

# Pseudocontingencies derived from categorically organized memory representations

Tobias Vogel · Peter Freytag · Florian Kutzner · Klaus Fiedler

Published online: 6 June 2013  
© Psychonomic Society, Inc. 2013

**Abstract** Pseudocontingencies (PCs) allow for inferences about the contingency between two variables  $X$  and  $Y$  when the conditions for genuine contingency assessment are not met. Even when joint observations  $X_i$  and  $Y_i$  about the same reference objects  $i$  are not available or are detached in time or space, the correlation  $r(X_i, Y_i)$  is readily inferred from base rates. Inferred correlations are positive (negative) if  $X$  and  $Y$  base rates are skewed in the same (different) directions. Such PC inferences afford useful proxies for actually existing contingencies. While previous studies have focused on PCs due to environmental base rates, the present research highlights memory organization as a natural source of PC effects. When information about two attributes  $X$  and  $Y$  is represented in a hierarchically organized categorical memory code, as category-wise base rates  $p(X)$  and  $p(Y)$ , the reconstruction of item-level information from category base rates will naturally produce PC effects. Three experiments support this contention. When the *yes* base rates of two respondents in four questionnaire subscales (categories) were correlated, recalled and predicted item-level responses were correlated in the same direction, even when the original responses to specific items within categories were correlated in the opposite direction.

**Keywords** Deductive reasoning · Memory organization · Categorization · Contingency learning

## Contingency assessment as a major adaptive function

A new research program on pseudocontingencies (PCs) challenges the common belief in contingencies as the

chief module of adaptive learning (Fiedler, 2010; Fiedler & Freytag, 2004; Fiedler, Freytag, & Meiser, 2009; Freytag, Bluemke, & Fiedler, 2011). Contingency learning is presupposed to underlie causal inferences (Cheng & Nowick, 1992), conditioning (Murphy & Baker, 2004), concept formation (Richardson & Bhavnani, 1984), stereotyping (McCauley & Stitt, 1978), multicue inferences (Lagnado, Newell, Kahan, & Shanks, 2006), and fast and frugal heuristics (Gigerenzer & Todd, 1999) in the context of Brunswik's (1952) lens model.

However, closer examination of learning environments reveals that the conditions for proper contingency assessment are hardly ever met. Causes and effects can rarely be observed in close temporal succession; multiple cues are often confounded and hard to disentangle; and the environment rarely provides us with the complete multivariate data and nonselective feedback that are necessary to identify all cue correlations. For example, a typical get-acquainted task calls for inferences about when and why a target person answers “yes” or “no” in response to different utterances or communicative actions. The person's positive or negative affective responses may depend on a multitude of cues, such as the attractiveness and the voice of communication partners, their status as friends or opponents, their gender, the hedonic quality of the utterances, their consistency with the person's own political attitudes, or their touching a threatening aspect of the person's self. However, for various reasons, the multivariate data matrix that would be necessary to assess the correlations between all these cues and the criterion is incomplete and impoverished. Only some utterances are, in reality, met with an overt and immediate feedback. Delayed feedback (e.g., the person's reaction on the next day, in a new context) is detached from a specific set of eliciting cues. Many cues are often missing (e.g., telephone calls do not reveal attractiveness, e-mails have no voice) or remain unknown (e.g., social status) or irrelevant (utterances are politically mute). Or information about different cues (e.g., reactions to political attitudes and to

---

T. Vogel (✉)  
University of Mannheim, Parkring 47,  
68159 Mannheim, Germany  
e-mail: tovogel@mail.uni-mannheim.de

P. Freytag · F. Kutzner · K. Fiedler  
University of Heidelberg, Heidelberg, Germany

gender) is learned on separate occasions, making it impossible to locate connected cue values within the same compact matrix. In other words, real life is replete with missing and uncoordinated data.

### Pseudocontingencies

The resulting impoverished data array, including many missing and disconnected data, nevertheless allows for PC inferences. This inference scheme does not depend on systematic joint observations of paired attributes  $X_i$  and  $Y_i$  pertaining to the same individual observations  $i$  to assess the correlation  $r(X, Y)$ . Under certain conditions to be explained below, the sign and, to some degree, the size of the correlation is inferred from information as simple and easily available as the attribute base rates  $p(X)$  and  $p(Y)$ . For example, all that is needed to infer a contingency between the person's responses and the political attitude cue is to notice that, say, there were clearly more positive than negative responses and clearly more utterances by friends than by opponents. This alignment of two skewed base rate distributions,  $p(\text{'Yes'}) > p(\text{'No'})$  and  $p(\text{friend}) > p(\text{opponent})$ , gives rise to the inference that friends tend to elicit a higher rate of yes and a lower rate of no responses than do opponents. More generally, PC inferences relate what is frequent in one aspect to what is frequent in the other aspect and, vice versa, what is infrequent in one to what is infrequent in the other aspect (Fiedler & Freytag, 2004; Fiedler et al., 2009).

### PC as a proxy for genuine contingency inferences

PC inferences afford a useful inference scheme when the demanding conditions for genuine contingency assessment are not met. Base rates can easily be assessed in two or more attributes. Even when observations in two or more attribute dimensions are incomplete and disconnected, gathered on separate occasions, the base rate trends are normally somehow discernible. Moreover, although PC inferences are logically unwarranted—granting that the more frequent  $Y$  level may be more likely given the more or less frequent  $X$  level—PCs provide a valid proxy of actual correlations most of the time. Monte Carlo simulations by Kutzner, Vogel, Freytag, and Fiedler (2011) revealed that the a posteriori probability that  $r(X, Y)$  is positive (negative) increases to the extent that the base rate distributions of high and low  $X$  and  $Y$  values are skewed in the same (different) directions.

The PC algorithm not only warrants many valid inferences, but also constitutes a fast and frugal heuristic (Fiedler et al., 2009). Given  $k$  dichotomous attributes, only  $2 \cdot k$  base rate estimates are required for PC inferences of all  $k(k-1)/2$  pairwise correlations. By comparison, a tensor with  $k(k-1)/2 \cdot 4$  cells would be necessary for genuine contingency

inferences. For  $k = 5$ , this amounts to  $10 = 2 \cdot 5$  versus  $32 = 2^5$  estimates. For  $k = 8$ , the PC parsimony is even more striking (i.e., 16 vs. 256). Thus, in terms of both memory demands and the size of necessary information samples, PCs would appear to provide a more parsimonious inference tool.

### Limitations of PC inferences

PCs are not confined to situations in which genuine contingency data are not available. They have also been shown to generalize to task settings that provide participants with simultaneously presented attribute pairs that render contingencies maximally visible. If such a conveniently presented contingency points in the direction opposite to the PC, the PC will typically override the conflicting contingency (Fiedler, 2010; Fiedler & Freytag, 2004; Kutzner, Freytag, Vogel, & Fiedler, 2008), probably because base rates are easier to discern than bivariate or multivariate joint frequencies.

An apparent limitation, to be sure, is that PCs are applicable only to unequal base rate distributions. However, even when  $X$  and  $Y$  are symmetrically distributed in the universe, the samples drawn from such a universe are often sufficiently skewed to allow for PC inferences (Kutzner et al., 2011), especially when sample size is small.  $X$  and  $Y$  samples tend to be skewed in the same direction if  $r(X, Y)$  is positive but in opposite directions if  $r(X, Y)$  is negative. This nice property of the probabilistic world increases the applicability of the PC algorithm.

### PCs reflecting categorically organized memory processes

One moderator of PC effects that motivates the present research is the number of ecologies or categories. For simplicity, we have so far referred to the case of a single category, in which the unequal base rates of high and low base rates of two attributes  $X$  and  $Y$  are aligned or not. For the case of two or more categories, the antecedent condition for PCs can be described more generally as an ecological correlation (cf. Fiedler et al., 2009; Robinson, 1950) between the two variables'  $\geq 2$  category base rates. Because the correlation over two measurement pairs is always perfect, PCs have been found to be particularly strong when there are two contrasting ecologies (Fiedler & Freytag, 2004). Such a contrast of two ecologies is easy to represent in memory, due to its simple structure. For instance, it is easy to realize that in one ecology (e.g., private life), where most utterances are by friends, the modal message is *yes*, whereas in another ecology (e.g., professional life), where most utterances are by opponents, the modal message is *no*. The example illustrates how a simplified ecological correlation can facilitate the formation of a sensible memory representation.

When the number of ecologies or categories increases, learning and discerning ecological correlations (between base rates across many ecologies) may itself become a demanding memory task that may ultimately restrict PC inferences. If, in an episodic memory task, a complex distribution of newly acquired statistical information has to be represented in memory, an inconsistent pattern of base rates for several attributes across several categories may interfere with the assessment of category base rates and, hence, the occurrence of base-rate-driven PCs. A long tradition of memory research suggests that a coherent categorical organization of memory is the key to performance on complex tasks (Mandler, 2011). When the stimulus information can be organized hierarchically and consistently, so that an extended list of specific stimulus items need not be learned at item level but can be derived from higher-order principles encoded at the level of superordinate categories, the amount of information that can be kept in memory can increase dramatically. For example, the proportion of items that can be recalled from a typical episodic-memory list can be doubled at least when stimulus list lends itself to categorical coding (Cinan, 2003; Cohen, 1966; Fiedler, 1986; Tulving & Donaldson, 1972; Yamaguchi, 1983), allowing for the construction of a consistent memory organization.

An intriguing implication of this analysis, which has not been targeted so far, is that PC inferences are not only triggered by skewed base rates in the external environment. They may also reflect the consistency of the resulting representation of base rate trends in categorically organized memory. Whenever the trends encoded at a superordinate level can be used to reconstruct specific-item information at a subordinate level, the unequal base rates that underlie these trends may trigger PC inferences. Having encoded, say, that  $p(\text{'Yes'}|\text{private life}) \sim p(\text{Friend}|\text{private life}) \sim 75\%$ , in contrast to  $p(\text{'Yes'}|\text{professional life}) \sim p(\text{Friend}|\text{professional life}) \sim 25\%$ , an ecological correlation (linking *yes* to *friends*) can be used to reconstruct the specific-item responses—provided the base rates are extracted correctly and the memory organization features a consistent structure that allows for an effortless top-down reconstruction process.

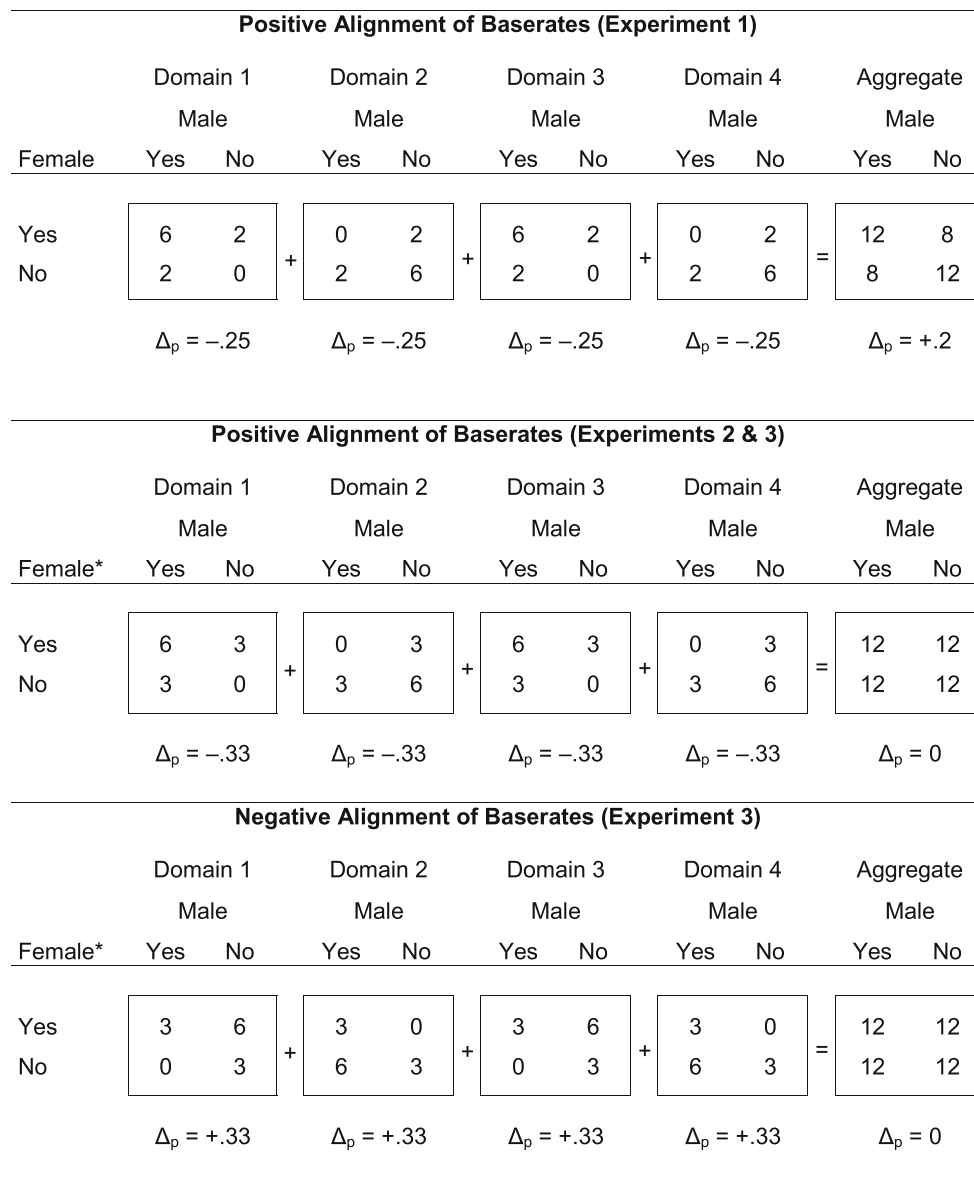
In any case, PCs necessarily entail an unwarranted inference that an ecological correlation observed at the level of category base rates also holds for specific item-level information. Even when *yes* responses correlate positively with *friends* at the ecological level (i.e., *yes* responses and *friends* tend to be jointly present or absent), they may actually correlate negatively within ecologies (e.g., within the private-life ecology, the *yes* rate may be higher for the few opponents than for the many friends). Still, a genuine PC effect would also be manifested in higher estimates of the *yes* rate for friends than for opponents *within ecologies*. Thus, rather than just reproducing the same category base rate  $p(\text{'Yes'}|\text{private life}) = 75\%$  for friends and opponents, the within-category correlation (i.e., the differential *yes* rate for friends and opponents) should be biased toward the ecological correlation.

Note that this prediction is at variance with a common finding in category-based induction research (Murphy & Ross, 2010). According to this approach, the recall and estimation of the *yes* rate should reflect the within-category correlation, beyond the mere base rates. That is, if the conjunction of *yes* and *friends* is actually less likely than *yes* and *opponents* in the private-life category, they should actually reproduce a negative rather than a positive correlation. However, this contradiction is more apparent than real. Closer inspection of the findings by Murphy, Ross, and colleagues shows that a precondition for the sensitivity to attribute conjunctions (such as *yes* and *friends*) is the formation of exemplar-based memory codes that are sensitive to item-level information beyond category-level information. Exemplar-based memories are exactly the opposite of the base-rate-driven memories that constitute the domain of PC effects. Thus, the task situations that give rise to (low-level) exemplar-based and to (high-level) base-rate-driven inferences complement and delimit each other from a metatheoretical perspective.

## Overview of the experiments

To investigate this novel variant of PC effects derived from categorically organized memory, we pitted ecological correlations against discrepant individuating correlations. Using a similar task setting as for the illustration above, participants in Experiments 1 were presented *yes* and *no* responses of a couple, Carlos and Lilly, to a list of 40 partner questionnaire items, 10 of which pertained to each of four categories (intimacy, joint activities, household, and arguing). Participants were asked to take the role of a partner therapist whose task is to assess the match or mismatch between Carlos's and Lilly's responses to the same items. Within all four categories, the correlation between his and her responses was negative (e.g.,  $r = -.25$ ); Carlos's *yes* responses were more frequent when Lilly responded *no* rather than *yes* (cf. Fig. 1). The ecological (category-level) correlation between the rates of *yes* and *no* responses provided by Carlos and Lilly across categories was perfectly positive, such that both provided mostly *yes* responses in two categories and mostly *no* responses in the other two categories. The pooled correlation across all 40 responses, disregarding categories, was close to zero.

We expected that participants will form a categorical memory organization of the 40 interview questions. Consequently, their correlation inferences and recall responses should follow the ecological correlation between the base rates of Carlos's and Lilly's *yes* (vs. *no*) responses across categories, rather than the correlations within categories (i.e., the partial correlations). Even though the two major dependent measures—recalling Carlos's and Lilly's original responses to all  $4 \times 10$  questions and predicting their responses to new items from each



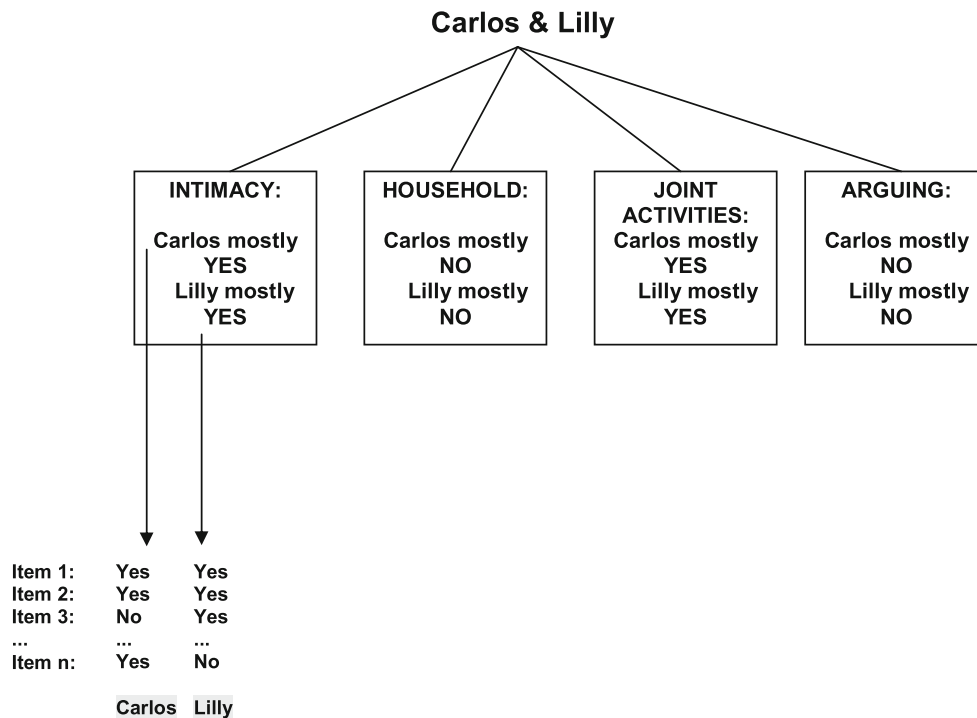
\* In Experiment 3, the actual labels were Politician X and Politician Y instead of Female and Male Partners.

**Fig. 1** Frequency distributions underlying stimulus presentation at the domain level and at the aggregate level in Experiments 1, 2, and 3

category—clearly call for item-level responses, the abstract category-level base rates should override the specific item-level information. Figure 2 illustrates this category-driven recall process. Since Carlos’s and Lilly’s base rates in the four categories exhibit the same trends (i.e., their ecological correlation is positive), the specific item responses that are derived from the categorical code should reproduce the base rates and, hence, the ecological correlation.

The expected top-down influence of category information on item-level inferences, much in line with the CAM model proposed by Duffy, Huttenlocher, Hedges, and Crawford (2010), should reflect a systematic PC strategy based on a categorical memory code. Ironically, the strength of the

erroneous tendency to reproduce and produce positively correlated responses by Carlos and Lilly should be most pronounced in participants with the strongest memory for category base rates, because base rate knowledge should trigger the reconstruction of item-level responses. Moreover, assuming that the memory code is formed at category level anyway, it should not matter if instructions explicitly ask for the correlation at category level or at item level. In either case, superordinate base rates should dominate contingency inferences. Also, it should matter little whether the four subsets of interview responses are presented blockwise or in mixed order. Finally, a manipulation of relationship quality—whether yes responses prevail in pleasant categories (intimacy, joint



**Fig. 2** Graphical illustration of the category-driven bivariate recall of Carlos’s and Lilly’s item-level responses from a categorical memory code

activities) or in unpleasant categories (household work, arguing)—should also have little influence on inferences determined by a categorical code.

An intriguing question, though, is whether the responses to individual items merely reflect a base-rate-driven response strategy (e.g., always or mostly predicting the most prevalent value per category) or whether the ecological correlation also induces the subjective expectancy of a correlation between Carlos’s and Lilly’s item responses. In Experiment 2, we therefore include distinct frequency estimates for item-level correlations. Furthermore, in Experiment 2, we used a slightly different stimulus distribution. That is, the item-level contingency in Experiment 2 was set to exactly zero in order to rule out the possibility that the effects in Experiment 1 reflected an amplification of the pooled item-level contingency. Experiment 3 served to eliminate an alternative explanation in terms of the stereotypical expectation of positive correlations in a couple. To this end, we modified the task setting such that the stimulus series no longer referred to questionnaire responses of two partners, and we manipulated the positive or negative alignment of base rates supposed to trigger positive and negative PCs. We expected this manipulation to systematically moderate the size or even the sign of the resulting PC.

**Experiment 1**

The first experiment used the stimulus distribution portrayed in the upper part of Fig. 1. In addition to the within-subjects

variation of target persons (Carlos vs. Lilly) and category reference, we manipulated three task features between participants. First, the interview items pertaining to the different categories were presented either blockwise or in a mixed order. Second, explicit instructions emphasized the need to assess the contingency either at item or at category level. Finally, the allocation of domains or categories to high versus low base rates of *yes* responses was counterbalanced. *Yes* rates were high for pleasant categories (i.e., joint activities and intimacy) and low for unpleasant categories (i.e., household and arguing) in a good-relationship condition or vice versa in a bad-relationship condition.

**Method**

*Participants and design*

Sixty-four undergraduate psychology students (32 women and 32 men; mean age = 25.47, *SD* = 6.82) were recruited for a study on clinical judgment in partial fulfillment of a course requirement. Under an equal-*n* constraint, participants were randomly assigned to the eight experimental conditions resulting from the orthogonal variation of the between-participants factors of instruction (item focus vs. scale focus), presentation order (alternating vs. blockwise), and relationship quality (good vs. bad). The within-participants factors of target gender (male vs. female) and categories (joint activities vs. household vs. intimacy vs. arguing) allowed us to calculate

various measures of contingency assessment at different levels of aggregation (see the [Results](#) section).

### Materials and procedures

Written instructions informed participants that disagreement between partners may point to potential sources of conflict in romantic relationships and that successful partner therapy depends on the accuracy with which counselors can extract the covariation of partners' experience to different aspects of life. Participants learned that their first task was to study the answers provided by a couple to a relationship inventory assessing four domains of life, with the explicit goal of monitoring agreement either at the level of individual items (itemwise condition) or at the level of domains (scalewise condition). Participants then selected one pair of envelopes (from a set of eight) containing the questionnaires filled in by the female and male partners of a couple attending partner therapy.

Each subscale of the questionnaire comprised 10 items that could be endorsed or rejected (by checking "agree" or "disagree"). In the blockwise condition, items were presented in one block per subscale, in a fixed random order (i.e., joint activities, household, intimacy, arguing). In the mixed-order condition, each subset of four consecutive items came from four different categories or subscales. Two different versions of the questionnaire were used, a self-assessment version filled in by the female partner and a partner evaluation version filled in by the male partner. The two versions differed only in the perspective. For instance, the first intimacy item read "I use to kiss him hello" in the self-assessment version, but "She uses to kiss me hello" in the partner evaluation version.

The relationship quality variable determined whether or not endorsement rates were high in pleasant (i.e., joint activities and intimacy) and low in unpleasant (i.e., household and arguing) categories, as shown in Fig. 1, or vice versa. Participants were instructed to study the female partner's responses first and to put her questionnaire back into the envelope before studying the male partner's responses. Once they had returned the questionnaires to the experimenter, participants received a new questionnaire with the dependent measures.

*Extrapolation task* The first measure involved an extrapolation task calling for inferences of the female's and male's responses (in the same format) to four new items per subscale. As in the learning phase, the female partner's form always preceded the male partner's form. The presentation order (blockwise or alternating) matched the original order.

*Cued-recall task* Upon completion of the extrapolation task, all participants were asked to reproduce the original

responses, using a copy of the stimulus questionnaire. Again, recall of the female partner's responses preceded the recall of the male partner's responses.

*Base rate estimates* As a measure of base rate sensitivity, participants were asked to estimate the base rates of endorsing responses separately for each target person and subscale (e.g., "In the domain of intimacy, how many of the 12 items had been endorsed by the female partner?"). This measure was postponed to avoid demand effects facilitating the utilization of base rates in the completion of the primary item-level dependent measures. After furnishing some demographical data, participants were debriefed, thanked, and dismissed.

### Results

#### *Base rate estimates*

The accurate extraction of skewed category base rates is crucial for the emergence of PCs. The subjective ecological correlation between the partners' endorsement base rates across the four categories affords a reasonable index of the degree to which this premise was met. Subjective ecological correlations  $r_{eco}$  were computed within participants between the base rate estimates of *yes* responses by the female and the male partners, respectively, in each of the four content categories. The high average correlations in the top row of Table 1 (between .49 and .76) testify to the accuracy with which base rates and their intercorrelations were extracted in all experimental conditions. A two-factorial analysis of variance (ANOVA) of the subjective ecological correlations with the between-participants factors presentation order (blockwise vs. mixed order) and instruction (focus on item vs. category) did not reveal any differences between conditions,  $F_s(1, 60) < 1.7$ .

#### *Extrapolating inferences*

The correlation between the participants' extrapolating inferences of the female and male partners' responses to new questionnaire items provide a measure of the subjectively inferred contingency. First, subjective ecological correlations were computed within participants between the base rates of *yes* responses generated for the female and the male partners, respectively, in each of the four content categories. As is evident from the second row in Table 1, the categorical encoding of base rate information was clearly manifested in strong ecological correlations (ranging from .48 to .76).

Second, and crucial for a test of the hypothesis that ecological correlations affect individuating correlations, we examined the correlation between extrapolating inferences about the female and male partners at the level of individual

**Table 1** Mean ecological correlations ( $r_{eco}$ ), itemwise correlations pooled across categories ( $r_{item}$ ), and partial correlations within categories ( $r_{partial}$ ) derived from recall and prediction of partners’ responses in the four subscales of the partnership questionnaire (Experiment 1)

Contingency Measure	Objective	Presentation Condition			
		Mixed Instruction Focus on		Blockwise Instruction Focus on	
		Items	Categories	Items	Categories
Base rate estimates					
$r_{eco}$	1.0	.49 (.61)	.68 (.46)	.76 (.37)	.70 (.36)
Extrapolation task					
$r_{eco}$	1.0	.67 (.49)	.56 (.54)	.76 (.32)	.48 (.60)
$r_{item}$	.20	.38 (.47)	.52 (.21)	.53 (.21)	.48 (.36)
$r_{partial}$	-.25	.47 (.29)	.43 (.39)	.42 (.43)	.58 (.26)
Cued recall					
$r_{eco}$	1.0	.79 (.50)	.90 (.12)	.84 (.25)	.88 (.21)
$r_{item}$	.20	.48 (.16)	.55 (.19)	.41 (.25)	.46 (.34)
$r_{partial}$	-.25	.29 (.38)	.22 (.26)	.16 (.21)	.15 (.21)
Cued-recall accuracy					
$p(\text{correct})$ overall		.80 (.09)	.77(.05)	.79 (.10)	.78 (.12)

Note. Numbers in parentheses are standard deviations.

items. Recall that the objective correlation was +.20 across categories and -.25 within categories.

As was predicted, the item-level contingency between extrapolating inferences exceeded objective values in all conditions. This is evident from a series of *t*-tests of the item-level correlations  $r_{item}$  against the objective values. Pooling over categories,  $r_{item}$  exceeded the objective value of .20 for each experimental condition [i.e., blockwise/item focus, mean  $r_{item}$  = .53,  $t(15) = 6.32, p < .001$ ; blockwise/category focus, .48,  $t(15) = 3.08, p = .004$ ; alternating/item focus, .38,  $t(15) = 1.50, p = .077$ ; alternating/category focus, .52,  $t(15) = 4.15, p < .001$ ].<sup>1</sup> A two-factorial ANOVA of the pooled  $r_{item}$  values yielded no significant influence of conditions,  $F_s(1, 60) < 1.20$ .

The mean within-category  $r_{partial}$  values also exceeded the objective value of -.25, and they were even reliably higher than zero (see Table 1). This PC effect was significant not only across all 64 participants, mean  $r_{partial}$  = .48,  $SD = .35, t(63) = 10.91, p < .001$ , but also within each of the four experimental groups (with mean  $r_{partial}$  values ranging from .42 to .58, all  $t_s > 3.9$ ). No significant differences between conditions emerged in an ANOVA (all  $F_s < 1.3$ ).

*Cued recall*

To substantiate the claim that subjective contingency inferences are rooted in the memory representation of the original stimulus list, we computed analogous correlations from

participants’ cued-recall responses. As is apparent from Table 1, the ecological correlations  $r_{eco}$  between the base rates of the cued-recall responses for Carlos and Lilly were again very high (condition-wise averages ranging from .79 to .90) across all conditions (all  $F_s < 1$ ).

Turning to item-level contingencies, the pooled  $r_{item}$  scores computed across all categories ranged from .41 to .55 in the four experimental conditions. No reliable differences between conditions were found in an ANOVA, all  $F_s(1, 60) < 1.67$ . Moreover, a series of *t*-tests against the objective value of .20, conducted separately for each experimental condition, revealed that the item-level  $r_{item}$  values reproduced in the cued-recall task (means ranging from .41 to .55; cf. Table 1) were reliably higher than the objective values, all  $t_s > 3.11, p_s < .01$ .

Similarly, the mean within-category  $r_{partial}$  values exceeded the objective value of -.25. Across conditions,  $r_{partial}$  scores ranged from .15 to .29 (all  $t_s > 2.9$ ), with the grand mean being .21,  $t(63) = 13.23, p < .001$ , for a test against -.25. Moreover, the more conservative test against zero revealed that even the sign of the  $r_{partial}$  scores followed the PC logic,  $t(63) = 5.95, p < .001$ .

*Recall accuracy*

The strong PC bias did not prevent cued-recall responses from being generally accurate. The average rate of recall responses matching the original item responses lay between .77 and .80 across all categories (cf. Table 1). The ANOVA did not reveal any differences between conditions (all  $F_s < 1$ ). Individual differences in recall accuracy correlated strongly with

<sup>1</sup> Since hypotheses were directed, all *t*-tests and correlations reported in this article were one-tailed.

individual differences in the (correctly) extracted ecological correlation between Carlos's and Lilly's average responses to the items of the four categories,  $r_{\text{eco}}(62) = .74, p < .001$ , consistent with the assumption that sensitivity to category base rates is the key to good memory for categorically organized lists.

Moreover, consistent with the notion that category-level memory codes trigger the production of PC effects at a specific item level, individual differences in ecological correlations (between average recall responses to the four categories) correlated positively with individual differences in PC biases in cued recall,  $r(62) = .34, p = .003$ , and in the extrapolation task  $r(62) = .52, p < .001$ .

As a consequence of the double relation between (veridical) ecological correlations to recall accuracy and (illusory) PC effects, the latter two variables tended to correlate positively. High recall accuracy did not eliminate or decrease PC biases but, if anything, led to increased PC effects in cued recall,  $r = .10$ , n.s., and in extrapolating inferences,  $r(62) = .31, p = .007$ .

## Discussion

Altogether, these results corroborate the contention that categorical memory codes induce PC inferences in complex task settings involving multiple categories. When it is unlikely or impossible to encode all specific item information, an adaptive cognitive strategy is to encode, at a superordinate level, existing trends in category base rates, such as “Carlos tends to say ‘Yes’ to housework items” or “Lilly tends to say ‘No’ to joint activity items.” These base rate trends are captured easily, and the ecological correlations between Carlos's and Lilly's category base rates are recognized accurately.

The ecological correlations between category base rates can then be used to reconstruct the specific item-level information that is nested within categories. Thus, to recall Carlos's and Lilly's responses to all 40 items and to predict their responses to new items, participants can use their base rate knowledge for adaptive response strategies (cf. Fig. 2). Applying a maximizing strategy (i.e., inferring the more prevalent response per category for all items) or a probability-matching strategy (i.e., inferring each response at a rate that matches its base rate), they manage to reconstruct information for which no specific memory traces are available. As a result of this category-driven process, the implicitly reconstructed correlation between Carlos's and Lilly's itemwise responses is biased toward the ecological correlations.

Consistent with this explanation, the reproductions and predictions of both partners' item responses were clearly more positively correlated than the actually observed correlation across all categories and clearly less negatively correlated than the actually presented negative (partial) correlation within categories. The strength of this (nonveridical) PC illusion is positively related to individual differences in the (veridical)

extraction of ecological correlations between Carlos's and Lilly's category base rates. This pattern corroborates the interpretation that PC illusions arise as a natural product of categorical memory codes.

However, if this interpretation is correct, it raises an intriguing question about the nature of the underlying cognitive process: Is the PC illusion just the result of two separately applied response strategies (one for Carlos and one for Lilly), without any cognitive inference about their correlation? Or does the PC effect entail an inference from an ecological to an individuating correlation? To illustrate this point, assume that *yes* and *no* responses occur at rates of a .8 and .2, respectively. A probability-matching strategy applied to both Carlos and Lilly would thus yield joint probabilities of  $.8 \cdot .8 = .64$  ‘Yes’ $\wedge$ ‘Yes’,  $.8 \cdot .2 = .16$  ‘Yes’ $\wedge$ ‘No’,  $.2 \cdot .8 = .16$  ‘No’ $\wedge$ ‘Yes’, and  $.2 \cdot .2 = .04$  ‘No’ $\wedge$ ‘No’ responses. Because each partner's *yes* prevalence is the same given *yes* and *no* responses of the other (i.e., because the  $.64 / .16 = .16 / .04$ ), the resulting within-category correlation is zero. However, because *yes* rates and joint *yes* rates are high in some and low in other categories, the same probability-matching strategy also predicts a positive pooled correlation (of  $+.48$ ) across all items from all categories.

Thus, blind probability matching alone, applied separately to Carlos and Lilly, may account for the positive item-level correlations obtained in Experiment 1 across all categories. It is therefore unclear whether participants actually inferred a positive correlation between Carlos's and Lilly's item-level responses from the positive ecological correlation at category level, or whether they merely applied a twofold probability-matching strategy. To gain more substantive evidence for a carryover of ecological to individuating correlations, we therefore included an explicit measure of subjective item-level correlations in a second experiment, in addition to the implicit measures entailed in the cued-recall and extrapolation tasks.

It would also be important to rule out an alternative account in terms of the pooled item correlation, which was slightly positive ( $\Delta_p = .20$ ) in Experiment 1, leaving open the possibility that the causal origin of positive correlation inferences was this pooled item correlation rather than the ecological correlation. Finally, the blockwise presentation of items from the four categories may have helped participants to detect the contingencies in the stimulus series.

## Experiment 2

In our attempt to deal with these problems, we based a second experiment on a refined procedure. In particular, we modified the stimulus distribution such that the pooled correlation across all 48 items was exactly 0, while the (partial) correlation within categories attained an even more negative value of  $-.33$ . We manipulated the ease with which



participants could coordinate both partners' responses on the extrapolation task, in that inferences of male and female responses were made simultaneously on the same form or successively on separate forms (as in Experiment 1). The reliability of the extrapolation measure was increased by including twice as many items as in Experiment 1. We included an explicit measure of subjective item-level correlations by asking participants to estimate the proportion of congruent and incongruent responses by Carlos and Lilly. Finally, participants were asked to indicate their use of several response strategies while working on the indirect measures.

## Method

### *Participants and design*

Forty undergraduate psychology students (31 women and 9 men; mean age = 23.25,  $SD = 5.82$ ) were recruited for a study on clinical judgment in partial fulfillment of a course requirement. Under an equal  $n$  constraint, participants were randomly assigned to one of the four conditions resulting from the orthogonal variation of the between-participants variables extrapolation task (simultaneous vs. successive) and relationship quality (good vs. bad). As in Experiment 1, several contingency measures were administered within participants.

### *Materials and procedure*

The same general procedure was used as in Experiment 1, with a few exceptions. First, the pooled item-level correlation across all categories was set to zero, and the within-category correlations were set to  $-.33$ . This was accomplished by changing the within-category distributions from 6 2 2 0 to 6 3 3 0 (see the middle panel of Fig. 1). Note in passing that the resulting skewness of base rates was slightly weaker, which might lead to less pronounced PCs, as compared with Experiment 1.

Second, the number of items per domain in the extrapolation task was increased to eight, and the extrapolation task for Lilly and Carlos was administered either simultaneously within the same form or successively in separate forms.

Third, participants provided conditional percentage estimates for the likelihood of Carlos endorsing items Lilly had endorsed and for the likelihood of Carlos endorsing items Lilly had denied. These percentage estimates correspond to the proportion of observations falling in the left cell of the upper and lower rows, respectively, in the  $2 \times 2$  table of a given domain. Subtracting the latter estimate from the former thus yields an analog of  $\Delta_p$ .

Finally, participants completed several ratings assessing the utilization of different strategies when working on the

indirect contingency measures. Using 6-point rating scales from 1 (*completely incorrect*) to 6 (*completely correct*), participants indicated the degree to which they believed they had used the following four strategies: (1) base rate (reproduction of the base rate for partner and domain; "I checked the 'agree' option at a rate corresponding to the proportion of agreeing answers provided by a partner in a given domain"), (2) convergence (reproduction of convergent answers per domain; "I checked the agree option for both partners at a rate corresponding to the proportion of converging answers in a given domain"), (3) mode (invariant utilization of the modal value for each partner and domain; "I always checked the option corresponding to the more prevalent answer provided by a partner in a given domain"), and (4) alignment (pseudocontingent alignment of answers within each domain; "I checked the same option for both partners if they had shown the same response tendency in a domain").

## Results

### *Base rate estimates*

Subjective ecological correlations  $r_{eco}$  were computed within participants between the base rate estimates for *yes* responses by the female and the male partners, respectively, in each of the four categories. One participant in the simultaneous/bad condition was removed from the data set because of complete insensitivity to base rate information. The remaining participants again extracted the base rates accurately (see top row of Table 2). An extrapolation task forms  $\times$  relationship quality ANOVA of the  $r_{eco}$  scores revealed a marginally significant main effect of relationship quality,  $F(1, 38) = 3.21$ ,  $p = .081$ , but no other effects. Table 2 reveals that the subjective ecological correlations of the four pairs of category base rates were slightly reduced in the simultaneous/bad condition. However, the average  $r_{eco}$  of .83 indicates that the covariation of base rates was clearly noticed.

### *Extrapolating inferences*

The correlations computed from the extrapolating inferences corroborate this conclusion. These  $r_{eco}$  scores were again computed within participants. Category base rates of *yes* responses to the eight new items per category were highly correlated (simultaneous/good,  $r_{eco} = .87$ ; simultaneous/bad,  $r_{eco} = .57$ ; successive/good,  $r_{eco} = .83$ ; successive/bad,  $r_{eco} = .46$ ). In the ANOVA, only a relationship quality main effect emerged,  $F(1, 38) = 12.90$ ,  $p < .001$  (all other  $F$ s  $< 1$ ).

At item level, the contingency  $r_{item}$  between extrapolating inferences for the female and male partners was computed pooling across categories. An extrapolation task  $\times$  relationship quality ANOVA yielded only a relationship quality main

**Table 2** Mean ecological correlations ( $r_{\text{eco}}$ ), itemwise correlations pooled across categories ( $r_{\text{item}}$ ), and partial correlations within categories ( $r_{\text{partial}}$ ) derived from recall and prediction of partners' responses in the four subscales of the partnership questionnaire (Experiment 2)

Contingency Measure	Objective	Extrapolation Task Form			
		Simultaneous Relationship Quality		Successive Relationship Quality	
		Good	Bad	Good	Bad
Base rate estimates					
$r_{\text{eco}}$	1.0	.93 (.13)	.83 (.18)	.96 (.06)	.92 (.08)
Extrapolation task					
$r_{\text{eco}}$	1.0	.87 (.10)	.57 (.28)	.83 (.17)	.46 (.47)
$r_{\text{item}}$	.00	.39 (.15)	.26 (.24)	.38 (.20)	.17 (.28)
$r_{\text{partial}}$	-.33	.03 (.18)	.16 (.31)	.03 (.28)	.09 (.28)
Cued recall					
$r_{\text{eco}}$	1.0	.95 (.05)	.94 (.08)	.96 (.04)	.92 (.10)
$r_{\text{item}}$	.00	.26 (.15)	.25 (.16)	.22 (.19)	.23 (.11)
$r_{\text{partial}}$	-.33	-.03 (.21)	.02 (.23)	-.06 (.22)	-.01 (.19)
Contingency based on conditional Probability estimates					
	.00	.30 (.28)	.27 (.28)	.32 (.36)	.31 (.19)
Cued-recall accuracy					
$p(\text{correct})$		.77 (.11)	.71 (.21)	.75 (.17)	.72 (.17)

Note. Numbers in parentheses are standard deviations.

effect,  $F(1, 38) = 5.38, p = .026$ . The pooled contingency  $r_{\text{item}}$  was reduced in the bad-relationship condition (cf. Table 2) but still exceeded the objective value of zero [good,  $t(19) = 10.12, p < .001$ ; bad,  $t(18) = 3.59, p = .001$ ].

The mean within-category  $r_{\text{partial}}$  values (ranging between .03 and .16) reliably exceeded the objective correlation of  $-.33$  (all  $t_s > 4.08, p_s < .01$ ), with the grand mean  $r_{\text{partial}}$  being .07 ( $SD = .26$ ). An extrapolation task  $\times$  relationship quality ANOVA of  $r_{\text{partial}}$  values revealed no significant effects, all  $F_s < 1.28, n.s.$

#### Cued recall

Ecological correlations  $r_{\text{eco}}$  were again computed within participants between the base rates of *yes* responses generated in cued recall for the female and the male partners, respectively, in each of the four content categories. The category base rates of recalled female and male responses correlated highly positively, reflecting accurate utilization of the existing ecological correlation (simultaneous/good,  $r_{\text{eco}} = .95$ ; simultaneous/bad,  $r_{\text{eco}} = .94$ ; successive/good,  $r_{\text{eco}} = .96$ ; successive/bad,  $r_{\text{eco}} = .92$ ). In an extrapolation task  $\times$  relationship quality ANOVA, no significant effects emerged, all  $F_s < 1.53, n.s.$

Although the objective item-level correlation across all categories was exactly zero, the subjective correlations of the cued-recall responses for the female and male partners were significantly positive, all  $t_s > 3.75, p_s < .01$ . Average

$r_{\text{item}}$  scores ranged from .22 to .26 (see Table 2). The lack of significant effects in an extrapolation task  $\times$  relationship quality ANOVA suggests a generalized, similarly strong PC effect in all conditions, all  $F_s < 1, n.s.$

As is evident from Table 2, the mean within-category  $r_{\text{partial}}$  values reliably exceeded the objective value of  $-.33$ , all  $t_s > 3.93, p_s < .01$ . An extrapolation task  $\times$  relationship quality ANOVA of the within-category  $r_{\text{partial}}$  values revealed no significant effects, all  $F_s < 1, n.s.$

#### Direct contingency estimates

Consistent with the extrapolating inferences and cued-recall responses, the mean item-level contingencies derived from conditional probability estimates per category were clearly positive (see Table 2). Participants erroneously rated the rate of Carlos's *yes* responses to be higher when Lilly's responses were also *yes*, rather than *no*. Significant *t*-test results were obtained when the explicit contingency estimates in the four experimental conditions were tested against  $-.33$ , all  $t_s > 5.53, p_s < .01$ , or against zero, all  $t_s > 2.70, p_s < .05$ . All ANOVA comparisons between conditions were negligible, all  $F_s < 1$ .

#### Recall accuracy

Recall accuracy (i.e., overall proportions of correct reproductions) was moderately high, varying between .71 and

.77. An ANOVA did not reveal any differences between conditions, all  $F_s < 1$ .

**Strategy endorsement** The four self-attributed strategy attributions (i.e., base rate vs. convergence vs. mode vs. alignment) were treated as a repeated measures factor in a 2 (extrapolation task form)  $\times$  2 (relationship quality)  $\times$  4 (strategies) ANOVA. The only significant result was a strategies main effect,  $F(3, 105) = 6.29, p < .001$ . Self-reports were concentrated on the two strategies that entailed the utilization of base rate information, base rate ( $M = 3.59, SD = 1.46$ ) and alignment ( $M = 3.21, SD = 1.36$ ), as compared with markedly lower ratings for base-rate-independent strategies, convergence ( $M = 2.62, SD = 1.16$ ) and mode ( $M = 2.72, SD = 1.47$ ). Thus, the base rate influence was visible not only in indirect and direct contingency measures, but also in introspective strategy reports.

## Discussion

Experiment 2 replicated the results and corroborates the conclusions drawn from Experiment 1. Even when the overall correlation between Carlos's and Lilly's responses was exactly zero and when the within-category correlations were even more negative ( $-.33$ ) than in Experiment 1 ( $-.25$ ), participants inferred positive correlations at all levels. They not only correctly recognized the ecological correlation between Carlos's and Lilly's response rates per category. They also incorrectly inferred positive item-level correlations in the pooled analysis across items from all categories and nonnegative correlations within categories. These persistent findings did not depend on the presentation format of the extrapolation task. The self-reported strategies corroborated the base-rate-driven recall and judgment process.

Explicit percentage estimates of one partner's *yes* responses given different responses by the other partner allowed us to derive a measure of subjective contingency at the level of items within categories. This measure enabled us to test whether the positive PC inferences came along with the subjective experience of positive item-level correlations, rather than merely reflecting a twofold base-rate-driven response strategy. Indeed, estimates of Carlos's endorsement rates were higher for items that Lilly had endorsed than for items that Lilly had denied. This cannot be due to participants' applying the same probability-matching or maximizing strategy to inferences about Carlos and Lilly across categories.

Altogether, these findings provide consistent support for PC inferences derived from categorical memory codes. However, before we can draw any final conclusions, we need to rule out an obvious and very simple explanation in

terms of stereotypical expectancies. In both experiments reported so far, the ecological correlation was positive, with both Carlos and Lilly responding predominantly *yes* or *no* in the same categories, and the PC illusion consisted in the erroneous tendency to recall and predict Carlos's and Lilly's item-level responses as if they were also positively correlated. This illusory inference may, of course, be facilitated by the stereotypical expectation that responses of couples are positively correlated. We therefore made an additional attempt to demonstrate the independence of a PC of such a partnership stereotype.

## Experiment 3

Indeed, we ran another experiment using the same questionnaire paradigm and stimulus distributions, but in a way that PCs could no longer be due to partnership stereotypes. For this purpose, we introduced two changes in the experimental task. First, we modified the nature and contents of the two sets of questionnaire responses. Participants saw the positive versus negative responses of two politicians, X and Y, to 12 political opinion topics in each of four categories (education, cultural integration, security, environment). Thus, in Experiment 3, the reconstruction of a positive versus negative contingency between the two sets of responses cannot be driven by partnership-stereotypical expectancies.

Second, a manipulation of the positive versus negative alignment of the category base rates should affect the sign and size of the expected PC effects. By inverting the binary values of one of the two response sets, the base rates (of *yes* vs. *no* responses) were either positively aligned, as in Experiments 1 and 2 (i.e., the same response option was frequently chosen in both vectors), or negatively aligned (i.e., opposite response options were frequently chosen). Within categories, the (partial) correlation was always opposite to the ecological correlation. That is, item-level responses were negatively correlated ( $\Delta_p = -.33$ ) in the positive alignment condition, as before (see middle part of Fig. 1), but positively correlated in the negative alignment condition ( $\Delta_p = +.33$ ; see bottom part of Fig. 1). Pooling across categories, the item-level correlation was always zero.

Although prior research has shown positively and negatively aligned base rates to result in opposite PC effects (cf. Fiedler et al., 2009), we did not necessarily expect equally strong illusions for both conditions of the present experiments. The overall consistency should be higher and the inductive inference task should be easier when the dominant response per category is always the same than when two modal responses per category are inconsistent. It may thus be easier to reconstruct the same item-level responses for both response sets from a category with a single modal response than to reconstruct divergent responses from a

categorical code that entails divergent modes for different response sets. Note that such an asymmetry would be independent of any stereotypical expectancy. However, in any case, we expected the alignment manipulation to exert a significant influence on the size of the ecological correlation, regardless of whether positive and negative PCs were equally strong.

Method

Participants and design

Forty-eight students (36 female, 11 male, 1 undisclosed; mean age = 23.42, *SD* = 3.99) participated in Experiment 3. They were randomly allocated to one of two experimental groups representing the positive and negative PC conditions. Within participants, we analyzed the same measures of ecological and item-level correlations as in the first two experiments. Only the recall task and direct contingency estimates, but no extrapolating inferences to new items, were included.

Materials and procedure

Participants were asked to become familiar with how two politicians, X and Y, responded to a politics questionnaire. They were provided with two politicians’ responses to various items grouped by content (i.e., cultural integration, education, environment, security). For instance, politician X could have indicated his agreement or disagreement with statements such as “The state should invest more money in renewable energies” or “All pupils should have free access to learning materials (e.g., books).”

The same stimulus distribution was used as in Experiment 2 (see bottom part of Fig. 1), except for the following modifications. In the negative alignment condition, responses were inverted for one of the two politicians (either X or Y, depending on the counterbalancing condition).<sup>2</sup> Dependent measures were the same as in Experiment 2, but the extrapolation task was omitted. The presentation format was identical to the successive presentation mode in Experiment 2. However, for a more conservative test of participants’ spontaneous use of category information, the categories were given no explicit labels. Category labels were therefore also missing from the cued-recall task. Also, items were presented in a previously drawn random succession, instead of a regular order.

<sup>2</sup> Between participants, we counter-balanced in which domains politician X (politician Y, respectively) predominantly responded *yes*. Data are collapsed across this factor because none of the terms including this factor reached conventional levels of significance.

Results and discussion

Base rate estimates

Let us again first consider the ecological correlations  $r_{eco}$  computed from the base rate estimates of both variables across the four categories. The mean  $r_{eco}$  for the positive and negative alignment conditions (.27 and  $-.28$ ) differed significantly from each other,  $t(44) = 2.93, p = .003$ , and each one differed from zero,  $t(23) = 2.02, p = .028$ , and  $t(21) = -2.15, p = .022$ , respectively (cf. Table 3).

PC effects in cued-recall responses

Ecological correlations that were extracted from the cued-recall responses also differed between alignment conditions (.58 vs.  $-.72$ ),  $t(43) = 7.91, p < .001$ , which is in support of the notion that participants exploit category base rates when reproducing item-level responses.

The individual participants’ item-level correlations  $r_{item}$  of recalled responses for politicians X and Y were highly correlated with individual variation in  $r_{eco}$  ( $r = .76, p < .001$ ). It is thus not surprising that the mean  $r_{item}$  scores for the positive and negative alignment groups (.08 vs.  $-.19$ ) also differed between conditions,  $t(44) = 4.81, p < .001$ . Separate *t*-tests against zero indicate that the alignment manipulation determined the sign, but the size of the effect was weaker in the

**Table 3** Mean ecological correlations ( $r_{eco}$ ), itemwise correlations pooled across categories ( $r_{item}$ ), and partial correlations within categories ( $r_{partial}$ ) derived from recall of politicians’ responses in the four subscales of the politics questionnaire (Experiment 3)

Contingency Measure	PC Condition			
	Positive Alignment		Negative Alignment	
	Objective	Observed	Objective	Observed
Base rate estimates				
$r_{eco}$	1.0	.27 (.66)	-1.0	-.28 (.62)
Extrapolation Task				
$r_{eco}$	1.0	.58 (.59)	-1.0	-.73 (.51)
$r_{item}$	.0	.08 (.21)	.0	-.19 (.18)
$r_{partial}$	-.33	-.05 (.17)	.33	-.00 (.17)
Contingency based on estimates	.0	.14 (.36)	.0	-.20 (.26)
Cued-recall accuracy				
$p(\text{correct})$ overall		.67 (.12)		.65 (.12)

*Note.* True partial correlations were  $-.33$  in the positive alignment condition, but  $+.33$  in the negative alignment condition. Thus, the objective partial correlations had different signs than the ecological correlations in both conditions. Numbers in parentheses are standard deviations.

positive,  $t(23) = 1.83, p = .040$ , than in the negative,  $t(23) = -5.22, p < .001$ , alignment group.

The analysis of the within-category  $r_{\text{partial}}$  scores in Experiment 3 revealed that the mean  $r_{\text{partial}}$  score for the positive ( $-.05$ ) and negative ( $-.00$ ) alignment conditions did not differ significantly,  $t(46) = -1.14, p = .130$ . However, recall that the objective values were  $-.33$  in the positive alignment condition, but  $+.33$  in the negative alignment condition. Indeed,  $r_{\text{partial}}$  clearly deviated from the objective partial contingency ( $-.33$  vs.  $+.33$ ), as expected by the PC logic [ $t(23) = 7.89, p < .001$ , for positive alignment, and  $t(23) = -9.49, p < .001$ , for negative alignment].

Also, we computed the proportion of correct responses as an indicator of accuracy. This index did not differ between alignment conditions ( $.67$  and  $.65$ ),  $t < 1$ , n.s. Interestingly, accuracy was higher for participants who reconstructed the ecological correlation, as is evident from the between-participants correlation with  $r_{\text{eco}}$ . That is, higher accuracy went together with more positive ecological correlations in the positive-alignment condition,  $r = .48, p = .008$ , but with more negative ecological correlations in the negative-alignment condition,  $r = -.43, p = .026$ . Furthermore, higher recall accuracy did not prevent biased contingency scores. If anything, accuracy predicted biases in  $r_{\text{item}}$  in the positive alignment condition,  $r = .19$ , as well as the negative alignment condition,  $r = -.21$ , although these trends did not reach conventional levels of significance,  $ps > .05$ .

#### Direct contingency estimates

Let us finally consider the direct contingency estimates at item level. On this measure, positive and negative PC effects appear as positive versus negative difference scores, respectively. Direct contingency estimates were not only higher in the positive than in the negative alignment group ( $.14$  vs.  $-.20$ ),  $t(45) = 3.69, p < .001$ . A  $t$ -test against zero was significant for positive alignment,  $t(22) = 1.88, p = .037$ . Moreover, contingency estimates in the negative alignment condition were significantly lower than zero,  $t(23) = -3.69, p < .001$ .

Altogether, the findings from Experiment 3 further corroborate the notion of categorical memory as a source of contingency inferences. Findings in the positive alignment condition replicate the results of Experiments 1 and 2, demonstrating that the previous results cannot be due to the stereotypical expectation that closely related partners provide similar questionnaire responses. Furthermore—and clearly supporting the PC hypothesis—the between-participants manipulation of the ecological correlation exerted a significant influence: Across measures and experiments, positively aligned base rates led to more positive contingency estimates than did negatively aligned base rates. Although the results obtained in the experiment were not perfectly symmetric—PCs' effects were weaker in the positive than in the negative alignment condition—these asymmetries

merely suggest that peculiarities of stimulus contents may also affect the reproduction of questionnaire responses. However, taken together, the experiment provides substantial support to the PC account.

#### General discussion

Going beyond previous findings that PC inferences are often correct and more feasible in reality than are genuine contingency assessment, the present research sought to highlight another reason for the PC rationale. To the extent that higher-order memory codes are used for the economic encoding and storage of complex and extensive stimulus information, the reconstruction of item-level information from abstract category codes will be naturally determined by the category base rates, thus leading to PC inferences.

On the basis of a single conversation with a stranger, we easily find out that he or she has many hobbies related to arts but hardly any hobbies in sports, even when we cannot recollect that person's responses to all hobbies discussed (Fiedler, 1986). As a consequence, we rate this person high in arts and low in sports with high confidence. When asked to guess individual items under uncertainty, we will produce a high endorsement rate in arts and a low rate in sports, even though many individual guesses will be wrong. This reconstructive process is well represented in the literature on categorized word lists and memory organization (Cohen, 1966; Tulving & Donaldson, 1972).

When this framework is expanded to cover memory for bivariate or multivariate information, it is but one step further to predict that inferences from category base rates for several attributes ( $X, Y$ , etc.) will lead to PC illusions. If the relative rates of positive  $X$  and  $Y$  values in different categories covary in the same direction, exhibiting a positive ecological correlation, inferences about specific  $X$  and  $Y$  items will also be positively correlated. If the ecological correlation at category level is negative, item-level  $X$  and  $Y$  inferences will be negatively correlated. Thus, independently of the fact that the external stimulus and feedback environment rarely provides us with joint multivariate information, and independently of the unrealistic memory demands of genuine contingency algorithms, the very organization of memory may impose a structure on our world knowledge that strongly fosters PC inferences.

We do not want to construct a theoretical conflict between category-based PC inferences and exemplar-based memory processes (Juslin, Jones, Olsson, & Winman, 2003; Nosofsky & Johansen, 2000). Categorical codes and exemplar codes do not contradict but complement each other. It is appropriate to say, though, that the domain of exemplar-based memory delimits the domain on PC inferences. Future research will have to demarcate these domains and to figure out the precise conditions under which

memory representations specify or abstract from individual stimuli. In the context of contingency assessment, when the focus is on the rule between two variables rather than the peculiarities of individual observations, we suspect that exemplar codes will often be of minor importance—for example, when assessing the contingency between weather and well-being, income and health, conditional and unconditional stimuli, or prime meaning and target meaning in a conditioning experiment.

Category-level summary information is more likely to be reliable and to entail less noise and less empty design cells or missing data than are specific stimulus items. Moreover, categories are not only useful and reliable for diagnostic inferences (Duffy et al., 2010); categories of a reasonably high level of abstractness also represent a natural level for interventions. Affirmative action is not meant to help individual persons but disadvantaged ethnic or societal groups. Education aims at improving performance in certain areas, rather than correct responses to specific items. Or, with respect to Carlos and Lilly, the purpose of partner therapy is to induce mutual understanding in domains such as joint activities, household, intimacy, or arguing.

Reliance on information about superordinate categories can be highly functional and normatively sound (Duffy et al., 2010; Olivola & Todorov, 2010). The PC heuristic exploits these assets of base rate information, and it can be expected to be particularly useful in complex multivariate situations (Fiedler, 2010), when the environmental input is replete with missing data (White & Koehler, 2004), and when contingencies are changing dynamically (Speekenbrink & Shanks, 2010). When these boundary conditions are met, PC inferences can have a strong impact on judgments and behavior, and PC illusions can account for stereotyping, superstition, and similar everyday illusions.

**Authors' Note** Tobias Vogel, University of Mannheim, Parkring 47, 68159 Mannheim, Germany; Peter Freytag, Florian Kutzner, and Klaus Fiedler, Department of Psychology, University of Heidelberg, Hauptstrasse 47-51, 69117 Heidelberg, Germany. The research reported in this article was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG) to the second and the last authors. The authors are grateful to Oliver Arnold, Hannah Helfrich, Almuth Randoll, Julica Rang, Stephanie Tremmel, and Sarah Weisang for their assistance in conducting this study.

## References

- Brunswik, E. (1952). *The conceptual framework of psychology*. *International encyclopedia of unified science*. Chicago: University of Chicago Press.
- Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, *99*, 365–382.
- Cinan, S. (2003). Executive processing in free recall of categorized lists. *Learning and Motivation*, *34*, 240–261.
- Cohen, B. H. (1966). Some-or-none characteristics of coding behavior. *Journal of Verbal Learning & Verbal Behavior*, *5*, 182–187.
- Duffy, S., Huttenlocher, J., Hedges, L. V., & Crawford, L. (2010). Category effects on stimulus estimation: Shifting and skewed frequency distributions. *Psychonomic Bulletin & Review*, *17*, 224–230.
- Fiedler, K. (1986). Person memory and person judgments based on categorically organized information. *Acta Psychologica*, *61*, 117–135.
- Fiedler, K. (2010). Pseudocontingencies can override genuine contingencies between multiple cues. *Psychonomic Bulletin & Review*, *17*, 504–509.
- Fiedler, K., & Freytag, P. (2004). Pseudocontingencies. *Journal of Personality and Social Psychology*, *87*, 453–467.
- Fiedler, K., Freytag, P., & Meiser, T. (2009). Pseudocontingencies: An integrative account of an intriguing cognitive illusion. *Psychological Review*, *116*, 187–206.
- Freytag, P., Bluemke, M., & Fiedler, K. (2011). An adaptive-learning approach to affect regulation: Strategic influences on evaluative priming. *Cognition and Emotion*, *25*, 426–439.
- Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In G. Gigerenzer & P. M. Todd (Eds.), *Simple heuristics that make us smart* (pp. 3–34). New York, NY US: Oxford University Press.
- Juslin, P., Jones, S., Olsson, H., & Winman, A. (2003). Cue abstraction and exemplar memory in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 924–941.
- 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*, 23–32.
- Kutzner, F., Vogel, T., Freytag, P., & Fiedler, K. (2011). Contingency inferences driven by baserates: Valid by sampling. *Judgment and Decision Making*, *6*, 211–221.
- Lagnado, D. A., Newell, B. R., Kahan, S., & Shanks, D. R. (2006). Insight and strategy in multiple cue learning. *Journal of Experimental Psychology: General*, *135*, 162–183.
- Mandler, G. (2011). From association to organization. *Current Directions in Psychological Science*, *20*, 232–235.
- McCausley, C., & Stitt, C. L. (1978). An individual and quantitative measure of stereotypes. *Journal of Personality and Social Psychology*, *36*, 929–940.
- Murphy, R. A., & Baker, A. G. (2004). A role for CS-US contingency in Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *30*, 229–239.
- Murphy, G. L., & Ross, B. H. (2010). Category vs. object knowledge in category-based induction. *Journal of Memory and Language*, *63*(1), 1–17.
- Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of “multiple-system” phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, *7*, 375–402.
- Olivola, C. Y., & Todorov, A. (2010). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology*, *46*, 315–324.

- Richardson, K. K., & Bhavnani, K. K. (1984). How a concept is formed: Prototype or contingency abstraction? *British Journal of Psychology*, *75*, 507–519.
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*, 351–357.
- Speekenbrink, M., & Shanks, D. R. (2010). Learning in a changing environment. *Journal of Experimental Psychology: General*, *139*, 266–298.
- Tulving, E., & Donaldson, W. (1972). *Organization of memory*. Oxford England: Academic Press.
- White, C. M., & Koehler, D. J. (2004). Missing information in multiple-cue probability learning. *Memory & Cognition*, *32*, 1007–1018.
- Yamaguchi, H. (1983). The effects of rehearsal on recall and organization in categorized and noncategorized lists. *Japanese Journal of Psychology*, *54*, 29–35.