group A given assignment of item to Town X; $g^{group}_{|Y}$ = probability of guessing Group A given assignment to Town Y.

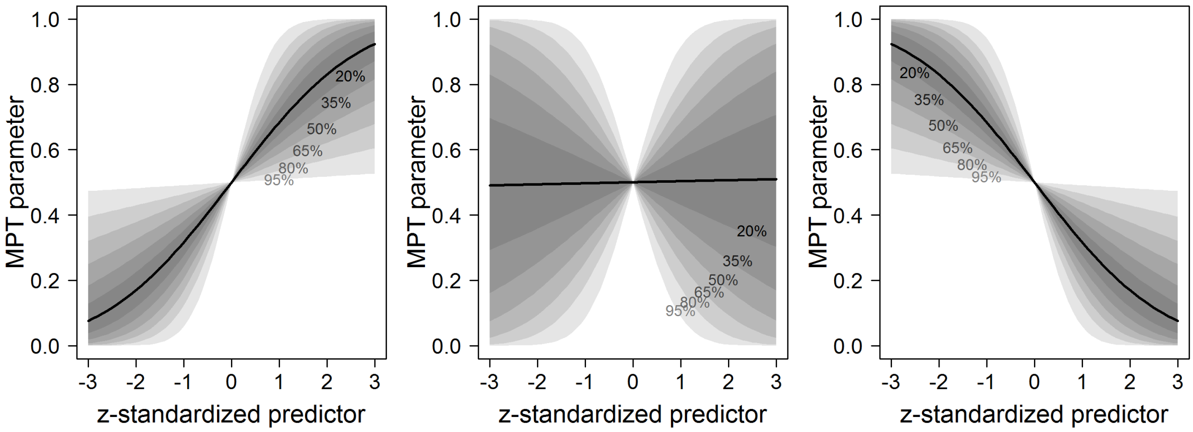

Figure 2. Median (solid black line) and prediction intervals (areas shaded in gray) of the prior predictive distribution for a probit regression of an MPT parameter on a covariate as implied by the prior distribution on the slope parameter, $\beta \sim \mathcal{N}(\mu = 0, \sigma = \sqrt{2}/2)$. The left and right panels show the prior predictive distribution when testing a directional hypothesis ($\mathcal{H}_1$: slope parameter $\beta > 0$ or $\beta < 0$, respectively). The middle panel shows the prior predictive distribution when testing a non-directional hypothesis ($\mathcal{H}_1$: $\beta \neq 0$).

the prior predictive distribution that is implied when regressing an MPT parameter on a covariate.[3]

We ensured convergence of all models by visual inspection of the trace plots. Parameter convergence was evaluated by means of Gelman-Rubin statistics of $\hat{R} \leq 1.1$.[4] The model goodness of fit was assessed graphically as well as by posterior predictive p-values. We tested the model fit to the means and covariance matrix of the observed category frequencies across participants by using the T_1 and T_2 test statistics (Klauer,

[3] We additionally ran the analyses using a normal distribution $\mathcal{N}(\mu = 0, \sigma = 1)$ as prior on the slope parameters. As this prior is less informative, the Bayes factors generally indicated more evidence in favor of the null hypothesis (i.e., for the absence of an effect of a covariate on an MPT parameter). Beyond these differences, the results lead to substantive conclusions identical to those reported in the main text below. Detailed results using the less informative prior can be found on OSF (`https://osf.io/a6fcz/`).

[4] Except for a few ρ parameters (correlations between MPT model parameters), all parameters actually met the standard criterion of $\hat{R} \leq 1.05$.

2010), respectively. T_1 quantifies the discrepancy between the observed and the expected means of response frequencies, whereas T_2 relates to the discrepancy between the observed and the expected covariance matrix of individual response frequencies. Both are conceptually similar to Pearson's X^2. If the corresponding posterior-predictive p-values p_{T_1} and p_{T_2} are close to zero, this indicates model misfit.[5]

First of all, we fitted two separate hierarchical MPT models for the two experimental conditions without versus with explicit information about the base rates of Group A and Group B in either town. Model fit was adequate as indicated by $p_{T1} = 0.245$ and $p_{T2} = 0.264$ as well as $p_{T1} = 0.090$ and $p_{T2} = 0.201$, for the condition without and with explicit base rate information, respectively. Second, as there was little support for an effect of experimental condition in the previous analyses, we fitted one single hierarchical MPT model to the whole data set, thus assuming that the means and covariances of all parameters were not affected by this manipulation. The joint model showed less adequate fit to the average observed frequencies than the two separate models, $p_{T_1} = .021$, but still accounted for variances and covariances satisfactorily, $p_{T_2} = .311$. In a final step, we restricted the parameters representing source memory for group membership, d^{group} and e^{group}, to be equal. Thereby, the assumption that memory for group membership may be stochastically dependent on memory for town of residence is removed to yield a more parsimonious model. Again, the T_1 and T_2 test statistics revealed a less adequate fit of the restricted model to means, $p_{T1} = .025$, than to variances and covariances, $p_{T2} = .322$. In light of these results, all validation analyses reported below were conducted using the restricted, more parsimonious model while controlling for effects of the experimental condition on theoretically relevant MPT model parameters. More precisely, we accounted for mean differences in the parameters by adding the discrete predictor experimental

[5] The aim of using posterior-predictive p-values was to reduce the number of free parameters and MPT models in order to use a model as parsimonious as possible for the analyses of main interest. However, if model selection among MPT models was of central theoretical interest, it would be preferable to compute Bayes factors, for instance, using bridge sampling for nested hierarchical MPT models (Gronau, Wagenmakers, Heck, & Matzke, 2019). Below, we report Bayes Factors for testing our key theoretical questions regarding the selective influence of continuous covariates on MPT parameters.

condition in the latent probit regressions for all guessing parameters (i.e., g^{town}, $g^{group}_{|X}$, $g^{group}_{|Y}$). This allowed us to account for the reduced model fit to the observed mean frequencies as indicated by the posterior-predictive checks above, while fitting a single model to the full data set instead of separate models per condition. As we included experimental condition only as control factor, we do not report the corresponding Bayes factors.

Table 3 displays the estimated parameters of the restricted hierarchical MPT model for crossed source dimensions. To compare parameter estimates for positive versus negative behavioral statements, we computed the corresponding differences in item-memory and source-memory parameters at the group level. Looking at the posterior means, memory for negative behavioral statements was, if at all, slightly higher than memory for positive behavioral statements: $\Delta D = -0.08$ (with a 95% Bayesian credibility interval of $[-0.11, -0.05]$), $\Delta d^{town} = -0.08$ $[-0.18, 0.01]$, and $\Delta d^{group} = -0.04$ $[-0.10, 0.02]$. Moreover, Table 3 shows that the parameters of source memory for group membership were low thus replicating previous findings (Meiser & Hewstone, 2004, 2006).

The parameters of greater interest to research on pseudocontingency inference are the guessing parameters g^{town}, $g^{group}_{|X}$, and $g^{group}_{|Y}$ as they are assumed to reflect informed guessing. If participants guessed the source of a statement at chance, guessing parameters should be at 50%. However, if the parameters are to mirror inferred non-zero item-source contingencies, they should at least differ depending on the statements' desirability. Thus, consistent with the genuine item-source contingency between behavioral statements and town of residence in the stimulus distribution, we tested whether the probability g^{town} of guessing Town X was higher for positive behavioral statements than for negative behavioral statements. Additionally, in line with a pseudocontingency between behavioral statements and group membership within each town of residence, we tested whether the probabilities $g^{group}_{|X}$ and $g^{group}_{|Y}$ of guessing Group A were higher for positive statements in comparison to negative statements. If participants instead based their source guessing on the true contingency between group membership and type of behavior within each town, they should estimate positive behaviors to be less likely for Group A than negative behaviors (i.e., $g_{|X}^{group}$ and $g_{|Y}^{group}$ should be higher for negative statements). We computed the differences between the estimated guessing parameters for positive items and negative items at the group level. The probability of guessing Town X was higher for positive items than for negative items, $\Delta g^{town} = 0.27$ $[0.20, 0.34]$. Likewise, the probability of guessing Group A was higher for positive statements as compared to negative statements which were assigned to Town X, $\Delta g_{|X}^{group} = 0.13$ $[0.03, 0.23]$, and for those that were assigned to Town Y, $\Delta g_{|Y}^{group} = 0.12$ $[0.03, 0.21]$.

Taken together, the direct rating scores and frequency estimates reflected a pseudocontingency effect on source evaluations: whereas judgments of towns corresponded to the genuine contingency between town of residence and behavior, the biased judgments of groups within towns indicated a reliance on the base rates of behavioral statements and group membership within each town. Evaluative judgments were in favor of Town X and in favor of Group A as compared to Town Y and Group B, respectively. Similarly, the (conditional) source-guessing parameters of the hierarchical MPT model differed depending on the desirability of the behavioral statements: replicating previous MPT analyses of aggregate data (Meiser & Hewstone, 2004, 2006), positive behavioral statements were more likely attributed to Town X and Group A. The estimated guessing probabilities paralleled differences in the rating scores and frequency estimates. Therefore, the results corroborate the interpretation of the guessing parameters as reconstructive processes mirroring inferred evaluative contingencies.

Table 3

MPT Model Parameter Estimates (Posterior Means) and 95% Bayesian Credibility Intervals (BCI)

	Group-Level Median (probability scale)		Standard Deviation (probit-scale)	
	Posterior Mean	95% BCI	Posterior Mean	95% BCI
Dn	0.58	[0.55, 0.61]	0.38	[0.30, 0.46]
Dp	0.50	[0.46, 0.53]	0.38	[0.31, 0.47]
bn	0.25	[0.21, 0.29]	0.56	[0.45, 0.69]
bp	0.22	[0.18, 0.26]	0.67	[0.54, 0.80]
dn^{town}	0.50	[0.43, 0.57]	0.43	[0.17, 0.73]
dp^{town}	0.42	[0.35, 0.49]	0.28	[0.03, 0.56]
dn^{group}	0.05	[0.00, 0.11]	0.57	[0.03, 1.40]
dp^{group}	0.01	[0.00, 0.05]	1.03	[0.10, 1.90]
gn^{town}	0.36	[0.32, 0.41]	0.47	[0.35, 0.60]
gp^{town}	0.63	[0.59, 0.68]	0.46	[0.34, 0.58]
$gn^{group}_{\vert X}$	0.52	[0.44, 0.59]	0.84	[0.66; 1.05]
$gp^{group}_{\vert X}$	0.65	[0.60, 0.70]	0.61	[0.48; 0.75]
$gn^{group}_{\vert Y}$	0.40	[0.36, 0.45]	0.55	[0.44; 0.67]
$gp^{group}_{\vert Y}$	0.53	[0.46, 0.59]	0.66	[0.50; 0.85]

Note. D = probability of item recognition; b = probability of guessing that an item is old; d^{town} = probability of remembering town of residence; d^{group} = probability of remembering group membership, whether source memory for town failed or not ($d^{group} = e^{group}$); g^{town} = probability of guessing Town X; $g^{group}_{|X}$ = probability of guessing Group A given that the item was assigned to Town X; $g^{group}_{|Y}$ = probability of guessing Group A given that the item was assigned to Town Y. n and p denote parameter estimates for negative items and positive items, respectively.

3 Parameter Validation Using Continuous Variables

The framework of hierarchical MPT models allows to test this psychological interpretation. Construct validity of MPT model parameters can be tested via functional dissociations in terms of selective relations of specific model parameters with other measures of psychological constructs. So far, the MPT model for crossed source dimensions used here has been validated by fitting it to aggregate data and showing selective influences on specific model parameters. For instance, Meiser and Bröder (2002) reported selective effects of varying item location and item font size on the source-memory parameters d for location and font size, respectively. Meiser and Hewstone (2006), using similar stimuli as in the present experiment, demonstrated that differences between Town X and Town Y with regard to the relative frequency of desirable behaviors in the stimulus distribution were reflected in the guessing probability g^{town}. Furthermore, a quasi-experimental comparison between participants who inferred the true contingency between group membership and behavior versus those who inferred the pseudocontingency indicated that the guessing parameters g^{group}, too, echoed evaluations (Meiser & Hewstone, 2004).

To further establish the parameters' validity, we applied the new approach of parameter validation using hierarchical MPT models. We first analyzed the relations of interindividual differences in source-guessing parameters with interindividual differences in direct measures of source evaluations. For this purpose, we computed differences in individual source evaluations. Differences in rating scores between Town X and Town Y were computed by first averaging individual rating scores across Group A and Group B per town, before subtracting rating scores for Town Y from rating scores for Town X (i.e., reflecting a main effect of town). Differences in rating scores between Group A and Group B were computed separately for Town X and for Town Y (i.e., reflecting simple effects of group for either town) according to the conditional specification of guessing parameters for group membership in the MPT model. The same three evaluative differences were calculated for the estimated proportions of negative behaviors, resulting in a total of six covariates. We fitted separate latent-trait MPT models for each covariate which was included as predictor of the memory parameters and guessing parameters. As higher rating scores indicate more positive evaluations and source-guessing parameters are assumed to reflect source evaluations, we expected interindividual differences in rating scores to have a positive effect on source-guessing probabilities for positive behavioral statements and a negative effect for source guessing of negative statements. The opposite was expected for interindividual differences in frequency estimates, since they reflect estimated proportions of negative behaviors. Put differently, differences in evaluations of Town X and Town Y should be selectively related to the guessing probabilities gn^{town} and gp^{town}. Evaluative differences between Group A and Group B in Town X should specifically predict the guessing parameters $gn^{group}_{|X}$ and $gp^{group}_{|X}$, while differences between groups in Town Y should specifically predict the probabilities $gn^{group}_{|Y}$ and $gp^{group}_{|Y}$. We expected no effects on item memory parameters, source-memory parameters, and the probability of guessing "old". Moreover, we included experimental condition (i.e., whether participants were informed about base rates or not) as additional predictor for the source-guessing parameters to control for any effects of the experimental manipulation. As discussed above, this allowed us to estimate a joint model for both conditions while controlling for effects of experimental condition on the theoretically relevant guessing parameters (see Section 2.2.2). Posterior estimates of the guessing parameters per experimental condition and per estimated MPT model are reported in Table 4.

Furthermore, the framework of hierarchical MPT models, in general, allows to also test parameters' construct validities in a wider context of continuous measures of psychological constructs. Convergent validity is supported if interindividual differences in specific MPT parameters match interindividual differences in related psychological constructs, while discriminant validity is indicated by the specificity of such associations. Therefore, in a final step, we aimed at further establishing construct validity of the MPT model parameters in the context of pseudocontingency research. In exploratory analyses, we analyzed whether interindividual differences in MPT parameters are accounted for by cognitive performance measures which may reflect cognitive processes that are potentially relevant for source monitoring: memory performance, fluid intelligence, and quantitative thinking. Again, we fitted separate latent-trait MPT models for each of the eight co-

Table 4

Posterior Estimates of Source Guessing Parameters per Experimental Condition (1 and 2) and Estimated MPT Model.

	gn^{town}		gp^{town}		$gn^{group}_{\|X}$		$gp^{group}_{\|X}$		$gn^{group}_{\|Y}$		$gp^{group}_{\|Y}$	
MPT model with predictor:	1	2	1	2	1	2	1	2	1	2	1	2
Rating_{XY}	0.36	0.36	0.65	0.61	0.44	0.60	0.68	0.62	0.41	0.39	0.56	0.49
Frequency_{XY}	0.36	0.36	0.65	0.62	0.44	0.61	0.68	0.61	0.41	0.39	0.55	0.50
$\text{Rating}_{AB\|X}$	0.36	0.37	0.65	0.61	0.45	0.59	0.67	0.63	0.41	0.39	0.56	0.49
$\text{Frequency}_{AB\|X}$	0.36	0.37	0.66	0.61	0.45	0.59	0.67	0.63	0.41	0.40	0.56	0.49
$\text{Rating}_{AB\|Y}$	0.35	0.37	0.66	0.61	0.44	0.60	0.68	0.62	0.41	0.40	0.56	0.48
$\text{Frequency}_{AB\|Y}$	0.36	0.37	0.66	0.61	0.44	0.60	0.68	0.62	0.41	0.39	0.56	0.49
Long-term memory	0.35	0.37	0.68	0.59	0.45	0.59	0.65	0.65	0.44	0.37	0.58	0.47
Verbal short-term memory	0.35	0.38	0.68	0.58	0.46	0.58	0.65	0.66	0.44	0.37	0.58	0.47
Fluid intelligence	0.35	0.37	0.68	0.59	0.45	0.59	0.65	0.65	0.44	0.3	0.58	0.47
Figural inductive reasoning	0.35	0.37	0.67	0.59	0.45	0.59	0.64	0.66	0.44	0.37	0.58	0.47
Verbal deductive reasoning	0.35	0.37	0.67	0.59	0.45	0.59	0.65	0.65	0.44	0.37	0.58	0.47
Quantitative reasoning	0.35	0.37	0.68	0.59	0.45	0.59	0.65	0.65	0.44	0.37	0.58	0.47
Computational estimation	0.35	0.37	0.68	0.59	0.45	0.59	0.65	0.65	0.44	0.37	0.58	0.47
Arithmetical competence	0.35	0.37	0.69	0.58	0.45	0.59	0.65	0.65	0.44	0.37	0.58	0.47

Note. XY denotes differences between Town X and Town Y, while $AB|X$ denotes differences between Group A and Group B in Town X and $AB|Y$ denotes differences between Group A and Group B in Town Y. Participants in experimental condition 1 received no information about the groups' base rates per town, while participants in experimental condition 2 did receive base rate information.

variates (four factors and four of their subscales) and included the factor experimental condition as additional predictor of source-guessing parameters to control for potential effects.

Besides estimating the regression slopes, for all models, we computed Bayes factors quantifying the relative support for positive slope parameters, negative slope parameters, or slope parameters unequal to zero against the hypothesis of slope parameters equal to zero. The resulting Bayes factors are indicated by $B_{>,0}$, $B_{<,0}$, and $B_{\neq,0}$, respectively, and were computed using the Savage-Dickey ratio (Heck, 2019; Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010).

3.1 Results

3.1.1 Direct Measures of Source Evaluation as Covariates. Tables 5 and 6 summarize the validation results using the direct measures of source evaluation (i.e., trait ratings and frequency estimates) as predictors. Figures 3 to 5 display the relationships of trait ratings and frequency estimates with respective model parameters. In line with our prediction, we found strong evidence for the relation of differences in rating scores between towns with the guessing probabilities gn^{town} and gp^{town}, $B_{<,0} = 6.60 \cdot 10^4$ and $B_{>,0} = 109.84$, respectively. Contrarily, the rating score differences between towns did not predict either of the other parameters as indicated by Bayes factors between $B_{0,\neq} = 2.28$ and $B_{0,\neq} = 15.16$, with a median of $Md_{B_{0,\neq}} = 6.89$. Likewise, interindividual differences in gn^{town} and gp^{town} were related to interindividual differences in frequency estimates of negative behaviors given towns, $B_{>,0} = 26.32$ and $B_{<,0} = 4.38 \cdot 10^3$, respectively. No other parameters were affected, $1.95 \leq B_{0,\neq} \leq 15.91$, $Md_{B_{0,\neq}} = 5.98$.

Moreover, there was strong support for rating score differences between Group A and Group B in Town X to specifically predict the guessing parameters $gn^{group}_{|X}$, $B_{<,0} = 3.71$, and $gp^{group}_{|X}$, $B_{>,0} = 90.18$. All other parameters were not affected, $1.64 \leq B_{0,\neq} \leq 16.23$, $Md_{B_{0,\neq}} = 7.35$. Likewise, the difference in frequency estimates of negative behaviors between Group A and Group B in Town X only affected $gn^{group}_{|X}$ and $gp^{group}_{|X}$, $B_{>,0} = 54.23$ and $B_{<,0} = 2.24 \cdot 10^5$, respectively. Equivalent results were found for

Table 5

Estimated Slope Parameters, 95% Bayes Credibility Intervals, and Bayes Factors for Rating Score Differences (Rating) and Differences in Estimated Frequencies (Frequency) as Predictors of Source Guessing Parameters

	gn^{town}	gp^{town}	$gn^{group}_{\mid X}$	$gp^{group}_{\mid X}$	$gn^{group}_{\mid Y}$	$gp^{group}_{\mid Y}$
Rating_{XY}	**-0.28 [-0.39, -0.17]**	**0.20 [0.09, 0.31]**	-0.10 [-0.30, 0.09]	-0.02 [-0.15, 0.11]	-0.01 [-0.12, 0.11]	0.02 [-0.15, 0.19]
	$\boldsymbol{B_{<,0} = 6.60 \cdot 10^4}$	$\boldsymbol{B_{>,0} = 109.84}$	$B_{\neq,0} = 0.23$	$B_{\neq,0} = 0.10$	$B_{\neq,0} = 0.08$	$B_{\neq,0} = 0.13$
Frequency_{XY}	**0.18 [0.07, 0.30]**	**-0.25 [-0.36, -0.14]**	-0.07 [-0.27, 0.12]	0.08 [-0.05, 0.21]	-0.01 [-0.13, 0.11]	-0.09 [-0.26, 0.08]
	$\boldsymbol{B_{>,0} = 26.32}$	$\boldsymbol{B_{<,0} = 4.38 \cdot 10^3}$	$B_{\neq,0} = 0.18$	$B_{\neq,0} = 0.20$	$B_{\neq,0} = 0.09$	$B_{\neq,0} = 0.21$
$\text{Rating}_{AB\mid X}$	-0.07 [-0.18, 0.05]	0.10 [-0.02, 0.21]	**-0.21 [-0.40, -0.03]**	**0.23 [0.10, 0.36]**	-0.03 [-0.15, 0.10]	-0.01 [-0.17, 0.14]
	$B_{\neq,0} = 0.16$	$B_{\neq,0} = 0.33$	$\boldsymbol{B_{<,0} = 3.71}$	$\boldsymbol{B_{>,0} = 90.18}$	$B_{\neq,0} = 0.09$	$B_{\neq,0} = 0.11$
$\text{Frequency}_{AB\mid X}$	0.09 [-0.02, 0.20]	-0.06 [-0.17, 0.05]	**0.29 [0.12, 0.47]**	**-0.34 [-0.46, -0.21]**	-0.02 [-0.14, 0.09]	0.10 [-0.05, 0.25]
	$B_{\neq,0} = 0.19$	$B_{\neq,0} = 0.09$	$\boldsymbol{B_{>,0} = 54.23}$	$\boldsymbol{B_{<,0} = 2.24 \cdot 10^5}$	$B_{\neq,0} = 0.07$	$B_{\neq,0} = 0.18$
$\text{Rating}_{AB\mid Y}$	-0.02 [-0.13, 0.09]	0.05 [-0.06, 0.16]	0.13 [-0.05, 0.32]	-0.05 [-0.18, 0.08]	**-0.27 [-0.38, -0.15]**	**0.27 [0.11, 0.42]**
	$B_{\neq,0} = 0.08$	$B_{\neq,0} = 0.11$	$B_{\neq,0} = 0.36$	$B_{\neq,0} = 0.12$	$\boldsymbol{B_{<,0} = 5.11 \cdot 10^3}$	$\boldsymbol{B_{>,0} = 68.34}$
$\text{Frequency}_{AB\mid Y}$	0.06 [-0.05, 0.17]	-0.07 [-0.18, 0.04]	-0.07 [-0.26, 0.12]	-0.03 [-0.17, 0.10]	**0.19 [0.08, 0.31]**	**-0.23 [-0.38, -0.07]**
	$B_{\neq,0} = 0.14$	$B_{\neq,0} = 0.16$	$B_{\neq,0} = 0.17$	$B_{\neq,0} = 0.11$	$\boldsymbol{B_{>,0} = 29.00}$	$\boldsymbol{B_{<,0} = 12.54}$

Note. XY denotes differences between Town X and Town Y, while $AB|X$ denotes differences between Group A and Group B in Town X and $AB|Y$ denotes differences between Group A and Group B in Town Y. Bayesian credibility intervals are reported in brackets. Reported Bayes factors quantify the relative support for positive slope parameters, negative slope parameters or slope parameters unequal to zero against slope parameters equal to zero ($B_{>,0}$, $B_{<,0}$, and $B_{\neq,0}$, respectively).

Table 6

Estimated Slope Parameters, 95% Bayes Credibility Intervals, and Bayes Factors for Rating Score Differences (Rating) and Differences in Estimated Frequencies (Frequency) as Predictors of Memory Parameters and Item Guessing

	Dn	Dp	dn^{town}	dp^{town}	dn^{group}	dp^{group}	bn	bp
Rating_{XY}	-0.02 [-0.10, 0.06]	-0.07 [-0.15, 0.02]	0.05 [-0.12, 0.22]	-0.04 [-0.22, 0.13]	0.10 [-0.29, 0.46]	0.10 [-0.54, 0.62]	-0.04 [-0.17, 0.08]	0.09 [-0.04, 0.23]
	$B_{\neq,0} = 0.07$	$B_{\neq,0} = 0.20$	$B_{\neq,0} = 0.15$	$B_{\neq,0} = 0.14$	$B_{\neq,0} = 0.31$	$B_{\neq,0} = 0.44$	$B_{\neq,0} = 0.11$	$B_{\neq,0} = 0.25$
Frequency_{XY}	-0.01 [-0.10, 0.07]	0.04 [-0.05, 0.13]	0.00 [-0.17, 0.17]	0.07 [-0.11, 0.25]	-0.11 [-0.51, 0.33]	-0.19 [-0.74, 0.45]	0.05 [-0.07, 0.17]	-0.07 [-0.21, 0.06]
	$B_{\neq,0} = 0.06$	$B_{\neq,0} = 0.09$	$B_{\neq,0} = 0.13$	$B_{\neq,0} = 0.17$	$B_{\neq,0} = 0.34$	$B_{\neq,0} = 0.51$	$B_{\neq,0} = 0.12$	$B_{\neq,0} = 0.17$
$\text{Rating}_{AB\|X}$	-0.02 [-0.10, 0.07]	-0.01 [-0.09, 0.08]	-0.13 [-0.31, 0.05]	0.01 [-0.16, 0.17]	0.21 [-0.28, 0.77]	-0.26 [-0.82, 0.30]	0.04 [-0.08, 0.16]	0.07 [-0.07, 0.21]
	$B_{\neq,0} = 0.06$	$B_{\neq,0} = 0.06$	$B_{\neq,0} = 0.34$	$B_{\neq,0} = 0.12$	$B_{\neq,0} = 0.51$	$B_{\neq,0} = 0.61$	$B_{\neq,0} = 0.11$	$B_{\neq,0} = 0.16$
$\text{Frequency}_{AB\|X}$	0.03 [-0.05, 0.11]	-0.01 [-0.10, 0.07]	0.01 [-0.15, 0.17]	-0.12 [-0.29, 0.05]	-0.23 [-0.76, 0.18]	0.20 [-0.32, 0.75]	-0.04 [-0.16, 0.08]	-0.09 [-0.22, 0.03]
	$B_{\neq,0} = 0.08$	$B_{\neq,0} = 0.06$	$B_{\neq,0} = 0.12$	$B_{\neq,0} = 0.30$	$B_{\neq,0} = 0.55$	$B_{\neq,0} = 0.51$	$B_{\neq,0} = 0.11$	$B_{\neq,0} = 0.26$
$\text{Rating}_{AB\|Y}$	-0.01 [-0.10, 0.07]	-0.03 [-0.12, 0.05]	0.03 [-0.15, 0.20]	0.03 [-0.14, 0.21]	0.21 [-0.25, 0.71]	0.22 [-0.39, 0.77]	0.05 [-0.07, 0.17]	0.05 [-0.09, 0.18]
	$B_{\neq,0} = 0.06$	$B_{\neq,0} = 0.08$	$B_{\neq,0} = 0.13$	$B_{\neq,0} = 0.13$	$B_{\neq,0} = 0.49$	$B_{\neq,0} = 0.54$	$B_{\neq,0} = 0.12$	$B_{\neq,0} = 0.12$
$\text{Frequency}_{AB\|Y}$	0.04 [-0.04, 0.13]	0.04 [-0.05, 0.13]	0.10 [-0.08, 0.30]	0.00 [-0.18, 0.17]	-0.12 [-0.59, 0.31]	0.08 [-0.45, 0.63]	-0.11 [-0.23, 0.02]	-0.12 [-0.27, 0.02]
	$B_{\neq,0} = 0.10$	$B_{\neq,0} = 0.09$	$B_{\neq,0} = 0.24$	$B_{\neq,0} = 0.13$	$B_{\neq,0} = 0.36$	$B_{\neq,0} = 0.39$	$B_{\neq,0} = 0.36$	$B_{\neq,0} = 0.42$

Note. XY denotes differences between Town X and Town Y, while $AB|X$ denotes differences between Group A and Group B in Town X and $AB|Y$ denotes differences between Group A and Group B in Town Y. Bayesian credibility intervals are reported in brackets. Reported Bayes factors quantify the relative support for slope parameters unequal to zero against slope parameters equal to zero.

differences between Group A and Group B in Town Y: the rating score difference between Group A and Group B in Town Y selectively predicted $gn^{group}_{|Y}$, $B_{<,0} = 5.11 \cdot 10^3$, and $gp^{group}_{|Y}$, $B_{>,0} = 68.34$, but no other model parameters, $1.87 \leq B_{0,\neq} \leq 16.31$, $Md_{B_{0,\neq}} = 8.44$. Finally, the difference in estimated frequencies between Group A and Group B in Town Y predicted the probability to guess Group A in Town Y for negative behaviors ($gn^{town}_{|Y}$), $B_{>,0} = 29.00$, as well as for positive behaviors ($gp^{town}_{|Y}$) , $B_{<,0} = 12.54$. All other parameters were again not affected, $2.40 \leq B_{0,\neq} \leq 11.66$, $Md_{B_{0,\neq}} = 6.02$.

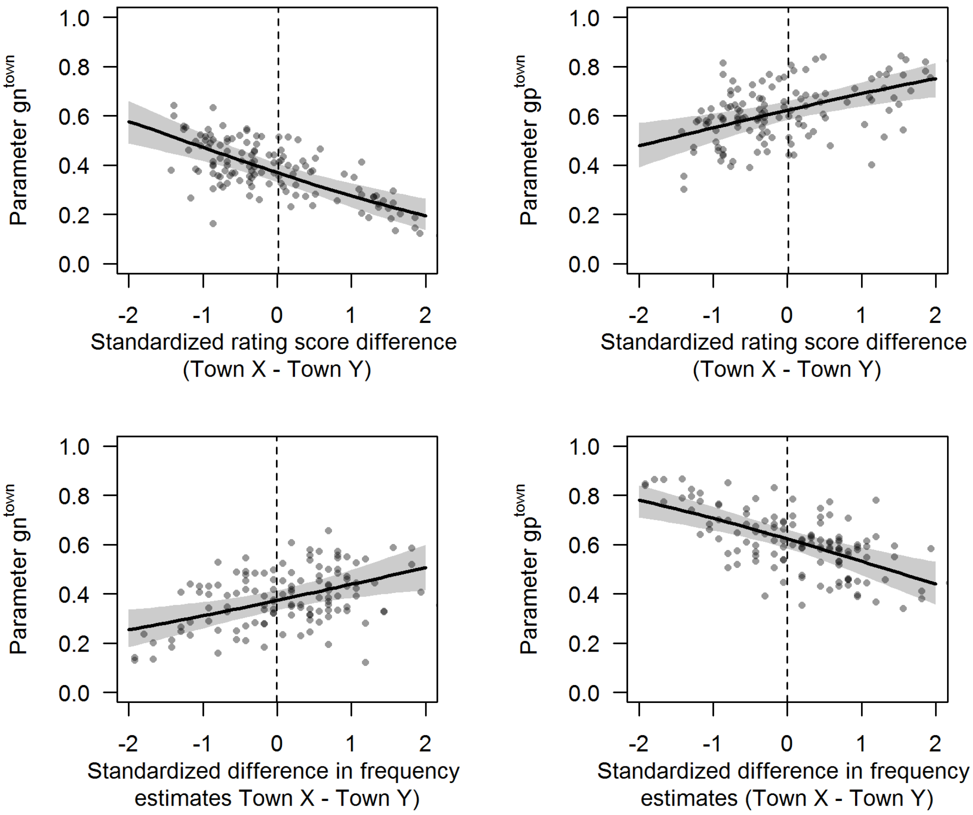

Figure 3. Regression of MPT model parameters gn^{town} and gp^{town} on direct measures of source evaluations: The upper panels show the covariations with trait rating differences between Town X and Town Y, while the lower panels show the covariations with the difference between towns in estimated frequencies of negative behavior. The solid lines depict the posterior median of the prediction function including the respective direct measure as predictor of gn^{town} or gp^{town}. Corresponding 95% Bayesian credibility intervals are shown in gray. Individual-level predictions are depicted by the gray points. Vertical dashed lines show the group mean of the respective covariate.

Overall, the results show selective covariations of interindividual differences in (conditional) source-guessing parameters with interindividual differences in continuous measures of evaluations of either towns or groups. All remaining MPT model parameters, including other source-guessing parameters, were not affected, thus indicating not only convergent validity, but also discriminant validity. This suggests that the source-guessing parameters can indeed be interpreted as indicators of inferred evaluative contingencies.

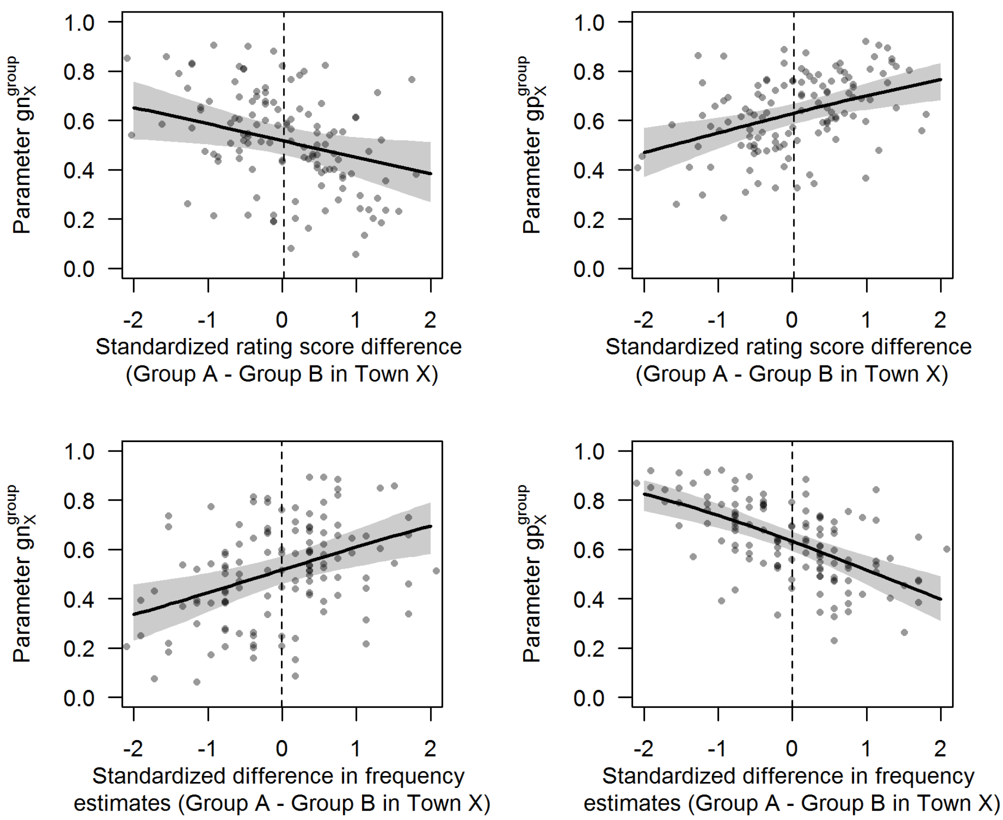

Figure 4. Regression of MPT model parameters $gn_{|X}^{group}$ and $gp_{|X}^{group}$ on direct measures of source evaluations: The upper panels show the covariations with trait rating differences between Group A and Group B in Town X, while the lower panels show the covariations with differences in estimated frequencies of negative behavior between Group A and Group B in Town X. The solid lines depict the posterior median of the prediction function including the respective direct measure as predictor of $gn_{|X}^{group}$ or $gp_{|X}^{group}$. Corresponding 95% Bayesian credibility intervals are shown in gray. Individual-level predictions are depicted by the gray points. Vertical dashed lines show the group mean of the respective covariate.

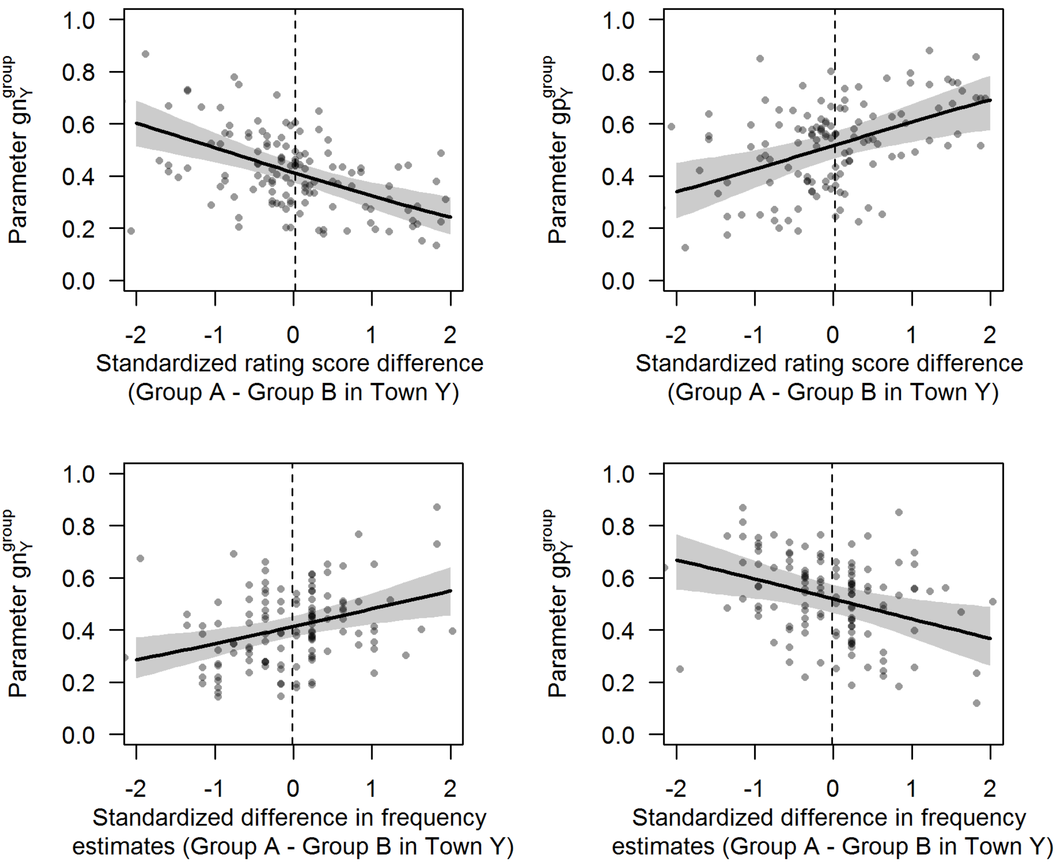

Figure 5. Regression of MPT model parameters $gn_{|Y}^{group}$ and $gp_{|Y}^{group}$ on direct measures of source evaluations: The upper panels show the covariations with trait rating differences between Group A and Group B in Town Y, while the lower panels show the covariations with differences in estimated frequencies of negative behavior between Group A and Group B in Town Y. The solid lines depict the posterior median of the prediction function including the respective direct measure as predictor of $gn_{|Y}^{group}$, or $gp_{|Y}^{group}$. Corresponding 95% Bayesian credibility intervals are shown in gray. Individual-level predictions are depicted by the gray points. Vertical dashed lines show the group mean of the respective covariate.

3.1.2 Cognitive Performance Measures as Covariates. Tables 7 and 8 summarize the results of testing interindividual differences in the cognitive performance measures as predictors of differences in MPT model parameters. First, analyzing memory performance, there was only support for an effect of long-term memory on item memory for positive behavioral statements (parameter Dp), $B_{\neq,0} = 5.33$, and for an effect of verbal short-term memory on the probability to remember the source dimension town of residence for negative items (parameter dn^{group}), $B_{\neq,0} = 22.52$. All other $B_{\neq,0} \leq 1.11$ and $B_{\neq,0} \leq 1.04$ for long-term memory and verbal short-term memory, respectively, indicating no support for an influence on any other memory parameter or guessing parameter. Second, we fitted hierarchical MPT models with figural inductive reasoning scores and verbal deductive reasoning scores as well as a total score of fluid intelligence as predictors of model parameters. In our analyses, we found no evidence for an effect of figural inductive reasoning, $B_{\neq,0} \leq 0.52$, verbal deductive reasoning, $B_{\neq,0} \leq 1.26$, or the total score of fluid intelligence, $B_{\neq,0} \leq 0.78$. Finally, we also found no support that quantitative thinking or its subscales predict any MPT model parameter, $B_{\neq,0} \leq 0.58$.

4 Discussion

Traditionally, parameters of a multinomial processing tree (MPT) model are experimentally validated by means of discrete experimental manipulations: the selective influence of certain experimental conditions on a specific model parameter is tested by setting this parameter equal across experimental conditions. Parameter validity is then analyzed by subsequently comparing the resulting model to a model with separate parameters for each condition (e.g., Bayen et al., 1996). Alternatively or additionally, correlational analyses can be conducted based on parameter estimates that are obtained separately for each individual in a first step: selective covariations can then be assessed between the parameter estimates and measures of the assumed psychological process or other theoretically relevant constructs or behavior (e.g., Meissner & Rothermund, 2013). This two-step procedure requires, however, that model parameters are estimated at the level of individual participants with sufficient reliability which is often not feasible due to the small number of observations per participant. The rationale of selective covariations can also be applied in the recently developed framework of hierarchical MPT models that accommodate parameter heterogeneity across participants by assuming random effects for individuals. Arnold et al. (2013), for instance, estimated a source-monitoring model using a hierarchical beta-MPT approach (Smith & Batchelder, 2010) to account

Table 7

Estimated Slope Parameters, 95% Bayes Credibility Intervals, and Bayes Factors for Cognitive Performance Measures as Predictors of Source Guessing Parameters

	gn^{town}	gp^{town}	$gn^{group}_{\|X}$	$gp^{group}_{\|X}$	$gn^{group}_{\|Y}$	$gp^{group}_{\|Y}$
Long-term memory	0.01 [-0.11, 0.12]	-0.02 [-0.12, 0.09]	0.10 [-0.08, 0.28]	-0.01 [-0.13, 0.11]	0.05 [-0.06, 0.17]	-0.07 [-0.23, 0.09]
	$B_{\neq,0} = 0.08$	$B_{\neq,0} = 0.08$	$B_{\neq,0} = 0.23$	$B_{\neq,0} = 0.09$	$B_{\neq,0} = 0.13$	$B_{\neq,0} = 0.17$
Verbal short-term memory	-0.02 [-0.15, 0.10]	0.04 [-0.06, 0.15]	0.16 [-0.03, 0.34]	-0.04 [-0.17, 0.09]	-0.03 [-0.15, 0.08]	0.01 [-0.15, 0.17]
	$B_{\neq,0} = 0.10$	$B_{\neq,0} = 0.11$	$B_{\neq,0} = 0.54$	$B_{\neq,0} = 0.11$	$B_{\neq,0} = 0.10$	$B_{\neq,0} = 0.11$
Fluid intelligence	-0.07 [-0.19, 0.04]	0.02 [-0.09, 0.14]	-0.01 [-0.20, 0.17]	-0.03 [-0.16, 0.11]	-0.01 [-0.13, 0.10]	0.07 [-0.09, 0.22]
	$B_{\neq,0} = 0.18$	$B_{\neq,0} = 0.09$	$B_{\neq,0} = 0.14$	$B_{\neq,0} = 0.10$	$B_{\neq,0} = 0.09$	$B_{\neq,0} = 0.16$
Figural inductive reasoning	-0.01 [-0.13, 0.11]	-0.01 [-0.12, 0.11]	0.00 [-0.18, 0.19]	-0.06 [-0.20, 0.07]	0.03 [-0.09, 0.15]	0.02 [-0.14, 0.17]
	$B_{\neq,0} = 0.09$	$B_{\neq,0} = 0.08$	$B_{\neq,0} = 0.13$	$B_{\neq,0} = 0.15$	$B_{\neq,0} = 0.09$	$B_{\neq,0} = 0.11$
Verbal deductive reasoning	-0.10 [-0.22, 0.01]	0.04 [-0.07, 0.15]	0.01 [-0.18, 0.20]	0.02 [-0.11, 0.15]	-0.04 [-0.16, 0.08]	0.08 [-0.09, 0.24]
	$B_{\neq,0} = 0.37$	$B_{\neq,0} = 0.10$	$B_{\neq,0} = 0.14$	$B_{\neq,0} = 0.10$	$B_{\neq,0} = 0.11$	$B_{\neq,0} = 0.18$
Quantitative reasoning	0.00 [-0.12, 0.11]	0.07 [-0.04, 0.18]	0.02 [-0.16, 0.20]	-0.01 [-0.14, 0.12]	-0.03 [-0.15, 0.09]	-0.08 [-0.24, 0.07]
	$B_{\neq,0} = 0.08$	$B_{\neq,0} = 0.17$	$B_{\neq,0} = 0.13$	$B_{\neq,0} = 0.10$	$B_{\neq,0} = 0.10$	$B_{\neq,0} = 0.19$
Computational estimation	0.00 [-0.11, 0.12]	0.03 [-0.08, 0.14]	0.07 [-0.10, 0.25]	0.00 [-0.14, 0.13]	-0.04 [-0.16, 0.07]	-0.11 [-0.27, 0.04]
	$B_{\neq,0} = 0.08$	$B_{\neq,0} = 0.10$	$B_{\neq,0} = 0.18$	$B_{\neq,0} = 0.10$	$B_{\neq,0} = 0.11$	$B_{\neq,0} = 0.31$
Arthimetical competence	-0.02 [-0.14, 0.09]	0.10 [0.00, 0.21]	0.00 [-0.18, 0.18]	-0.03 [-0.16, 0.10]	-0.01 [-0.13, 0.10]	-0.04 [-0.20, 0.12]
	$B_{\neq,0} = 0.09$	$B_{\neq,0} = 0.47$	$B_{\neq,0} = 0.13$	$B_{\neq,0} = 0.10$	$B_{\neq,0} = 0.09$	$B_{\neq,0} = 0.13$

Note. Bayesian credibility intervals are reported in brackets. Reported Bayes factors quantify the relative support for slope parameters unequal to zero against slope parameters equal to zero.

Table 8

Estimated Slope Parameters, 95% Bayes Credibility Intervals, and Bayes Factors for Cognitive Performance Measures as Predictors of Memory Parameters and Item Guessing

	Dn	Dp	dn^{town}	dp^{town}	dn^{group}	dp^{group}	bn	bp
Long-term memory	0.10 [0.02, 0.19]	**0.13 [0.05, 0.21]**	-0.13 [-0.31, 0.03]	0.01 [-0.16, 0.18]	0.24 [-0.14, 0.67]	0.17 [-0.35, 0.70]	-0.02 [-0.14, 0.11]	0.02 [-0.12, 0.16]
	$B_{\neq,0} = 1.11$	$\boldsymbol{B_{\neq,0} = 5.33}$	$B_{\neq,0} = 0.41$	$B_{\neq,0} = 0.12$	$B_{\neq,0} = 0.57$	$B_{\neq,0} = 0.46$	$B_{\neq,0} = 0.09$	$B_{\neq,0} = 0.10$
Verbal short-term memory	0.07 [-0.02, 0.15]	0.07 [-0.01, 0.16]	**0.25 [0.10, 0.41]**	0.11 [-0.05, 0.27]	0.39 [-0.23, 0.92]	0.26 [-0.37, 0.84]	0.02 [-0.10, 0.14]	0.00 [-0.14, 0.13]
	$B_{\neq,0} = 0.20$	$B_{\neq,0} = 0.28$	$\boldsymbol{B_{\neq,0} = 22.52}$	$B_{\neq,0} = 0.30$	$B_{\neq,0} = 1.04$	$B_{\neq,0} = 0.62$	$B_{\neq,0} = 0.09$	$B_{\neq,0} = 0.10$
Fluid intelligence	0.05 [-0.03, 0.14]	0.05 [-0.03, 0.14]	0.04 [-0.13, 0.21]	0.10 [-0.06, 0.27]	0.30 [-0.14, 0.78]	0.02 [-0.53, 0.56]	0.01 [-0.11, 0.13]	-0.01 [-0.15, 0.12]
	$B_{\neq,0} = 0.13$	$B_{\neq,0} = 0.12$	$B_{\neq,0} = 0.14$	$B_{\neq,0} = 0.25$	$B_{\neq,0} = 0.78$	$B_{\neq,0} = 0.39$	$B_{\neq,0} = 0.09$	$B_{\neq,0} = 0.10$
Figural inductive reasoning	-0.01 [-0.09, 0.07]	0.05 [-0.04, 0.13]	0.15 [-0.02, 0.33]	-0.01 [-0.18, 0.16]	0.21 [-0.24, 0.63]	-0.09 [-0.64, 0.43]	0.01 [-0.11, 0.13]	-0.01 [-0.14, 0.13]
	$B_{\neq,0} = 0.06$	$B_{\neq,0} = 0.11$	$B_{\neq,0} = 0.52$	$B_{\neq,0} = 0.12$	$B_{\neq,0} = 0.49$	$B_{\neq,0} = 0.40$	$B_{\neq,0} = 0.09$	$B_{\neq,0} = 0.10$
Verbal deductive reasoning	0.09 [0.01, 0.18]	0.03 [-0.06, 0.11]	-0.05 [-0.22, 0.12]	0.18 [0.02, 0.35]	0.22 [-0.18, 0.61]	0.14 [-0.36, 0.62]	0.01 [-0.11, 0.13]	-0.02 [-0.16, 0.11]
	$B_{\neq,0} = 0.64$	$B_{\neq,0} = 0.08$	$B_{\neq,0} = 0.14$	$B_{\neq,0} = 1.26$	$B_{\neq,0} = 0.51$	$B_{\neq,0} = 0.41$	$B_{\neq,0} = 0.09$	$B_{\neq,0} = 0.10$
Quantitative reasoning	-0.01 [-0.09, 0.07]	0.00 [-0.09, 0.08]	0.05 [-0.10, 0.21]	-0.04 [-0.20, 0.12]	0.06 [-0.36, 0.53]	-0.24 [-0.79, 0.29]	0.01 [-0.12, 0.13]	-0.02 [-0.16, 0.11]
	$B_{\neq,0} = 0.06$	$B_{\neq,0} = 0.06$	$B_{\neq,0} = 0.14$	$B_{\neq,0} = 0.13$	$B_{\neq,0} = 0.32$	$B_{\neq,0} = 0.58$	$B_{\neq,0} = 0.09$	$B_{\neq,0} = 0.10$
Computational estimation	-0.01 [-0.09, 0.08]	0.01 [-0.07, 0.10]	0.05 [-0.11, 0.21]	0.04 [-0.13, 0.20]	0.03 [-0.44, 0.48]	-0.21 [-0.79, 0.32]	0.02 [-0.11, 0.14]	0.00 [-0.14, 0.13]
	$B_{\neq,0} = 0.06$	$B_{\neq,0} = 0.06$	$B_{\neq,0} = 0.14$	$B_{\neq,0} = 0.13$	$B_{\neq,0} = 0.33$	$B_{\neq,0} = 0.53$	$B_{\neq,0} = 0.09$	$B_{\neq,0} = 0.10$
Arthimetical competence	0.01 [-0.08, 0.09]	0.00 [-0.09, 0.09]	0.03 [-0.13, 0.20]	-0.14 [-0.31, 0.02]	0.09 [-0.39, 0.58]	-0.22 [-0.76, 0.30]	0.00 [-0.13, 0.12]	-0.06 [-0.20, 0.08]
	$B_{\neq,0} = 0.06$	$B_{\neq,0} = 0.06$	$B_{\neq,0} = 0.13$	$B_{\neq,0} = 0.48$	$B_{\neq,0} = 0.36$	$B_{\neq,0} = 0.53$	$B_{\neq,0} = 0.09$	$B_{\neq,0} = 0.14$

Note. Bayesian credibility intervals are reported in brackets. Reported Bayes factors quantify the relative support for slope parameters unequal to zero against slope parameters equal to zero.

for interindividual differences in model parameters. In a second step, they investigated the correlation of estimated model parameters with manifest measures of perceived item-source contingencies.

Meanwhile, hierarchical MPT models (Klauer, 2010) enable researchers to estimate covariations between model parameters and extraneous measures simultaneously with the MPT model parameters in a one-step procedure (for illustrations, see Klein et al., 2017 and Michalkiewicz et al., 2018). Thereby, using hierarchical MPT models even allows the assessment of convergent and discriminant parameter validity by testing specific continuous variables as selective predictors of interindividual differences in specific MPT model parameters.[6] By integrating cognitive modeling through MPT models and psychometric assessment of interindividual differences, the novel approach to parameter validation provides an instantiation of the perspective of "cognitive psychometrics" that was envisaged by William Batchelder (1998). In the present article, we applied this new procedure of parameter validation to data collected in an experiment on biased stereotype formation based on pseudocontingency inferences.

In the experiment, positive behavioral statements and negative behavioral statements were presented in combination with information on two source dimensions, group membership (Group A vs. Group B) and town of residence (Town X vs. Town Y). Base rates of behavioral statements and group membership were skewed and co-varied across the source dimension town of residence, which should give rise to the inference of a pseudocontingency between behavioral statements and group membership. Source evaluations were assessed by trait ratings, frequency estimates, and behavior assignments. Replicating earlier results (Meiser & Hewstone, 2004, 2006), participants' trait ratings and frequency estimates for Town X and Town Y mirrored the genuine item-source contingency as indicated by more positive evaluations of Town X. Furthermore, indicating

[6] With this approach, differences in model parameters between individuals are explained by the continuous variable and additional between-subjects variance. Alternatively, one could drop the random effect and assume that all of the variance in individual MPT parameters is explained by the covariates (Coolin, Erdfelder, Bernstein, Thornton, & Thornton, 2015). Yet, in our case, it seemed theoretically implausible that the covariates perfectly explained interindividual differences in MPT model parameters.

a pseudocontingency, Group A was evaluated more positively than Group B within both towns of residence. The data on behavior assignments were analyzed with a multidimensional source-monitoring MPT model (Meiser & Bröder, 2002) using the latent-trait hierarchical MPT approach (Klauer, 2010) implemented in the R-package TreeBUGS (Heck, Arnold, & Arnold, 2018). The analyses revealed better item memory for negative items than for positive items. Moreover, not only the probability to guess Town X versus Town Y was higher for positive items as compared with negative items, but also the probability to guess Group A versus Group B, even though the genuine contingency between items and the source dimension group membership was opposite in sign.

Contrary to traditional research on source monitoring, the present research mainly focused on guessing processes as opposed to memory processes: the guessing parameters are assumed to reflect evaluative biases originating in inferred item-source (pseudo-) contingencies. Thus, besides the validity of the memory parameters, the substantive interpretation of source-guessing parameters is essential. Assessing construct validity, we tested the selective relationship between direct measures of source evaluations (i.e., rating scores and frequency estimates) and corresponding guessing parameters. To this end, we included the participants' evaluations as predictors of all hierarchical MPT model parameters. Indicating convergent validity, differences in evaluations between Town X and Town Y only accounted for differences in the probability of guessing Town X versus Town Y. Likewise, differences in evaluations between Group A and Group B per town of residence only predicted differences in the probability of guessing Group A versus Group B given the respective town. Each of the remaining guessing and memory parameters was not associated with the continuous covariates, thereby indicating discriminant validity.

In a further step, we analyzed whether interindividual differences in MPT model parameters may be accounted for by interindividual differences in cognitive performance measures. Apart from long-term memory predicting item memory for negative items and short-term memory predicting source memory for town of residence of negative items, there were no effects of cognitive measures on the MPT model parameters. The results suggest that at least guessing processes based on pseudocontingency inferences are rather independent of general cognitive performance as assessed by the cognitive tests.

To conclude, following William Batchelder's idea of "cognitive psychometrics" (Batchelder, 1998, 2010; Riefer et al., 2002), hierarchical MPT models provide a new approach to parameter validation through the combination of cognitive modeling in experimental psychology and the psychometric assessment of relevant constructs. Being a powerful tool in itself, MPT models thereby offer additional means of measurement validation and theory testing in a nomological network. Last but not least, assessing parameter validity with continuous measures of interindividual differences in criterion constructs may also be fruitful for gaining a better theoretical and empirical understanding of the cognitive processes and psychological phenomena that are captured by MPT parameters.

References

Anders, R., Oravecz, Z., & Alario, F.-X. (2017). Improved information pooling for hierarchical cognitive models through multiple and covaried regression. *Behavior Research Methods*, 1–22. doi: 10.3758/s13428-017-0921-7

Arendasy, M., Hornke, L. F., Sommer, M., Häusler, J., Wagner-Menghin, M., Gittler, G., ... Wenzl, M. (2009). *INSBAT – Intelligenz-Struktur-Batterie.* Mödling, Austria: Schuhfried.

Arnold, N. R., Bayen, U. J., Kuhlmann, B. G., & Vaterrodt, B. (2013). Hierarchical modeling of contingency-based source monitoring: A test of the probability-matching account. *Psychonomic Bulletin & Review*, *20*(2), 326–333. doi: 10.3758/s13423-012-0342-7

Batchelder, W. H. (1998). Multinomial processing tree models and psychological assessment. *Psychological Assessment*, *10*(4), 331-344. doi: 10.1037/1040-3590.10.4.331

Batchelder, W. H. (2010). Cognitive psychometrics: Using multinomial processing tree models as measurement tools. In S. E. Embretson & S. E. Embretson (Ed) (Eds.), *Measuring psychological constructs: Advances in model-based approaches.* (pp. 71–93). Washington, DC, US: American Psychological Association. doi: 10.1037/12074-004

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, *97*, 548-564. doi: 10.1037/0033-295X.97.4.548

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*(1), 57–86. doi: 10.3758/BF03210812

Bayen, U. J., & Kuhlmann, B. G. (2011). Influences of source-item contingency and schematic knowledge on source monitoring: Tests of the probability-matching account. *Journal of Memory and Language*, *64*(1), 1–17. doi: 10.1016/j.jml.2010.09.001

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology:*

Learning Memory and Cognition, *22*(1), 197–215. doi: 10.1037/0278-7393.22.1.197

Bayen, U. J., Nakamura, G. V., Dupuis, S. E., & Yang, C.-L. (2000). The use of schematic knowledge about sources in source monitoring. *Memory & Cognition*, *28*(3), 480-500. doi: 10.3758/BF03198562

Boehm, U., Steingroever, H., & Wagenmakers, E. J. (2018). Using Bayesian regression to test hypotheses about relationships between parameters and covariates in cognitive models. *Behavior Research Methods*, *50*(3), 1248–1269. doi: 10.3758/s13428-017-0940-4

Bröder, A., & Meiser, T. (2007). Measuring source memory. *Zeitschrift für Psychologie/Journal of Psychology*, *215*(1), 52-60. doi: 10.1027/0044-3409.215.1.52

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81-105. doi: 10.1037/h0046016

Coolin, A., Erdfelder, E., Bernstein, D. M., Thornton, A. E., & Thornton, W. L. (2015). Explaining individual differences in cognitive processes underlying hindsight bias. *Psychonomic Bulletin & Review*, *22*(2), 328–348. doi: 10.3758/s13423-014-0691-5

Ehrenberg, K., & Klauer, K. C. (2005). Flexible use of source information: Processing components of the inconsistency effect in person memory. *Journal of Experimental Social Psychology*, *41*(4), 369-387. doi: 10.1016/j.jesp.2004.08.001

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Journal of Psychology*, *217*(3), 108–124. doi: 10.1027/0044-3409.217.3.108

Erdfelder, E., & Buchner, A. (1998). Decomposing the hindsight bias: A multinomial processing tree model for separating recollection and reconstruction in hindsight. *Journal of Experimental Psychology: Learning Memory and Cognition*, *24*(2), 387–414. doi: 10.1037/0278-7393.24.2.387

Ernst, H. M., Kuhlmann, B. G., & Vogel, T. (2019). The origin of illusory correlations. *Experimental Psychology*, *66*(3), 195–206. doi: 10.1027/1618-3169/a000444

Fiedler, K., & Freytag, P. (2004). Pseudocontingencies. *Journal of Personality and Social*

Psychology, *87*(4), 453-467. doi: 10.1037/0022-3514.87.4.453

Fiedler, K., Freytag, P., & Meiser, T. (2009). Pseudocontingencies: An integrative account of an intriguing cognitive illusion. *Psychological Review*, *116*(1), 187–206. doi: 10.1037/a0014480

Fiedler, K., Kutzner, F., & Vogel, T. (2013). Pseudocontingencies: Logically unwaranted but smart inferences. *Current Directions in Psychological Science*, *22*(4), 325-329. doi: 10.1177/0963721413480171

Fleig, H., Meiser, T., Ettlin, F., & Rummel, J. (2017). Statistical numeracy as a moderator of (pseudo)contingency effects on decision behavior. *Acta Psychologica*, *174*, 68–79. doi: 10.1016/j.actpsy.2017.01.002

Gronau, Q. F., Wagenmakers, E. J., Heck, D. W., & Matzke, D. (2019). A simple method for comparing complex models: Bayesian model comparison for hierarchical multinomial processing tree models using warp-III bridge sampling. *Psychometrika*, *84*(1), 261–284. doi: 10.1007/s11336-018-9648-3

Heck, D. W. (2019). A caveat on the Savage-Dickey density ratio: The case of computing Bayes factors for regression parameters. *British Journal of Mathematical and Statistical Psychology*, *72*, 316-333. doi: 10.1111/bmsp.12150

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, *50*(1), 264–284. doi: 10.3758/s13428-017-0869-7

Heck, D. W., Thielmann, I., Moshagen, M., & Hilbig, B. E. (2018). Who lies? A large-scale reanalysis linking basic personality traits to unethical decision making. *Judgment and Decision Making*, *13*(4), 356–371.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, *14*(4), 382–401. doi: 10.1214/ss/1009212519

Hütter, M., & Klauer, K. C. (2016). Applying processing trees in social psychology. *European Review of Social Psychology*, *27*(1), 116–159. doi: 10.1080/10463283.2016.1212966

Klauer, K. C. (2006). Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika*, *71*(1), 7–31. doi: 10.1007/S11336-004-1188-3

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, *75*(1), 70–98. doi: 10.1007/s11336-009-9141-0

Klauer, K. C., & Meiser, T. (2000). A source-monitoring analysis of illusory correlations. *Personality and Social Psychology Bulletin*, *26*(9), 1074–1093. doi: 10.1177/01461672002611005

Klauer, K. C., & Wegener, I. (1998). Unraveling social categorization in the "who said what?" paradigm. *Journal of Personality and Social Psychology*, *75*(5), 1155–1178. doi: 10.1037/0022-3514.75.5.1155

Klein, S. A., Hilbig, B. E., & Heck, D. W. (2017). Which is the greater good? A social dilemma paradigm disentangling environmentalism and cooperation. *Journal of Environmental Psychology*, *53*, 40-49. doi: 10.1016/j.jenvp.2017.06.001

Kroneisen, M., & Heck, D. W. (in press). Interindividual differences in the sensitivity for consequences, moral norms, and preferences for inaction: Relating basic personality traits to the CNI model. *Personality and Social Psychology Bulletin*. doi: 10.1177/0146167219893994

Kuhlmann, B. G., Bayen, U. J., Meuser, K., & Kornadt, A. E. (2016). The impact of age stereotypes on source monitoring in younger and older adults. *Psychology and Aging*, *31*(8), 875–889. doi: 10.1037/pag0000140

Matzke, D., Dolan, C. V., Batchelder, W. H., & Wagenmakers, E.-J. (2015). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika*, *80*(1), 205-235. doi: 10.1007/s11336-013-9374-9

Meiser, T. (2003). Effects of processing strategy on episodic memory and contingency learning in group stereotype formation. *Social Cognition*, *21*(2), 121–156. doi: 10.1521/soco.21.2.121.21318

Meiser, T., & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning Memory and Cognition*, *28*(1), 116–137. doi: 10.1037/0278-7393.28.1.116

Meiser, T., & Hewstone, M. (2004). Cognitive processes in stereotype formation: The role of correct contingency learning for biased group judgments. *Journal of Personality and Social Psychology*, *87*(5), 599–614. doi: 10.1037/0022-3514.87.5.599

Meiser, T., & Hewstone, M. (2006). Illusory and spurious correlations: Distinct phenomena or joint outcomes of exemplar-based category learning? *European Journal of Social Psychology*, *36*(3), 315–336. doi: 10.1002/ejsp.304

Meiser, T., & Hewstone, M. (2010). Contingency learning and stereotype formation: Illusory and spurious correlations revisited. *European Review of Social Psychology*, *21*(1), 285–331. doi: 10.1080/10463283.2010.543308

Meiser, T., Rummel, J., & Fleig, H. (2018). Pseudocontingencies and choice behavior in probabilistic environments with context-dependent outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(1), 50–67. doi: 10.1037/xlm0000432

Meissner, F., & Rothermund, K. (2013). Estimating the contributions of associations and recoding in the implicit associal test: The ReAL model of the IAT. *Journal of Personality and Social Psychology*, *104*(1), 45-69. doi: 10.1037/a0030734

Michalkiewicz, M., Arden, K., & Erdfelder, E. (2018). Do smarter people employ better decision strategies? The influence of intelligence on adaptive use of the recognition heuristic. *Journal of Behavioral Decision Making*, *31*(3), 3-11. doi: 10.1002/bdm.2040

Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of Bayes factors for common designs [Computer software manual]. Retrieved from `https://cran.r-project.org/package=BayesFactor`

R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.r-project.org/`

Riefer, D. M., Knapp, B. R., Batchelder, W. H., Bamber, D., & Manifold, V. (2002). Cognitive psychometrics: Assessing storage and retrieval deficits in special populations with multinomial processing tree models. *Psychological Assessment*, *14*, 184–201. doi: 10.1037/1040-3590.14.2.184

Rouder, J. N., Lu, J., Sun, D., Speckman, P., Morey, R., & Naveh-Benjamin, M. (2007). Signal detection models with random participant and item effects. *Psychomtrika*, *72*(4), 621-642. doi: 10.1007/s11336-005-1350-6

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*, 356-374. doi: 10.1016/j.jmp.2012.08.001

Smith, J. B., & Batchelder, W. H. (2010). Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology*, *54*(1), 167–183. doi: 10.1016/j.jmp.2009.06.007

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*, 158-189. doi: 10.1016/j.cogpsych.2009.12.001

Wegener, I., & Klauer, K. C. (2004). Inter-category versus intra-category fit: When social categories match social context. *European Journal of Social Psychology*, *34*(5), 567-593. doi: 10.1002/ejsp.217

Appendix

Bayes Factors of Bayesian ANOVA analyzing rating scores and frequency estimates

See Tables A1 and A2.

Table A1

Results of Bayesian 2 (condition)× 2 (town of residence) × 2 (group membership) mixed ANOVAs analyzing mean ratings scores

Model	BayesFactor ($B_{model,null}$)	posterior probability
group	$1.75 \cdot 10^{5}$	$1.74 \cdot 10^{-30}$
condition	0.11	$1.08 \cdot 10^{-36}$
group + condition	$1.94 \cdot 10^{4}$	$1.85 \cdot 10^{-31}$
group + condition + group:condition	$3.39 \cdot 10^{3}$	$3.19 \cdot 10^{-32}$
town	$3.94 \cdot 10^{27}$	$3.94 \cdot 10^{-08}$
group + town	$6.26 \cdot 10^{34}$	0.62
condition + town	$4.43 \cdot 10^{26}$	$4.15 \cdot 10^{-09}$
group + condition + town	$6.62 \cdot 10^{33}$	0.06
group + condition + group:condition + town	$1.29 \cdot 10^{33}$	0.01
group + town + group:town	$2.57 \cdot 10^{34}$	0.25
group + condition + town + group:town	$2.77 \cdot 10^{33}$	0.03
group + condition + group:condition + town + group:town	$5.12 \cdot 10^{32}$	$5.14 \cdot 10^{-03}$
condition + town + condition:town	$7.15 \cdot 10^{25}$	$6.44 \cdot 10^{-10}$
group + condition + town + condition:town	$9.42 \cdot 10^{32}$	$9.13 \cdot 10^{-03}$
group + condition + group:condition + town + condition:town	$2.01 \cdot 10^{32}$	$1.85 \cdot 10^{-03}$
group + condition + town + group:town + condition:town	$4.29 \cdot 10^{32}$	$4.99 \cdot 10^{-03}$
group + condition + group:condition + town + group:town + condition:town	$8.79 \cdot 10^{31}$	$7.77 \cdot 10^{-04}$
group + condition + group:condition + town + group:town + condition:town + group:condition:town	$3.96 \cdot 10^{31}$	$3.92 \cdot 10^{-04}$

Note. The Bayes factors $B_{model,null}$ quantify the relative evidence for the respective model against the intercept-only model. *Group* denotes the within-participant factor group membership, *town* denotes the within-participant factor town of residence, and *condition* codes whether participants did versus did not receive explicit information about the groups' base rates per town (between-participants factor).

Table A2

Results of Bayesian $2 \times 2 \times 2$ mixed ANOVAs analyzing frequency estimates

Model	BayesFactor ($B_{model,null}$)	posterior probability
group	$9.50 \cdot 10^{3}$	$2.51 \cdot 10^{-16}$
condition	0.14	$3.81 \cdot 10^{-21}$
group + condition	$1.32 \cdot 10^{3}$	$3.36 \cdot 10^{-17}$
group + condition + group:condition	$2.06 \cdot 10^{2}$	$6.00 \cdot 10^{-18}$
town	$4.47 \cdot 10^{14}$	$1.57 \cdot 10^{-05}$
group + town	$2.48 \cdot 10^{19}$	0.66
condition + town	$6.55 \cdot 10^{13}$	$1.79 \cdot 10^{-06}$
group + condition + town	$3.54 \cdot 10^{18}$	0.09
group + condition + group:condition + town	$8.07 \cdot 10^{17}$	0.02
group + town + group:town	$6.40 \cdot 10^{18}$	0.16
group + condition + town + group:town	$9.34 \cdot 10^{17}$	0.02
group + condition + group:condition + town + group:town	$1.49 \cdot 10^{17}$	$4.07 \cdot 10^{-03}$
condition + town + condition:town	$1.61 \cdot 10^{13}$	$4.69 \cdot 10^{-07}$
group + condition + town + condition:town	$9.40 \cdot 10^{17}$	0.03
group + condition + group:condition + town + condition:town	$1.44 \cdot 10^{17}$	$7.21 \cdot 10^{-03}$
group + condition + town + group:town + condition:town	$2.21 \cdot 10^{17}$	$6.53 \cdot 10^{-03}$
group + condition + group:condition + town + group:town + condition:town	$4.21 \cdot 10^{16}$	$1.13 \cdot 10^{-03}$
group + condition + group:condition + town + group:town + condition:town + group:condition:town	$1.19 \cdot 10^{16}$	$3.13 \cdot 10^{-04}$

Note. The Bayes factors $B_{model,null}$ quantify the relative evidence for the respective model against the intercept-only model. *Group* denotes the within-participant factor group membership, *town* denotes the within-participant factor town of residence, and *condition* codes whether participants did versus did not receive explicit information about the groups' base rates per town (between-participants factor).

Normative Accounts of Illusory Correlations

Franziska M. Bott

University of Mannheim

David Kellen

Syracuse University

Karl Christoph Klauer

Albert-Ludwigs-Universität Freiburg

Author Note

Franziska M. Bott and David Kellen contributed equally to this paper. Franziska M. Bott was supported by the DFG grant 2277, Research Training Group “Statistical Modeling in Psychology" (SMiP). Karl Christoph Klauer was supported by DFG Reinhart-Koselleck grant DFG Kl 614/39-1. Scripts and data are made available at `https://osf.io/7sdgn/?view_only=5574464a07fd40eaa17d0bfd8a048f1f`.

Correspondence: Franziska M. Bott, Department of Psychology, University of Mannheim (f.bott@uni-mannheim.de) or David Kellen, Department of Psychology, Syracuse University (davekellen@gmail.com).

Abstract

When learning about the joint occurrence of different variables, individuals often manifest biases in the associations they infer. In some cases they infer an association when none is present in the observed sample. Other times they infer an association that is contrary to the one that is in fact observed. These *illusory correlations* are often interpreted as being the byproduct of selective processing or as the outcome of an 'illogical' pseudocontingency heuristic. More recently, a normative account of illusory correlations has been proposed, according to which they result from an application of Laplace's Rule of Succession. The present work will discuss the empirical and theoretical limitations associated with this normative account, and argue for its dismissal. As an alternative, we propose a normative account that casts illusory correlations as the expected outcome of a Bayesian reasoner relying on marginal frequencies. We show that this account succeeds in capturing the qualitative patterns found in a corpus of published studies.

Keywords: probability, rationality, biases, illusory correlation, pseudocontingencies

Normative Accounts of Illusory Correlations

Assessments of explanatory sufficiency play an essential role in theoretical debates. At a minimum, any candidate theory or model needs to be able to accommodate an established body of empirical results (i.e., "save the phenomena", Bogen & Woodward, 1988). Such arguments are particularly impressive when they are associated with normative frameworks, as norms often have limited descriptive power. Take the case of the Kolmogorov axioms of probability (e.g., DeGroot, 1975): people's choices and judgments often deviate from them, for instance, when the probability of the conjunction of two events (e.g., "*Linda is a feminist and a secretary*") is judged to be larger than the probability of any single event (e.g., "*Linda is a secretary*").

Erroneous inferences of this kind are commonly perceived to be "irrational" and have long motivated the development of alternative theoretical accounts that do not conform to probability theory (e.g., *representativeness heuristic*, Tversky & Kahneman, 1974; *configural weighting*, Nilsson, Winman, Juslin, & Hansson, 2009) or modal logic (e.g., *model theory*, Johnson-Laird, Khemlani, & Goodwin, 2015). In fact, the non-negligible failure rates of normative accounts have been one argument of proposals to dismiss normative accounts in favor of more descriptive research programs (see Elqayam & Evans, 2011). Another reaction has been to call for the adoption of extended or alternative normative frameworks (e.g., Buchak, 2013; Busemeyer & Bruza, 2012; Dzhafarov & Kujala, 2016; Oaksford & Chater, 2007; Skovgaard-Olsen, Kellen, Hahn, & Klauer, 2019; Spohn, 2012). Yet another group of researchers has argued that certain normative axioms can be upheld when assuming that they are realized by systems that are subjected to different sources of noise (Costello & Watts, 2014, 2016; Dougherty, Gettys, & Ogden, 1999; Fiedler, 1996; Hilbert, 2012).

In this paper, we will focus on one specific class of effects in probability judgments, which we will broadly refer to as *illusory correlations*. These illusory correlations are shown to occur when individuals attempt to infer the probabilities associated with different variables based on the observation of skewed samples (e.g., Fiedler, 2000; Hamilton & Gifford, 1976; Kareev, 1995; McConnell, Sherman, & Hamilton, 1994; Sherman et al., 2009). Consider the 2×2 contingency subtable reported at the top of Table 1. This subtable consists of M joint realizations of two binary random variables X and Y, each taking on values 0 and 1. These joint realizations are assumed to be samples $\boldsymbol{k} = (k_{00}, k_{01}, k_{10}, k_{11})$ taken from a multinomial distribution with the (unknown) probability vector $\boldsymbol{p} = (p_{00}, p_{01}, p_{10}, p_{11})$. The association between the two variables can be quantified by the Phi coefficient ϕ with $-1 \leq \phi \leq 1$:

$$\phi = \frac{k_{11}k_{00} - k_{10}k_{01}}{\sqrt{k_{1\bullet}k_{0\bullet}k_{\bullet 0}k_{\bullet 1}}} \tag{1}$$

The value $\phi = 0$ indicates that there is no association between X and Y in the sample, whereas positive/negative ϕ values indicate a positive/negative relation between X and Y. Another way of quantifying associations is through the Δp rule (Cheng & Novick, 1990), which compares the probability distribution of a binary random variable conditional on the two possible values of another binary random variable. In the case of variables X and Y in Table 1, we can define

$$\Delta p = P(Y = 0 \mid X = 0) - P(Y = 0 \mid X = 1). \tag{2}$$

As in the case of ϕ, Δp will be zero when there is no association, and positive/negative when there is a positive/negative relationship.

Illusory Correlations

We will classify the illusory correlations discussed into two types, *Type-1* and *Type-S*. This distinction is a direct reference to the idea of Type-1 and Type-S inference errors discussed in the statistical literature (Gelman & Carlin, 2014). Type-1 illusory correlations (IC_1) occur when individuals infer an association between variables when no association whatsoever can be found in the observed sample, while *Type-S illusory correlations* (IC_S) correspond to cases in which individuals infer an association that has the *opposite sign* of the one found in the observed sample.

Table 1

General contingency table (with joint frequencies/probabilities k_{ij}/p_{ij} and marginal frequencies $k_{i\bullet}$ and $k_{\bullet j}$) and examples of tables used to elicit Type-1 and Type-S illusory correlations

		2 × 2 Contingency Table		
		$Y = 0$	$Y = 1$	
	$X = 0$	k_{00} (p_{00})	k_{01} (p_{01})	$k_{0\bullet}$
	$X = 1$	k_{10} (p_{10})	k_{11} (p_{11})	$k_{1\bullet}$
		$k_{\bullet 0}$	$k_{\bullet 1}$	
		Type-1 Illusory Correlation (IC_1)		
		Outcome		
		+	−	
Group	A	27	9	36
	B	9	3	12
		36	12	($M = 48$)
		Type-S Illusory Correlation (IC_S)		
		Outcome		
		+	−	
Group	A	16	8	24
	B	8	0	8
		24	8	($M = 32$)

Type-1 Illusory Correlations

First, consider subtable IC_1 in Table 1, which lists the joint and marginal frequencies of variables ‘Group’ and ‘Outcome’. Inspection of the marginal frequencies shows that the majority of the observations comes from Group A, and that ‘+’ is the most common outcome (each majority event is found in 36 of the 48 total instances). Although most observations correspond to outcomes ‘+’ coming from Group A, there is no association between Group and Outcome ($\phi = 0$; note that $P(+|A) = \frac{k_{00}}{k_{0\bullet}} = \frac{27}{36} = .75$ and $P(+|B) = \frac{k_{10}}{k_{1\bullet}} = \frac{9}{12} = .75$). When people encounter a sample of instances such as the ones summarized in that table, they tend to infer a group-outcome association at the population level, such that the ‘+’ outcome is judged as being more probable under the majority group A than under the minority group B ($\hat{\phi} > 0$ and $\Delta\hat{p} > 0$). As an example, consider Experiment 2 of Hamilton and Gifford (1976): participants were shown a sample of group-outcome instances, with the outcomes being desirable and undesirable behaviors.

Group A was the majority group and undesirable behavior the most common outcome (each majority event was found in 24 out of 36 total instances). Although the relative frequency of undesirable behaviors was $\frac{2}{3}$ in both groups, the estimated probabilities of undesirable behaviors in Groups A and B were $\hat{p}_A = .66$ and $\hat{p}_B = .45$, respectively, with $\hat{\phi} = .20$ and $\Delta\hat{p} = .21$.

Pseudocontingency Heuristic

Early proposals of illusory correlations include the differential processing of frequent and infrequent events (e.g., distinctiveness accounts; see Chapman & Chapman, 1969; Hamilton & Gifford, 1976; E. A. Wasserman, Dorner, & Kao, 1990), but have not found considerable empirical support (e.g., Bulli & Primi, 2006; Fiedler, Russer, & Gramm, 1993; Klauer & Meiser, 2000; Meiser, 2003; Meiser & Hewstone, 2001, 2004, 2006). A prominent alternative explanation for the occurrence of illusory correlations is that individuals' correlational inferences are based on the observed *marginal frequencies* (e.g., Fiedler, Freytag, & Meiser, 2009; Fiedler, Kutzner, & Vogel, 2013). When observing skewed marginal frequencies of two variables, participants are assumed to infer heuristically that the most frequent observations in each variable are associated with each other (e.g., Group A and + in subtable IC_1 of Table 1) as well as the infrequent observations (e.g., Group B and −). This *pseudocontingency heuristic* account is supported by results showing that the inferred associations reflect the skewness found in the marginals, independent of whether joint observations are actually presented (e.g., Fiedler & Freytag, 2004; Meiser, 2006; Meiser & Hewstone, 2004; Meiser, Rummel, & Fleig, 2018; Vadillo, Blanco, Yarritu, & Matute, 2016; Vogel, Kutzner, Fiedler, & Freytag, 2013).

Type-S Illusory Correlations

According to the pseudocontingency account, the inference of an illusory correlation should also not depend on the true contingency between the two focal variables. In line, Fiedler (2010) demonstrated that the inferred associations between two variables were based on the marginal frequencies, independent of their actual contingencies, which were zero (IC_1), consistent with the inference based on marginals, or inconsistent with the inference based on marginals (IC_S). Additionally, consider, for example, the IC_S subtable in Table 1: once again, Group A is the majority group and '+' is the most common outcome. In this case, however, there is a negative group-outcome association ($\phi = -.33$, $\Delta p = -.33$) due to the fact that the proportion of '+' outcomes is larger in Group B ($\frac{8}{8} = 1$) than in Group A ($\frac{16}{24} = .67$). Yet, participants tend to judge the probability of '+' under majority group A to be higher than under minority group B. For example, Meiser et al. (2018, Experiment 1, within-subject condition Casino X) found the average probability estimates to be $\hat{p}_A = .57$ and $\hat{p}_B = .46$, which indicate a positive association ($\Delta\hat{p} = .11$).

Illusory correlations have been shown to increase in strength with decreasing working memory capacity (Eder, Fiedler, & Hamm-Eder, 2011), the more salient marginal frequencies are (Meiser et al., 2018) and the more attention is paid to them (Fleig, Meiser, Ettlin, & Rummel, 2017). Accordingly, many studies on IC_S effects use more complex stimulus distributions including a third (context) variable: Typically, the focal variables' marginal frequencies are skewed within one context and co-vary across the contexts. By implication, the focal variables are each associated with the context variable. The use of contrasting contexts is assumed to increase the salience of the co-occurrence of skewed marginal frequencies and to thereby increase illusory correlation effects (Fiedler & Freytag, 2004).[1]

[1] As pointed out by a reviewer, it is possible that this association between focal and context variables is behind the observed Type-S illusory correlations. However, there are good reasons to believe that this is not the case. First, we can find the IC_S effect in published studies in which there is no association between focal and context variables. For example, Fiedler (2010) found an IC_S effect in a study in which participants were presented with fictitious students described in terms of four binary variables (`sex`, `city`, `university major`, and `hobby`) and subsequently asked to judge conditional probabilities associated with each pair of variables (e.g., "*What is the % of persons coming from Mannheim, given that they study psychology?*"). Moreover, an unpublished study by the first author found an IC_S effect in an experimental design in which there was no association between focal and contextual variables. In this study, the two contexts (`Machine X` and `Machine Y`) involved two exclusive and distinct pairs of buttons (A/B and C/D, respectively), with one context having a negative button/outcome association ($P(+|A,X) = \frac{18}{24} = .75$ and $P(+|B,X) = \frac{10}{12} = .83$), whereas the other context had no such association ($P(+|C,Y) = \frac{4}{18} = .22$ and $P(+|D,Y) = \frac{4}{18} = .22$). While `Machine Y` trials were only used as filler items, an IC_S effect was found in `Machine X`, with $\Delta\hat{p} = .04$, $t(100) = 2.04$, $p = .044$.

Two Opposing Accounts of Illusory Correlations

Both types of illusory correlations have been shown to occur across numerous studies using a wide variety of stimulus materials (for reviews, see Fiedler, 2000; Fiedler et al., 2009). Several of these studies involve real-world groups (e.g., occupational groups, Hamilton & Rose, 1980), such that stereotypes are in all likelihood affecting the observed estimates. However, illusory correlations have also been shown to occur with abstract stimuli (e.g., geometric shapes and colors), for which stereotypes are unlikely to exist (e.g., Primi & Agnoli, 2002, Experiment 2). These results suggest that illusory correlations are by and large a byproduct of the general way in which individuals infer associations between sampled variables.

The assumed use of marginal frequencies to infer contingencies is described by some of its proponents as "*logically unwarranted but useful and smart*" (Fiedler et al., 2013, p.328). Simply put, joint frequencies cannot be deduced from marginal frequencies. Yet, the heuristic of associating frequent categories and infrequent categories turns out to have considerable adaptive value: it often succeeds in capturing *actual* contingencies while relying on a simpler data format (Kutzner, Vogel, Freytag, & Fiedler, 2011). Like with many other heuristics, the case made for the inferences based on marginals is couched in the notion that they are highly successful in real-world environments in which they are applied as compared to in experiments (i.e., the notion of *ecological rationality*, e.g., Gigerenzer, 2019; Goldstein & Gigerenzer, 2002).

In comparison, a normative account of illusory correlations has recently been proposed by Costello and Watts (2019). Rather than reflecting some heuristic use of marginal frequencies, the occurrence of Type-1 illusory correlations is attributed to the updating of prior information as captured by the *Rule of Succession* proposed by Laplace (1820/1951). When outcome '+' is observed k times in N independent samples coming from a given group, the inferred probability p of '+' occurring is:

$$\hat{p} = \frac{k+1}{N+2}. \tag{3}$$

Applying this rule to the example in subtable IC_1 of Table 1, $\hat{p}_A = \frac{27+1}{36+2} = .74$ and $\hat{p}_B = \frac{9+1}{12+2} = .71$ result as probability estimates, with $\Delta\hat{p} = \hat{p}_A - \hat{p}_B = 0.03$ indicating the presence of an association between group membership and outcomes. Costello and Watts (2019) reported simulations showing that the application of the Rule of Succession to probability estimation outperforms the use of sample proportions or observed relative frequencies in producing Type-1 illusory correlations. Consequently, they argue that Type-1 illusory correlations are in fact the “mathematically correct” response when estimating probabilities based on observed frequencies.

These recent developments present two distinct interpretations of illusory correlations. On the one hand, we have a heuristic account of illusory correlations suggesting that the effects cannot be accommodated within the tenets of probability theory. On the other hand, the newly proposed normative account makes the opposite case, with Type-1 illusory correlations being framed as rational responses that follow the presumably optimal Rule of Succession. In the remainder of this paper, we will argue against these two interpretations. First, we will show that the Rule of Succession has some important limitations, theoretical as well as empirical. Second, we will show that the presumably heuristic pseudocontingency inferences can be seen as a rational consequence of a Bayesian updating of beliefs when the updates are based on (skewed) marginal frequencies. We will then show that a Bayesian Marginal Model that realizes this theoretical insight is able to accommodate the patterns found in the data at large. Finally, we show that the model is fruitful in generating new predictions, some of which can be tested on existing data and turn out to hold when inspecting previously-published data.

Deconstructing the Rule of Succession

> *No other formula in the alchemy of logic has exerted more astonishing powers. For it has established the existence of God from total ignorance, and it has measured with numerical precision the probability that the sun will rise tomorrow.* (Keynes, 1921, p. 89)

Consider a scenario in which only two events are possible, namely a ‘success’ and a 'failure' event. Assume that you encounter k success events in a total of N independent observations. What is the probability that a success will be observed in the $(N+1)$th observation? Laplace (1820/1951) showed that, when assuming no prior preference for any probability value, one can reach the solution known as the Rule of Succession:

$$\hat{p} = \frac{\int_0^1 p^{k+1}(1-p)^{N-k}\,\mathrm{d}p}{\int_0^1 p^{k}(1-p)^{N-k}\,\mathrm{d}p} = \frac{k+1}{N+2}. \tag{4}$$

According to Costello and Watts (2019), the Rule of Succession presents a rationalization of IC_1 effects. Specifically, when asked to judge the probability of a success in a population based on a sample, the Rule of Succession provides a mathematically correct way to estimate the probability p. IC_1 effects are then a byproduct of this rule: the +1 and +2 terms in the numerator and denominator, respectively, will regress the inferred probability towards $\frac{1}{2}$ for each group. The more observations one has of one group, the smaller the deviation from $\frac{k}{N}$, which is caused by the +1 and +2 terms. This means that the estimates for a majority group A will be less regressive towards $\frac{1}{2}$ than estimates for a minority group B. Hence, the estimates will conform to the inequality $\hat{p}_A > \hat{p}_B > \frac{1}{2}$ when $p > \frac{1}{2}$, or $\hat{p}_A < \hat{p}_B < \frac{1}{2}$ when $p < \frac{1}{2}$. In both cases, the estimates imply a group-outcome contingency when none is present in the data. In other words, Type-1 illusory correlations are predicted. Going beyond, Costello and Watts (2019) developed different computational accounts that implement the Rule of Succession (Algorithms 1 and 2) and showed that they can approximate the IC_1 effects reported by Hamilton and Gifford (1976, Experiment 1) and Van Rooy, Vanhoomissen, and Van Overwalle (2013, Experiment 1).

When discussing the Rule of Succession, it is useful to frame it as a specific solution within the broader context of *Bayesian inference* (Jeffreys, 1939): Beliefs are represented by probability distributions that are updated in light of the data using Bayes' Theorem, with the estimated probability $\hat{p}$ corresponding to a *point-summary* of the posterior beliefs. In their presentation of the Rule of Succession, Costello and Watts (2019) state, for instance, that it is not reasonable for participants to "*conclude*" that $\hat{p} = 0$ when

$k = 0$ in a total of $N = 2$ observations (p. 439). Indeed, also when framing the Rule of Succession within the context of Bayesian inference, it would be unreasonable for someone to place all probability mass on 0 given a sample size of $N = 2$. However, under no set of reasonable priors is that the case. The point estimate $\hat{p} = 0$, in whatever way it is computed, is only a *summary* of a distribution that places non-zero probability mass over the unit interval. It does not follow that all other values are deemed impossible or implausible, especially when $N = 2$. Yet, it does not follow, either, that $\hat{p} = 0$ might not be a reasonable point estimate.[2]

Elaborating on the framing of the Rule of Succession within the context of Bayesian inference, let p be the rate parameter of a Binomial distribution. The parameter p indicates the probability of the occurrence of a 'success' event among N independent and identically-distributed observations. Assuming a prior distribution over p, we can capture a reasoner's preconceptions regarding the probability of a success. We will assume a Beta distribution as a prior given its flexibility in capturing beliefs through its shape parameters α and β. The parameters' values can be interpreted as '*pseudocounts*' in terms of 'successes' and 'failures'. The Beta distribution is a *conjugate prior* of the Binomial distribution. This implies that the posterior beliefs, resulting from the application of Bayes' Theorem when observing k successes out of N observations, will follow a Beta distribution with parameters $\alpha^* = \alpha + k$ and $\beta^* = \beta + N - k$. The mean of this posterior distribution is $\frac{\alpha+k}{\alpha+\beta+N}$. Consequently, when $\alpha = \beta = 1$, we are assuming the prior observations of one success and one failure. This prior distribution is uniform over the $[0, 1]$ interval, indicating no preference for any specific range of probability values. The *posterior mean* under this prior is $\frac{\alpha+k}{\alpha+\beta+N} = \frac{1+k}{2+N}$, which corresponds to the Rule of Succession (see Equation 4).

The shape of the prior distribution will affect the posterior mean. Specifically, the

[2] This statement is a good reflection of Costello and Watts' (2019) ambiguity regarding the concept of 'probability' that they are relying on. The Bayesian foundation of the Rule of Succession explicitly calls for a *subjectivist* interpretation, according to which the strength of our beliefs are represented by probability distributions. However, Costello and Watts' focus on point estimates and argument against specific values such as $\hat{p} = 0$ speak directly against that. Also of note is the fact that their argument for the rationality of the Rule of Succession (which we will discuss in detail later on) focuses on the *frequentist* properties of $\hat{p}$.

posterior mean is shifted towards the prior mean, an effect known as *shrinkage*. Figure 1 illustrates the updating under different priors such as *Jeffrey's prior* ($\alpha = \beta = \frac{1}{2}$). The amount of shrinkage will depend on the prior distribution as well as on the amount of data. The fewer data the greater the shrinkage will be, or put differently, by virtue of a larger sample size, the data are more successful in overwhelming the prior. The ability of the Rule of Succession to account for Type-1 illusory correlations rests on this fact: Because the minority group B is less often observed than the majority group A, the respective posterior mean for Group B will be closer to the prior mean of $\frac{1}{2}$ (cf. middle panels in Figure 1).

The possibility to vary the prior distribution allows for a generalization of the Rule of Succession. Without loss of generality, let $\alpha = \lambda \cdot \pi_p$ and $\beta = \lambda \cdot (1 - \pi_p)$, with $0 \leq \pi_p \leq 1$ and $\lambda \geq 0$. The posterior mean of p under this parametrization is

$$\hat{p} = \frac{\lambda \cdot \pi_p + k}{\lambda + N}. \tag{5}$$

Parameter λ describes the number of pseudocounts $\alpha+\beta$ implied by the prior and thus, its informativeness. Parameter π_p introduces skew into the prior distribution to the extent to which it deviates from 0.5.[3] This generalization of the Rule of Succession corresponds to *Carnap's Method of Updating* (Carnap, 1952), which has been proposed in the decision-making literature to account for a number of phenomena (e.g., Viscusi, 1989; Wakker, 2002). Viscusi (1989), for instance, showed that its incorporation into the Expected Utility model (Savage, 1954; Von Neumann & Morgenstern, 1944) enables the model to accommodate the paradoxes of Allais. More recently, it has been used to characterize decisions based on experienced options and the 'description-experience gap' (Aydogan, 2019; Hertwig, Barron, Weber, & Erev, 2004; see also Wakker, 2002).

The Rational Status of the Rule of Succession

Costello and Watts (2019) motivated the use of the Rule of Succession by arguing that it minimizes errors when estimating the probability of a feature or outcome based

[3] In the special case of the Rule of Succession, $\alpha = \beta = 1$ and hence, $\lambda = 2$, $\pi_p = .5$.

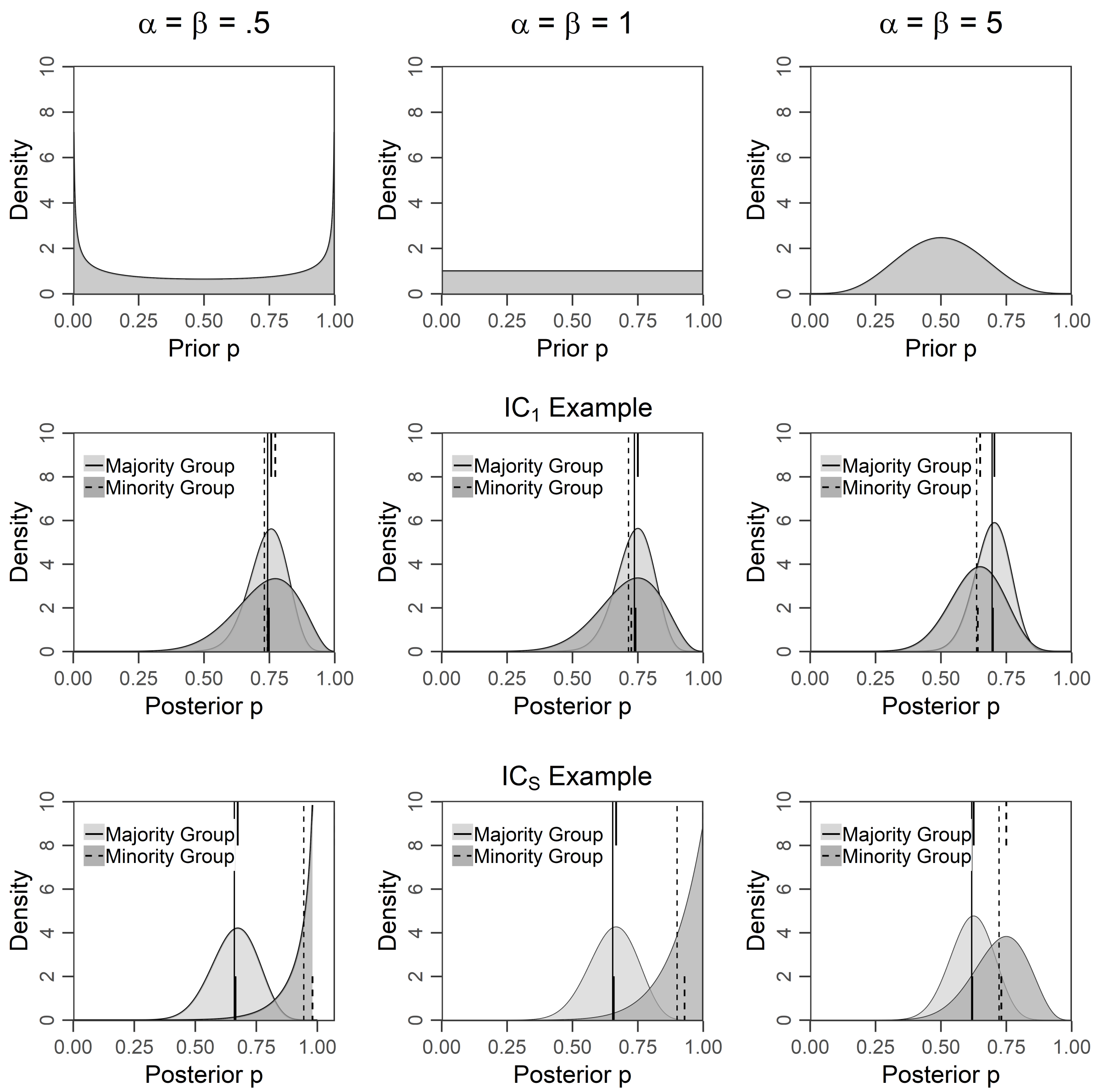

Figure 1. Posterior Beta distributions (middle and lower panels) that result from updating the prior distributions (upper panels), parametrized by different shape parameters α and β, when observing the frequencies in Table 1 for Type-1 and Type-S illusory correlations. The vertical lines represent posterior means. While the lower segment lines represent posterior medians, the upper segment lines represent posterior modes.

on a limited sample. Under the notion that a rational approach is one that minimizes errors, they justify the occurrence of a Type-1 illusory correlation by stating that "*the mathematically correct and hence rational response is to judge the rare feature as more likely in the Minority than the Majority population*" (p. 447). This argument is based on (a) a simulation (Costello & Watts, 2019, Algorithm 1) that computes the posterior mean as an estimate of p given a uniform prior, leading to the conclusion that a "*reasoner who makes this estimate for the underlying probability will be, on average, closest to the true population probability that generated the observed sample*" (p. 440) and on (b) the observation that the use of the posterior mean predicts Type-1 illusory correlations due to shrinkage.

Even when taking for granted that accuracy is a yardstick for rationality, we have to take a closer look at Algorithm 1 they use to establish the rationality of the Rule of Succession, which goes as follows:

1. Set integers k and N, with $k \leq N$.
2. Generate a probability p from a uniform distribution.
3. Generate binomially distributed data with probability p and sample size N. Whenever k successes occur in the binomial data, record p.
4. Repeat steps 2-3 many times and compute the average of the recorded ps.

The average values coming from this simulation very closely match the values obtained with the Rule of Succession, that is $\frac{k+1}{N+2}$. Costello and Watts (2019) interpret this result as support for the rationality of the rule. However, the simulation algorithm effectively samples from the posterior distribution of p given $\frac{k}{N}$ and a flat prior over p. The average of these posterior samples has to match the Rule of Succession (excusing sampling variability) given that the latter is nothing more than the closed-form expression of the posterior mean of p under the same data and prior. In other words, the two values being compared refer to the exact same quantity.

Moreover, note that the mean of the posterior distribution is only one of many legitimate summary descriptions of a posterior distribution's central tendency, given the observed sample. For instance, it would have been perfectly legitimate to use the *posterior mode* or *maximum a posteriori estimate* (MAP), which in the case of the Beta distribution corresponds to $\frac{\alpha+k-1}{\alpha+\beta+N-2}$. Under the prior parameters $\alpha = \beta = 1$, that yield the Rule of Succession, the posterior mode equals the sample proportion $\frac{k}{N}$, which means that no

IC_1 effect would be predicted, an absence that Costello and Watts (2019) consider to be incorrect.

The justification for the use of a given estimator and the discussion of its merits in terms of error minimization requires us to establish how errors are quantified. In other words, it requires a *loss function*. Let $C(\theta, \hat{\theta})$ be a loss function quantifying the discrepancy between a true value θ and its estimate $\hat{\theta}$ (for an overview, see L. Wasserman, 2004, Chap. 12). Among the many possible loss functions, we find the following popular options (with $q > 0$ being a scaling parameter):

- **Quadratic Loss**: $C(\theta, \hat{\theta}) = q \cdot (\theta - \hat{\theta})^2$,
- **Linear Loss**: $C(\theta, \hat{\theta}) = q \cdot \mid \theta - \hat{\theta} \mid$,
- **All-or-Nothing Loss**: $C(\theta, \hat{\theta}) = \begin{cases} 0, & \text{if } \theta = \hat{\theta}, \\ q, & \text{otherwise.} \end{cases}$

Under a quadratic loss function, the optimal estimator is the posterior mean. The only case in which illusory correlations are not expected using posterior means is when the completely uninformative, so-called *Haldane prior* ($\alpha = \beta = 0$) is used. In the case of linear loss, the optimal estimate corresponds to the posterior median, which is approximated by $\frac{\alpha+k-1/3}{\alpha+\beta+N-2/3}$ (Kerman, 2011). If $\alpha = \beta < \frac{1}{3}$, a *reversed* IC_1 effect is predicted, such that the estimated posterior median of the majority group is closer to $\frac{1}{2}$ than the minority group's (.751 and .754 respectively, in the IC_1 example in Table 1 when $\alpha = \beta = \frac{1}{4}$). Finally, an all-or-nothing loss function is optimized by the posterior mode, which predicts a reversed IC_1 for $\alpha = \beta < 1$ and no IC_1 effect under the uniform prior $\alpha = \beta = 1$. Type-1 illusory correlations are only expected when $\alpha = \beta > 1$. Based on these cases, we see that the Rule of Succession and the illusory correlations that it can predict are only rational/optimal if one takes for granted that $\hat{p}$ corresponds to the posterior mean and a quadratic loss function is assumed. In fact, the estimator $\hat{p}$ deemed optimal differs according to the assumed prior and loss function, with illusory correlations being expected under some but not others.

The optimality results above describe a reasoner who moves in an ecology in which s/he randomly encounters N samples from groups with different true probabilities p that follow the assumed prior distribution. In this context, optimality refers to a minimization of the average estimation error *across* groups. This, however, is at odds with the conceptual characterizations that Costello and Watts (2019) allude to. They conceptualize the prior and posterior distributions as epistemic probability distributions or as descriptions of the reasoner's prior and posterior beliefs, respectively, about the unknown value p underlying the observed group frequencies. Accordingly, the group in question is *fixed* and the true probability that generates the data is the unknown p value in force for that group. Let us call this fixed value p_G. The uniform prior describes the assumption that in the absence of data, the reasoner considers each admissible value for the unknown true p_G value as equally likely a priori. An estimate $\hat{p}_G$ is in error to the extent to which it deviates from the true value p_G of the group under scrutiny, irrespective of one's prior beliefs. The expected estimation error or loss must be computed as what is to be expected when sampling N observations from group G, that is relative to the binomial distribution with parameters N and p_G. If the sample proportion $\frac{k}{N}$ is taken as estimate, then the expected quadratic loss corresponds to its variance under the binomial distribution, namely $\frac{1}{N}p_G(1-p_G)$. In the case of the Rule of Succession, $\frac{k+1}{N+2}$, the expected quadratic loss is $\frac{N^2}{(N+2)^2}\frac{1}{N}p_G(1-p_G)+\frac{1}{(N+2)^2}(2p_G-1)^2$. The latter expectation is larger (i.e., worse) than the former when $|p_G-\frac{1}{2}|>\sqrt{\frac{N+1}{8N+4}}$, and smaller otherwise. This means that the question of which of these estimators is the best depends on both p_G and N and cannot be answered in the absence of knowledge about p_G.

The Empirical Status of the Rule of Succession

Even if one cannot make a compelling case for the Rule of Succession or its generalization à la Carnap, theoretically, they might nevertheless succeed in capturing the effects found in the data at large. In this section, we will show that this is not the case.

The Case of Type-S Illusory Correlations. Costello and Watts (2019) discussed the Rule of Succession within the context of IC_1 effects. However, there is no reason to not consider IC_S effects as well, especially given that the difference between both types can be boiled down to the relative frequencies observed in each group (namely, whether they are the same or not). It turns out that the Rule of Succession has trouble capturing IC_S effects: First, consider the simulation reported in Figure 2 that displays the probability inferred by the Rule of Succession as a function of the number of observations across the different groups. Note that in order to illustrate these predictions, we replaced k with $N \cdot p$ (see also Viscusi, 1989; Wakker, 2002). The top panel of Figure 2 shows that the Rule of Succession generally expects IC_1 effects to occur. But as shown in the center and bottom panels, an IC_S effect is only expected under the Rule of Succession for very specific frequencies, and/or extremely small samples (e.g., when $P(+|A) = .75$, $P(+|B) = .83$, and $N_A < 4$). Otherwise, according to the Rule of Succession, the estimate $\hat{p}$ is expected to be higher for the minority group(s), in line with the genuine group-outcome relationship.

Qualitative Misses in Illusory Correlations. The ability to capture an *effect*, a difference between two estimated values, does not imply the ability to accommodate the values from which the effect is computed. Considering the absolute values of people's probability judgments more closely, it turns out that beyond the result $\hat{p}_A > \hat{p}_B$, data on IC_1 and IC_S effects show a recurring pattern of $\hat{p}_B < \frac{1}{2}$: In other words, the majority event for Group B is in fact perceived as the minority event for that group. These recurring qualitative misses cannot be explained by any regression-to-the-mean account with prior mean $\frac{1}{2}$, as they are unable to cross over the $\frac{1}{2}$-midpoint. We checked the occurrence of these qualitative misses in a corpus of previously-published studies, which are listed in Table 2.

The study corpus was built by searching the databases *PSYNDEX* and *PsycInfo* (on 08/20/2019) using the terms "illusory correlation" and "pseudocontingencies" and restricting the search to manuscripts published in academic journals in the period 2000-2019. Of the articles identified, we only considered experiments that additionally met the following criteria: (1) the experimental stimuli consisted of binary variables, (2) participants had to judge the frequency or probability of an outcome given one group, (3) the groups were artificial, without any stereotypical information associated with them, (4)

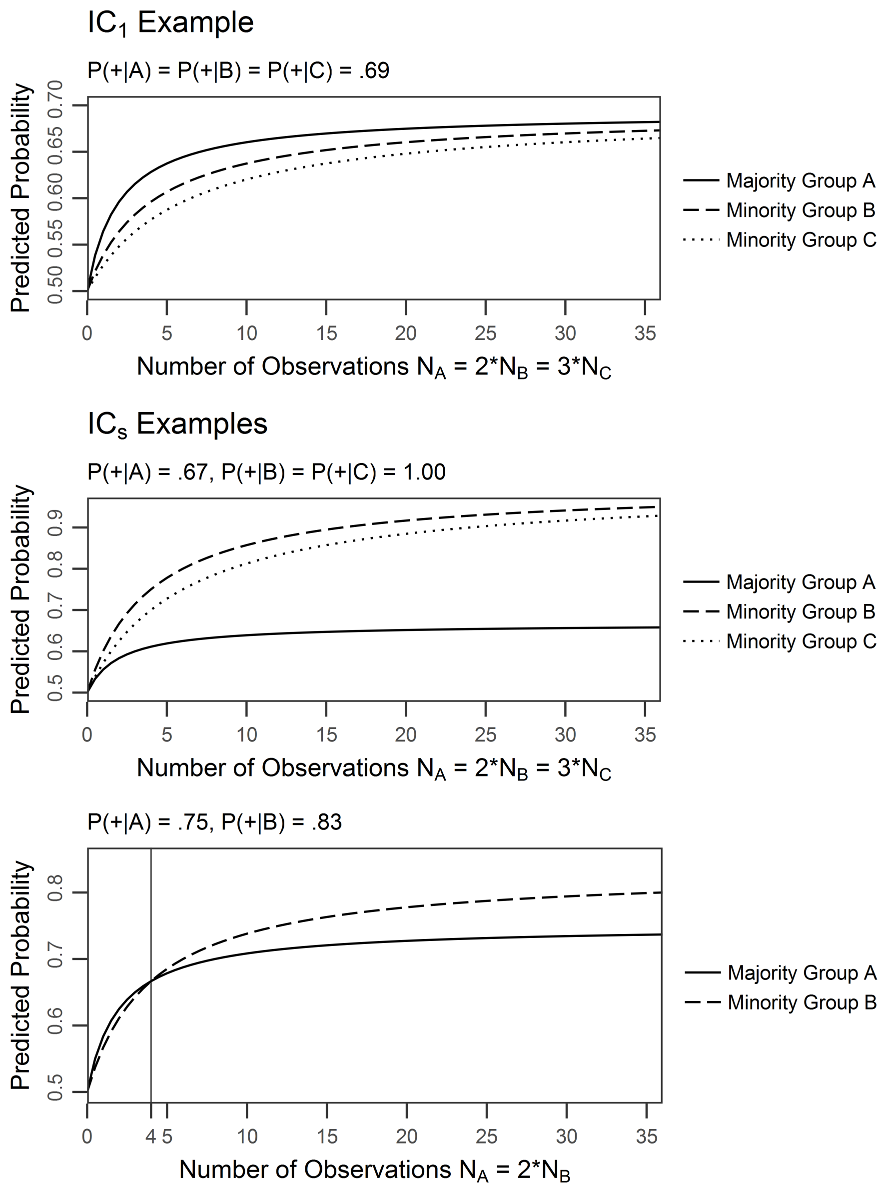

Figure 2. Probability of success predicted by Rule of Succession as a function of number of observations (N) per option.

the frequency estimates or probability estimates are reported and were not converted into a Δp measure, (5) the effect addressed (i.e., IC_1 or IC_S) was present in the data and determined to be reliable using some statistical test,[4] and (6) only non-clinical participants took part in the study. The application of those inclusion criteria amounted to a set of 43 experimental conditions in 14 articles. We additionally included data from Fiedler et

[4] In several cases, application of this criterion resulted in including only one of two within-subject conditions.

al. (1993), Spears, van der Pligt, and Eiser (1985), and Hamilton and Gifford (1976).[5] The aim was not to provide an exhaustive literature review, but rather to illustrate the general pattern of empirical findings in contemporary research.

Checking the data corpus for the occurrence of qualitative misses described above ($\hat{p}_B < \frac{1}{2}$), we found these misses in 55% of the cases. The magnitude of the qualitative misses is additionally displayed in the right panel of Figure 3: in more than half of the cases, participants' estimates for the minority group do not only reflect an underestimation, but a reversal in which outcome is most frequent. Importantly, note that the qualitative misses are found in both IC_1 and IC_S studies, as can be seen in Table 2, so that they cannot be attributed to the different designs used to study Type-1 and Type-S illusory correlations (see Footnote 1).

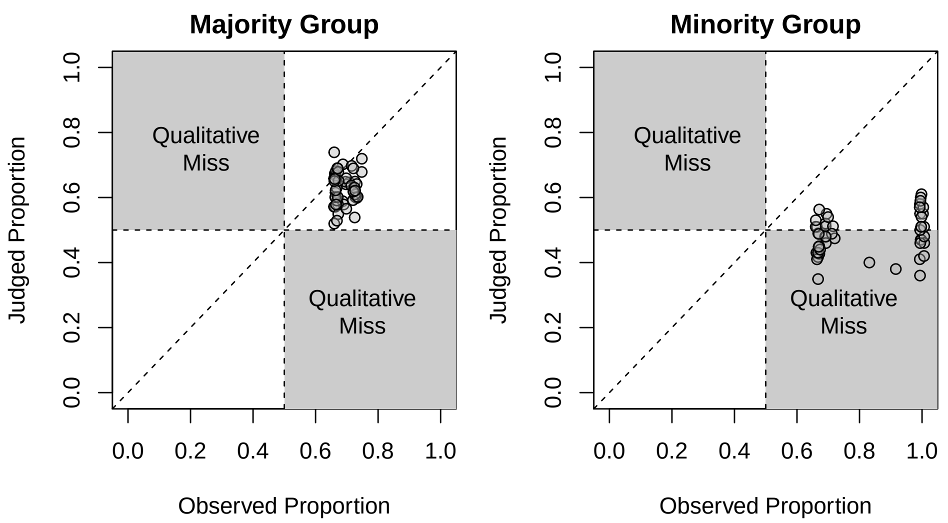

Figure 3. Observed and estimated relative frequencies in the majority (A) and minority (B) groups in the data corpus. In order to improve the discriminability between data points, we jittered the observed proportions.

Fitting the Rule of Succession to the Data Corpus. To corroborate the above assessments, we fitted the Rule of Succession as well as Carnap's generalization to the corpus of studies reported in Table 2. In the case of the Rule of Succession, for each study, we applied the Rule of Succession to the observed frequencies k and N for both groups A and B. In the case of Carnap's generalization, we fixed π to $\frac{1}{2}$ and estimated λ

[5] Note that applying the stated criteria resulted in not including the two data sets used by Costello and Watts (2019).

freely. For each study, we assumed a single λ parameter for both Group A and Group B when computing estimates. Models were estimated by minimizing squared errors, with the constraint that IC_S and qualitative misses were to be predicted if they were present, and the model able to produce them. These fits tell us how well the models can capture the qualitative patterns in the data, irrespective of their magnitude.[6] The best-fitting predictions are illustrated in Figure 4.

It is clear that the Rule of Succession cannot provide a reasonable account of the data, especially when it comes to IC_S effects and qualitative misses.[7] The fact that it has no free parameters also compromises the model's ability to successfully capture differences between studies in which participants observed the same joint events. The generalized rule performs somewhat better, although some of the effects are barely captured at the expense of extreme prior parameters (e.g., $\alpha = \beta = 50$).[8]

[6] A purely quantitative comparison between observed and predicted values is not really useful here as we first and foremost want to know whether the model can capture the qualitative patterns (reproducing a qualitative pattern is not guaranteed when simply minimizing misfit). Also, one should keep in mind that these data are an amalgamation of different strategies and beliefs that are unlikely to be successfully captured by any single model. In this situation it is best to focus on the qualitative aspects of models (for discussions, see Kellen, 2019; Navarro, 2019).

[7] As a sanity check, we investigated the possibility that qualitative misses may be caused by sampling noise and/or forgetting. We simulated probability estimates produced by the Rule of Succession including forgetting rates for the two example distributions in Table 1 and the less extreme IC_S-event distribution from Meiser et al. (2018, Experiment 3). Following Algorithm 2 of Costello and Watts (2019), each observation is recalled with probability $1 - f$. This is achieved by comparing a number randomly drawn from a uniform distribution between 0 and 1, against a fixed value f. If the random number is larger than f, the observation is recalled, otherwise forgotten. We simulated 1000 probability estimates for each combination of forgetting rate and outcome probability. Overall, qualitative misses were found to be extremely unlikely, only occurring (at a low rate) when forgetting rates were extremely high.

[8] Equivalent results were observed when introducing random forgetting via a probability parameter f (see Costello & Watts, 2019) instead of estimating the weight λ on the prior. The reason for the similar results is that both λ and f serve the same role of modulating the weight of the prior probability relative to the observed sample and therefore the regressive effect toward $\frac{1}{2}$.

Table 2

Overview of the Data Corpus

Study	N	Effect	Qualitative Miss
1 Berndsen, McGarty, Van Der Pligt, and Spears (2001, E2)	22	IC_1	✖
2 Bulli and Primi (2006)	158	IC_1	✓
3 Eder et al. (2011, E1, online-prediction)	24	IC_1	✓
4 Eder et al. (2011, E2, no-load)	33	IC_1	✓
5 Eder et al. (2011, E2, load)	33	IC_1	✓
6 Eder et al. (2011, E3, negative)	21	IC_1	✓
7 Eder et al. (2011, E3, positive)	21	IC_1	✓
8 Fiedler et al. (1993, no-valence-group)	15	IC_1	✓
9 Fiedler et al. (1993, no-valence-person)	15	IC_1	✖
10 Fiedler et al. (1993, valence-group)	15	IC_1	✖
11 Fiedler et al. (1993, valence-person)	15	IC_1	✓
12 Hamilton and Gifford (1976, E2)	70	IC_1	✓
13 Klauer and Meiser (2000, E2)	20	IC_1	✖
14 Madey and Chasteen (2004, younger)	22	IC_1	✖
15 Meiser and Hewstone (2001, group)	20	IC_1	✓
16 Meiser and Hewstone (2001, town)	20	IC_1	✖
17 Meiser (2003, E1, impression)	33	IC_1	✖
18 Meiser (2003, E1, memory)	33	IC_1	✓
19 Meiser and Hewstone (2006, (a), town X)	35	IC_1	✓
20 Meiser and Hewstone (2006, (a), town Y)	35	IC_1	✓
21 Mutter (2000, distraction-older)	24	IC_1	✓
22 Mutter (2000, distraction-younger)	24	IC_1	✓
23 Primi and Agnoli (2002, E1)	253	IC_1	✖
24 Primi and Agnoli (2002, E2)	106	IC_1	✓
25 Rodríguez-Ferreiro and Barberia (2017)	240	IC_1	✓
26 Spears et al. (1985)	141	IC_1	✖
Subtotal (Qualitative Misses)			$\frac{17}{26}$ (65%)
27 Fleig et al. (2017, E1, negative)	56	IC_S	✖
28 Fleig et al. (2017, E1, positive)	56	IC_S	✓
29 Fleig et al. (2017, E2, base-rate, positive)	36	IC_S	✖
30 Fleig et al. (2017, E2, standard, positive)	39	IC_S	✓
31 Meiser (2003, E2, impression, negative)	38	IC_S	✖
32 Meiser (2003, E2, impression, positive)	38	IC_S	✖
33 Meiser (2003, E2, memory, negative)	38	IC_S	✖
34 Meiser (2003, E2, memory, positive)	38	IC_S	✖
35 Meiser and Hewstone (2004, E1, negative)	39	IC_S	✖
36 Meiser and Hewstone (2004, E1, positive)	39	IC_S	✓
37 Meiser and Hewstone (2004, E2, group, negative)	39	IC_S	✖
38 Meiser and Hewstone (2004, E2, group, positive)	39	IC_S	✖
39 Meiser and Hewstone (2004, E2, town, negative)	39	IC_S	✖
40 Meiser and Hewstone (2004, E2, town, positive)	39	IC_S	✓
41 Meiser and Hewstone (2004, E3, negative)	66	IC_S	✖
42 Meiser and Hewstone (2004, E3, positive)	66	IC_S	✖
43 Meiser et al. (2018, E1, positive)	66	IC_S	✓
44 Meiser et al. (2018, E2, visualization, negative)	30	IC_S	✖
45 Meiser et al. (2018, E2, visualization, positive)	30	IC_S	✓
46 Meiser et al. (2018, E3, pc, positive)	39	IC_S	✓
47 Meiser et al. (2018, E3, pc-salience, positive)	40	IC_S	✓
48 Meiser et al. (2018, E4, computer, positive)	64	IC_S	✓
49 Meiser et al. (2018, E4, self, positive)	63	IC_S	✓
Subtotal			$\frac{10}{23}$ (43%)
Total			$\frac{27}{49}$ (55%)

Note. Column "N" reports the number of participants in a given study. Column "Effect" indicates the specific illusory correlation being tested. Finally, column "Qualitative Miss" indicates whether participants estimated the majority outcome in the minority group as being the minority outcome in that group.

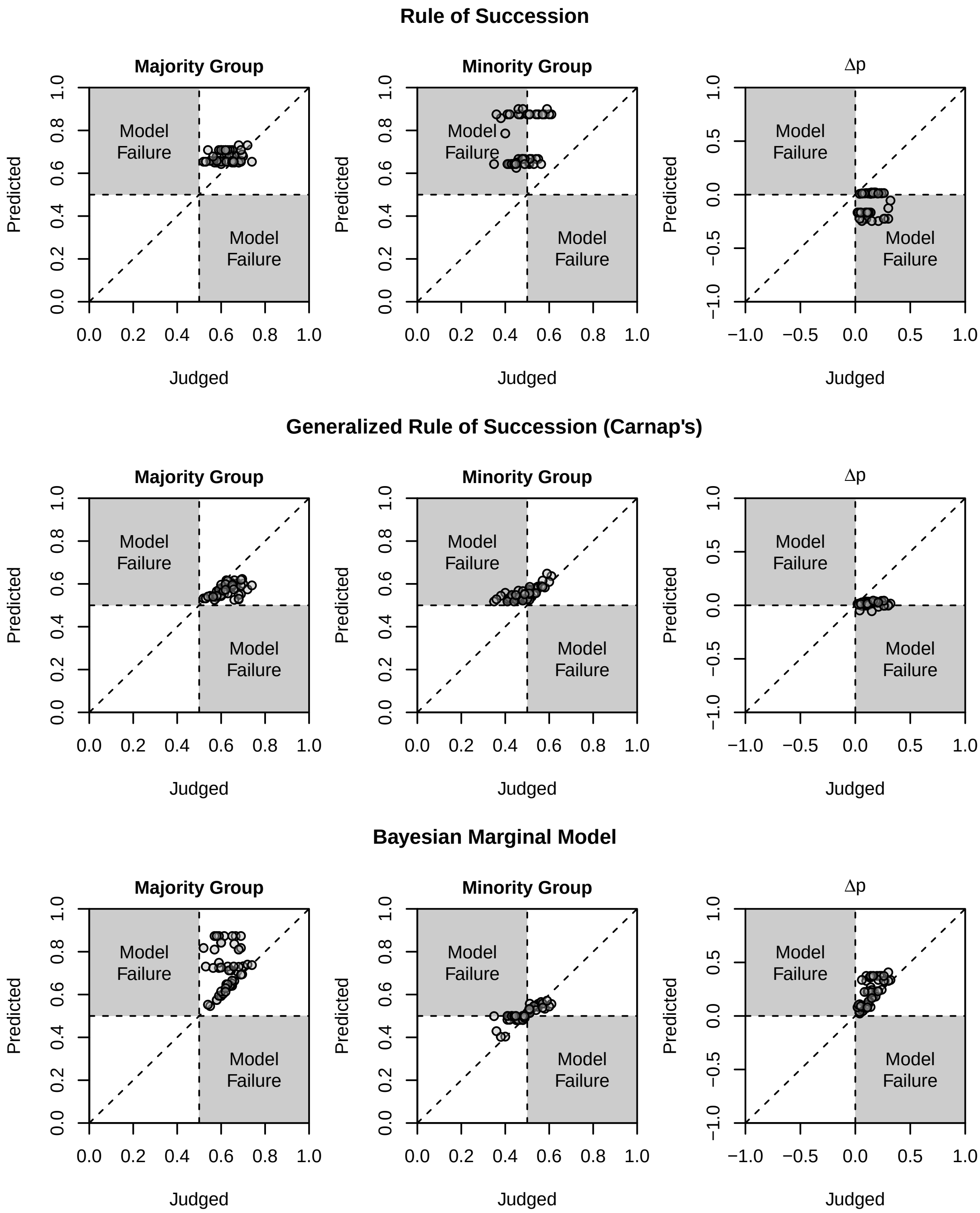

Figure 4. Model fits for the studies listed in Table 2. The 'model failure' quadrants correspond to regions in which there is a qualitative mismatch between model predictions and participants' judgements.

Interim Discussion

The case for the Rule of Succession as a successful normative account stands on two legs. The first one is the notion that it constitutes an optimal solution to an inference problem. The second is its ability to capture the observed effects in the data. What our investigation shows is that the optimality of the Rule of Succession and the prediction of illusory correlations only holds under very specific circumstances and conceptualizations that are not well justified. The Rule of Succession often failed to account for Type-S illusory correlations. Indeed, we applied the rule to cases outside the domain of IC_1 effects for which it was originally proposed. Yet, it is not clear what kind of rationale would permit a demarcation between IC_1 and IC_S effects. Additionally, the rule also fails to account for the qualitative misses that are often found in the minority-group estimates. These misses show that illusory correlations involve something more than a regression to the $\frac{1}{2}$-midpoint. Altogether, the theoretical and empirical issues discussed here raise the question whether a normative account of illusory correlations is possible. In the section below, we will answer this question affirmatively.

A Bayesian Marginal Account

As previously discussed, Fiedler and colleagues proposed that the observed illusory correlations are due to the use of a simple *pseudocontingency heuristic* that can be applied to marginal frequencies (for an overview, see Fiedler et al., 2013): *If two things occur often then assume that they are associated.* Because most instances come from Group A and most outcomes are '+', Group A is thus assumed to be associated with the occurrence of '+'. According to Fiedler et al. (2009), this heuristic is broadly used and is claimed to underlie several phenomena, like the perception of ecological correlations (Hammond, 1973). The case for this heuristic typically revolves around its merits under the lens of ecological rationality: Even though one cannot *conclude* the presence of an association between the variables based on the marginal frequencies, this heuristic succeeds in capturing the sign of actual contingencies found in different environments (Kutzner et al., 2011).

The goal of this section is to show that one can make a case for this so-called heuristic within the tenets of Bayesian inference and to demonstrate its ability to overcome the limitations of the Rule of Succession. For this purpose, we will elaborate on an early proposal by Klauer (2015), which we will refer to as the *Bayesian Marginal Model*. We will explore the model's viability vis-a-vis the corpus of data collected. The reason for using Bayesian inference as a normative framework is that it provides us with a way to coherently represent beliefs regarding latent quantities such as population-level probabilities, and how these beliefs should be informed by incoming data (Griffiths & Tenenbaum, 2005; Knill & Richards, 1996; Oaksford & Chater, 1994; Tauber, Navarro, Perfors, & Steyvers, 2017). Importantly, note that the assumed reliance on Bayesian inference is predicated on the assumption that the reasoning processes used when generating probability estimates approximates the process of taking samples from a distribution (for an overview, see Sanborn & Chater, 2016). Also note that Bayesian accounts like the one developed below are rational in the sense that they lead to *coherent* sets of beliefs that cannot be exploited by an outsider through 'Dutch-Book' scenarios (Teller, 1973). This coherence is independent of whether they are optimal or not, which is something that will ultimately hinge on the match between the priors and the environment, among other aspects.

The model proposed explicitly adopts a subjectivist view of probability, representing the relative strength of (rational/coherent) beliefs in terms of probability *distributions*. This contrasts with the Rule of Succession as advocated by Costello and Watts (2019), who focused on point estimates and utilized their relative accuracy in known environments as a yardstick for rationality. These differences are reflected in our discussion of the Bayesian Marginal Model's IC predictions, which focuses on the emergence of IC effects at the level of the posterior distributions of p_A and p_B. Thus, if participants base their probability estimates on a sample from the posterior distribution (and their frequency estimates on a sample of the posterior predictive distribution for the cell frequencies), then our predictions apply for the observed data without the need to postulate that people use a loss function of any kind or that groupwise proportions p in people's environment follow a uniform distribution.

The model, its properties, and its performance

According to the pseudocontingency account (e.g., Fiedler et al., 2009), contingencies are inferred based on the association of skewed marginal frequencies. Even though marginal frequencies do not determine cell frequencies and thus the contingency, they do restrict the range of possible values (Duncan & Davis, 1953). Therefore, our model jointly considers the four cells of a 2 × 2 contingency table (see Table 1). It takes all possible cell distributions, with cell frequencies $\boldsymbol{k} = (k_{00}, k_{01}, k_{10}, k_{11})$ and cell probabilities $\boldsymbol{p} = (p_{00}, p_{01}, p_{10}, p_{11})$, into account that are in line with the observed marginal frequencies. These potential cell distributions and thus contingencies are weighted depending on the marginal frequencies and the reasoner's prior beliefs. The resulting output is a (posterior) belief about the contingency, or broadly speaking, the four cells will be reconstructed by the marginals.

Like any Bayesian account, our model can be broken down into three main components:

- A prior distribution $P(\boldsymbol{p})$ over the probability vector $\boldsymbol{p}$,
- a likelihood, here given by the multinomial probability mass function $P(\boldsymbol{k} \mid \boldsymbol{p})$, and
- a posterior distribution $P(\boldsymbol{p} \mid \boldsymbol{k})$ of the probability vector $\boldsymbol{p}$ given the data $\boldsymbol{k}$.

The prior distribution captures our beliefs regarding the probabilities $\boldsymbol{p}$ before encountering data $\boldsymbol{k}$. Although infinitely many priors are possible, we will focus our discussion on 'uninformed' priors that do not establish a predominance of any of the four table cells. Given the assumption that the data follow a multinomial distribution, we use a Dirichlet distribution to capture prior beliefs:

$$g(\boldsymbol{p}) = \frac{\Gamma(\sum \alpha_{ij})}{\prod \Gamma(\alpha_{ij})} \prod p_{ij}^{\alpha_{ij}-1}, \tag{6}$$

where $\boldsymbol{\alpha} = (\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11})$, with all $\alpha_{ij} > 0$, is a vector of concentration parameters that can be interpreted as pseudocounts on each of the table cells. The expected prior

p_{ij} under this distribution is $\frac{\alpha_{ij}}{\sum \alpha_{ij}}$. Setting all α_{ij} to the same value for all i and j yields a prior distribution in which all joint outcomes are expected to be equiprobable. The larger this common value is, the stronger the prior beliefs will be (just as in the case of α and β parameters in the Beta distribution). The use of prior beliefs that are unbiased towards one group and/or outcome are reasonable, when studies use variables for which little to no prior information is given, as is the case of the studies included in our corpus. The motivation for using the Dirichlet distribution is twofold: First, it is an extremely flexible distribution, able to capture a multitude of beliefs via $\boldsymbol{\alpha}$. Second, while in the Rule of Succession estimates are computed per group, our model considers all four cell frequencies jointly and thus contingencies. Therefore, the generalization of the Beta distribution discussed earlier is needed. The Dirichlet distribution is the conjugate prior of the multinomial distribution. This means that when encountering multinomial-distributed data with probability mass function

$$P(\boldsymbol{k} \mid \boldsymbol{p}) = \binom{M}{k_{00}, k_{01}, k_{10}, k_{11}} \prod p_{ij}^{k_{ij}}, \tag{7}$$

the updated posterior beliefs $P(\boldsymbol{p} \mid \boldsymbol{k})$ will also be Dirichlet distributed, with parameter vector $\boldsymbol{\alpha}^* = \boldsymbol{\alpha} + \boldsymbol{k}$.

The prior and posterior Dirichlet distributions implicitly capture the reasoner's beliefs about the presence of an association between the variables (i.e., ϕ). The left panel of Figure 5 shows the case of $\boldsymbol{\alpha} = (1, 1, 1, 1)$, which yields a prior ϕ distribution whose mass is dispersed over the [-1,1] range but still more concentrated around zero. The center and right panels of Figure 5 illustrate the posterior ϕ distributions that are obtained when individuals update their beliefs based on the joint frequencies reported in Table 1. In the case of the middle panel (IC_1 example), we see that the posterior distribution is more peaked around $\phi = 0$, consistent with the fact that $P(+|A) = \frac{27}{36} = P(+|B) = \frac{9}{12}$. In the right panel ($\text{IC}_S$ example), most of the posterior mass is on the negative range, consistent with the fact that $P(+|A) = \frac{16}{24} < P(+|B) = \frac{8}{8}$ and that the sample's ϕ is -0.33.

Now, let us turn to the updating of beliefs based on the observation of *marginal frequencies*. Although marginals are insufficient to determine the underlying joint fre-

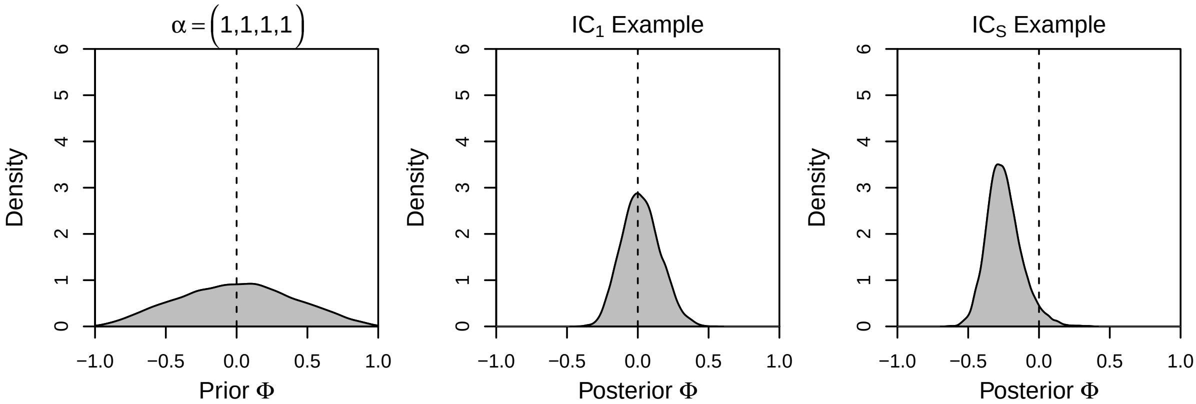

Figure 5. Prior ϕ under $\boldsymbol{\alpha} = (1, 1, 1, 1)$ and the subsequent posterior distributions when updating the prior with the joint frequencies reported in Table 1.

quencies, they nevertheless *constrain the range of possibilities.* To see this, note that given some set of marginal frequencies, the joint frequencies are completely determined if one of the joint frequencies is known. For instance, if we know k_{00}, then $k_{01} = k_{0\bullet} - k_{00}$, $k_{10} = k_{\bullet 0} - k_{00}$, and $k_{11} = M - k_{0\bullet} - k_{\bullet 0} + k_{00}$. It follows that we can generate all possible sets of joint frequencies by varying k_{00} between $\max(0, k_{0\bullet} + k_{\bullet 0} - M)$ and $\min(k_{0\bullet}, k_{\bullet 0})$. Given this relationship between joint and marginal frequencies, the probability of any set of marginal frequencies corresponds to:

$$P(k_{0\bullet}, k_{1\bullet}, k_{\bullet 0}, k_{\bullet 1} \mid \boldsymbol{p}) = \sum_{k_{00}=\max(0,k_{0\bullet}+k_{\bullet 0}-M)}^{\min(k_{0\bullet},k_{\bullet 0})} P(k_{00}, k_{0\bullet}-k_{00}, k_{\bullet 0}-k_{00}, M-k_{0\bullet}-k_{\bullet 0}+k_{00} \mid \boldsymbol{p}), \tag{8}$$

where the probabilities on the right side are given by Equation 7.

Table 3 reports all possible joint frequencies under the marginals of the examples in Table 1. The associations (quantified by the ϕ-statistic) observed in most possible sets of joint frequencies turn out to be positive and the range of possible associations to be asymmetric, as associations can be as high as $\phi = 1$ but cannot go below $\phi = -0.33$. As shown in Figure 6, this asymmetry in possible associations across the range of possible joint frequencies will be *greater* the more skewed the base rates are (see also Fiedler et al., 2013; Kareev, 1995).

Table 3
Possible joint frequencies and associations (ϕ) under the marginals of the IC_1 and IC_S examples in Table 1

Type-1 Illusory Correlation Example (True $\phi = 0$)					
Index	k_{00}	k_{01}	k_{10}	k_{11}	ϕ
1	24	12	12	0	-0.33
2	25	11	11	1	-0.22
3	26	10	10	2	-0.11
4	**27**	**9**	**9**	**3**	**0.00**
5	28	8	8	4	0.11
6	29	7	7	5	0.22
7	30	6	6	6	0.33
8	31	5	5	7	0.44
9	32	4	4	8	0.56
10	33	3	3	9	0.67
11	34	2	2	10	0.78
12	35	1	1	11	0.89
13	36	0	0	12	1.00
Type-S Illusory Correlation Example (True $\phi = -0.33$)					
Index	k_{00}	k_{01}	k_{10}	k_{11}	ϕ
1	**16**	**8**	**8**	**0**	**-0.33**
2	17	7	7	1	-0.17
3	18	6	6	2	0.00
4	19	5	5	3	0.17
5	20	4	4	4	0.33
6	21	3	3	5	0.50
7	22	2	2	6	0.67
8	23	1	1	7	0.83
9	24	0	0	8	1.00

Using Bayes' Theorem, we can combine Equations 6 and 8 to obtain the density of $\boldsymbol{p}$ *conditional* on the marginal frequencies and prior belief distribution:

$$g(\boldsymbol{p} \mid k_{0\bullet}, k_{1\bullet}, k_{\bullet 0}, k_{\bullet 1}, \boldsymbol{\alpha}) \propto \sum_{k_{00}=\max(0, k_{0\bullet}+k_{\bullet 0}-M)}^{\min(k_{0\bullet}, k_{\bullet 0})} \binom{M}{k_{00}, k_{0\bullet}-k_{00}, k_{\bullet 0}-k_{00}, M-k_{0\bullet}-k_{\bullet 0}+k_{00}} \times \frac{\prod \Gamma(\alpha_{i,j}+k_{i,j})}{\Gamma(\sum(\alpha_{i,j})+M)} \times g(\boldsymbol{p} \mid \boldsymbol{\alpha}^*(k_{00})). \quad (9)$$

This posterior density is a mixture of the posterior Dirichlet distributions $g(\boldsymbol{p} \mid \boldsymbol{\alpha}^*)$ with $\boldsymbol{\alpha}^*(k_{00}) = \boldsymbol{\alpha} + \boldsymbol{k}'(k_{00})$, where $\boldsymbol{k}'(k_{00})$ corresponds to the vector of joint frequencies in the contingency table consistent with the given marginals and the given k_{00}. Each mixture

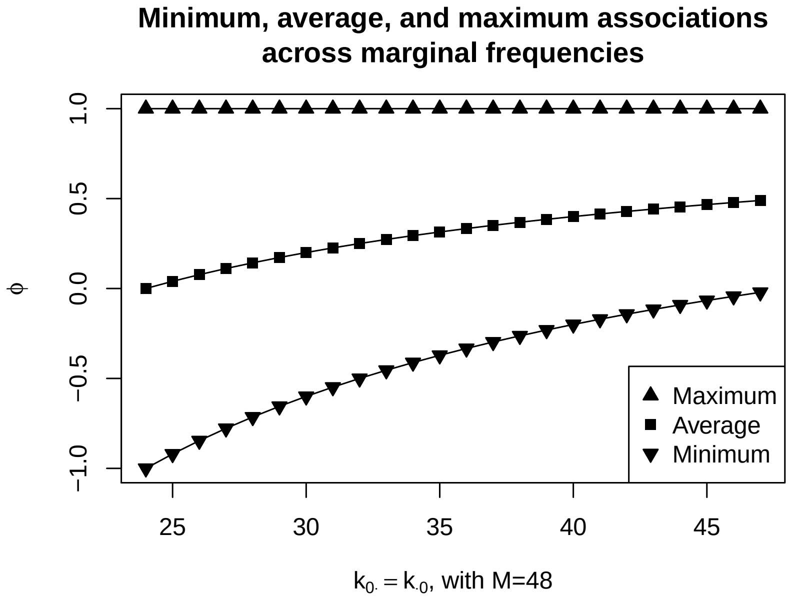

Figure 6. Minimum, average, and maximum associations (quantified by the ϕ statistic) across different marginal frequencies ($k_{0\bullet} = k_{\bullet 0}$, with $M = 48$). When computing the average association for a given set of marginals, we attributed equal weights to all possible ϕ values.

weight (which is proportional to the first two multiplicative terms) is a function of the marginal frequencies and prior $\boldsymbol{\alpha}$. The second row of panels in Figure 7 illustrates the weights associated with the IC_1 example in Table 3 under different priors. The different joint-frequency sets are equally weighted when all $\alpha_{i,j} = 1$. Moving away from this prior leads to differential weighting. For instance, when all $\alpha_{i,j} < 1$, the more 'extreme' joint frequencies are overweighted, with particular emphasis on the joint-frequencies that produce the largest ϕ. The reason behind this asymmetric overweighting is the fact low $\alpha_{i,j}$ values push most mass to extreme probabilities (Figure 1 provides an illustrative example, namely a Beta distribution with $\alpha = \beta = \frac{1}{2}$). Because the marginal frequencies are skewed, there is only one possible table that is consistent with these extreme probabilities, hence its overweighting.

Taking a single sample from this mixture density is straightforward: We first sample from a discrete distribution, with probabilities corresponding to the mixture weights. The sampled integer determines the component Dirichlet distribution from which to subsequently sample probabilities from. Finally, one can use these probabilities to sample

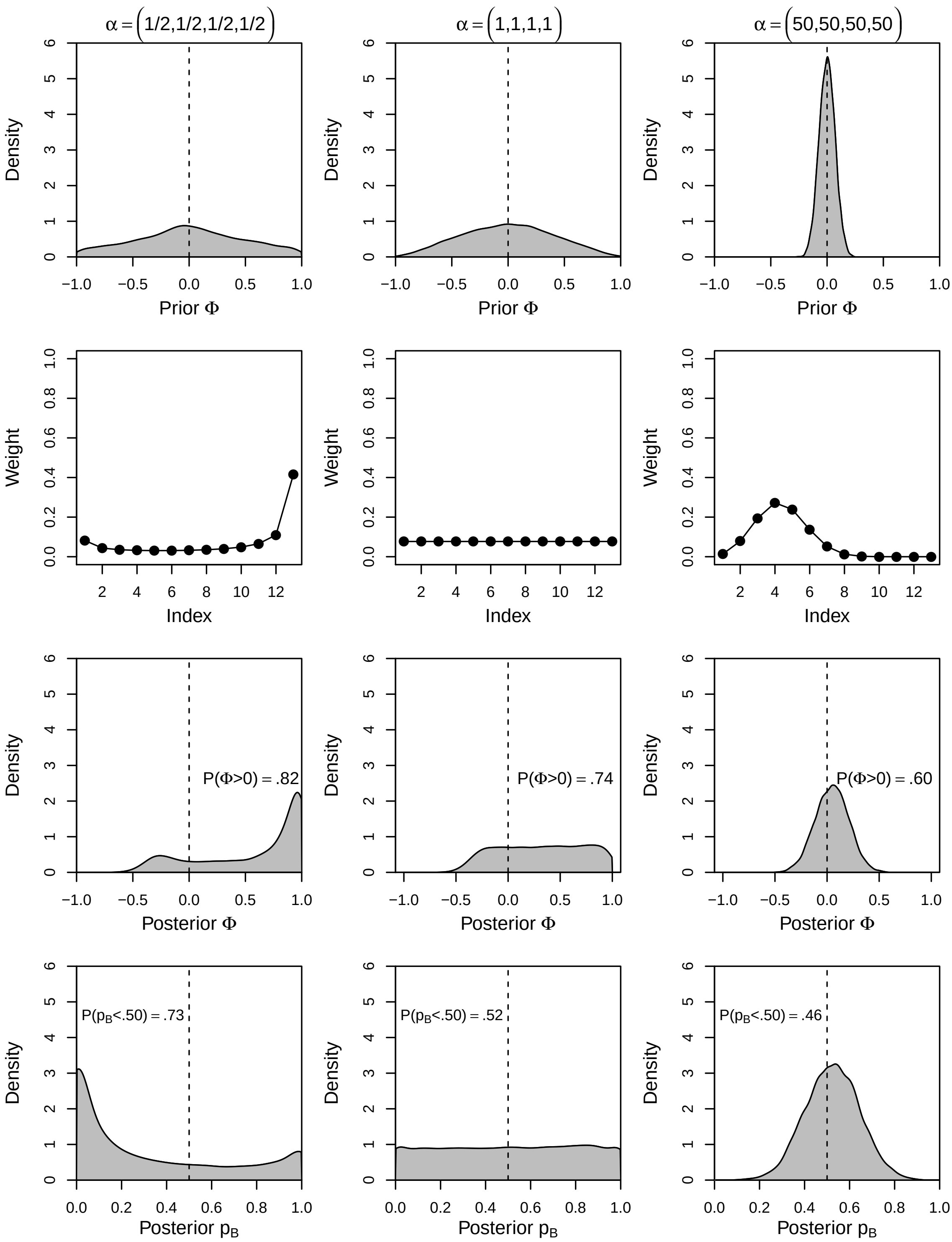

Figure 7. Illustration of the Bayesian Marginal Model and its predictions under different priors for the IC_1 example given in Table 1.

a set of joint frequencies of size M. Multiple samples can be taken by repeating these steps.[9] The third row of panels in Figure 7 illustrates the posterior ϕ distributions for the IC_1 example in Table 1 under different priors. We see that illusory correlations will be most prevalent in cases where the reasoner does not have strong prior beliefs against an association (e.g., when all $\alpha_{ij} = 1$). When such beliefs do in fact exist (e.g., when all $\alpha_{ij} = 50$), the propensity and magnitude of illusory correlations will be reduced. For example, consider a reasoner who learns that a) most students in a classroom perform well in tests, and b) that most of these students do their homework. Under the notion that an association is not implausible, the reasoner is likely to infer a positive association. In contrast, it seems somewhat implausible a priori that either of these observations are associated with students liking chocolate, which would justify a prior ϕ that is more strongly peaked on zero. Under such prior, it is unlikely that one would subsequently infer an association between student performance and liking chocolate, even after learning that most students like chocolate. Moreover, note that the occurrence of qualitative misses is far from rare. Instead, its occurrence is quite robust across priors (see the bottom row of Figure 7 depicting posterior probabilities for the minority group B). This follows from the fact that $\frac{k_{10}}{k_{1\bullet}} < \frac{1}{2}$ in 46% of the possible joint frequencies that are consistent with marginals found in the IC_1 example (see Table 3). A 'problematic' phenomenon under the Rule of Succession turns out to be an expected result under the Bayesian Marginal Model.

Given the potential of the Bayesian Marginal Model, we investigated its ability to capture the illusory correlations and qualitative misses observed in our study corpus. We restricted the α_{ij} priors to take on the same value, which means that *the model only has one free parameter*. This is analogous to the constraint imposed in the generalization of the Rule of Succession when fitting it to the same study corpus. In fact, our parameter restriction can be represented in the exact same way, with $\alpha_{ij} = \lambda \cdot \pi_{ij}$, with $\pi_{ij} = \frac{1}{4}$

[9] Note that the output of the Bayesian Marginal Model is a posterior distribution, not a point estimate. As the model is concerned with epistemic rationality as compared with optimality, it is not committed to the use of a specific loss function or any loss function at all. Also note, we focus our model evaluation on the ability to capture qualitative patterns rather than the relative precision of point predictions, wherefore our analysis is also not committed to some specific point estimator.

for $i = 1, 2$ and $j = 1, 2$, where λ is a free parameter quantifying the concentration or peakedness of the prior.[10]

The results depicted in Figure 4 show that, overall, the Bayesian Marginal Model performs better than the Rule of Succession and its generalization in the sense that it can capture qualitative misses. Taken together, the assumption that individuals estimate probabilities based on marginal frequencies is able to account for the qualitative patterns found in the data. Put differently, one can make a rational as well as an empirical case for the pseudocontingency account proposed by Fiedler and colleagues.

Dependencies Across Probability Estimates, Skewed Beliefs, and New Data Formats

One of the main virtues of normative models is the fact that they often introduce new testable predictions that would have not been considered otherwise. We explore a number of such predictions concerning dependencies between the frequency estimates, the role of skewed beliefs, and new data formats.

Dependencies. As previously stated, the samples taken under the Bayesian Marginal Model come from a mixture of Dirichlet distributions, each associated with one possible contingency table. Across these tables, the joint frequencies are negatively correlated as they sum to M, such that whenever k_{00} is small/large, k_{10} is likely to be large/small. As an example, we generated posterior samples for the IC_1 example in Table 1. We used a different $\boldsymbol{\alpha}$ prior per sample (once again, with all α_{ij} being equal), which was taken from a mixture between a Gamma distribution with shape and rate parameters 1.26 and 1.16, respectively, and a uniform distribution between 0 and 50. The mixture weights were .63 and .37, respectively. This mixture distribution was found to be the one best fitting the $\boldsymbol{\alpha}$ estimates obtained with the study corpus. We found the resulting $\hat{p}_A$ and $\hat{p}_B$ samples to be negatively correlated (Spearman's $r = -0.67$).

In contrast, the probability estimates $\hat{p}$ coming from the Rule of Succession are

[10] For the sake of comparability, we fitted the Bayesian Marginal Model in exactly the same way as the Rule of Succession, using a constrained least-squares algorithm based on the model's predictions for posterior means.

assumed to be computed separately for each group, with no presumption of any kind of crosstalk. This means that $\hat{p}_A$ and $\hat{p}_B$ should be *stochastically independent* (i.e., the expected correlation is 0). We tested these differing predictions by fitting a regression model to the IC_S data coming from Klauer and Meiser (2000) and Meiser et al. (2018, see Table 2).[11] In addition to having $\hat{p}_A$ as a predictor of $\hat{p}_B$, we also introduced the predictor 'Experiment'. The estimated coefficient was $\beta_{\hat{p}_A} = -0.25$ ($SE = 0.08$, $p = .001$), supporting the predictions made by the Bayesian Marginal Model.[12]

Skewed Beliefs. So far, we assumed that all parameter values of prior $\boldsymbol{\alpha}$ are equal, which is reasonable when participants know nothing about the relative frequency of groups and/or outcomes beforehand. However, when skewed frequencies are expected *a priori*, illusory correlations are only expected under certain conditions. Figure 8 shows the posterior ϕ distributions under different priors that respect the 3:1 ratio found in the IC_1 example in Table 1. The figure contains examples involving skewed outcome priors with outcome '+' cells being overweighted relative to outcome '−' cells (e.g., $\boldsymbol{\alpha} = (3, 1, 3, 1)$) and skewed group-membership priors (e.g., $\boldsymbol{\alpha} = (3, 3, 1, 1)$). As can be seen, the positive bias in the posterior ϕ distribution, indicating an IC_1 effect, vanishes the more informative the $\boldsymbol{\alpha}$ becomes. Consequently, illusory correlations are expected to be reduced or even absent whenever individuals have clear prior beliefs regarding the actual skewness of the event distributions that is to be experienced. Similarly, no IC_1 effect is expected, when prior beliefs correspond to the actual frequencies (e.g., $\boldsymbol{\alpha} = (9, 3, 3, 1)$). Finally, Figure 8 also depicts posterior ϕ distributions under priors that overweight the 'wrong' cells (e.g., $\boldsymbol{\alpha} = (1, 1, 3, 3)$ and $\boldsymbol{\alpha} = (1, 3, 1, 3)$) which are expected to increase illusory correlations. Therefore, future work should consider systematically investigating the role of prior beliefs in the occurrence of illusory correlations.

[11] We thank Thorsten Meiser for making the data from Meiser et al. (2018) available.

[12] As a caveat, note that the prediction assumes (a) equal $\alpha_{i,j}$ and (b) that respondents do not differ in their propensities or response biases to produce generally high or low frequency estimates. The first assumption is plausible where groups and outcomes are abstract or novel. The second assumption is more plausible when looking at intraindividual correlations across several IC tasks administered to the same person.

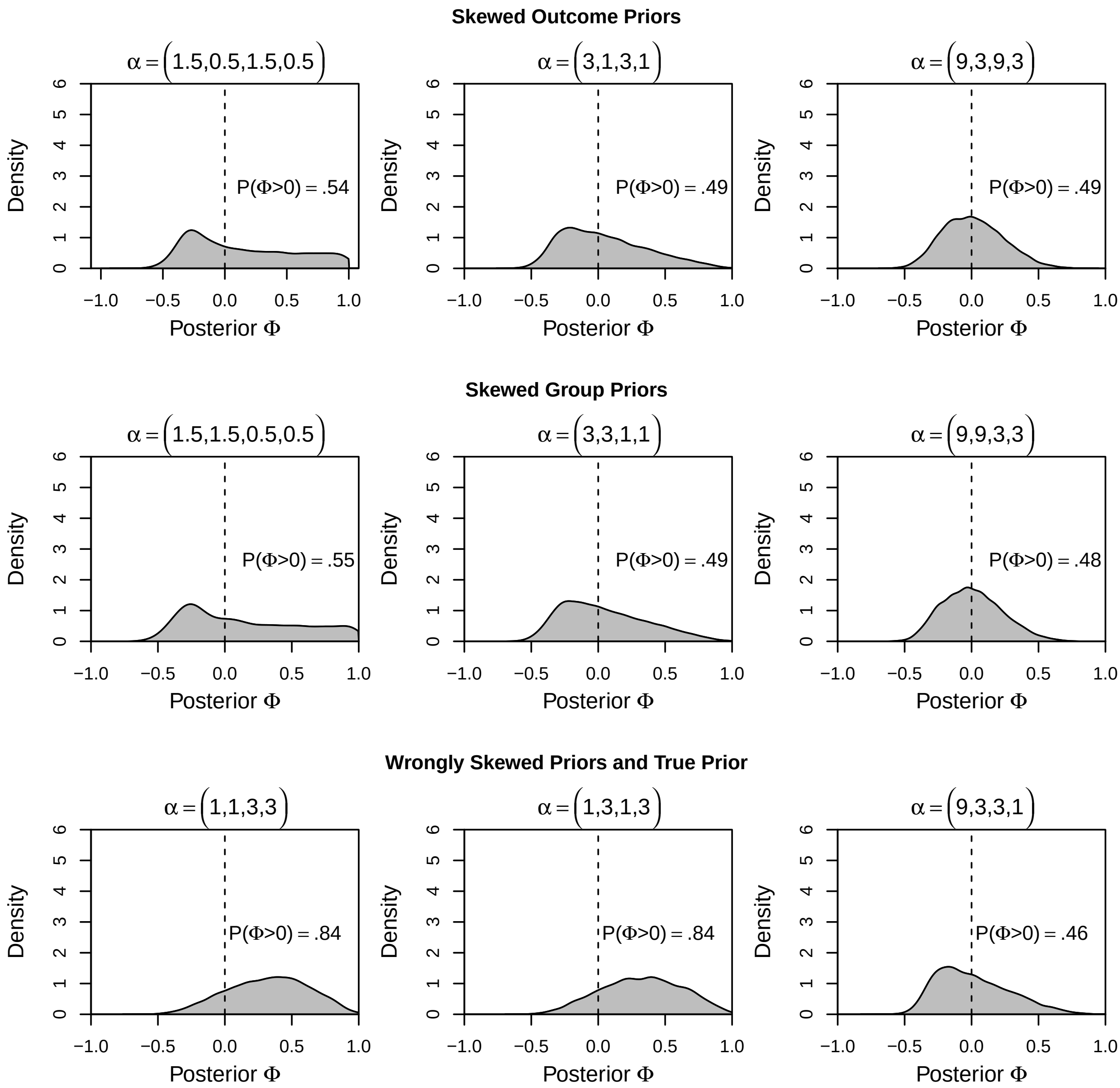

Figure 8. Posterior ϕ for the IC_1 example in Table 1 according to the Bayesian Marginal Model under different skewed priors.

Asymmetry. Asymmetric illusory correlation effects are documented in the literature, with IC effects tending to be greater in contexts that are mostly comprised of positive outcomes (e.g., Fleig et al., 2017; Meiser et al., 2018). These asymmetries are observed in experimental designs in which the overall positivity/negativity is manipulated within participants. In such scenarios, there are two contexts (e.g., Town X and Town Y). The marginal frequencies of groups and outcomes are skewed within each context, and co-vary across contexts: what is frequent within one context is infrequent within the other context and vice versa. One possible explanation for this asymmetry is that the overall positivity/negativity of each context is detected quickly by the participants, which leads them to subsequently make inferences about each group in each context using skewed priors: In both contexts, the most frequent outcome will be overweighted relative to the cells corresponding to the infrequent outcome. In terms of the Bayesian Marginal Model, such priors would reduce the magnitude of the IC_S effects (see the first row of Figure 8). Given people's tendency to attend more to losses than gains (Yechiam & Hochman, 2013), one would expect a greater skew in the context where negative outcomes are most frequent, which in turn would produce the observed asymmetry.

New Response Formats. Given that the output of the Bayesian Marginal Model is a posterior distribution, the model readily generates predictions for alternative response formats for which the Rule of Succession, relying on one particular point estimate, remains silent. For example, participants should not only be able to generate point estimates for the probabilities $P(+|A)$ and $P(+|B)$, they should also be able to estimate the probabilities that each probability falls into given ranges, and these range estimates should correspond to the model's predictions derived from the posterior distribution. Again, algorithmically, the range estimate could be based on mental sampling from the posterior distribution.

New Data Formats. A final prediction illustrates that the model is readily extended to account for findings using new data formats. For instance, the model can be adapted to the case where data from just one cell of the contingency table is presented. Consider a scenario in which participants learn that 21 of 49 cases under study are desirable behaviors of Group A members. This information necessitates neither skewed marginals nor a contingency. Yet, under flat priors, it should bias posterior beliefs towards an illusory correlation.

Taken together, the Bayesian Marginal Model does not only show explanatory sufficiency (it accounts for the major qualitative patterns in the data corpus) and a normative backing, it is also a fruitful theory, readily generating new predictions, including the effects of prior beliefs (see Coenen et al., 2018).

General Discussion

Theoretical work on illusory correlations has relied on one of two main interpretations. On the one hand, illusory correlations have been argued to reflect biased or at least normatively unjustified information processing. On the other hand, Costello and Watts (2019) argued that illusory correlations reflect optimal, mathematically correct and hence rational responses. The present work raises a number of issues that ultimately speak against both theoretical interpretations: First, a sensible account of a specific illusory correlation (e.g., IC_1) should also be able to accommodate closely-related variants (e.g., IC_S). Second, any successful theoretical account needs to go beyond a single effect and consider other qualitative patterns in the data. By focusing on illusory correlation effects in isolation, one is ignoring the fact that the data often show qualitative misses that cannot be accommodated by models that assume some kind of regression towards the mean/prior. Third, we nevertheless disagree that illusory correlations reflect an irrational, normatively unjustified kind of information processing: The Bayesian Marginal Model provides a normative reconstruction of the pseudcontingency heuristic, predicting IC_1 and IC_S effects based on the use of marginal frequencies following the norms of Bayesian belief updating. Furthermore, this model turned out to be able to account for the qualitative patterns in the study corpus and successfully predict the presence of dependencies across individual probability estimates. The results demonstrate that a successful normative account of illusory correlations is possible.

The occurrence of illusory correlations in Figure 7 should be sufficient to dispel the notion that the pseudocontingency heuristic proposed by Fiedler and colleagues (e.g., Fiedler et al., 2009) is logically invalid. Marginal frequencies do carry valid information about the joint frequencies, otherwise, in the Bayesian Marginal Model, the posterior beliefs would match the priors. Under weakly informative priors, skewed marginal frequencies are expected to result (via Bayes' theorem) in the belief that both variables are associated.

Between the two contrary interpretations of illusory correlations as irrational and unjustified versus 'mathematically correct' and rational, we therefore argue for a third possibility, namely that they reflect an instance of *bounded rationality* (Simon, 1990). According to this view, limitations of the human information-processing system introduce constraints that (a) prevent human reasoners and decision makers from reaching overall optimal solutions, but (b) do not prevent them from finding normatively adequate solutions within the smaller search space of solutions consistent with these constraints. In the present case, an optimally tuned information-processing system would make full use of the presented joint frequencies instead of only the marginal frequencies. A human decision maker's memory limitations, however, constrain the number of cases in which the realization of two variables in an event are bound together (e.g., Bays, Wu, & Husain, 2011; Treisman, 1998; Wheeler & Treisman, 2002), which might discourage individuals to use them in inferential judgments. Eder et al. (2011) have reported results supporting this hypothesis, in which they show that the impairment of working memory exacerbates the occurrence of illusory correlations. In other cases, a reasoner might often encounter the different variables separately, sometimes even at different points in time. Given these limitations, the Bayesian Marginal Model holds that boundedly rational reasoners fall back on the marginal frequencies from which illusory correlations are predicted as an outcome of a normatively adequate, Bayesian use of the information in these marginal frequencies.

According to Marr's (1982) classification, the Bayesian Marginal Model provides a characterization at the *computational level*. It is assumed that individuals have the capacity to perform or mimic the normatively appropriate computations and use the normative prescriptions to guide their choices and behaviors. At no point does it make any claims at the *algorithmic level*, that is, claims regarding the mechanisms by which the aforementioned computations are performed or mimicked (for a relevant discussion, see Van Rooij, 2008). We find modeling at the computational level to be the most appropriate here given the nature of the data available: In illusory-correlation experiments, each participant contributes very few judgments. This means that one does not have sufficient degrees of freedom to test theoretical accounts that make fine-grained algorithm-level claims. To make matters worse, aggregating data across participants is very likely to introduce distortions that can spuriously reject models and lead researchers astray (for recent discussions, see Kellen & Klauer, 2019; Regenwetter & Robinson, 2017).

Addressing Concerns on Model Flexibility, Evaluation, and Comparison

The minimal requirement for any candidate model is that it succeeds in capturing the data at hand (Bogen & Woodward, 1988). Given that we are dealing with aggregate data, it is reasonable to focus our evaluation on the models' ability to capture the qualitative patterns found in the data (for discussions, see Kellen, 2019; Navarro, 2019; Shiffrin & Nobel, 1997). It turns out that the Rule of Succession (as well as its generalization) cannot fulfill this requirement. This reason alone is why the Rule of Succession has to be dismissed; not because our comparison between the Rule of Succession and the Bayesian Marginal Model might place the former model at a disadvantage due to not taking potential differences in flexibility into account (e.g., Kellen, Klauer, & Bröder, 2013; Myung & Pitt, 2018). Note also that in our fits of the Rule of Succession and the Bayesian Marginal Model to the data corpus, we used only one free parameter, governing the peakedness of the prior.

The Bayesian Marginal Model succeeded where the Rule of Succession failed. However, Bayesian models are notorious for their extreme flexibility, due to the possibility of establishing priors and likelihoods in so many different, arbitrary ways. This may render their eventual success as nothing short of certain (Bowers & Davis, 2012; see also Coenen, Nelson, & Gureckis, 2019). We argue that the development history of the Bayesian Marginal Model successfully addresses this concern. First, the model is a formal instantiation of the pseudocontingency heuristic proposed much earlier by Fiedler and colleagues (e.g., Fiedler et al., 2009, 2013). This means that the likelihood function of the model was constrained from the onset. As sketched out by Klauer (2015), the model is able to yield IC_1 effects under non-informative priors, introducing the notion that the occurrence of illusory correlations does not require any kind of bias in people's prior beliefs. In the present paper, we went beyond Klauer's first proof of concept and tested the model against a large corpus of published studies while keeping the basic components of the model fixed. The only part of the model that was 'free to vary' across studies was a *single* concentration parameter λ, which determines the strength of the prior (here non-informative beliefs, i.e., prior beliefs that are kept unbiased towards any specific group or outcome).[13] Moreover, the Bayesian Marginal Model *successfully predicted* the presence of (a) qualitative misses and (b) negative dependencies across probability estimates. The model is fruitful in generating a host of new predictions, as illustrated in the section "Dependencies Across Probability Estimates, Skewed Beliefs, and New Data Formats".

Decisions from Experience and Skewed Sampling

The learning phase in illusory-correlation experiments shares important similarities with some of the paradigms used in the study of *experience-based decisions* (e.g., Hertwig et al., 2004; Kellen, Pachur, & Hertwig, 2016). As an example, consider the following options:

$$\mathcal{A} = \left(\begin{smallmatrix} \$120 & \$0 \\ .70 & .30 \end{smallmatrix} \right) \quad \mathcal{B} = \left(\begin{smallmatrix} \$100 \\ 1 \end{smallmatrix} \right)$$

Option $\mathcal{A}$ is a lottery stating the yields \$120 with probability .70, otherwise nothing. Option $\mathcal{B}$ always yields \$100. Note that $\mathcal{B}$ has a greater expected value than $\mathcal{A}$ (\$100 and \$84, respectively). In a typical *sampling paradigm*, both options' outcomes and outcome probabilities are learned by sampling random draws from each of them before the participant makes a final, consequential choice (Hertwig et al., 2004).

If option $\mathcal{A}$ is sampled more often than option $\mathcal{B}$, one might infer an association between the majorly-sampled option and monetary gains that would increase $\mathcal{A}$'s relative attractiveness. Note that a preference for $\mathcal{A}$ over $\mathcal{B}$ would deviate from people's choices when options are described (e.g., Hertwig et al., 2004). The potential effects of skewed sampling in experience-based choices has received minimal attention by researchers so far (e.g., Bott & Meiser, in press).[14] This situation is partly due to participants being

[13] Note that the use of non-informative priors was accompanied by a focus on studies in which such priors are deemed highly plausible. As previously discussed, the studies included in the corpus involved variables for which little to no prior information is available.

[14] Mostly, investigated effects referred to how representative or veridical the options' sampled probabilities are (e.g., Hau, Pleskac, Kiefer, & Hertwig, 2008; Hilbig & Glöckner, 2011), rather than testing the effect of sample sizes per option on preferences.

allowed to freely sample both options as often and in any order they desire. However, this freedom ignores the fact that in realistic settings, some options are experienced more often than others, without any control on the decision-maker's part.

Final Thoughts

Finally, let us highlight the fact that the Bayesian Marginal Model proposed here attributes a passive observer role to the reasoner. This situation implies that many phenomena attributed to the actions of the reasoner in a given environment, such as the way in which different options are actively sampled, or to the environment explored, are outside its current scope (for an overview on the topic of information search, see Todd, Hills, & Robbins, 2012). For instance, Le Mens and Denrell (2011) considered the case in which two options are objectively equally attractive, but asymmetric with regard to information: We can learn about one of the options only when choosing it (*selective access*), whereas we can learn about the other irrespective of whether it is chosen or not (*systematic access*). This asymmetry will lead to a preference for the latter option, even if the decision maker is rational and knows the structure of the environment.

Likewise in naturalistic settings, different groups differ in how informational access is controlled. For some, we are bombarded with relevant samples in the news media almost everyday; for others, we have to make an effort to gather the relevant bits of information. Such informational asymmetries and active sampling biases (e.g., Fiedler & Wänke, 2009) likely interact with the formation of illusory correlations. The integration of different Bayesian accounts thereof into a single, comprehensive account constitutes an important theoretical challenge that is yet to be tackled.

References

Aydogan, I. (2019). *The role of prior beliefs and their updating in decisions under experienced ambiguity.* Manuscript submitted for publication.

Bays, P. M., Wu, E. Y., & Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia*, *49*(6), 1622–1631.

Berndsen, M., McGarty, C., Van Der Pligt, J., & Spears, R. (2001). Meaning-seeking in the illusory correlation paradigm: The active role of participants in the categorization process. *British Journal of Social Psychology*, *40*(2), 209–233. doi: 10.1348/014466601164821

Bogen, J., & Woodward, J. (1988). Saving the phenomena. *The Philosophical Review*, *97*(3), 303–352.

Bott, F. M., & Meiser, T. (in press). Pseudocontingency inference and choice: The role of information sampling. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in Psychology and Neuroscience. *Psychological Bulletin*, *138*(3), 389–414. doi: 10.1037/a0026450

Buchak, L. (2013). *Risk and rationality.* Oxford, UK: Oxford University Press.

Bulli, F., & Primi, C. (2006). Illusory correlation and cognitive processes: A multinomial model of source-monitoring. *Review of Psychology*, *13*(2), 95–102.

Busemeyer, J. R., & Bruza, P. D. (2012). *Quantum models of cognition and decision.* Cambridge, UK: Cambridge University Press.

Carnap, R. (1952). *The continuum of inductive methods.* Chicago, IL: Chicago University Press.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, *74*(3), 271–280. doi: 10.1037/h0027592

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

Coenen, A., Nelson, J. D., & Gureckis, T. M. (2019). Asking the right questions about the psychology of human inquiry: Nine open challenges. *Psychonomic Bulletin and Review*, *26*, 1548–1587. doi: 10.3758/s13423-018-1470-5

Costello, F., & Watts, P. (2014). Surprisingly rational: Probability theory plus noise explains biases in judgment. *Psychological Review*, *121*(3), 463–480. doi: 10.1037/a0037010

Costello, F., & Watts, P. (2016). People's conditional probability judgments follow probability theory (plus noise). *Cognitive Psychology*, *89*, 106–133. doi: 10.1016/j.cogpsych.2016.06.006

Costello, F., & Watts, P. (2019). The rationality of illusory correlation. *Psychological Review*, *126*(3), 437–450. doi: 10.1037/rev0000130

DeGroot, M. H. (1975). *Probability and statistics*. Reading, Mass.: Addison-Wesley.

Dougherty, M. R., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review*, *106*(1), 180–209. doi: 10.1037/0033-295X.106.1.180

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

Dzhafarov, E. N., & Kujala, J. V. (2016). Context–content systems of random variables: The Contextuality-by-Default theory. *Journal of Mathematical Psychology*, *74*, 11–33. doi: 10.1016/j.jmp.2016.04.010

Eder, A. B., Fiedler, K., & Hamm-Eder, S. (2011). Illusory correlations revisited: The role of pseudocontingencies and working-memory capacity. *Quarterly Journal of Experimental Psychology*, *64*(3), 517–532. doi: 10.1080/17470218.2010.509917

Elqayam, S., & Evans, J. S. B. T. (2011). Subtracting ought from is: Descriptivism versus normativism in the study of human thinking. *Behavioral and Brain Sciences*, *34*(5), 233–248. doi: 10.1017/S0140525X1100001X

Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environments. *Psychological Review*, *103*(1), 193–214. doi: 10.1037/0033-295X.103.1.193

Fiedler, K. (2000). Illusory correlations: A simple associative algorithm provides a convergent account of seemingly divergent paradigms. *Review of General Psychology*, *4*(1), 25–58. doi: 10.1037/1089-2680.4.1.25

Fiedler, K. (2010). Pseudocontingencies can override genuine contingencies between multiple cues. *Psychonomic Bulletin & Review*, *17*(4), 504–509. doi: 10.3758/PBR.17.4.504

Fiedler, K., & Freytag, P. (2004). Pseudocontingencies. *Journal of Personality and Social Psychology*, *87*(4), 453-467. doi: 10.1037/0022-3514.87.4.453

Fiedler, K., Freytag, P., & Meiser, T. (2009). Pseudocontingencies: An integrative account of an intriguing cognitive illusion. *Psychological Review*, *116*(1), 187–206. doi: 10.1037/a0014480

Fiedler, K., Kutzner, F., & Vogel, T. (2013). Pseudocontingencies: Logically unwarranted but smart inferences. *Current Directions in Psychological Science*, *22*(4), 324–329. doi: 10.1177/0963721413480171

Fiedler, K., Russer, S., & Gramm, K. (1993). Illusory correlations and memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 111–136.

Fiedler, K., & Wänke, M. (2009). The cognitive-ecological approach to rationality in social psychology. *Social Cognition*, *27*(5), 699–732.

Fleig, H., Meiser, T., Ettlin, F., & Rummel, J. (2017). Statistical numeracy as a moderator of (pseudo)contingency effects on decision behavior. *Acta Psychologica*, *174*, 68–79. doi: 10.1016/j.actpsy.2017.01.002

Gelman, A., & Carlin, J. (2014). Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on Psychological Science*, *9*(6), 641–651. doi: 10.1177/1745691614551642

Gigerenzer, G. (2019). Axiomatic rationality and ecological rationality. *Synthese*. doi: 10.1007/s11229-019-02296-5

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*(1), 75–90. doi: 10.1037/0033-295x.109.1.75

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive psychology*, *51*(4), 334–384.

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology*, *12*(4), 392–407. doi: 10.1016/S0022-1031(76)80006-6

Hamilton, D. L., & Rose, T. L. (1980). Illusory correlation and the maintenance of stereotypic beliefs. *Journal of Personality and Social Psychology*, *39*(5), 832–845. doi: 10.1037/0022-3514.39.5.832

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

Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description-experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making*, *21*(5), 493–518. doi: 10.1002/bdm.598

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*(8), 534–539. doi: 10.1111/j.0956-7976.2004.00715.x

Hilbert, M. (2012). Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making. *Psychological Bulletin*, *138*(2), 211–237. doi: 10.1037/a0025940

Hilbig, B. E., & Glöckner, A. (2011). Yes, they can! Appropriate weighting of small probabilities as a function of information acquisition. *Acta Psychologica*, *138*(3), 390–396. doi: 10.1016/j.actpsy.2011.09.005

Jeffreys, H. (1939). *Theory of Probability* (1st ed.). Oxford: The Clarendon Press.

Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015). Logic, probability, and human reasoning. *Trends in Cognitive Sciences*, *19*(4), 201–214. doi: 10.1016/j.tics.2015.02.006

Kareev, Y. (1995). Positive bias in the perception of covariation. *Psychological Review*, *102*(3), 490–502. doi: 10.1037/0033-295X.102.3.490

Kellen, D. (2019). A model hierarchy for psychological science. *Computational Brain &*

Behavior, *2*(3), 160–165. doi: 10.1007/s42113-019-00037-y

Kellen, D., & Klauer, K. C. (2019). Theories of the wason selection task: A critical assessment of boundaries and benchmarks. *Computational Brain & Behavior*, 1–13.

Kellen, D., Klauer, K. C., & Bröder, A. (2013). Recognition memory models and binary-response ROCs: A comparison by minimum description length. *Psychonomic Bulletin and Review*, *20*, 693–719. doi: 10.3758/s13423-013-0407-2

Kellen, D., Pachur, T., & Hertwig, R. (2016). How (in) variant are subjective representations of described and experienced risk and rewards? *Cognition*, *157*, 126–138.

Kerman, J. (2011). *A closed-form approximation for the median of the beta distribution.* Retrieved from `https://arxiv.org/abs/1111.0433v1`

Keynes, J. M. (1921). *A treatise on probability.* Dover Publications.

Klauer, K. C. (2015). Mathematical modeling. In B. Gawronski & G. V. Bodenhausen (Eds.), *Theory and explanation in social psychology* (pp. 371–389). New York, NY, US: Guilford Press.

Klauer, K. C., & Meiser, T. (2000). A source-monitoring analysis of illusory correlations. *Personality and Social Psychology Bulletin*, *26*(9), 1074–1093. doi: 10.1177/01461672002611005

Knill, D. C., & Richards, W. (1996). *Perception as bayesian inference.* Cambridge, Mass: MIT Press.

Kutzner, F., Vogel, T., Freytag, P., & Fiedler, K. (2011). Contingency inferences driven by base rates: Valid by sampling. *Judgment and Decision Making*, *6*(3), 211–221.

Laplace, P. S. (1820/1951). *Philosophical Essays on Probabilities.* New York: Dover.

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.

Madey, S. F., & Chasteen, A. L. (2004). Age-related health stereotypes and illusory correlation. *International Journal of Aging and Human Development*, *58*(2), 109–126. doi: 10.2190/81XB-QPYN-9ADU-5L11

Marr, D. (1982). *Vision.* San Francisco, CA: Freeman.

McConnell, A. R., Sherman, S. J., & Hamilton, D. L. (1994). Illusory correlation in the perception of groups: An extension of the distinctiveness-based account. *Journal of Personality and Social Psychology*, *67*(3), 414–429. doi: 10.1037/0022-3514.67.3.414

Meiser, T. (2003). Effects of processing strategy on episodic memory and contingency learning in group stereotype formation. *Social Cognition*, *21*(2), 121–156. doi: 10.1521/soco.21.2.121.21318

Meiser, T. (2006). Contingency learning and biased group impressions. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition.* (pp. 183–209). Cambridge University Press.

Meiser, T., & Hewstone, M. (2001). Crossed categorization effects on the formation of illusory correlations. *European Journal of Social Psychology*, *31*(4), 443–466. doi: 10.1002/ejsp.55

Meiser, T., & Hewstone, M. (2004). Cognitive processes in stereotype formation: The role of correct contingency learning for biased group judgments. *Journal of Personality and Social Psychology*, *87*(5), 599–614. doi: 10.1037/0022-3514.87.5.599

Meiser, T., & Hewstone, M. (2006). Illusory and spurious correlations: Distinct phenomena or joint outcomes of exemplar-based category learning? *European Journal of Social Psychology*, *36*(3), 315–336. doi: 10.1002/ejsp.304

Meiser, T., Rummel, J., & Fleig, H. (2018). Pseudocontingencies and choice behavior in probabilistic environments with context-dependent outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(1), 50–67. doi: 10.1037/xlm0000432

Mutter, S. A. (2000). Illusory correlation and group impression formation in young and older adults. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *55*(4), 224–237. doi: 10.1093/geronb/55.4.P224

Myung, J. I., & Pitt, M. A. (2018). Model comparison in psychology. In E.-J. Wagenmakers (Ed.), *Stevens' handbook of experimental psychology and cognitive neuroscience.*

Volume 5: Methodology (4th ed., p. 85-118). Hoboken, New Jersey: Wiley. doi: 10.1002/9781119170174.epcn503

Navarro, D. J. (2019). Between the devil and the deep blue sea: Tensions between scientific judgement and statistical model selection. *Computational Brain & Behavior*, *2*(1), 28–34.

Nilsson, H., Winman, A., Juslin, P., & Hansson, G. (2009). Linda is not a bearded lady: Configural weighting and adding as the cause of extension errors. *Journal of Experimental Psychology: General*, *138*(4), 517–534. doi: 10.1037/a0017351

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*(4), 608-631.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning* (1st ed.). Oxford [u.a.]: Oxford University Press.

Primi, C., & Agnoli, F. (2002). Children correlate infrequent behaviors with minority groups: A case of illusory correlation. *Cognitive Development*, *17*(1), 1105–1131. doi: 10.1016/S0885-2014(02)00076-X

Regenwetter, M., & Robinson, M. M. (2017). The construct–behavior gap in behavioral decision research: A challenge beyond replicability. *Psychological Review*, *124*(5), 533-550. doi: 10.1037/rev0000067

Rodríguez-Ferreiro, J., & Barberia, I. (2017). The moral foundations of illusory correlation. *PLoS ONE*, *12*(10), 1–10. doi: 10.1371/journal.pone.0185758

Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in cognitive sciences*, *20*(12), 883–893.

Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.

Sherman, J. W., Kruschke, J. K., Sherman, S. J., Percy, E. J., Petrocelli, J. V., & Conrey, F. R. (2009). Attentional processes in stereotype formation: A common model for category accentuation and illusory correlation. *Journal of Personality and Social Psychology*, *96*(2), 305–323. doi: 10.1037/a0013778

Shiffrin, R. M., & Nobel, P. A. (1997). The art of model development and testing. *Behavior Research Methods, Instruments, & Computers*, *29*(1), 6–14.

Simon, H. A. (1990). Bounded rationality. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *Utility and probability* (pp. 15–18). London: Palgrave Macmillan UK. Retrieved from https://doi.org/10.1007/978-1-349-20568-4_5 doi: 10.1007/978-1-349-20568-4_5

Skovgaard-Olsen, N., Kellen, D., Hahn, U., & Klauer, K. C. (2019). Norm conflicts and conditionals. *Psychological Review*, *126*(5), 611-633. doi: 10.1037/rev0000150

Spears, R., van der Pligt, J., & Eiser, J. R. (1985). Illusory correlation in the perception of group attitudes. *Journal of Personality and Social Psychology*, *48*(4), 863–875. doi: 10.1037/0022-3514.48.4.863

Spohn, W. (2012). *The laws of beliefs.* Oxford, UK: Oxford University Press.

Tauber, S., Navarro, D. J., Perfors, A., & Steyvers, M. (2017). Bayesian models of cognition revisited: Setting optimality aside and letting data drive psychological theory. *Psychological review*, *124*(4), 410-441.

Teller, P. (1973). Conditionalization and observation. *Synthese*, *26*(2), 218–258.

Todd, P. M., Hills, T. T., & Robbins, T. W. (2012). *Cognitive search: Evolution, algorithms, and the brain.* Boston, Mass.: MIT press.

Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *353*(1373), 1295–1306.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. doi: 10.1126/science.185.4157.1124

Vadillo, M. A., Blanco, F., Yarritu, I., & Matute, H. (2016). Single-and dual-process models of biased contingency detection. *Experimental Psychology*, *63*(1), 3–19. doi: 10.1027/1618-3169/a000309

Van Rooy, D., Vanhoomissen, T., & Van Overwalle, F. (2013). Illusory correlation, group size and memory. *Journal of Experimental Social Psychology*, *49*(6), 1159–1167. doi: 10.1016/j.jesp.2013.05.006

Van Rooij, I. (2008). The tractable cognition thesis. *Cognitive science*, *32*(6), 939–984.

Viscusi, W. K. (1989). Prospective reference theory: Toward an explanation of the paradoxes. *Journal of Risk and Uncertainty*, *2*, 235–264. doi: 10.1007/BF00209389

Vogel, T., Kutzner, F., Fiedler, K., & Freytag, P. (2013). How majority members become associated with rare attributes: Ecological correlations in stereotype formation. *Social Cognition*, *31*(4), 427–442. doi: 10.1521/soco_2012_1002

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

Wakker, P. P. (2002). Decision-principles to justify Carnap's updating method and to suggest corrections of probability judgments. *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, 544–551.

Wasserman, E. A., Dorner, W. W., & Kao, S. F. (1990). Contributions of specific cell information to judgments of interevent contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 509–521. doi: 10.1037/0278-7393.16.3.509

Wasserman, L. (2004). *All of statistics.* New York: Springer.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*(1), 48-64.

Yechiam, E., & Hochman, G. (2013). Losses as modulators of attention: Review and analysis of the unique effects of losses over gains. *Psychological Bulletin*, *139*, 497-518.

Pseudocontingency Inference and Choice: The Role of Information Sampling

Franziska M. Bott & Thorsten Meiser

University of Mannheim

The final article is available at https://doi.org/10.1037/xlm0000840

Author Note

This research was funded by the Deutsche Forschungsgemeinschaft (DFG) grant 2277, Research Training Group "Statistical Modeling in Psychology" (SMiP).

Parts of this research were presented at the 61st Conference of Experimental Psychologists (TeaP) in London, UK. Data and analysis scripts are provided on OSF (`https://osf.io/rx6qa/`).

Correspondence to: Franziska Bott, Department of Psychology, University of Mannheim, L13 15, D-68161 Mannheim, Germany. E-mail: f.bott@uni-mannheim.de

Abstract

Pseudocontingencies are inferences of correlations between variables, like two options and two outcomes, drawn on the basis of their skewed base rates covarying across a third variable (e.g., two contexts). Here, we investigated the effect of pseudocontingency inference on choice behavior. When choices between two options are not based on the actual contingency between options and outcomes, but instead on a pseudocontingency, the latter may override the existing contingency resulting in potentially suboptimal choice behavior. Whereas research has mainly focused on investigating the pseudocontingency effect by presentation of predetermined learning trials, we examined the role of free information sampling for the pseudocontingency effect as compared with predetermined learning. Experiment 1 replicated previous findings of a pseudocontingency effect in choice behavior. In Experiment 2, we compared predetermined information and free information sampling in a bivariate decision scenario with only two options and two outcomes. Experiments 3 and 4 aimed at investigating the inference of a pseudocontingency when sampling information by context or by context and option in the trivariate scenario. The results revealed an asymmetry between positive contexts with predominantly gains and negative contexts with predominantly losses. Within a negative context we found no differences between options, neither during information sampling nor for subsequent choices. Within the positive context, when information sampling was self-determined, participants sampled skewed base rates of options and preferred the predominant option. The findings underline the influence of self-determined information sampling on the pseudocontingency effect on choice behavior.

Keywords: pseudocontingency, information sampling, choice, contingency inference, decision making

Pseudocontingency Inference and Choice: The Role of Information Sampling

Most decisions are based on some kind of evaluation of the choices available. In many laboratory studies, decision making is investigated via choice between (two) options typically resulting in gains and/or losses with different probabilities. In research on risky choice, for example, paradigms often comprise two options differing in their outcomes' magnitudes and probabilities of occurrence: The task is to choose between one option, very likely or for sure resulting in a medium outcome, and another option yielding a rare extreme outcome (Ashby & Rakow, 2016; Barron & Erev, 2003; Wulff, Hills, & Hertwig, 2015). The options' winning probabilities may be stated or experienced over the course of several experimental trials. Either way, in those tasks, options' evaluations and thereby choices are ideally derived from the underlying winning probability of each option. In line with this reasoning, provided that the options' expected values do differ, the option with the higher expected value is chosen in the majority of cases (e.g., Hilbig & Glöckner, 2011). At least, the option that has the higher expected value based on an observed sample is predominantly chosen, regardless whether the sample properly maps the options' actual winning probabilities (Fiedler, 2000; Hertwig, Barron, Weber, & Erev, 2006). In other words, the stronger an option is associated with positive outcomes (e.g., gains), the more likely it is chosen. Correspondingly, evaluations can also be considered as inferred contingencies between options and their outcomes.

In the following series of experiments, we were interested in which information is considered for contingency learning and subsequent choice. We investigated information sampling behavior and so called pseudocontingency inference across various experimental conditions.

Contingency Between Binary Variables

In order to correctly determine a contingency between two binary variables mathematically, information on event co-occurrences is needed. In terms of a two-by-two contingency table, all four cell frequencies have to be taken into account to express a contingency. Take, for instance the 2 x 2 contingency table labeled *Context C1* in Table 1. It contains 34 events of joint occurrences of the two binary variables option X and outcome Y, taking values $X1$ and $X2$ (option) or $Y1$ (gain) and $Y2$ (loss; outcome). One way to measure the contingency between options and outcomes is Δp which compares two conditional (winning) probabilities (Allan, 1980):

$$\Delta p = P(Y = \text{``}gain\text{''}|X = X1) - P(Y = \text{``}gain\text{''}|X = X2) \tag{1}$$

In the example, $P(\text{``}gain\text{''}|X1) = \frac{15}{15+9} = .625$ and $P(\text{``}gain\text{''}|X2) = \frac{8}{8+2} = .80$. Therefore, $\Delta p = -.175$ indicating an association between options and outcomes with a lower probability of option X1 to result in a gain as compared with option X2.

The ability to assess contingencies has been widely recognized as essential for everyday life. Whether, for instance, in conditioning, between conditioned and unconditioned stimuli (e.g., Rescorla & Wagner, 1972), or in causal reasoning, between causes and effects (Cheng & Novick, 1992; Spellman & Mandel, 1999; Waldmann & Holyoak, 1992), researchers are interested in the cognitive process of assessing the contingency. Not surprisingly, various measures or rules have been proposed on how individuals integrate information into a contingency judgment all of which suggest that (estimated) cell frequencies or conditional probabilities constitute the input (Perales, Catena, Cándido, & Maldonado, 2017). Accordingly, for a long time, it was assumed that individuals infer contingencies in a similar way as the Δp-measure is computed, by taking joint frequencies (i.e., cell frequencies) into account. Yet, contingency inference does not merely reflect co-occurrences or covariation knowledge, but instead subjective causal models may guide which covariation information is used and how (Waldmann, Hagmayer, & Blaisdell, 2006). However, under certain conditions, decision makers might also employ simplifying strategies.

Pseudocontingencies in Judgment and Choice

Empirical evidence suggests that cell frequencies are not necessarily used for contingency inference. As a consequence, options' evaluations may be based on information other than cell frequencies, like marginal frequencies or base rates of options and outcomes (e.g., Fiedler, Freytag, & Meiser, 2009). When a contingency between two options and two outcomes is indeed inferred based on their skewed base rates, the phenomenon is referred to as *pseudocontingency* (Fiedler & Freytag, 2004; Fiedler et al., 2009; Fiedler, Freytag, & Unkelbach, 2007; Kutzner, Freytag, Vogel, & Fiedler, 2008). Broadly speaking, provided that there is one frequent category per variable, frequent categories may be associated with each other as well as infrequent categories.

Pseudocontingency inferences are usually investigated with an experimental paradigm comprising not only two options and two outcomes but also two contexts. Typically, one option and one outcome are the frequent categories in one context, but the infrequent categories in the other context. In that case, the options' and outcomes' base rates are skewed within a context and covary across the contexts. By implication, the option variable and the outcome variable are each correlated with the context variable. This gives rise to the inference of a pseudocontingency between options and outcomes within each context (Fiedler & Freytag, 2004; Meiser & Hewstone, 2004). Basically, the frequent option is associated with the frequent outcome as well as the infrequent option with the infrequent outcome within each context. Concisely, pseudocontingencies between two variables are inferred on the basis of their skewed base rates or their pairwise contingencies with a third variable (Fiedler et al., 2009, 2007; Fleig, Meiser, Ettlin, & Rummel, 2017; Meiser, Rummel, & Fleig, 2018).

The prefix *pseudo* does not mean that the inferred contingency is necessarily wrong, instead, it merely indicates that the inference is based on aggregate information, like base rates, rather than pairs of observations. Actually, pseudocontingency inferences are quite adaptive and useful. For instance, when no information about the co-occurrence of two options and two outcomes is available, but instead only their base rates, inferences about their contingency can still be estimated based on the pseudocontingency heuristic. Additionally, skewed base rates of two variables restrict the range of possible contingencies between them (Duncan & Davis, 1953; Kline, 2011, p. 50; Meiser, 2006). Hence, pseudocontingencies can facilitate detecting the sign of true contingencies (Kutzner, Vogel, Freytag, & Fiedler, 2011a). This is corroborated by a Bayesian algorithm with unbiased priors that provides increased posterior probabilities for contingency coefficients that are in accordance with inferred pseudocontingencies when feeding in information on skewed bases rates only (Klauer, 2015). Therefore, pseudocontingencies prove to be fairly good estimates for genuine contingencies (Fiedler, Kutzner, & Vogel, 2013).

Nevertheless, in a contingency table, base rates do not determine cell frequencies, wherefore the exact same marginal distribution can result from various cell frequencies. Accordingly, an inferred pseudocontingency may differ and even be opposite in sign relative to the true contingency. Such scenarios, in which pseudocontingency and genuine contingency contradict, are usually created for experimental studies investigating pseudocontingency inferences. Several experiments using various scenarios have demonstrated that pseudocontingencies are inferred from skewed or covarying base rates and may indeed override existing contingencies (e.g., Fiedler, 2010; Fiedler & Freytag, 2004; Fiedler et al., 2007; Meiser & Hewstone, 2004; Vogel, Freytag, Kutzner, & Fiedler, 2013). More recently, it has been shown that the inference of a pseudocontingency may lead to suboptimal choice behavior even in situations of direct relevance for the respondent and free choice (Fleig et al., 2017; Kutzner, Vogel, Freytag, & Fiedler, 2011b; Meiser et al., 2018). These results suggest that decision makers perceive pseudocontingencies as solid basis for their choice.

Those results, however, do not suggest that individuals cannot engage in genuine contingency assessment. Cheng and Novick (1990, 1992), for instance, already argued that covariations between multiple variables, like outcomes, options, and contexts, can be assessed when considering the proportions of outcomes given all combinations of options and contexts (i.e., joint frequencies). Indeed, individuals are capable of assessing genuine conditional probabilities or contingencies and may also consider a relevant third (context) variable when being aware of its moderating role and if cognitively manageable (Spellman, Price, & Logan, 2001; Waldmann & Hagmayer, 2001). Likewise, in studies on pseudocontingencies discussed, participants are able to correctly infer contingencies between outcomes and contexts and also conditionalize their assessment of the contingency between options and outcomes on the context variable (e.g., Fleig et al., 2017; Meiser et al., 2018). Still, under certain conditions, individuals fall back to using a more heuristic account and infer pseudocontingencies, in the example between options and outcomes within each context, even if the cell frequencies could in theory be used.

Context-Dependent Inference of Pseudocontingencies

In some experiments, an asymmetry in the strength of the inferred pseudocontingency was observed between predominantly positive and predominantly negative domains (Fleig et al., 2017; Meiser & Hewstone, 2004; Meiser et al., 2018). The experiments by Fleig et al. (2017) and Meiser et al. (2018) too comprised two options resulting in either a positive outcome or a negative outcome within two contexts. Option X1 and gains were predominant within one context, with their base rates covarying across contexts, so that option X1 and gains were the infrequent categories within the second context. This suggested the inference of a pseudocontingency between options and outcomes which, however, contradicted the genuine contingency within each context.

In the context of predominantly positive outcomes, Fleig et al. (2017) and Meiser et al. (2018) could consistently observe the inference and use of the pseudocontingency, resulting in participants preferring the option with the lower underlying winning probability. Yet, in the second context of predominantly negative outcomes, the pseudocontingency effect was weaker or even absent, so that participants preferred neither option. This asymmetry might be explained by participants considering the predominantly negative context as less important for subsequent choices as they would avoid choosing within that context at all. Whenever participants learned that negative outcomes result predominantly within a context, further regularities within that context, like the options' base rates, would hardly be learned. The genuine contingency between options and outcomes would be detected even less likely. On the contrary, the asymmetry between predominantly positive contexts and predominantly negative contexts might also arise from allocating more attention to the task at hand when negative outcomes are possible or likely (Yechiam & Hochman, 2013, 2014). This might increase performance with the result that the genuine contingency between options and outcomes is learned within the negative context and competes with the inferred pseudocontingency. Anyway, the asymmetry in the strength of the inferred pseudocontingency may indicate that attention is allocated to available information differently in a predominantly negative context and that contingencies are learned differently in comparison with a mostly positive context.

Information Sampling

One possibility to examine which information is actually considered for contingency learning is to leave information sampling to the participants themselves. So far, pseudocontingency inferences have primarily been investigated by presenting predetermined information to participants during a learning phase. Subsequently, in a decision phase, participants had to evaluate and/or choose between the options that were presented during the learning phase. In contrast, in typical experiments on information sampling, different types of experimental paradigms have been employed (Wulff, Mergenthaler-Canseco, & Hertwig, 2018). In the *sampling paradigm*, usually there are two options representing two different payoff structures. To investigate self-directed information sampling participants are free to sample any number of single events of any option in any order they desire. No draw is incentivized except one final choice to be made after information sampling (e.g., Hertwig, Barron, Weber, & Erev, 2004). A learning phase, in which participants have the opportunity to familiarize themselves with the options' winning probabilities is thus separated from a consequential decision phase. Hence, the basic structure of the information sampling task is similar to experimental tasks investigating the inference of (pseudo-)contingencies in which participants observe sequences of learning trials before choosing between options.

However, research on free information sampling suggests that typically small samples are drawn when learning trials are self-determined. Hertwig et al. (2004) reported a median total sample size of $Md = 15$ for sampling two options. Hau, Pleskac, Kiefer, and Hertwig (2008) found even smaller total sample sizes of $Md = 11$. This in itself may already be sufficient for a sample of options to become a less veridical representation of underlying winning probabilities (e.g., Hau et al., 2008; Hilbig & Glöckner, 2011).

Although choices during free information sampling do not have direct consequences, information sampling may still depend on various factors, like the valence of potential events or outcomes. On the one hand, it has been shown that larger total samples of all options are drawn the higher the outcomes' magnitudes (Hau et al., 2008), the more likely losses are (Lejarraga, Hertwig, & Gonzalez, 2012; Wulff et al., 2018), and the more similar the options' expected values are (Wulff et al., 2018). On the other hand, research is inconclusive whether individual options are also sampled to varying degrees. Although it has been found that two options are sampled equally often independent of how risky they are (Hau et al., 2008; Hertwig et al., 2004), Lejarraga et al. (2012) reported higher samples of risky options as compared with safe options if the variance of a risky option was encountered. Finally, Fiedler, Wöllert, Tauber, and Hess (2013) found that already during learning, larger samples are drawn from positively evaluated options. Likewise, Wulff et al. (2018) showed that over the course of trials participants were more likely to sample the one option which they ultimately selected for their final, consequential choice.

Consequently, even for a rather simple scenario with only two options, results are inconclusive. It is therefore unclear whether both options would be sampled to varying degrees and whether thereby conditions for pseudocontingency inferences, in terms of skewed base rates, would be created by self-determined, free information sampling. By implication, it is an open question whether a freely drawn sample gives rise to the inference of pseudocontingencies and whether they will contradict genuine contingencies. Most of aforementioned research on information sampling uses two options that vary in their outcomes' magnitudes and probabilities of occurrence, but have similar expected payoffs (nondominated choice problems). Studies on free information sampling reported that if the two options do differ in their expected value (dominated choice problem) the option with the higher expected value is preferred after learning provided that it also has the higher expected value in the drawn sample (e.g., Hertwig et al., 2006; Hilbig & Glöckner, 2011). Contrarily, research on pseudocontingency inference suggests that this option may only be preferred if it was the most frequently observed option when gains are more likely than losses or if it was the least frequently observed option when losses are most common (cf. e.g., Fiedler et al., 2009). Fiedler et al. (2007), for example, restricted the options to sample from. As a result, participants sampled skewed base rates and inferred a pseudocontingency. Still, for a pseudocontingency to arise, skewed base rates do not only have to be sampled, but they also have to be used for contingency inference. Alternatively, sampling options freely, without any restrictions, may draw the focus of attention away from base rates toward joint occurrences. In that case, free information sampling might foster utilizing cell frequencies and genuine contingency assessment, similarly to an experiment by Fleig et al. (2017) where participants had to make a tally recording cell frequencies. Moreover, to the extent to which information is sampled based on evaluations, these might already be the result of inferred (pseudo-) contingencies based on the current sample. For those reasons, a systematic investigation of the interplay between information sampling and contingency inference is needed.

Overview of Present Research

For this purpose, the present research investigated information sampling behavior as well as its impact on the inference of pseudocontingencies in a decision scenario that induces pseudocontingency inferences. All experiments were implemented in OpenSesame (Mathôt, Schreij, & Theeuwes, 2012) using the mousetrap plug-in to record mouse clicks (Kieslich & Henninger, 2017). In Experiment 1, we implemented a new experimental paradigm to replicate previous findings of pseudocontingency inference in self-relevant choice (Fleig et al., 2017; Meiser et al., 2018). We tested whether a pseudocontingency is inferred and used for actual and voluntary choices, thereby setting the stage for investigating effects of free information sampling on pseudocontingency inference in Experiments 2 to 4. More specifically, in Experiment 2, differences between predetermined information and self-determined information sampling were investigated in a bivariate scenario comprising two options and two outcomes. Besides choices during a learning phase, we analyzed choice behavior after learning, also as a function of the individual sample drawn during learning. A trivariate decision scenario with the addition of two contexts was used in Experiments 3 and 4. Experiment 3 aimed at the asymmetry between predominantly positive contexts and predominantly negative contexts reflected by stronger pseudocontingency inferences in predominantly positive contexts (Fleig et al., 2017; Meiser & Hewstone, 2004; Meiser et al., 2018). We realized a sampling by context condition, in which participants were allowed to choose which context to observe in each learning trial, in order to test whether predominantly positive contexts are more or less frequently observed than predominantly negative contexts. Due to such a potential bias in attentional focus a (pseudo-)contingency might be learned to varying degrees. Experiment 4, finally, extends Experiments 2 and 3. By implementing a sampling by context and option condition we investigated whether a pseudocontingency is even inferred and used when free information sampling is completely self-determined.

Experiment 1

The first experiment aimed at replicating the pseudocontingency effect on choice behavior in relevant decisions. The experimental task was to evaluate two shares (options X1 and X2). They were traded at two times of a day (contexts C1 and C2) and yielded either a gain or a loss (outcomes Y1 and Y2) with different probabilities varying over the time of day. Similar to earlier studies on pseudocontingencies, the underlying base rates of options and outcomes were skewed within each context and covaried across contexts resulting in pairwise contingencies with the context variable. After a series of learning trials, participants traded the shares themselves while playing for points. If a pseudocontingency between options and outcomes is inferred and used, participants should prefer the predominant option within the predominantly positive context and the infrequent option within the predominantly negative context.

Method

Participants. Forty-three participants were recruited at a German University to take part in Experiment 1 (37 female, $M_{age} = 25.98, SD_{age} = 10.33$). In return, they received a compensation of €2.

Material and Procedure. Computerized instructions informed the participants that the experiment was about assessment and trading of shares and that it comprised a learning phase as well as a subsequent decision phase. In the learning phase, participants observed a total of 68 trials. In each trial, one context (i.e., either labeled as *morning* or labeled as *evening*), one option (i.e., share *hat* or share *pig*), and the respective outcome (i.e., gain or loss of 10 points) were presented. The two contexts and two options were displayed at different positions on the screen and, additionally, were distinguishable by their appearance (see Figure 1).

Table 1 shows the frequencies of occurrence for all single events during the learning phase. The respective winning probabilities were assigned to contexts and options counterbalanced across participants. Over the course of the 68 learning trials, within one context (C1) gains were more frequent than losses (23 × gains, 11 × losses) and vice versa within the second context (C2; 11 × gains, 23 × losses). Thus, the contingency between contexts and outcomes was $\Delta p_{YC} = \frac{23}{23+11} - \frac{11}{11+23} = .353$. Addi-

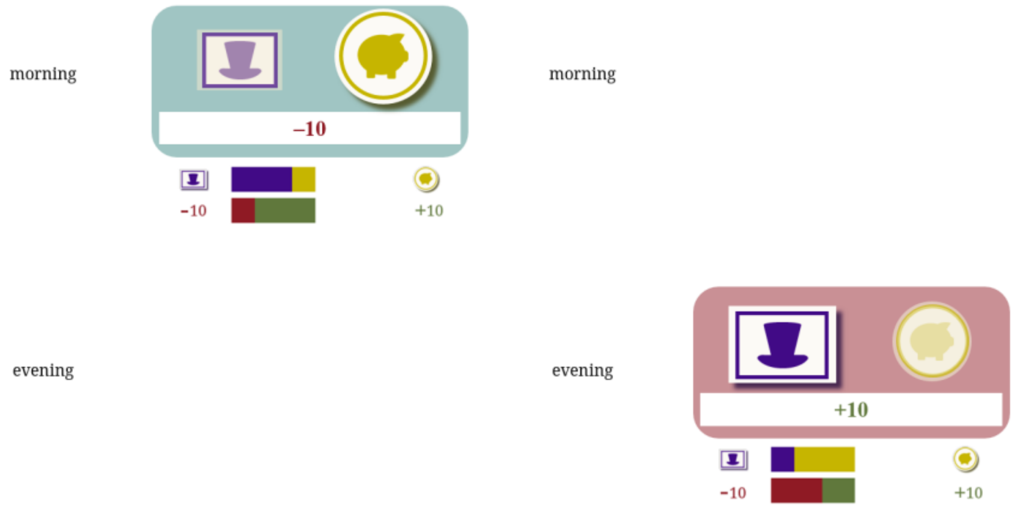

Figure 1. Two sample displays during the learning phase of Experiments 1, 3, and 4. The left example shows an event of the right share "pig" in the context "morning" that results in a loss (-10). The right example displays an event of the left share "hat" in the context "evening" that results in a gain (+10). The bar charts visualize context-dependent base rates of options (upper bar chart per context) and outcomes (lower bar chart per context) observed on previous trials within the respective context. The display during the decision phase was similar, with the exception of neither showing context-dependent base rates nor displaying the outcomes of each individual trial.

tionally, one option (X1) was more frequent than the other option (X2) within context C1 (24 × X1, 10 × X2) and vice versa within context C2 (10 × X1, 24 × X2). The contingency between contexts and options was $\Delta p_{XC} = \frac{24}{24+10} - \frac{10}{10+24} = .412$. Aggregated across contexts, a zero correlation between options and outcomes resulted ($\Delta p_{YX} = \frac{15+2}{15+2+9+8} - \frac{8+9}{8+9+2+15} = .00$). However, within each context, the contingency between options and outcomes was $\Delta p_{YX|C1} = \frac{15}{15+9} - \frac{8}{8+2} = -.175$ and $\Delta p_{YX|C2} = \frac{2}{2+8} - \frac{9}{9+15} = -.175$, respectively. Therefore, conditional on context, option X2 had a higher winning probability than option X1. However, the options' and outcomes' base rates were skewed within each context and covaried across contexts suggesting instead the association of option X1 with gains and the association of option X2 with losses within each context according to a pseudocontingency.

The 68 learning trials summarized in Table 1 were presented in random order over the course of the learning phase. In an individual learning trial, first, the context (morning vs. evening) was displayed. After 650 ms an option (pig or hat) was highlighted by increasing its size and decreasing its transparency (cf. Figure 1). After another 650 ms the respective outcome (gain vs. loss) was additionally presented. Context, option, and

Table 1

2 x 2 Contingency Tables Representing Probability Distributions and Contingencies Between Contexts, Options, and Outcomes Underlying All Experiments.

	Context C1			Context C2		
	Option X1	Option X2		Option X1	Option X2	
Outcome Y1	15 (.44)	8 (.24)	23 (.68)	2 (.06)	9 (.26)	11 (.32)
Outcome Y2	9 (.26)	2 (.06)	11 (.32)	8 (.24)	15 (.44)	23 (.68)
	24 (.71)	10 (.29)		10 (.29)	24 (.71)	

Note. Absolute frequencies are shown, as well as context-specific probabilities in parentheses. $\Delta p_{YX|C1} = -.175$; $\Delta p_{YX|C2} = -.175$; $\Delta p_{XC} = .412$; $\Delta p_{YC} = .353$; $\Delta p_{YX} = .00$

outcome were displayed together for another 650 ms. After an intertrial interval of 500 ms the next learning trial started. In addition to the individual events of contexts, options, and outcomes, context-dependent base rates of options and outcomes within each context were visualized and updated in each learning trial. Below each context, one pair of bars depicted the current frequencies of option X1 events versus option X2 events within the respective context. A second pair of bars represented the total number of gains and losses observed within the context so far (cf. Figure 1). Visualizing context-dependent base rates might enhance the salience of skewed base rates varying across contexts and thereby the predicted pseudocontingency effect (Meiser et al., 2018).

Following the learning phase, participants had to repeatedly choose between the options. The first half of 24 choice trials were forced choice trials, the second half were free choice trials. In each of the 12 forced choice trials, the experimental program determined the context which then was displayed including the options. Then, participants had to choose between the two options while aiming at maximizing points. After participants selected one of the options via mouse click, the next decision trial started. Each option could either result in a gain of 10 points or a loss of 10 points. The options' winning probabilities within each context were identical to the learning phase (cf. Table 1), but participants did not receive feedback on the obtained outcome or their performance until the very end of the experiment. In the 12 subsequent free choice trials, participants had the additional opportunity to skip a trial by clicking on a field labeled *skip* instead of an option after the context was set by the experimental program. In the forced choice trials as well as in the free choice trials, both contexts were presented six times each, in randomized order.

At the end of the experiment, participants were asked to estimate the contexts' and options' winning probabilities. They should indicate the winning probability of each option within each context as well as the overall winning probabilities of the contexts and the options across contexts. For this purpose, first, each context-option combination was presented in random order. Afterwards, each option and each context were presented separately and again in random order. Participants were asked to state the winning probability on a scale from 0 to 100.

Based on previous research on pseudocontingencies, we expected participants to infer the genuine contingency between contexts and outcomes. Therefore, they should skip more free choice trials given context C2 of predominantly losses as compared with context C1 of predominantly gains and give higher winning probability estimates for context C1 than for context C2. Moreover, we expected participants to also take the context into account when inferring the contingency between options and outcomes. If a pseudocontingency was inferred on the basis of the options' and outcomes' base rates within each context, participants should prefer option X1 to option X2, even though this pseudocontingency contradicted the genuine contingency in each context (i.e., $P(\text{“}gain\text{”}|X1) < P(\text{“}gain\text{”}|X2)$; $\Delta p_{YX|C1} = \Delta p_{YX|C2} = -.175$). As Wulff et al. (2015) showed that repeatedly choosing between options may alter choice behavior in order to maximize gains in the long run as compared with one single choice, we expected option X1 to be chosen more frequently than option X2 as the participants' task was to maximize points while no feedback on the outcome of each individual trial was given. Equivalently to choice, participants should estimate the winning probability of option X1 to be higher than the winning probability of option X2, even though option X2 had a higher true winning probability.

Results

Choice Behavior. First, to analyze the relative frequencies of decisions to play a trial within a given context in free choice trials, we conducted a paired t-test predicting the relative frequencies by context. In accordance with the contexts' underlying winning probabilities, participants made more choices in the predominantly positive context C1 as compared with the predominantly negative context C2, $t(42) = 5.49$, $p < .001$, $d = 0.84$. On average, participants played in $M = 5.35$ ($SD = 1.02$) out of six free choice trials in the predominantly positive context, but in only $M = 3.28$ ($SD = 2.12$) trials in the predominantly negative context. Six participants skipped all six free choice trials in the predominantly negative context C2.

For all remaining free choice trials as well as for all forced choice trials, according to our hypotheses, we expected a pseudocontingency effect on choice behavior resulting in a preference for option X1 within both contexts. Therefore, the primary dependent variable was dichotomous (choosing option X1 coded as 1, choosing option X2 coded as 0). Furthermore, owing to the additional opportunity to skip a free choice trial, we faced different numbers of observations per participant for free choices between options. On these grounds, we used generalized linear mixed models to analyze the probability to choose option X1 over option X2 in individual trials, as they use the available data through a full-information approach and are appropriate for modeling binary dependent variables (e.g., McCulloch, Searle, & Neuhaus, 2008; Snijders & Bosker, 2012).[1]

In Experiment 1, we analyzed the probability that participant i chooses option X1 in trial j as a function of the trial-level predictors context and trial type. The predictor context contrast-coded the decision context with the predominantly positive context C1 coded as 1 and the predominantly negative context C2 coded as -1. For the contrast-coded predictor trial type, forced choice trials were coded as 1 and free choice trials were coded as -1. Additionally, we accounted for between-participants differences in the propensity to choose option X1 through a random intercept. Differences in the effects of context and trial type were analyzed by specifying random slopes. The probability to choose option X1 and the resulting linear prediction model, with γ denoting fixed effects and u denoting random effects, were related with a logit link function:[2]

$$\begin{aligned}\ln\left(\frac{p(Y_{ij} = \text{“}X1\text{”})}{1 - p(Y_{ij} = \text{“}X1\text{”})}\right) &= \gamma_{00} + (\gamma_{10} + u_{1j}) \cdot context + (\gamma_{20} + u_{2j}) \cdot trial \\ &\quad + (\gamma_{30} + u_{3j}) \cdot context \cdot trial + u_{0j}\end{aligned} \tag{2}$$

All analyses were conducted in R (R Core Team, 2018) using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) for estimating generalized linear mixed models.

[1] In all experiments, two ANOVAs predicting relative frequencies of choices in forced choice trials and in free choice trials separately provided results equivalent to those of the generalized linear mixed models.

[2] The random effects were uncorrelated. A model estimating correlations among the random effects failed to converge. Although its results should be interpreted with caution, they were equivalent to the reported model without estimating correlations among random effects.

Table 2 summarizes relative frequencies of choices aggregated across choice trials and participants. The descriptive statistics imply a preference for option X1, as opposed to option X2. In line, the generalized linear mixed model yielded a significant positive intercept, $\gamma_{00} = 0.68$, $SE = 0.21$, $p = .001$, indicating an above chance probability to choose option X1 in individual choice trials. Neither context nor trial type did have significant effects, $p \geq .439$. All parameter estimates of the generalized linear mixed model are reported in Table A1 in the Appendix.

Probability Estimations. With regard to the estimated winning probabilities of options within contexts, a within ANOVA yielded a significant main effect of context, $F(1, 42) = 16.75, p < .001$, $\eta^2 = 0.06$. In line with participants' choice behavior, the results indicated higher winning probability estimates within the predominantly positive context C1 (cf. Table 2). The main effect of option fell short of significance, $F(1, 42) = 3.27$, $p = .078$, $\eta^2 = 0.02$, but was in the direction of the pseudocontingency, preferring option X1. The interaction between context and option was nonsignificant, $F(1, 42) < 1$, $p = .655$. Analyzing the estimated unconditional winning probabilities of contexts, likewise, participants gave significantly higher winning probability estimates for the predominantly positive context C1 ($M = 63.88$, $SD = 13.49$) as compared with the predominantly negative context C2 ($M = 41.23$, $SD = 17.82$), $t(42) = 5.70$, $p < .001$, $d = 0.87$. Finally, a paired t-test comparing the estimated winning probabilities of options across contexts showed significantly higher estimates for option X1 ($M = 54.53, SD = 13.96$) as compared with option X2 ($M = 43.95, SD = 14.33$), $t(42) = 3.23$, $p = .002$, $d = 0.49$.

Discussion

Experiment 1 aimed at replicating the pseudocontingency effect on choice behavior. To this end, participants had to evaluate two options, each presented within two contexts. A pseudocontingency effect should be reflected in the preference of the frequent option within the predominantly positive context and a preference of the infrequent option within the predominantly negative context (i.e., option X1 within both contexts). Taken

Table 2

Relative Frequencies of Choices and Winning Probability Estimates of Options Within Contexts in Experiment 1

	Context C1				Context C2			
	Option X1		Option X2		Option X1		Option X2	
	M	SD	M	SD	M	SD	M	SD
Relative choice frequency	61.33	32.31	38.67	32.31	60.07	31.03	39.93	31.03
Estimated probabilities	54.95	20.36	46.95	22.31	42.60	21.32	37.84	19.42

Note. Data in percent.

together, the results of Experiment 1 demonstrate that participants indeed inferred and used the pseudocontingency between options and outcomes as basis for their judgments and choices. Even though they correctly identified the context with the higher underlying winning probability, participants did not learn the genuine contingency between options and outcomes within a context. Instead, they preferred the option with the actually lower winning probability within each context, independent of whether in forced choices or free choices.

Yet, literature on information sampling might render learning situations like, for instance, in Experiment 1 artificial. When learning trials are not predetermined, free information sampling might create learning situations in which decision makers do not (need to) fall back on a pseudocontingency to make a choice. Instead, they might be able to detect genuine contingencies between options and outcomes in order to inform their decisions. If pseudocontingencies were also inferred and used when engaging in self-determined information sampling, however, this could be a further indication that pseudocontingencies are a robust and even more relevant effect. Hence, building on Experiment 1, we investigated free information sampling behavior and its impact on pseudocontingency inference and choice behavior, in a bivariate scenario in Experiment 2 and in a trivariate scenario in Experiments 3 and 4.

Experiment 2

Research on free information sampling suggests that the total sample size drawn might be influenced by certain factors such as the options' expected values (e.g., Wulff et al., 2018). Yet, research is inconclusive whether there is a bias in sample sizes per option or whether a drawn sample comprises all options evenly. In addition, regarding choice in a subsequent decision phase, studies on free information sampling reported a preference of the option with a higher expected value (Hertwig et al., 2006; Hilbig & Glöckner, 2011). By contrast, in accordance with literature on pseudocontingency inference, a preference for the option with the higher underlying winning probability would only be expected if it was the predominant option within a decision scenario of mostly positive outcomes or if it was the infrequent option when mostly negative outcomes result and just as long as only the base rates are taken into account (e.g., Fiedler et al., 2009).

Consequently, Experiment 2 used the decision scenario introduced in Experiment 1 and investigated whether two options are sampled to varying degrees if information sampling is free and not predetermined. Most research on free information sampling compares two options, wherefore we reduced the decision scenario to only two options within one context for Experiment 2. The variable context (predominantly positive context C1 vs. predominantly negative context C2) was realized as a between-participants factor. Moreover, because literature on free information sampling is inconclusive whether options will be sampled to varying degrees and whether skewed base rates of options will result, we analyzed sampling behavior and pseudocontingency inference in a free information sampling condition as compared with a condition with predetermined learning trials. When participants are allowed to engage in information sampling, they may sample the option with the lower underlying winning probability more frequently, less frequently, or as often as the other option. Therefore, the skewness of base rates depends on the individual sampling strategy. Additionally, it is an open question whether subsequent choices will also be based on (skewed) base rates (i.e., a pseudocontingency) when learning trials are not predetermined. While preferences will depend on the individual sample in the sampling conditions, we expected participants in the no sampling conditions to prefer option X1, with a lower underlying winning probability than option X2, as in Experiment 1 and in line with a pseudocontingency inference.

Method

Participants and Design. To investigate differences between free information sampling and predetermined learning trials within a predominantly positive context as well as within a predominantly negative context, we implemented 2 x 2 between-participants conditions with information sampling procedure (no sampling vs. sampling by option) and context (predominantly positive C1 vs. predominantly negative C2) as independent variables. Additionally, option (X1 vs. X2) was realized as within-participant factor. For Experiment 2, a total of 128 participants were recruited at the University of Mannheim (104 female, $M_{age} = 21.14$, $SD_{age} = 4.43$). They participated in return for a compensation of either €1 or partial course credit and were randomly allocated to the no sampling-positive context condition ($n = 34$), the no sampling-negative context condition ($n = 32$), the sampling-positive context condition ($n = 29$), and the sampling-negative context condition ($n = 31$). Two participants had to be excluded, because they did not sample each option at least once during learning.

Material and Procedure. Materials and experimental procedure of Experiment 2 were similar to Experiment 1. However, because context was manipulated as between-participants factor, participants only observed a total of 34 trials during the learning phase. The two options were again presented next to each other, but for Experiment 2 at the center of the screen. The total number of trials was fixed for all experimental conditions in order to ensure the same total amount of information per participant as well as to prevent biases in samples only due to small sample sizes. For participants in the no sampling conditions, the experimental procedure was identical to Experiment 1: in each learning trial, one option and the respective outcome were presented. Participants in the no sampling-positive context condition only observed the learning trials corresponding to the absolute frequencies of context C1 depicted in Table 1. Likewise, participants in the no sampling-negative context condition observed the 34 events of context C2 in random order. In the two sampling by option conditions, participants engaged in free information sampling. There, in each learning trial, the two options were presented on the screen. Then, participants selected one of the two options. After 650 ms, the respective outcome was presented. They had to sample 34 events, but were free to sample from any option and in any order they desired without each draw contributing to their individual earnings. In the sampling by option-positive context condition, the outcomes were randomly drawn from the distribution of each option within context C1 in Table 1. In the sampling by option-negative context condition, the outcomes were drawn in accordance with the event distributions of context C2 in Table 1 in random order. As in Experiment 1, the respective winning probabilities were assigned to options counterbalanced across participants in all experimental conditions.

In the following decision phase, participants had to choose between the two options 20 times. As in Experiment 1, the first half of trials were forced choice trials whereas the second half were free choice trials. The options' winning probabilities were identical to the learning phase. That is, both options were more likely to result in a gain in the positive context conditions than in the negative context conditions and the underlying winning probability of option X2 was higher than the underlying winning probability of option X1. Again, in the decision phase, participants did not receive feedback on the obtained outcome of each choice. At the end of the experiment, participants had to estimate the options' winning probabilities on a scale from 0 to 100.

Results

Means of relative frequencies of choices during the sampling phase and the decision phase as well as winning probability estimates are displayed in Table 3. Sampling behavior and choice behavior were analyzed in generalized linear mixed models whereas participants' winning probability estimates were analyzed by means of a mixed ANOVA. Whereas investigating sampling behavior on an aggregate level would inform about overall preferences during sampling, the generalized linear mixed model analysis allows to predict information sampling behavior in individual trials as a function of individual previous observations.

Table 3

Relative Frequencies of Samples, Relative Frequencies of Choices, and Winning Probability Estimates of Options in Each Experimental Condition of Experiment 2

	Option X1		Option X2	
	M	SD	M	SD
Relative frequency of samples				
sampling, C1	46.55	17.47	53.45	17.47
sampling, C2	49.72	11.94	50.28	11.94
Relative frequency of choices				
no sampling, C1	56.73	22.27	43.27	22.27
no sampling, C2	47.66	19.36	52.34	19.36
sampling, C1	42.22	24.71	57.78	24.71
sampling, C2	46.26	25.99	53.74	25.99
Estimated winning probabilities				
no sampling, C1	59.88	15.31	53.62	19.05
no sampling, C2	42.50	19.09	45.09	18.17
sampling, C1	56.90	20.07	61.03	20.21
sampling, C2	36.94	15.38	49.55	19.82

Note. Data in percent. The experimental conditions in Experiment 2 result from the 2 x 2 between-subject factors sampling procedure (no sampling vs. sampling by options) and context (predominantly positive C1 vs. predominantly negative C2).

Information Sampling Behavior. In order to analyze the probability to sample option X1 in an individual learning trial, a generalized linear mixed model was specified that included a random intercept as well as fixed slopes for previous option (prevX) and previous outcome (prevY) at the trial-level. When a participant sampled option X1 in the previous trial the predictor previous option had the value 1. When option X2 was sampled in the preceding trial, it had the value -1. The predictor previous outcome was contrast-coded, too (gain vs. loss, coded as 1 and -1, respectively). The contrast-coded predictor context (predominantly positive context C1 coded as 1, predominantly negative context C2 coded as -1) was entered at the person-level. Additionally, the predictors' interactions were added:[3]

$$\begin{aligned}\ln\left(\frac{p(Y_{ij} = \text{“}X1\text{”})}{1 - p(Y_{ij} = \text{“}X1\text{”})}\right) &= \gamma_{00} + \gamma_{01} \cdot context + \gamma_{10} \cdot prevX + \gamma_{20} \cdot prevY \\ &\quad + \gamma_{30} \cdot context \cdot prevX + \gamma_{40} \cdot context \cdot prevY \\ &\quad + \gamma_{50} \cdot prevX \cdot prevY + \gamma_{60} \cdot context \cdot prevX \cdot prevY + u_{0j}\end{aligned} \tag{3}$$

Parameter estimates of the generalized linear mixed model are reported in Table A2 in the Appendix. The nonsignificant intercept reflected no overall preference for any option during sampling, $\gamma_{00} = -0.04$, $SE = 0.07$, $p = .559$. However, as the effect of previous option was significant and positive, $\gamma_{10} = 0.67$, $SE = 0.06$, $p < .001$, the probability to sample option X1 in a learning trial increased if it had been sampled in the preceding learning trial and decreased if option X2 had been sampled. Moreover, when option X1 resulted in a gain or when option X2 resulted in a loss in the previous trial, the probability to sample option X1 rose, $\gamma_{50} = 0.29$, $SE = 0.06$, $p < .001$. Put differently, the probability to resample an option was high, especially when it led to a positive outcome in the preceding trial. None of the other effects was significant, $p \geq .172$.

Choice Behavior. Analyzing the relative frequencies to skip a free choice trial, a 2 (context: predominantly positive vs. predominantly negative) x 2 (sampling procedure: no sampling vs. sampling by option) between-participants ANOVA showed a significant main effect of context, $F(1, 122) = 19.72$, $p < .001$, $\eta^2 = 0.139$. In the positive context conditions, participants skipped less free choice trials ($M = 1.46, SD = 1.37$) than participants in the negative context conditions ($M = 3.16, SD = 2.68$). Four participants in the negative context conditions even skipped all free choice trials. The main effect of sam-

[3] A model including previous option and previous outcome as random effects failed to converge, but led to equivalent results although they should be interpreted with caution.

pling procedure and its interaction with context were nonsignificant, $F(1,122) \leq 0.41$, $p \geq .521$.

As in Experiment 1, choice between options was analyzed by means of a generalized linear mixed model predicting the probability to choose option X1 in each choice trial. The model comprised a random intercept as well as the predictor type of trial (trial) that contrast-coded whether the choice trial was a forced choice trial (coded as 1) or a free choice trial (coded as -1). Moreover, the experimental conditions were added at the person-level by including the two factors context and sampling procedure as well as their interaction. They were each contrast-coded, as well, with value 1 indicating context C1 or sampling by option, respectively, and value -1 indicating context C2 or no sampling. The full model also included all additional interaction terms:[4]

$$\begin{aligned}\ln\left(\frac{p(Y_{ij} = \text{“}X1\text{”})}{1 - p(Y_{ij} = \text{“}X1\text{”})}\right) \quad = \quad & \gamma_{00} + \gamma_{01} \cdot sampling + \gamma_{02} \cdot context + (\gamma_{10} + u_{1j}) \cdot trial \\ & + \gamma_{03} \cdot sampling \cdot context + \gamma_{11} \cdot sampling \cdot trial \\ & + \gamma_{12} \cdot context \cdot trial + \gamma_{13} \cdot sampling \cdot context \cdot trial + u_{0j}\end{aligned} \tag{4}$$

Table A3 in the Appendix reports all parameter estimates of the generalized linear mixed model. The intercept was nonsignificant, $\gamma_{00} = -0.14$, $SE = 0.10$, $p = .173$, showing no overall preference for one of the two options. Still, the probability to choose option X1 was lower in the sampling by option conditions than in the no sampling conditions, $\gamma_{01} = -0.21$, $SE = 0.10$, $p = .033$, indicating a preference for option X2 with the higher genuine winning probability in the sampling by option conditions. Furthermore, the probability to choose option X1 was higher in forced choice trials as compared with free choice trials, $\gamma_{10} = 0.13$, $SE = 0.05$, $p = .007$, especially in the conditions of the predominantly negative context C2 as indicated by a significant interaction $\gamma_{12} = -0.12$, $SE = 0.05$, $p = .014$. All other parameter estimates were non-significant, $p \geq .152$.

Effects of Sampling on Choice Behavior. For the purpose of analyzing the relationship between sampling behavior and choice behavior, we estimated a generalized

[4] A model estimating correlations among random effects failed to converge. Although its results should be interpreted with caution, they were equivalent to the reported model.

linear mixed model of choice behavior only in the two sampling by option conditions. Besides the between-participants condition context (predominantly positive context C1 coded as 1 vs. predominantly negative context C2 coded as -1), the relative frequency of sampling option X1 during the learning phase was entered as person-level predictor (relX1). The relative frequencies of option X1 samples were calculated by dividing the number of times a participant had sampled option X1 by the total number of learning trials (34) and subtracting the constant $c = .50$. Again, we included trial type (forced choice trials coded as 1 vs. free choice trials coded as -1) as trial-level predictor. Consequently, the full model was:[5]

$$\begin{aligned} \ln\left(\frac{p(Y_{ij} = \text{“}X1\text{”})}{1 - p(Y_{ij} = \text{“}X1\text{”})}\right) &= \gamma_{00} + \gamma_{01} \cdot context + \gamma_{02} \cdot relX1 + \gamma_{10} \cdot trial \\ &+ \gamma_{03} \cdot context \cdot relX1 + \gamma_{20} \cdot context \cdot trial \\ &+ \gamma_{30} \cdot relX1 \cdot trial + \gamma_{40} \cdot context \cdot relX1 \cdot trial + u_{0j} \end{aligned} \quad (5)$$

Parameter estimates are reported in Table A4 in the Appendix. The negative intercept, $\gamma_{00} = -0.26, SE = 0.12, p = .032$, reflects an overall preference for option X2, if both options were sampled equally often during the learning phase. Still, the more frequently option X1 was sampled during learning, the higher the probability to choose option X1 in the decision phase, $\gamma_{02} = 6.63$, $SE = 0.98$, $p < .001$. Again, as was observed for choice behavior of all participants, including the no sampling conditions, the probability to choose option X1 was higher in forced choice trials in the negative context condition and in free choice trials in the positive context condition as compared with forced choice trials in the positive context condition and free choice trials in the negative context condition, $\gamma_{20} = -0.14$, $SE = 0.07$, $p = .083$. None of the other effects was significant, $p \geq .083$.

Probability Estimations. We analyzed participants' winning probability estimates by a 2 (sampling procedure: no sampling vs. sampling by option) x 2 (context: predominantly positive context C1 vs. predominantly negative context C2) x 2 (option:

[5] A generalized linear mixed model including trial type as random effect failed to converge. Although its results should be interpreted cautiously, they were equivalent to those reported.

X1 vs. X2) mixed ANOVA. The winning probability estimates for the two options differed depending on context, as indicated by the main effect of context, $F(1, 122) = 45.81$, $p < .001$, $\eta^2 = 0.135$. In line with the genuine winning probabilities, participants in the predominantly positive context conditions gave higher winning probability estimates than participants in the predominantly negative context conditions. Moreover, according to a significant two-way interaction between sampling procedure and option, $F(1, 122) = 4.12$, $p = .045$, $\eta^2 = 0.019$, the options' estimated winning probabilities depended on the sampling procedure, too. Follow-up analyses on this interaction, using the Bonferroni-Holm method to control for the error rate, showed higher overall estimated winning probabilities for option X2 than option X1 in the sampling by option conditions, $t(122) = -2.30$, $p = .046$, whereas there was no significant difference between options for the no sampling conditions, $t(122) = 0.53$, $p = .598$. All other main effects and interactions were not significant $F(1, 122) \leq 2.97$, $p \geq .087$.

Discussion

In Experiment 2, we compared free information sampling by option with predetermined learning trials and their consequences for pseudocontingency inference and choice between two options. While participants in the no sampling conditions observed learning trials in random order, our results suggest that participants engaging in free information sampling used a win-stay/lose-shift strategy to decide which option to sample in each trial. Besides, in line with some research on free information sampling (e.g., Hau et al., 2008; Hertwig et al., 2004), overall, none of the two options was sampled more frequently than the other. When both options were sampled equally often, no pseudocontingency was inferred, but instead the option with the higher underlying winning probability was preferred. However, the more frequently an option was sampled, the more likely it was chosen in a subsequent decision phase, too, independent of whether it was the option with the higher or lower underlying winning probability. At least for a predominantly positive context, this means that pseudocontingencies may also be inferred if information sampling is self-determined. Compared with free information sampling, participants in the no sampling conditions were more likely to choose the option with the lower underlying winning probability after learning, although a pseudocontingency effect was also not apparent in the winning probability estimates.

Most research on pseudocontingencies uses a trivariate scenario as in Experiment 1, because pairwise contingencies of options and outcomes with a context factor may foster pseudocontingency inferences (cf. Fleig et al., 2017; Meiser et al., 2018). Still, asymmetries in the strength of the inferred pseudocontingency between predominantly positive contexts and predominantly negative contexts have been found (Fleig et al., 2017; Meiser & Hewstone, 2004; Meiser et al., 2018), as well. For these reasons, Experiments 3 and 4 aimed at extending research on information sampling to a trivariate decision scenario.

Experiment 3

Experiment 3 investigated whether attention is allocated to contexts to varying degrees and whether asymmetric attention allocation may result in asymmetric pseudocontingency effects. Contexts may be sampled to varying degrees as a function of their valences. On the one hand, as soon as a context is associated with losses it might be considered as irrelevant. As such a context would be avoided for decisions as possible, it might soon be disregarded during learning. Therefore, neither a pseudocontingency nor a genuine contingency between options and outcomes would be inferred within that context. On the other hand, decision makers might pay particular attention to and therefore draw larger samples of options within a predominantly negative context (Lejarraga et al., 2012; Wulff et al., 2018) in order to avoid negative outcomes and experiences in later decisions.

Consequently, in Experiment 3, we compared free information sampling by context to predetermined learning trials in a trivariate decision scenario with a predominantly positive context and a predominantly negative context. When learning trials are predetermined, skewed base rates of options and outcomes are presented within a context that covary across contexts (cf. Table 1). According to pseudocontingency research, a pseudocontingency should be inferred resulting in a preference of the option which is predominant within a mostly positive context and infrequent within a predominantly negative context (e.g, Fiedler et al., 2009; Fiedler, Kutzner, & Vogel, 2013; Meiser & Hewstone, 2004, 2010). Again, we expected suboptimal choice behavior indicated by a preference for option X1 instead of option X2 in the no sampling condition. Contrarily, if learning is self-determined, decision makers engaging in sampling by context may sample one context more extensively than the other. On the one hand, the focus on one context may allow for increased performance in contingency learning within that context and thereby choice. On the other hand, sampling by contexts may induce a general focus on contexts and thereby increase the salience of pairwise contingencies of options and outcomes with contexts that exist in the underlying event distribution. This in turn may lead to stronger pseudocontingency effects (Meiser et al., 2018) when sampling information by context in the trivariate scenario. Thus, we expected choice preferences and winning probability estimates to depend on information sampling.

Method

Participants and Design. Seventy-six participants (53 female, M_{age} = 23.39, SD_{age} = 6.06) were recruited at the University of Mannheim to participate in Experiment 3 in return for a compensation of either €2 or partial course credit. Information sampling procedure (sampling by context vs. no sampling) was manipulated as between-participants factor. The variables context (predominantly positive C1 vs. predominantly negative C2) and option (X1 vs. X2) were implemented as within-participant factors. Thus, the participants were randomly allocated to the no sampling condition (n = 37) and the sampling by context condition (n = 39).

Material and Procedure. In Experiment 3, the materials and experimental procedure were identical to Experiment 1 with the following exceptions. In the sampling by context condition, participants sampled from each of the two contexts freely, in any order they desired without each draw contributing to their earnings. In each learning trial, participants first selected one of the two contexts. After 650 ms of presenting only the selected context including both options, one option was highlighted by increasing its size and reducing its transparency. After another 650 ms, the outcome followed. Option and outcome were drawn by the experimental program and in accordance with the event distributions of the respective context in Table 1 in random order. In contrast, in the no sampling condition, context, option, and outcome were determined by the experimental program according to the event frequencies depicted in Table 1 and identical to Experiment 1. Again, the respective winning probabilities were assigned to contexts and options counterbalanced across participants in both experimental conditions. Besides, the total number of following choice trials amounted to 16 forced choice trials and 16 free choice trials.

Results

Information Sampling Behavior. In Experiment 3, the probability to sample the predominantly positive context C1 in each learning trial was analyzed by means of a generalized linear mixed model including a random intercept as well as previous context (prevC) and previous outcome (prevY) as trial-level predictors. These two predictors were contrast-coded and indicated the context (positive context C1 coded as 1 vs. negative context C1 coded as -1) and outcome (gain coded as 1 vs. loss coded as -1) of the previous trial:[6]

$$\ln\left(\frac{p(Y_{ij} = \text{“}C1\text{”})}{1 - p(Y_{ij} = \text{“}C1\text{”})}\right) = \gamma_{00} + \gamma_{10} \cdot prevC + \gamma_{20} \cdot prevY + \gamma_{30} \cdot prevC \cdot prevY + u_{0j} \tag{6}$$

On average, participants in the sampling by context condition sampled the predominantly positive context C1 in $M = 35.23$ ($SD = 5.15$) out of 68 learning trials. Accordingly, the intercept of the generalized linear mixed model was nonsignificant, $\gamma_{00} = 0.00$, $SE = 0.04$, $p = .909$, indicating no overall preference for one context during information sampling. Yet, the probability to sample the predominantly positive context C1 increased if it was sampled in the previous trial, $\gamma_{10} = 0.64$, $SE = 0.04$, $p < .001$, especially if it

[6] A model including previous context and previous outcome as random effects failed to converge, but led to equivalent results, although they should be interpreted with caution.

resulted in a gain, $\gamma_{30} = 0.18$, $SE = 0.04$, $p < .001$. The main effect of previous outcome per se was nonsignificant $\gamma_{20} = 0.05$, $SE = 0.04$, $p = .249$. All parameter estimates are depicted in Table A5 in the Appendix.

Choice Behavior. The frequencies of making a choice within each context over the course of the free choice trials were analyzed by a 2 (sampling procedure: no sampling vs. sampling by context) x 2 (context: predominantly positive vs. predominantly negative) mixed ANOVA with context as within-participant factor. A significant main effect of context, $F(1, 74) = 29.96$, $p < .001$, $\eta^2 = 0.182$ indicated that in the decision phase, participants skipped more trials in the predominantly negative context C2 as compared with the predominantly positive context C1. On average, participants skipped $M = 0.67$ ($SD = 1.20$) out of eight free choice trials in context C1 and $M = 2.58$ ($SD = 2.61$) out of eight free choice trials in context C2. Seven participants even skipped all free choice trials in context C2. Neither the main effect of sampling procedure nor its interaction with context was significant, $F(1, 74) \leq 0.42$, $p \geq 0.517$.

Mean relative frequencies of choosing the two options within each context are shown in Table 4. Again, the probability to choose option X1 in each choice trial was analyzed by using a generalized linear mixed model comprising a random intercept as well as the contrast-coded predictors context (predominantly positive context C1 coded as 1, predominantly negative context C2 coded as -1) and trial (forced choice trial coded as 1, free choice trial coded as -1) as random effects at trial-level and experimental condition at the level of person (sampling: no sampling coded as -1, sampling by context coded as 1):[7]

$$\begin{aligned} \ln\left(\frac{p(Y_{ij} = \text{“}X1\text{”})}{1 - p(Y_{ij} = \text{“}X1\text{”})}\right) &= \gamma_{00} + \gamma_{01} \cdot sampling + (\gamma_{10} + u_{1j}) \cdot context \\ &\quad + (\gamma_{20} + u_{2j}) \cdot trial + \gamma_{11} \cdot sampling \cdot context \\ &\quad + \gamma_{21} \cdot sampling \cdot trial + (\gamma_{30} + u_{3j}) \cdot context \cdot trial \\ &\quad + \gamma_{31} \cdot sampling \cdot context \cdot trial + u_{0j} \end{aligned} \tag{7}$$

[7] A generalized linear mixed model estimating correlations among the random effects failed to converge. Although its results should therefore be interpreted with caution, they were equivalent to the estimated model without estimating correlations among random effects.

Parameter estimates of the generalized linear mixed model are reported in Table A6 in the Appendix. The probability to choose option X1 was significantly different from chance in the predominantly positive context C1 as indicated by a significant positive intercept, $\gamma_{00} = 0.44$, $SE = 0.12$, $p < .001$, together with the significant effect of context, $\gamma_{10} = 0.57$, $SE = 0.17$, $p = .001$. As suggested by the effect of context outweighing the positive intercept, participants did not prefer any option within the predominantly negative context C2. All other effects were nonsignificant, $p \geq .150$.

Effects of Sampling on Choice Behavior. In order to analyze the effect of the relative frequency of sampling the predominantly positive context C1 during learning on choice behavior, we estimated a generalized linear mixed model of choice only in the sampling by context condition. The relative frequencies of context C1 samples were calculated by dividing the number of context C1 samples of each participant by the total number of learning trials (68) and subtracting the constant $c = .50$. The relative frequency (relC1) was entered as a person-level predictor in the generalized linear mixed model. Again, we included the trial-level predictor trial type (forced choice trials coded as 1 vs. free choice trials coded as -1).[8]

$$\begin{aligned}\ln\left(\frac{p(Y_{ij} = \text{“}X1\text{”})}{1 - p(Y_{ij} = \text{“}X1\text{”})}\right) &= \gamma_{00} + \gamma_{01} \cdot relC1 + (\gamma_{10} + u_{1j}) \cdot context \\ &\quad + (\gamma_{20} + u_{2j}) \cdot trial + \gamma_{11} \cdot relC1 \cdot context + \gamma_{21} \cdot relC1 \cdot trial \\ &\quad + \gamma_{30} \cdot context \cdot trial + \gamma_{40} \cdot relC1 \cdot context \cdot trial + u_{0j}\end{aligned} \tag{8}$$

The relative frequency of sampling the predominantly positive context C1 during learning did not affect the probability to choose option X1 in individual choice trials, $\gamma_{01} = 1.69$, $SE = 3.32$, $p = .611$. As for the generalized linear mixed model including all participants, the intercept was significant and positive, $\gamma_{00} = 0.64$, $SE = 0.25$, $p = .011$, indicating an above chance probability to choose option X1. Again, the probability to choose option X1 was higher within context C1 as compared with context C2, $\gamma_{10} = 0.67$, $SE = 0.33$, $p = .046$. All parameter estimates are shown in Table A7 in the Appendix.

[8] A generalized linear mixed model that included the interaction of context and trial as random effect and estimated correlations among random effects failed to converge. Although its results should therefore be interpreted with caution, they were equivalent to those of the reported model.

Table 4

Relative Frequencies of Choices and Winning Probability Estimates of Options Within Contexts in Each Experimental Condition of Experiment 3

	Context C1				Context C2			
	Option X1		Option X2		Option X1		Option X2	
	M	SD	M	SD	M	SD	M	SD
Relative choice frequency								
no sampling	65.49	22.59	34.51	22.59	46.30	23.44	53.70	23.44
sampling by context	66.54	33.16	33.46	33.16	50.75	32.98	49.25	32.98
Estimated probabilities								
no sampling	62.76	22.81	39.76	19.20	34.24	20.09	41.81	21.30
sampling by context	55.82	21.71	41.56	22.13	39.69	20.05	41.46	21.64

Note. Data in percent.

Probability Estimations. Table 4 displays mean winning probability estimates of the two options within each context. They were analyzed by a 2 (sampling procedure: no sampling vs. sampling by context) x 2 (context: predominantly positive vs. predominantly negative) x 2 (option: X1 vs. X2) mixed ANOVA with sampling procedure as between-participants factor. Significant main effects of context, $F(1, 74) = 36.61$, $p < .001$, $\eta^2 = 0.061$, and option, $F(1, 74) = 8.07$, $p = .006$, $\eta^2 = 0.027$, were qualified by a significant two-way interaction between these factors, $F(1, 74) = 13.32$, $p < .001$, $\eta^2 = 0.072$. In line with participants' choice behavior, participants gave higher winning probability estimates within the predominantly positive context C1. Moreover, follow-up analyses on the interaction, using the Bonferroni–Holm method to control for the error rate, indicated higher winning probability estimates for option X1 than for option X2 in the positive context C1, $t(138.92) = 4.63$, $p \leq .001$, but no difference in the predominantly negative context C2, $t(138.92) = -1.16$, $p = .248$.

Analyzing the unconditional winning probability estimates of contexts by means of a 2 (sampling procedure: no sampling vs. sampling by context) x 2 (context: predominantly positive vs. predominantly negative) mixed ANOVA, we found higher estimates for the positive context C1 ($M = 61.83$, $SD = 14.82$) than for the negative context C2 ($M = 40.20$, $SD = 16.68$), $F(1, 74) = 55.33$, $p < .001$, $\eta^2 = 0.325$. Finally, in line with the winning probability estimates of options within contexts, the estimated winning probabilities of options across contexts were higher for option X1 ($M = 52.75$, $SD = 16.59$) than for option X2 ($M = 44.16$, $SD = 17.53$), $F(1, 74) = 6.83$, $p = .011$, $\eta^2 = 0.061$.

Discussion

In Experiment 3, we compared free information sampling by context with predetermined information with regard to contingency learning and choice. Besides, we tested whether attention was allocated to contexts differing in their winning probabilities to varying degrees when sampling by context is self-determined. Overall, participants did not prefer one context during sampling, that is, both contexts were on average sampled roughly equally often. This may contradict previous findings in research on free information sampling reporting larger samples the more likely losses are (Lejarraga et al., 2012; Wulff et al., 2018). However, in Experiment 3, an important difference is that participants had to sample a fixed number of observations and did not have the opportunity to stop sampling at any point desired. This may have affected the allocation of samples to the contexts, even independent of the contexts' winning probabilities. Besides, similar to the win-stay/lose-shift strategy observed in Experiment 2, participants were more likely to resample the same context as in the preceding learning trial, especially if a positive outcome was observed. When engaging in free information sampling by context, on average, the amount of information per context does not differ from predetermined learning trials, whereas the order of information does differ from a random pattern.

Nevertheless, when making consequential choices, contexts' overall winning probabilities are included in decisions. Participants correctly identified and preferred the context in which gains result more often than losses, independent of whether learning was predetermined or self-determined. Moreover, with regard to the assessment of options within a context, participants in the sampling by context condition did not perform better than participants in the no sampling condition. Our results suggest that pseudocontingencies are inferred independent of the sampling condition and that they are used as basis for decision making in predominantly positive contexts, although leading to suboptimal choices. Yet again, the pseudocontingency effect was asymmetric. Within a predominantly negative context neither a pseudocontingency nor the genuine contingency was inferred, resulting in no preference at all.

Taken together, choices did not differ between free information sampling by context and predetermined information. Even if one context was sampled more frequently than the other, this seemed to neither affect contingency learning, judgments, nor choice between options. However, (pseudo-)contingency learning and choice in a trivariate decision scenario may deviate if decision makers do not only sample information by contexts, but also choose which option to be observed within which context in each learning trial. To this we turn next.

Experiment 4

Experiment 4 aimed at investigating pseudocontingency inferences conditional on free information sampling in a trivariate decision scenario when respondents sample by both, contexts and options. As for Experiments 2 and 3, literature on free information sampling is inconclusive whether and which information will be searched for more extensively. Nevertheless, according to pseudocontingency research a pseudocontingency should be inferred whenever base rates of options and outcomes are skewed within a context in the current sample and are taken into account (e.g., Fiedler et al., 2009). In consequence, in predominantly positive contexts, the most frequently observed option should be preferred, independent of whether it has the higher genuine winning probability. Contrarily, in predominantly negative contexts, a preference for the infrequently sampled option should result. If both options are equally frequent within a context, none should be preferred according to a pseudocontingency. Again, we expected contingency learning and choice to depend on the individual sampling strategy.

Method

Participants and Design. In order to compare free information sampling by context and option to predetermined learning in Experiment 4, we recruited 76 participants at the University of Mannheim (57 female, $M_{age} = 22.07, SD_{age} = 4.99$). They were randomly allocated to a no sampling condition ($n = 38$) and a sampling by context and option condition ($n = 38$) and received a compensation of either €2 or partial course credit. Again, the variables context (predominantly positive vs. predominantly negative) and option (X1 vs. X2) were implemented as within-participant factors.

Material and Procedure. Materials and procedure were identical to Experiment 3 except for the learning phase of the sampling by context and option condition: Participants did not only sample from each of the two contexts freely, but also from each of the two options within the chosen context. In each learning trial, they successively selected one of the two contexts and one of the two options within the selected context. After 650 ms the respective outcome was presented. The outcome was drawn from the stimulus distribution in Table 1 by the experimental program.

Results

Information Sampling Behavior. Mean relative frequencies of samples per option within each context are displayed in Table 5. As for Experiments 2 and 3, information sampling behavior was analyzed by a generalized linear mixed model. However, participants in the sampling by context and option condition sampled from each of the two contexts as well as from each of the two options within the chosen context. Therefore, a multivariate generalized linear mixed model was estimated to analyze the probability to sample the predominantly positive context C1 and the probability to sample option X1 simultaneously (McCulloch et al., 2008). We estimated separate regression parameters for each dependent variable by including two dummy-coded indicator variables, indC and indX. The indicator variable indC coded sampling choice between contexts as 1 and sampling choice between options as 0, whereas the variable indX indicated the dependent variable sampling choice between contexts by a value of 0 and sampling choice between options by a value of 1. Besides, the model comprised the contrast-coded predictors previous context (prevC: positive context C1 coded as 1, negative context C2 coded as -1), previous option (prevX: option X1 coded as 1, option X2 coded as -1), previous outcome (prevY: gain coded as 1, loss coded as -1), and context (positive context C1 coded as 1, negative context C2 coded as -1) on the trial level. Two separate intercepts were specified by estimating regression coefficients for each of the two indicator variables (γ_{100} and γ_{200}). In order to prevent rank-deficiency the overall intercept had to be omitted (Snijders & Bosker, 2012). Furthermore, separate slope parameters were specified conditional on the dependent variable by multiplying all predictors by the indicator variables:

$$\begin{aligned}\ln\left(\frac{p(Y_{hij}=1)}{1-p(Y_{hij}=1)}\right) &= \gamma_{100}\cdot indC + \gamma_{110}\cdot indC\cdot prevC + \gamma_{120}\cdot indC\cdot prevY \\ &\quad +\gamma_{130}\cdot indC\cdot prevC\cdot prevY + u_{10j} \\ &\quad +\gamma_{200}\cdot indX + \gamma_{210}\cdot indX\cdot prevX + \gamma_{220}\cdot indX\cdot prevY \\ &\quad +\gamma_{230}\cdot indX\cdot prevX\cdot prevY + \gamma_{240}\cdot indX\cdot context + u_{20j}\end{aligned} \tag{9}$$

Table A8 in the Appendix shows all parameter estimates of the multivariate generalized linear mixed model. The significant effects of indC, $\gamma_{100} = 0.26$, $SE = 0.08$, $p = .001$, and indX, $\gamma_{200} = -0.15$, $SE = 0.07$, $p = .030$, indicated overall higher probabilities to sample the predominantly positive context C1 as well as to sample option X2. On average, participants sampled context C1 in $M = 41.08$ ($SD = 9.19$) trials out of 68 learning trials in total and option X1 in $M = 29.97$ ($SD = 7.73$) out of 68 trials. As in Experiment 3, the probability to sample context C1 increased if it was sampled in the previous trial as well, $\gamma_{110} = 0.71$, $SE = 0.05$, $p < .001$, especially if it resulted in a gain, $\gamma_{130} = 0.22$, $SE = 0.05$, $p < .001$. Furthermore, as in Experiment 2, the probability to sample option X1 was higher if option X1 was also sampled in the preceding trial, $\gamma_{210} = 0.29$, $SE = 0.04$, $p < .001$, again, especially if it led to a gain, $\gamma_{230} = 0.28$, $SE = 0.04$, $p < .001$. All other effects were nonsignificant $p \geq .100$.

Choice Behavior. As for the previous experiments, frequencies of choices in the free choice trials were analyzed by means of a 2 (sampling procedure: sampling by context and option vs. no sampling) x 2 (context: predominantly positive vs. predominantly negative) mixed ANOVA. The probability to choose option X1 in individual choice trials was modeled in terms of a generalized linear mixed model.

In the decision phase, participants made more choices during the free choice trials given the predominantly positive context C1 ($M = 7.29$, $SD = 1.33$) as compared with the predominantly negative context C2 ($M = 4.21$, $SD = 2.90$) which was indicated by a significant main effect of context, $F(1, 74) = 76.69$, $p < .001$, $\eta^2 = 0.344$. This main effect was qualified by a significant two-way interaction of sampling procedure and context, $F(1, 74) = 13.45$, $p < .001$, $\eta^2 = 0.084$. Follow-up analyses showed that the difference between contexts was even greater in the sampling condition, $t(74) = 8.79$, $p < 0.001$, as compared to the no sampling condition, $t(74) = 3.60$, $p = 0.001$.

With regard to the individual choices between option X1 and option X2, mean relative frequencies of choices are shown in Table 5. The generalized linear mixed model included a random intercept as well as the contrast-coded predictors context (predominantly positive context C1 coded as 1, predominantly negative context C2 coded as -1) and trial (forced choice trials coded as 1, free choice trials coded as -1) at trial-level. The effect of context was set to be random.[9] Moreover, the model comprised experimental condition (sampling: no sampling coded as -1, sampling by context and option coded as 1) at person-level, so that the total model was:

$$\begin{aligned}\ln\left(\frac{p(Y_{ij} = \text{“}X1\text{”})}{1 - p(Y_{ij} = \text{“}X1\text{”})}\right) &= \gamma_{00} + \gamma_{01} \cdot sampling + (\gamma_{10} + u_{1j}) \cdot context + \gamma_{20} \cdot trial \qquad (10)\\ &+\gamma_{11} \cdot sampling \cdot context + \gamma_{30} \cdot sampling \cdot trial\\ &+\gamma_{40} \cdot context \cdot trial + \gamma_{50} \cdot sampling \cdot context \cdot trial + u_{0j}\end{aligned}$$

Parameter estimates of the generalized linear mixed model of choice behavior are displayed in Table A9 in the Appendix. The effect of sampling procedure was significant, $\gamma_{01} = -0.52$, $SE = 0.10$, $p < .001$, indicating a higher probability to choose option X1 in the no sampling condition as compared with the sampling by context and option condition. In addition, there was a significant interaction between sampling procedure

[9] A generalized linear mixed model that included both trial-level predictors and their interaction as random effects and estimated correlations among random effects failed to converge. Although it should therefore be interpreted with caution, the results were equivalent to those of the reported model.

and context, $\gamma_{11} = -0.64$, $SE = 0.14$, $p < .001$, indicating a more pronounced difference between the two experimental conditions within the predominantly positive context C1. Within the predominantly negative context C2, the results suggest that participants did not prefer any option, as the interaction offsets the main effect of sampling procedure. However, participants in the no sampling condition preferred option X1 within the predominantly positive context C1, whereas the probability to choose option X1 in the positive context C1 was below chance in the sampling by context and option condition. All other effects were nonsignificant, $p \geq .170$

Effects of Sampling on Choice Behavior. Furthermore, we analyzed the effect of relative frequency of option X1 samples within a context on the probability to choose option X1 within that context in the decision phase. For this reason, we estimated a multivariate generalized linear mixed model analyzing only data from the sampling by context and option condition. In order to be able to analyze the effect of sample size on choice probability within a specific context we included two dummy-coded indicator variables, indC1 and indC2. They indicate that a parameter is specific to the probability to choose option X1 either within the positive context C1 (coded as 1 in indC1 and coded as 0 in indC2) or within the negative context C2 (coded as 0 in indC1 and coded as 1 in indC2), respectively. Again, two separate intercepts were specified (γ_{10} and γ_{20}), wherefore the overall intercept had to be omitted (Snijders & Bosker, 2012). The model comprised two predictor variables relC1X1 and relC2X1 reflecting the relative frequencies of option X1 samples within the respective context. Relative frequencies were calculated by dividing the number of option X1 samples within a context by the total number of samples within that context and subtracting the constant $c = .50$. By multiplying the predictor variables by the indicator variables separate slope parameters are estimated for each context:

$$\begin{aligned} \ln\left(\frac{p(Y_{ij} = \text{“}X1\text{”})}{1 - p(Y_{ij} = \text{“}X1\text{”})}\right) &= (\gamma_{10} + u_{1j}) \cdot indC1 + (\gamma_{20} + u_{2j}) \cdot indC2 \\ &\quad + \gamma_{11} \cdot indC1 \cdot relC1X1 + \gamma_{21} \cdot indC2 \cdot relC2X1 \end{aligned} \tag{11}$$

The effect of indC1 was significant and negative, $\gamma_{10} = -0.86$, $SE = 0.21$, $p < .001$, indicating a below chance probability to choose option X1 given the predominantly

Table 5

Relative Frequencies of Samples, Relative Frequencies of Choices, and Winning Probability Estimates of Options Within Contexts in Each Experimental Condition of Experiment 4

	Context C1				Context C2			
	Option X1		Option X2		Option X1		Option X2	
	M	SD	M	SD	M	SD	M	SD
Relative sampling frequency								
sampling[a]	43.27	13.46	56.73	13.46	48.50	12.06	51.50	12.06
Relative choice frequency								
no sampling	68.05	28.03	31.95	28.03	45.62	30.03	54.38	30.03
sampling	27.89	22.47	72.11	22.47	49.72	25.07	50.28	25.07
Estimated probabilities								
no sampling	59.87	22.13	46.18	19.42	39.74	16.06	42.95	22.08
sampling	49.84	20.55	68.42	17.60	35.42	18.56	41.89	19.00

Note. Data in percent.

[a] Relative frequencies of sampled options conditional on sampled context.

positive context C1, hence a preference for option X2. Nonetheless, the more frequently option X1 was sampled within the positive context C1 during the learning phase, the more likely it was chosen within that context in the subsequent decision phase, $\gamma_{11} = 5.70$, $SE = 1.57$, $p < .001$. The probability to choose option X1 in the predominantly negative context C2 did not significantly differ from chance, $\gamma_{20} = 0.03$, $SE = 0.19$, $p = .866$, even though the probability to choose option X1 tended to increase with its sample size drawn within the negative context, $\gamma_{21} = 2.66$, $SE = 1.52$, $p = .080$. All parameter estimates are displayed in Table A10 in the Appendix.

Probability Estimations. Mean estimated winning probabilities of options within contexts per experimental condition are displayed in Table 5 and were analyzed by means of a 2 (sampling procedure: no sampling vs. sampling by context and option) x 2 (context: predominantly positive vs. predominantly negative) x 2 (option: X1 vs. X2) mixed ANOVA. The main effect of context was significant, $F(1,74) = 40.37$, $p < .001$, $\eta^2 = 0.148$, whereas the main effect of option fell short of significance, $F(1,74) = 3.86$, $p = .053$, $\eta^2 = 0.009$. Furthermore, main effects were qualified by a two-way interaction between sampling procedure and option, $F(1,74) = 22.92$, $p < .001$, $\eta^2 = 0.050$, as well as by a three-way interaction between sampling procedure, context, and option, $F(1,74) = 8.87$, $p = .004$, $\eta^2 = 0.034$. Follow-up analyses using the Bonferroni-Holm adjustment of p-values, showed that participants in the no sampling condition gave higher winning probability estimates for option X1 than for option X2 within the predominantly positive context C1, $t(138.27) = 3.16$, $p = .006$. Estimated winning probabilities of participants in the sampling by context and option condition were higher for option X2 as compared with option X1 within the predominantly positive context C1, $t(138.27) = -4.29$, $p < .001$. No differences between options were observed within the predominantly negative context C2, for the no sampling condition as well as for the sampling by context and option condition, $t(138.27) = -0.74$, $p = .460$ and $t(138.27) = -1.50$, $p = .274$, respectively.

In addition, the estimated winning probabilities of the two contexts reflected a preference for the predominantly positive context C1 ($M = 64.12$, $SD = 16.41$) as compared with the negative context C2 ($M = 37.28$, $SD = 13.59$). This was indicated by a significant main effect of context in a 2 (sampling procedure: no sampling vs. sampling by context and option) x 2 (context: predominantly positive vs. predominantly negative) mixed ANOVA, $F(1,74) = 113.43$, $p < .001$, $\eta^2 = 0.465$. This main effect was qualified by a significant two-way interaction with sampling procedure, $F(1,74) = 8.55$, $p = .005$, $\eta^2 = 0.061$. The difference between the predominantly positive context and the predominantly negative context was more pronounced in the sampling by context and option condition, $t(74) = 9.60$, $p < .001$, as compared to the no sampling condition, $t(74) = 5.46$, $p < .001$.

Finally, participants' winning probability estimates of the two options across contexts were analyzed using a 2 (sampling procedure: no sampling vs. sampling by context and option) x 2 (option: X1 vs. X2) mixed ANOVA. The main effects of sampling procedure and option were nonsignificant, $F(1, 74) < 1$, $p \geq .913$. Yet, a significant two-way interaction between sampling procedure and option indicated that participants gave different winning probability estimates for the two options depending on the experimental condition, $F(1, 74) = 17.83$, $p < .001$, $\eta^2 = 0.113$. Follow-up analyses using the Bonferroni-Holm method to control for the error rate showed that participants in the no sampling condition gave higher estimates for option X1 than for option X2, $t(74) = 3.05$, $p = .006$, whereas the mean winning probability estimation of option X1 was lower than the mean winning probability estimation of option X2 in the sampling by context and option condition, $t(74) = -2.92$, $p = .006$.

Discussion

Experiment 4 compared free information sampling by context and option to predetermined learning trials with regard to sampling behavior and choice. The results indicate that options are sampled depending on their winning probabilities when engaging in free information sampling. Although no draw was incentivized, participants sampled more frequently within a predominantly positive context as compared with a predominantly negative context. Additionally, the option with the higher underlying winning probability was more frequently sampled within the positive context, whereas there was no difference between options within the negative context.

With regard to choice in a decision phase, too, the results of Experiment 4 demonstrate that contingency learning differs depending on the contexts' overall winning probabilities. Whereas the contexts' winning probabilities are correctly inferred resulting in a preference for the context in which gains resulted more often than losses, inferred contingencies between options and outcomes within each context differ not only depending on the context's winning probability, but also depending on information sampling. An asymmetry between positive contexts and negative contexts resulted in no preference for any option within the negative context when both options were sampled equally often as well as when skewed base rates were observed during learning. By contrast, the probability to choose an option within a predominantly positive context increased with its sample size observed or drawn during the learning phase. On average, choice behavior within the positive context reflected a preference for the more frequent option independent of whether it had the objectively higher winning probability. In the no sampling condition with predetermined learning trials, the frequent option within the positive context always had the lower objective winning probability; in the sampling by context and option condition, in the majority of cases, the predominant option within the positive context corresponded to the option with the higher winning probability.

General Discussion

The present series of experiments extended research on pseudocontingency inference and research on free information sampling. We investigated information sampling behavior in a choice scenario to analyze its impact on the inference of pseudocontingencies. Pseudocontingency describes the phenomenon of inferring contingencies, for instance, between options and outcomes from statistical regularities like base rates (e.g., Fiedler et al., 2009). Accordingly, options are preferred that are frequent within contexts of predominantly positive outcomes as well as options that are infrequent within contexts where positive outcomes are rare. If options' and/or outcomes' base rates are uniform, no contingency should be inferred on the basis of pseudocontingencies resulting in no preference at all. Thus, empirical evidence suggests that options' evaluations and resulting choices may be based on skewed base rates of options and outcomes covarying across contexts. Pseudocontingency inferences are not necessarily wrong (Fiedler, Kutzner, & Vogel, 2013; Klauer, 2015; Kutzner et al., 2011a), but may be misleading in certain cases when the pseudocontingency is in discord with the genuine contingency. Such an event distribution was designed in Experiment 1, so that the potentially inferred pseudocontingency between options and outcomes within each context was of opposite sign to the genuine contingency. Over the course of learning trials, participants observed gains more frequently than losses within one context and vice versa within the second context. The options' base rates too were skewed within a context and covaried across contexts. Subsequently, participants chose between the observed options multiple times while aiming at maximizing points and had to estimate the options' winning probabilities. The same experimental procedure was realized as experimental conditions in Experiments 3 and 4. Replicating previous studies, we found the pseudocontingency effect on choice behavior: a preference for the frequently observed option within a predominantly positive context and a preference for the least frequently observed option within a predominantly negative context. However, in Experiments 3 and 4, our results revealed an asymmetry between positive contexts with mostly gains and negative contexts with mostly losses: whereas participants preferred the frequent option within the positive context, they preferred neither option within the negative context. Similar results have also been found in the literature (Fleig et al., 2017; Meiser & Hewstone, 2004; Meiser et al., 2018). On the one hand, the observed asymmetry might be the result of allocating more attention to the task at hand when processing losses (Yechiam & Hochman, 2013, 2014). This might have increased the focus on the genuine contingency as compared with the pseudocontingency within the negative context. On the other hand, participants might have disregarded the options within the negative context as irrelevant for subsequent choices with the aim of avoiding to choose options within the negative context anyway (cf. e.g., Fiedler, Wöllert, et al., 2013). Thus, neither the genuine contingency nor any other regularities like the skewness of base rates could be learned.

Information sampling behavior might shed light on which information is considered for contingency learning and choice, when information is not presented to, but actively sampled by decision makers. Research on information sampling primarily uses bivariate scenarios without the additional context variable. Likewise, in Experiment 2, we realized the context variable of our design as between-participants condition. Participants were free to sample any of the two options either within a predominantly positive context or within a predominantly negative context. They could sample information in any order they desired without each draw contributing to individual earnings during learning trials. In line with, for instance, Hau et al. (2008) or Hertwig et al. (2004), our results indicate that two options are sampled equally often. Independent of the decision scenario, thus regardless of whether participants sampled within a context of predominantly gains or within a context of predominantly losses, sample sizes did not significantly differ between the two options. In Experiments 3 and 4, we extended Experiment 2 and thereby research on information sampling to the trivariate choice scenario with two contexts, two options, and two outcomes. Comparing drawn samples within gain domains versus loss domains, Lejarraga et al. (2012) as well as Wulff et al. (2018) reported larger samples to be drawn from options when the task at hand involved losses. However, when realizing the decision domain as a third (within-participant) variable, such as the context variable in our trivariate choice scenario, results contradict. Depending on the variable sampled by we either found no effect of context on sample size or asymmetric exploration with larger samples drawn within a predominantly positive context as compared with a predominantly negative context. In Experiment 3, when participants sampled information by context, both contexts were sampled equally often. When participants engaged in free information sampling by context and option (Experiment 4), however, our results suggest that options within negative contexts are rather disregarded. The focus is instead on the options within the context of mostly gains. On average, participants moreover sampled the option with the higher underlying winning probability most frequently within the positive context.

Besides information sampling behavior itself and choice behavior after predetermined learning trials, we also investigated choice behavior when preceding learning was self-determined. Our results again indicated an asymmetry in the inferred contingency between predominantly positive contexts and predominantly negative contexts. Not only when learning trials were predetermined, but also when sampling information during learning, subsequent choice behavior reflected preferences for one of the two options only within positive contexts. In Experiments 2 and 4, when participants could influence the sample size per option during learning, on average, they sampled both options equally frequently within negative contexts. On the basis of pseudocontingencies, a contingency between options and outcomes should only be inferred when both, the options' base rate and the outcomes' base rate are skewed. In Experiment 4, indeed, we found no difference in choice probability within the negative context when both options were sampled equally often during learning. However, we neither found a difference between options within the negative context when predetermined skewed base rates were presented in Experiments 2 to 4. Moreover, in the negative context condition of Experiment 2, participants preferred the option with the higher winning probability when they sampled both options equally often. In a trivariate decision scenario, the indifference between options within negative contexts could thus be the result of subjective uncertainty with regard to the options' winning probabilities. Self-constructed feedback might have contributed. Even though no external feedback on the outcome of each individual choice trial was given, individuals may construct feedback themselves based on the outcome they expect to be most likely (Henriksson, Elwin, & Juslin, 2010). Within the negative context this corresponds to a loss for both options. Thus, both options might be equally devaluated by self-constructed (learning) experiences resulting in a similar non-preference for either option. By contrast, within positive contexts we observed the inference of a pseudocontingency between options and outcomes when learning trials were predetermined as well as when sampling information by context. Additionally, when sampling information by option or by context and option, participants' preference for an option within the positive context increased with its sampling frequency. Later choice preferences within the positive context were already discernible in the information sample drawn within the positive context, whether it was the option with the higher underlying winning probability or not. In the majority of cases, though, the frequently sampled and chosen option within the positive context corresponded to the option with the higher winning probability. Therefore, an inference of the pseudocontingency as well as an inference of the genuine contingency between options and outcomes within the positive context would have resulted in the same preference pattern given those self-generated samples: even if choices were based on a pseudocontingency the option with the higher underlying winning probability would be preferred.

Going beyond, contingency inference might have already influenced sampling behavior during the learning phase. Even though there were no direct consequences of sampling an option during free information sampling, we cannot preclude that current evaluations of options guided sampling behavior. Participants' sampling pattern was characterized by a stability in information sampling: in Experiments 2 to 4, we consistently found participants' tendency to repeat sampling the same event. This was in particular the case when a positive outcome was observed, indicating a win-stay/lose-shift strategy besides consistency in information sampling. Possibly, observing gains during a learning phase is already positive and reinforcing in itself, even if they have no direct consequences. As a result, individuals might be hedonically motivated to focus on positive events even in *free* information sampling and engage more in exploitation rather than exploration of the available options. First, this can lead to higher sampling frequencies of options with higher winning probabilities (i.e., options within the positive context, as observed in Experiment 4; cf. e.g., Denrell & Le Mens, 2011). Second, this might have contributed to sampling one option within the positive context quite constantly and therefore most frequently already during the learning phase. Yet, within the negative context, hedonic motivation might have resulted in repeatedly switching between options and sampling within the negative context less continuously. This in turn might impede learning and foster a subjective uncertainty with regard to the options' winning probabilities within a predominantly negative context. Future research should further investigate the interplay between information sampling and contingency learning to identify variables moderating when free information sampling is more or less dependent on evaluations resulting in lower or higher proportions of explorations versus exploitation during learning.

Moreover, the results suggest that active sampling from contexts and options that differ in their objective winning probabilities enables individuals to conditionalize winning and losing not only on contexts (i.e., learning outcomes' base rates within each context), but also on choice options. Active sampling can increase the attention towards single events and thereby foster genuine contingency assessment, at least in predominantly positive domains. In contrast, when passively observing predetermined information, assessing proportions of winning and losing conditional on options as well as contexts is rather unlikely.

The experiments reported herein indicate that the adopted information sampling procedure can on the one hand counteract the inference of a pseudocontingency that could result in suboptimal subsequent choice behavior. On the other hand, (a priori) expectations about the options' winning probabilities in combination with positive testing strategies or exploiting positively evaluated options can contribute to drawing skewed base rates and the inference of pseudocontingencies. Accordingly, it is possible that (pseudo-)contingencies inferred on the basis of the current sample already guide further sampling behavior even in free information sampling. A comparison with a partial feedback sampling paradigm in which each draw is incentivized would be worthwhile.

To conclude, the present research corroborates the notion that pseudocontingencies are used for decision making. Still, we report original evidence that even if decision makers rely on base rates, resulting choice behavior will often be in line with true contingencies when information is not predetermined, but may be actively sampled. The asymmetry in the strength of preferences between contexts of predominantly gains and contexts of predominantly losses might arise from a reduced focus on negatively evaluated options, an avoidance of negative outcomes (even in free information sampling), and/or from a resulting subjective uncertainty.

References

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, *15*(3), 147–149. doi: 10.3758/BF03334492

Ashby, N. J., & Rakow, T. (2016). Eyes on the prize? evidence of diminishing attention to experienced and foregone outcomes in repeated experiential choice. *Journal of Behavioral Decision Making*, *29*(2-3), 183–193. doi: 10.1002/bdm.1872

Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making*, *16*(3), 215–233. doi: 10.1002/bdm.443

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

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

Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, *99*(2), 365–382. doi: 10.1037/0033-295X.99.2.365

Denrell, J., & Le Mens, G. (2011). Seeking positive experiences can produce illusory correlations. *Cognition*, *119*(3), 313–324. doi: 10.1016/j.cognition.2011.01.007

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

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review*, *107*(4), 659–676. doi: 10.1037//0033-295X.107A659

Fiedler, K. (2010). Pseudocontingencies can override genuine contingencies between multiple cues. *Psychonomic Bulletin & Review*, *17*(4), 504–509. doi: 10.3758/PBR.17.4.504

Fiedler, K., & Freytag, P. (2004). Pseudocontingencies. *Journal of Personality and Social Psychology*, *87*(4), 453–467. doi: 10.1037/0022-3514.87.4.453

Fiedler, K., Freytag, P., & Meiser, T. (2009). Pseudocontingencies: An integrative account of an intriguing cognitive illusion. *Psychological Review*, *116*(1), 187–206. doi: 10.1037/a0014480

Fiedler, K., Freytag, P., & Unkelbach, C. (2007). Pseudocontingencies in a simulated classroom. *Journal of Personality and Social Psychology*, *92*(4), 665–677. doi: 10.1037/0022-3514.92.4.665

Fiedler, K., Kutzner, F., & Vogel, T. (2013). Pseudocontingencies: Logically unwarranted but smart inferences. *Current Directions in Psychological Science 2013*, *22*(4), 324–329. doi: 10.1177/0963721413480171

Fiedler, K., Wöllert, F., Tauber, B., & Hess, P. (2013). Applying sampling theories to attitude learning in a virtual school class environment. *Organizational Behavior and Human Decision Processes*, *122*(2), 222–231. doi: 10.1016/j.obhdp.2013.08.001

Fleig, H., Meiser, T., Ettlin, F., & Rummel, J. (2017). Statistical numeracy as a moderator of (pseudo)contingency effects on decision behavior. *Acta Psychologica*, *174*, 68–79. doi: 10.1016/j.actpsy.2017.01.002

Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description-experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making*, *21*(5), 493–518. doi: 10.1002/bdm.598

Henriksson, M. P., Elwin, E., & Juslin, P. (2010). What is coded into memory in the absence of outcome feedback? *Journal of Experimental Psychology: Learning Memory and Cognition*, *36*(1), 1–16. doi: 10.1037/a0017893

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*(8), 534–539. doi: 10.1111/j.0956-7976.2004.00715.x

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2006). The role of information sampling in risky choice. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition* (pp. 72–91). Cambridge: Cambridge University Press.

Hilbig, B. E., & Glöckner, A. (2011). Yes, they can! Appropriate weighting of small probabilities as a function of information acquisition. *Acta Psychologica*, *138*(3), 390–396. doi: 10.1016/j.actpsy.2011.09.005

Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods*, *49*(5), 1652–1667. doi: 10.3758/s13428-017-0900-z

Klauer, K. C. (2015). Mathematical modeling. In B. Gawronski & G. V. Bodenhausen (Eds.), *Theory and explanation in social psychology* (pp. 371–389). New York, NY, US: Guilford Press.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY [u.a.]: Guilford.

Kutzner, F., Freytag, P., Vogel, T., & Fiedler, K. (2008). Base-rate neglect as a function of base rates in probabilistic contingency learning. *Journal of the Experimental Analysis of Behavior*, *90*(1), 23–32. doi: 10.1901/jeab.2008.90-23

Kutzner, F., Vogel, T., Freytag, P., & Fiedler, K. (2011a). Contingency inferences driven by base rates: Valid by sampling. *Judgment and Decision Making*, *6*(3), 211–221.

Kutzner, F., Vogel, T., Freytag, P., & Fiedler, K. (2011b). A robust classic: Illusory correlations are maintained under extended operant learning. *Experimental Psychology*, *58*(6), 443–453. doi: 10.1027/1618-3169/a000112

Lejarraga, T., Hertwig, R., & Gonzalez, C. (2012). How choice ecology influences search in decisions from experience. *Cognition*, *124*(3), 334–342. doi: 10.1016/j.cognition.2012.06.002

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. doi: 10.3758/s13428-011-0168-7

McCulloch, C. E., Searle, S. R., & Neuhaus, J. M. (2008). *Generalized, linear, and mixed models* (2nd ed.). Hoboken, NJ: Wiley.

Meiser, T. (2006). Contingency learning and biased group impressions. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition.* (pp. 183–209). Cambridge: Cambridge University Press.

Meiser, T., & Hewstone, M. (2004). Cognitive processes in stereotype formation: The role of correct contingency learning for biased group judgments. *Journal of Personality and Social Psychology*, *87*(5), 599–614. doi: 10.1037/0022-3514.87.5.599

Meiser, T., & Hewstone, M. (2010). Contingency learning and stereotype formation: Illusory and spurious correlations revisited. *European Review of Social Psychology*, *21*(1), 285–331. doi: 10.1080/10463283.2010.543308

Meiser, T., Rummel, J., & Fleig, H. (2018). Pseudocontingencies and choice behavior in probabilistic environments with context-dependent outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(1), 50–67. doi: 10.1037/xlm0000432

Perales, J. C., Catena, A., Cándido, A., & Maldonado, A. (2017). Rules of causal judgments: Mapping statistical information onto causal beliefs. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 29–51). New York, NY: Oxford University Press.

R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.r-project.org/

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

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Los Angeles, Calif. [u.a.]: SAGE.

Spellman, B. A., & Mandel, D. R. (1999). When possibility informs reality: Counterfactual thinking as a cue to causality. *Current Directions in Psychological Science*, *8*(4), 120–123. doi: 10.1111/1467-8721.00028

Spellman, B. A., Price, C. M., & Logan, J. M. (2001). How two causes are different from one: The use of (un)conditional information in Simpson's paradox. *Memory and Cognition*, *29*(2), 193–208. doi: 10.3758/BF03194913

Vogel, T., Freytag, P., Kutzner, F., & Fiedler, K. (2013). Pseudocontingencies derived from categorically organized memory representations. *Memory & Cognition*, *41* (8), 1185–1199. doi: 10.3758/s13421-013-0331-8

Waldmann, M. R., & Hagmayer, Y. (2001). Estimating causal strength: The role of structural knowledge and processing effort. *Cognition*, *82* (1), 27–58. doi: 10.1016/S0010-0277(01)00141-X

Waldmann, M. R., Hagmayer, Y., & Blaisdell, A. P. (2006). Beyond the information given: Causal models in learning and reasoning. *Current Directions in Psychological Science*, *15* (6), 307–311. doi: 10.1111/j.1467-8721.2006.00458.x

Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General*, *121* (2), 222–236. doi: 10.1037/0096-3445.121.2.222

Wulff, D. U., Hills, T. T., & Hertwig, R. (2015). How short- and long-run aspirations impact search and choice in decisions from experience. *Cognition*, *144*, 29–37. doi: 10.1016/j.cognition.2015.07.006

Wulff, D. U., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin*, *144* (2), 140–176. doi: 10.1037/bul0000115

Yechiam, E., & Hochman, G. (2013). Loss-aversion or loss-attention: The impact of losses on cognitive performance. *Cognitive Psychology*, *66* (2), 212–231. doi: 10.1016/j.cogpsych.2012.12.001

Yechiam, E., & Hochman, G. (2014). Loss attention in a dual-task setting. *Psychological science*, *25* (2012), 494–502. doi: 10.1177/0956797613510725

Appendix

Parameter Estimates of the Generalized Linear Mixed Models

Table A1

Parameter Estimates of the Generalized Linear Mixed Model of Choice Behavior in Experiment 1

Parameter	Estimate	SE
Trial-level		
Intercept (γ_{00})	0.68**	0.21
Context (γ_{10})	0.03	0.23
Trial (γ_{20})	-0.05	0.09
Context × trial (γ_{30})	-0.07	0.09
Variances		
Intercept ($\sigma^2_{u_{0j}}$)	1.49	
Context ($\sigma^2_{u_{1j}}$)	1.87	
Trial ($\sigma^2_{u_{2j}}$)	0.00	
Context × trial ($\sigma^2_{u_{3j}}$)	0.00	

Note. Context (predominantly positive context C1 vs. predominantly negative context C2, coded as 1 and -1) and trial (forced choice trial vs. free choice trial, coded as 1 and -1) were contrast-coded.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table A2

Parameter Estimates of the Generalized Linear Mixed Model of Information Sampling Behavior in Experiment 2

Parameter	Estimate	SE
Trial-level		
Intercept (γ_{00})	-0.04	0.07
Previous option (γ_{10})	0.67***	0.06
Previous outcome (γ_{20})	0.05	0.06
Context $\times$ previous option (γ_{30})	-0.02	0.06
Context $\times$ previous outcome (γ_{40})	0.03	0.06
Previous option $\times$ previous outcome (γ_{50})	0.29***	0.06
Context $\times$ previous option $\times$ previous outcome (γ_{60})	-0.02	0.06
Person-level		
Context (γ_{01})	-0.09	0.07
Variances		
Intercept ($\sigma^2_{u_{0j}}$)	0.09	

Note. Context (predominantly positive context C1 vs. predominantly negative context C2, coded as 1 and -1), previous option (option X1 vs. option X2, coded as 1 and -1), and previous outcome (gain vs. loss, coded as 1 and -1) were contrast-coded.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table A3

Parameter Estimates of the Generalized Linear Mixed Model of Choice Behavior in Experiment 2

Parameter	Estimate	SE
Trial-level		
Intercept (γ_{00})	−0.14	0.10
Trial (γ_{10})	0.13**	0.05
Person-level		
Sampling (γ_{01})	−0.21*	0.10
Context (γ_{02})	0.08	0.10
Sampling × context (γ_{03})	−0.14	0.10
Sampling × trial (γ_{11})	−0.01	0.05
Context × trial (γ_{12})	−0.12*	0.05
Sampling × context × trial (γ_{13})	−0.02	0.05
Variances		
Intercept ($\sigma^2_{u_{0j}}$)	0.96	
Trial ($\sigma^2_{u_{1j}}$)	0.00	

Note. Sampling (no sampling vs. sampling by option, coded as -1 and 1), context (predominantly positive context C1 vs. predominantly negative context C2, coded as 1 and -1), and trial (forced choice trial vs. free choice trial, coded as 1 and -1) were contrast-coded.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table A4

Parameter Estimates of the Generalized Linear Mixed Model of Choice Behavior in the Sampling by Option Conditions of Experiment 2

Parameter	Estimate	SE
Trial-level		
Intercept (γ_{00})	-0.26*	0.12
Trial (γ_{10})	0.11	0.07
Context × trial (γ_{20})	-0.14*	0.07
Relative frequency of X1-samples × trial (γ_{30})	-1.12	0.65
Context × relative frequency of X1-samples × trial (γ_{40})	-0.49	0.65
Person-level		
Context (γ_{01})	-0.02	0.12
Relative frequency of X1-samples (γ_{02})	6.63***	0.98
Context × relative frequency of X1-samples (γ_{03})	-1.40	0.97
Variances		
Intercept ($\sigma^2_{u_{0j}}$)	0.55	

Note. Context (predominantly positive context C1 vs. predominantly negative context C2, coded as 1 and -1) and trial (forced choice trial vs. free choice trial, coded as 1 and -1) were contrast-coded. Relative frequencies of X1-samples during learning were calculated by dividing the number of X1-samples by 34 (total number of learning trials) and subtracting a constant $c = .50$.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table A5

Parameter Estimates of the Generalized Linear Mixed Model of Information Sampling Behavior in Experiment 3

Parameter	Estimate	SE
Trial-level		
Intercept (γ_{00})	0.00	0.04
Previous context (γ_{10})	0.64***	0.04
Previous outcome (γ_{20})	0.05	0.04
Previous context $\times$ previous outcome (γ_{30})	0.18***	0.04
Variances		
Intercept ($\sigma^2_{u_{0j}}$)	0.00	

Note. Previous context (predominantly positive context C1 vs. predominantly negative context C2, coded as 1 and -1) and previous outcome (gain vs. loss, coded as 1 and -1) were contrast-coded.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table A6

Parameter Estimates of the Generalized Linear Mixed Model of Choice Behavior in Experiment 3

Parameter	Estimate	SE
Trial-level		
Intercept (γ_{00})	0.44***	0.12
Context (γ_{10})	0.57**	0.17
Trial (γ_{20})	0.08	0.05
Context × trial (γ_{30})	−0.01	0.05
Person-level		
Sampling (γ_{01})	0.11	0.12
Sampling × context (γ_{11})	0.00	0.17
Sampling × trial (γ_{21})	0.01	0.05
Sampling × context × trial (γ_{31})	0.06	0.05
Variances		
Intercept ($\sigma^2_{u_{0j}}$)	0.80	
Context ($\sigma^2_{u_{1j}}$)	1.81	
Trial ($\sigma^2_{u_{2j}}$)	0.00	
Context × trial ($\sigma^2_{u_{3j}}$)	0.00	

Note. Sampling (no sampling vs. sampling by context, coded as -1 and 1), context (predominantly positive context C1 vs. predominantly negative context C2, coded as 1 and -1) and trial (forced choice trial vs. free choice trial, coded as 1 and -1) were contrast-coded.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table A7

Parameter Estimates of the Generalized Linear Mixed Model of Choice Behavior in the Sampling by Context Condition of Experiment 3

Parameter	Estimate	SE
Trial-level		
Intercept (γ_{00})	0.64*	0.25
Context (γ_{10})	0.67*	0.33
Trial (γ_{20})	0.09	0.08
Context × trial (γ_{30})	0.05	0.08
Relative frequency of C1-samples × context × trial (γ_{40})	0.96	1.16
Person-level		
Relative frequency of C1-samples (γ_{01})	1.69	3.32
Relative frequency of C1-samples × context (γ_{11})	0.72	4.38
Relative frequency of C1-samples × trial (γ_{21})	0.55	1.16
Variances		
Intercept ($\sigma^2_{u_{0j}}$)	1.94	
Context ($\sigma^2_{u_{1j}}$)	3.69	
Trial ($\sigma^2_{u_{2j}}$)	0.00	

Note. Context (predominantly positive context C1 vs. predominantly negative context C2, coded as 1 and -1) and trial (forced choice trial vs. free choice trial, coded as 1 and -1) were contrast-coded. Relative frequencies of C1-samples during learning were calculated by dividing the number of C1-samples by 68 (total number of learning trials) and subtracting a constant $c = .50$.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table A8

Parameter Estimates of the Multivariate Generalized Linear Mixed Model of Information Sampling Behavior in Experiment 4

Parameter	Estimate	SE
Trial-level		
indC (γ_{100})	0.26**	0.08
indC × previous context (γ_{110})	0.71***	0.05
indC × previous outcome (γ_{120})	0.04	0.05
indC × previous context × previous outcome (γ_{130})	0.22***	0.05
indX (γ_{200})	−0.15*	0.07
indX × previous option (γ_{210})	0.29***	0.04
indX × previous outcome (γ_{220})	0.00	0.04
indX × previous option × previous outcome (γ_{230})	0.28***	0.04
indX × context (γ_{240})	−0.07	0.04
(Co-)Variances		
indC ($\sigma^2_{u_{10j}}$)	0.15	
indX ($\sigma^2_{u_{20j}}$)	0.11	
indC, indX ($\sigma_{u_{10j},u_{20j}}$)	−0.05	

Note. The indicator variables indC and indX are dummy-coded and indicate that a parameter is specific to the probability to sample context C1 or the probability to sample option X1, respectively. Previous context (predominantly positive context C1 vs. predominantly negative context C2, coded as 1 and -1), previous outcome (gain vs. loss, coded as 1 and -1), previous option (X1 vs. X2, coded as 1 and -1), and context (predominantly positive context C1 vs. predominantly negative context C2, coded as 1 and -1) were contrast-coded.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table A9

Parameter Estimates of the Generalized Linear Mixed Model of Choice Behavior in Experiment 4

Parameter	Estimate	SE
Trial-level		
Intercept (γ_{00})	-0.14	0.10
Context (γ_{10})	0.03	0.14
Trial (γ_{20})	0.07	0.06
Sampling $\times$ trial (γ_{30})	0.05	0.06
Context $\times$ trial (γ_{40})	-0.03	0.06
Sampling $\times$ context $\times$ trial (γ_{50})	0.02	0.06
Person-level		
Sampling (γ_{01})	-0.52***	0.10
Sampling $\times$ context (γ_{11})	-0.64***	0.14
Variances		
Intercept ($\sigma^2_{u_{0j}}$)	0.55	
Context ($\sigma^2_{u_{1j}}$)	1.31	

Note. Sampling (no sampling vs. sampling by context and option, coded as -1 and 1), context (predominantly positive context C1 vs. predominantly negative context C2, coded as 1 and -1) and trial (forced choice trial vs. free choice trial, coded as 1 and -1) were contrast-coded. * $p < .05$, ** $p < .01$, *** $p < .001$.

Table A10

Parameter Estimates of the Multivariate Generalized Linear Mixed Model of Choice Behavior in the Sampling by Context and Option Condition of Experiment 4

Parameter	Estimate	SE
Trial-level		
indC1 (γ_{10})	−0.86***	0.21
indC2 (γ_{20})	0.03	0.19
Person-level		
indC1 × relative frequency of X1-samples in C1 (γ_{11})	5.70***	1.57
indC2 × relative frequency of X1-samples in C2 (γ_{21})	2.66	1.52
(Co-)Variances		
indC1 ($\sigma^2_{u_{1j}}$)	1.03	
indC2 ($\sigma^2_{u_{2j}}$)	0.87	
indC1, indC2 ($\sigma_{u_{1j},u_{2j}}$)	−0.52	

Note. The indicator variables indC1 and indC2 are dummy-coded and indicate that a parameter is specific to the probability to choose option X1 within context C1 or within context C2, respectively. Relative frequencies of X1-samples within context C1 and within context C2 during learning were calculated by dividing the number of X1-samples per context by the total number of learning trials within one context and subtracting the constant $c = .50$.

* $p < .05$, ** $p < .01$, *** $p < .001$.

This dissertation was funded by the Deutsche Forschungsgemeinschaft (DFG) grant 2277, Research Training Group “Statistical Modeling in Psychology" (SMiP).