



# Does context recollection depend on the base-rate of contextual features?

Marek Nieznański<sup>1</sup> · Michał Obidziński<sup>1</sup> · Daria Ford<sup>1</sup>

Received: 20 March 2023 / Accepted: 25 July 2023 / Published online: 11 September 2023  
© The Author(s) 2023

## Abstract

Episodic recollection is defined by the re-experiencing of contextual and target details of a past event. The base-rate dependency hypothesis assumes that the retrieval of one contextual feature from an integrated episodic trace cues the retrieval of another associated feature, and that the more often a particular configuration of features occurs, the more effective this mutual cueing will be. Alternatively, the conditional probability of one feature given another feature may be neglected in memory for contextual features since they are not directly bound to one another. Three conjoint recognition experiments investigated whether memory for context is sensitive to the base-rates of features. Participants studied frequent versus infrequent configurations of features and, during the test, they were asked to recognise one of these features with (vs. without) another feature reinstated. The results showed that the context recollection parameter, representing the re-experience of contextual features in the dual-recollection model, was higher for frequent than infrequent feature configurations only when the binding of feature information was made easier and the differences in the base-rates were extreme, otherwise no difference was found. Similarly, base-rates of features influenced response guessing only in the condition with salient differences in base-rates. The Bayes factor analyses showed that the evidence from two of our experiments favoured the base-rate neglect hypothesis over the base-rate dependency hypothesis; the opposite result was obtained in the third experiment, but only when high base-rate disproportion and facilitated feature binding conditions were used.

**Keywords** Context memory · Base-rate neglect · Conjoint recognition paradigm · Dual-recollection theory · Deep distortions

## Introduction

Dual-process models of memory postulate that recognition memory performance reflects the contribution of two distinct components referred to as recollection and familiarity (Yonelinas 2002). Recollection reflects the conscious reinstatement of details from a learning episode, including both target and contextual information, whereas familiarity reflects a more automatic and general activation of a memory trace. A variation of

the dual-process view of memory is the fuzzy-trace theory (e.g., Brainerd and Reyna 1990, 2002, 2004), which assumes that two qualitatively different types of representations, verbatim trace and gist trace, are encoded in parallel during a study experience. Verbatim trace stores perceptual item-specific information about a stimulus, whereas gist trace represents more general meaning-based information. Overall, recollection reflects verbatim trace retrieval, whereas familiarity is based on gist trace processing (e.g., Reyna 2012; cf. Nieznański et al. 2019).

Recently, Brainerd and colleagues (Brainerd et al. 2014a; Brainerd et al. 2015) have impugned the unitary view of recollection and proposed a model that distinguishes between the conscious recollection of contextual information and the vivid reinstatement of target information. In this model, target recollection derives from the retrieval of verbatim traces of old items, whereas context recollection is based—like familiarity—on gist trace processing. Most recently, however, Brainerd et al. (2022a) have acknowledged that contextual details may be stored in a type of memory trace

---

Editor: Valerio Santangelo (University of Perugia); Reviewers: Minyu Chang (McGill University), Charles Brainerd (Cornell University), and a third researcher who prefers to remain anonymous.

---

✉ Marek Nieznański  
m.nieznanski@uksw.edu.pl

<sup>1</sup> Institute of Psychology, Cardinal Stefan Wyszyński University, ul. Wóycickiego 1/3 Bud. 14, 01-938 Warsaw, Poland

that is separate from verbatim and gist, namely, a contextual trace. They argued that contextual details are typically associated with multiple old items, which makes them distinct from surface and semantic details specific to particular items. This three-dimensional structure was supported by a meta-analysis of conjoint recognition studies, which distinguished a semantic familiarity (gist trace-based) factor, a context recollection (contextual trace-based) factor, and a target recollection (verbatim trace-based) factor.

Our research stems from an assumption that the strength of the contextual trace can reflect the frequency of occurrence of a particular contextual feature among multiple old items. The more items share the same contextual feature, the stronger the contextual trace of this feature should be. We also hypothesize that the probability that a probe containing a particular contextual feature will evoke context recollection of another associated contextual feature is affected by the frequency of these two contextual features co-occurring. For example, context recollection that a cue word printed in a large font size was green should be higher when most of the presented large-font-size words were printed in green. In other words, we predict that context recollection is sensitive to the base-rate of contextual information experienced during study and reflects the frequency of context-context pairings. This *base-rate dependency account* finds some support in studies on multidimensional source recognition (e.g., Meiser and Bröder 2002) or in studies on ‘pattern completion’ (e.g., Horner et al. 2015; Horner and Burgess 2013). However, there are also some compelling arguments in favour of an alternative view—the *base-rate neglect account*, which refers to the phenomenon known from the judgment and decision-making literature that people have a strong tendency to favour diagnostic information over the base-rates when judging the probability of an event (e.g., Kahneman and Tversky 1973). The aim of the current study is to estimate the evidence in favour of the base-rate dependency hypothesis versus the base-rate neglect hypothesis in the recollection of correlated contextual features.

### Arguments in favour of base-rate dependency in memory

The dependency of context memory on the experienced base-rate of contextual features is consistent with the mutual cuing hypothesis (e.g., Arnold et al. 2019; Boywitt and Meiser 2012; Meiser 2014; Meiser and Bröder 2002) which claims that the successful retrieval of one contextual feature serves as a cue for the other contextual feature. The positive stochastic dependence among concurrent retrieval processes for multiple contextual features observed by Meiser and colleagues suggests that these features are integrated into coherent episodic trace (but see Starns and Hicks 2005; Vogt and Bröder 2007).

Encoding events into integrated traces facilitates the joint retrieval of the configurations of features. Importantly, such a dependence was observed when participants declared that they consciously *recalled* the contextual feature (the state of “remembering”) but not in the state based on familiarity (“knowing”). This supports our prediction that the context recollection process, which is defined in the dual recollection theory as a state of vivid reinstatement of contextual features (Brainerd et al. 2014a; Brainerd et al. 2015), is sensitive to the frequency of context-context configurations.

Since the mutual cuing hypothesis predicts that the successful retrieval of one contextual feature facilitates the retrieval of the other contextual feature, the more we can expect such facilitation to occur when one of these features does not need to be retrieved, but is provided to the subject. In such a case, the cueing of the second feature is not conditional on the successful retrieval of the first, but the provided feature is ready for use as a cue. Therefore, in our experiments, we introduced a manipulation of the reinstatement of one of the features as a condition that should enhance base-rate dependency.

The dependency of the retrieval of one element on the retrieval of another element was also demonstrated for elements that are not subordinates, that is, are not contextual features. In the Horner and Burgess (2013) experiments, participants were required to learn *location-person-object* triplets. The authors analysed how dependent the retrieval of one element (e.g., the person) is on the retrieval of another element (e.g., the object) when cued by a third element (e.g., the location), and they confirmed an interdependence in the ability to retrieve the different elements comprising the same event. Other studies also found support for the view that event elements are integrated into coherent ‘event engrams’ that enable episodic recollection (Horner et al. 2015). Incidental aspects of an event, as contextual details, are also retrieved along with other elements of a complete event. The retrieval of all these constituents of an event when presented with a partial cue is named ‘pattern completion’. The holistic recollection of event elements resulting from their associative structure is even regarded as the defining characteristic of episodic memory, and it was observed both for simultaneously and separately encoded event elements (Horner et al. 2015; James et al. 2020; but see Trinkler et al. 2006).

Research on mutual cuing hypothesis and pattern completion converge in their theoretical conclusions, but use quite different research paradigms, taking this into account, in our Experiment 1 we used a procedure more like that of source memory research (e.g., Meiser and Bröder 2002), while in Experiments 2 and 3 we also used a procedure like that of pattern completion research with colour-object-location triplets (e.g., Horner and Burgess 2013).

Important support for the base-rate dependency in memory also comes from Anderson and Schooler's (1991) environmental explanation of such memory phenomena as practice, retention, and spacing effects. They describe the memory system as making statistical inferences and reflecting the structure that exists in the environment. According to their observations, the memory system tries to make available those memories that are most likely to be useful in a given time and environment. Therefore, we can expect that memory will also mirror the frequencies of features configurations experienced during the study phase of a memory experiment. This should happen whether or not subjects consciously notice the frequency structure of features, just as awareness of the fact that an event is repeated is not needed for the practice effect to occur.

### Arguments in favour of base-rate neglect in memory

As Johnson et al. (1993) stated in their source monitoring framework, source attributions can be influenced by prior knowledge, schemas, or expectations. The strength of prior associations between features, especially when attentional resources are restricted, may influence item-context binding processes (Nieznański 2013). However, as demonstrated by Bayen et al. (2000), schema-based expectancy seems to influence guessing rather than the ability to remember the source. Source guessing is informed by (a) schema-based bias, which is cross-situational and based on general world knowledge, and (b) probability matching, which is based on situation-specific item-source contingency (e.g., Bell et al. 2020; Spaniol and Bayen 2002). The latter mechanism reflects base-rates experienced at encoding, so that, when source memory is not available, participants guess the source of detected-old items consistently with the proportion of sources associated with the particular type of items (e.g., Bayen and Kuhlmann 2011; Kuhlmann et al. 2012; Wulff et al. 2021). This line of research clearly indicates that specific contingencies of item types and sources influence guessing but not source detection, and this assertion is based on analyses using the two-high threshold multinomial model for source monitoring (Bayen et al. 1996), which enables the separation of the processes of item detection, source discrimination, and response bias. Therefore, the probability-matching account suggests that base-rate dependency appears in metamemory judgments rather than in object-level memory processes.

Base-rate neglect is well-known as one of the many errors and fallacies of human probability judgment, which were initially described in the Kahneman and Tversky research program (e.g., Kahneman and Tversky 1973; Tversky and Kahneman 1983; Tversky and Kohler 1994). In the domain of memory, some analogues of such fallacies were investigated by Brainerd and colleagues. For example, Brainerd et al.

(2014b) described conjunction illusions, that is, instances in which participants falsely remember that a target from a single source was presented in multiple sources (see also: Brainerd et al. 2017; Nakamura and Brainerd 2017). In recent reviews, instances when the structure of real-world events is not preserved by our memories were referred to by Brainerd (2021, 2022) as 'deep distortions'. The study of these phenomena has been inspired by the fuzzy-trace theory's idea of gist memory, which implies that the retrieval of gist traces supports the acceptance of items belonging to different reality states that are mutually incompatible, for example, a related distractor may be accepted when asked if it is a related new item, but also when asked if it is a target because the target and the related distractor share a gist. Deep distortions are a new family of false memories that operate at a higher level of measurement than surface distortions. Compared to traditional false memories, they are theoretically more fundamental and measurable by analysing relations between two or more memories. Emergent relations among these memories of events or sources, usually studied using the conjoint recognition paradigm, are confronted with certain normative principles and are classified as deep distortions when they violate the axioms and rules of logic or classical probability theory (Brainerd 2021, 2022). An interesting recent example of a violation of the laws of logic comes from the Brainerd et al. (2022b) experiments, which showed that old? and new? judgments do not produce equivalent recognition accuracy. Despite logical equivalence, accuracy levels differ for judgments that an item is old from judgments that it is not new, and judgments that an item is new differ from judgments that it is not old.

Our aim was to analyse relations between memories for frequent versus infrequent configurations of features. It is possible that base-rate neglect in context memory is another example of when the structure of an everyday experience is not preserved by our memories—in this case, our memory does not act on the logic of conditional probability. An attempt to demonstrate the base-rate neglect in source memory was made by Lu and Nieznański (2020), however, that study did not apply modelling analyses to separate the contribution of context recollection.

### Experiment 1: Context memory for equally versus unequally distributed features in neutral and reinstated test conditions

The general goal of Experiment 1 was to ascertain the presence or absence of an effect of an apparent correlation between contextual features on memory for one of these features. All the presented items differed in two dimensions of colour and size. The memory for the colour dimension was tested, and the distribution of colours by font size was

manipulated within-subjects. For small-size items, the colours were equally distributed, whereas for large-size items saliently more items were presented in one colour than another. The main question was whether the base-rates experienced during the study influence context memory or do they just affect the guessing bias. For evenly distributed features, the influence of the base-rate should result in the absence of differences in context memory performance, while for disproportionately distributed features, the impact of the base-rate should result in differences in context memory performance. Moreover, if context-to-context associations are encoded into an integrated memory trace, reinstating the item size should reactivate colour memory, resulting in better context memory test performance (e.g., Symeonidou and Kuhlmann 2021, but see Hicks and Starns 2016).

In the condition with the reinstated large or small font size at the test in comparison with the condition with the neutral (medium) font size, applying the (implicit or explicit) knowledge about the correlation between contextual features should be easier. In this condition, participants were directly informed that the font used to present the word at the test is the same size as the font used at the study, therefore, they can use their knowledge about the base-rates of colours in particular fonts (e.g., that words printed in green were often presented in large font and rarely in small font). Applying the learnt correlation between features is also possible in the neutral condition depending on the ability to spontaneously mentally reinstate the font size of the presented word (cf. Starns and Hicks 2013). However, since the study font size may be forgotten or falsely attributed, context memory in the neutral condition should be much less affected by the base-rates than in the condition with the reinstated font.

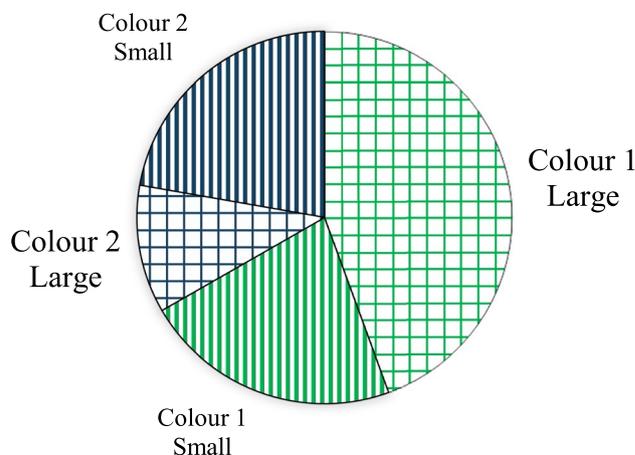
## Participants

In this experiment, 78 participants were recruited from among first and second-year psychology undergraduates. They received extra credits in their courses. One participant was excluded since he reported colour blindness. Participants' mean age was 20.93 years ( $SD = 3.46$ ), 18 were men.

## Stimuli

As the materials, we used 123 nouns in Polish taken from the dataset prepared by Imbir (2016). According to the ratings available in this dataset, the selected words were all low in arousal, of medium valence and frequency, and of medium or high imaginability. In detail, our materials met the following criteria: all were nouns, 4–6 letters long, with a valence rating (on a scale from 1 to 9) between 3 and 7, an arousal rating lower or equal to 3.6, imaginability higher or equal to 4; and a frequency of appearance in the language from 300 to 1500 (Mandera et al. 2014).

## BASE RATES OF CONTEXTUAL FEATURES



**Fig. 1** A circle diagram illustrating the proportions of contextual features presented at the study phase of Experiment 1. Green areas depict words in Colour 1, blue areas in Colour 2, lattices represent words in large font, and vertical lines represent words in small font

## Procedure and design

The participants were examined at individual workstations in the University Lab. The presentation of the stimuli and the response recording were controlled using the E-Prime 2.0 program (Psychology Software Tools, Pittsburgh, PA).

At study, 81 words were presented, two-thirds of them (54) were presented in font Colour 1, and one-third (27) in font Colour 2, thus, the base-rates were manipulated within subjects. For approximately half of the participants, Colour 1 was green and Colour 2 was blue, and vice versa for the other half. The participants were asked to try to remember words along with their colour and size. They were notified that some colours are more frequent in a particular font size than in another. The words were presented in a random order, at a rate of 4 s, with an interstimulus interval of 250 ms. Among 81 words, 45 were presented in large font size (96 pts) and 36 in small font size (24 pts); the font type was *Arial*, bold. Among 54 words in Colour 1, two-thirds (36) were presented in large and one-third (18) in small font size. Among the 27 words in Colour 2, one-third (9) were presented in large and two-thirds (18) in small font size. Overall, there were more Colour 1 than Colour 2 words, and Colour 1 words were more often in large font than small font, and the opposite was true for Colour 2 words. Figure 1 illustrates the proportions of words in each colour and font. Formally, the probability of a particular colour given a particular font size can be computed using the conditional probability formula, as follows:

$$P(C_1|L) = \frac{P(L \cap C_1)}{P(L)} = \frac{36/81}{45/81} = 0.8,$$

$$P(C_1|S) = \frac{P(S \cap C_1)}{P(S)} = \frac{18/81}{36/81} = 0.5,$$

$$P(C_2|L) = \frac{P(L \cap C_2)}{P(L)} = \frac{9/81}{45/81} = 0.2,$$

$$P(C_2|S) = \frac{P(S \cap C_2)}{P(S)} = \frac{18/81}{36/81} = 0.5.$$

where  $C_1$  = Colour 1,  $C_2$  = Colour 2,  $L$  = large font, and  $S$  = small font. Therefore, when a particular test probe is recognized as being presented at study in the large font, it is also expected that it was presented in Colour 1 rather than Colour 2 (the a priori hypothesis that it was Colour 1 is 4 times more probable than that it was Colour 2). However, when a word is presented in small font at test, it is equally probable that it was in Colour 1 or 2 at study.

At test, the studied words were presented intermixed with 42 distractors. Reinstatement of font size at test was manipulated between subjects. The words were presented in the same—large (96 pts) or small (24 pts)—font size for 37 participants, and in a new medium (48 pts) font size for 40 participants. In the reinstated condition, half of the distractors were presented in large font and the other half in small font. At test, the participants were informed that their task was to recognize if the word was presented and answer “yes” or “no” to the question that will be shown under the test word on a particular slide. There were three types of probe questions counterbalanced across participants and presented equally often with each type of test items: (a) Was this word presented in Colour 1?; (b) Was this word presented in Colour 2?; and (c) Was this word presented in either Colour 1 or 2? The slides were presented in random order. The test trials were participant-paced with the next trial appearing immediately after a response.

## Data analysis

Bayesian analyses were conducted in JASP (JASP Team 2019; jasp-stats.org, see: van Doorn et al. 2020). We used Bayes factor  $BF_{10}$  to compare the predictive performance of an alternative hypothesis over a null hypothesis. A Bayes factor between 1 and 3 is considered weak evidence, between 3 and 10 moderate evidence, and above 10 is considered strong evidence in favour of an alternative hypothesis. In symmetry, a  $BF_{10}$  lower than 1 supports a null hypothesis, a factor between 0.333 and 0.1 means moderate evidence, and below 0.1 is considered strong evidence for a null hypothesis. When the dependent variables were normally distributed

and the variances were homogenous across the groups, we performed Bayesian  $t$  tests, otherwise, we reported the Mann–Whitney  $U$ -test or the Wilcoxon rank-signed test. As priors we used default options in JASP, that is, the Cauchy distribution with  $r$  set to  $1/\sqrt{2}$ .

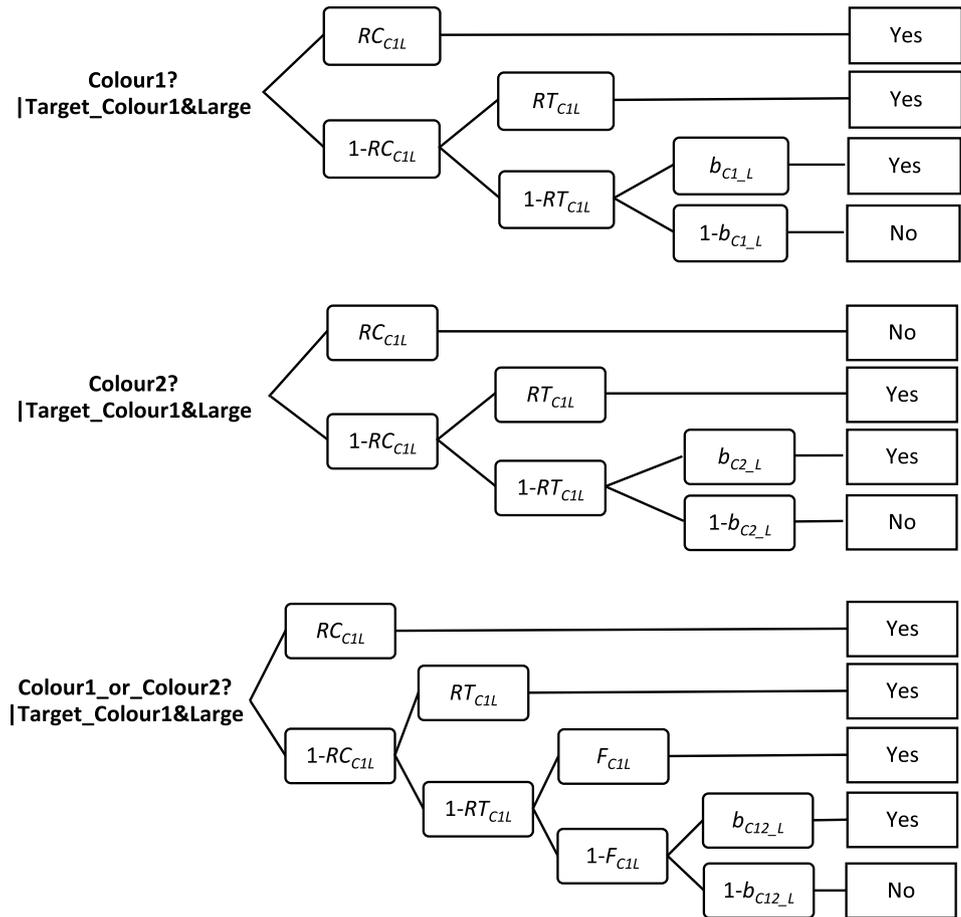
Multinomial modelling analyses were based on hierarchical Bayesian modelling using the latent-trait approach (Klauer 2010). This approach uses the multivariate normal distribution of the transformed individual parameters as the prior distribution on a group level. Monte Carlo Markov Chain sampling methods are employed to obtain the parameter posterior estimates (for more information about hierarchical multinomial processing tree models and examples of their application see: Arnold et al. 2019; Ernst et al. 2019; Heck et al. 2018; Klauer 2010). All hierarchical multinomial modelling analyses were conducted using the R package TreeBUGS (Heck et al. 2018).

## Multinomial model for conjoint recognition paradigm

In the present research, the multinomial dual-recollection model (Brainerd et al. 2015) was used as a measurement model. The original model was developed for a context memory experiment with targets presented on List 1 and List 2 at study, and with three types of probe questions presented during the test phase: “Was it on List 1?”, “Was it on List 2?” or “Was it on either List 1 or List 2?”. The model defines the following retrieval processes: (a) The  $RT_1$  ( $RT_2$ ) parameter (target recollection), which is the probability that a List 1 (List 2) target cue provokes the conscious reinstatement of its presentation during the study; (b) The  $RC_1$  ( $RC_2$ ) parameter (context recollection), which is the probability that a List 1 (List 2) target cue provokes the conscious reinstatement of the contextual details of List 1 (List 2) presentation; and (c) The  $F_1$  ( $F_2$ ) parameter (familiarity), which represents the probability that a List 1 (List 2) target cue provokes a sufficiently high familiarity to make the target be perceived as old. Moreover, two response bias parameters are also defined: one ( $b$ ) for accepting non-retrieved items (targets or distractors) for List 1? probe questions or List 2? probe questions, and another ( $b_{12}$ ) for accepting non-retrieved items for List 1 or 2? probe questions (see Brainerd et al. 2015, Table 2). In comparison with the original model, the present research replaced the List 1 targets and the List 2 targets with the targets presented in Colour 1 or Colour 2.

A part of the multinomial model applied in the current research is presented in Fig. 2. One tree can be depicted for each probe question and item type. In Fig. 2, only the model of processing of targets presented in Colour 1 and large font is shown as an example. On the left are the item types used at the test with the specified question probes (Colour 1?, Colour 2?, and Colour 1 or Colour 2?). On the right are the

**Fig. 2** A part of the multinomial dual-recollection model used in Experiment 1 (based on Brainerd et al. 2015, Fig. 1). Colour1?, Colour2?, and Colour1 or Colour2? refer to probe questions.  $RC$  is a context recollection parameter,  $RT$  is a target recollection parameter,  $F$  is a familiarity parameter, and  $b$  is a response bias parameter. Subscripts indicate the kind of target determined by the font size and colour



participants’ responses (accept or reject), which are connected with the question probes and the item types by the branches of the processing trees representing the latent cognitive processes postulated by the dual recollection theory. As can be seen in Fig. 2, when a target context is congruent with the question probe (C1?|Target\_C1), the target cues are accepted if the context recollection ( $RC_i$ ) or the target recollection ( $RT_i$ ) is successful and, if neither are successful, the response bias ( $b_i$ ) can produce a “yes” response. When a target context is incongruent with the question probe (C2?|Target\_C1), the target cues are rejected if the context recollection is successful but are accepted if the context recollection fails ( $1 - RC_i$ ) and the target recollection ( $RT_i$ ) is successful, and a “yes” response may also be produced by the response bias ( $b_2$ ). On the probes with the Colour 1 or Colour 2? question (C1or2?|Target\_C1), the participants respond “yes” if the context recollection, target recollection or familiarity ( $F_i$ ) are successful; and if all of these retrieval processes fail, the response bias ( $b_{i2}$ ) can produce acceptance. For distractors, only the response bias ( $b_1$  for C1?,  $b_2$  for C2?, and  $b_{i2}$  for C1or2?) can produce acceptance (cf. Brainerd et al. 2015). Separate models of this type were created for large and small font-size items.

## Results

### Results based on descriptive measures

Descriptive statistics concerning the mean acceptance rates and the mean corrected acceptance rates (CAR) (i.e., the probability of a “yes” response for targets minus the probability of a “yes” response for distractors) for particular colour/font configurations and for each type of probe question are presented in Tables 14 and 15 in the “Appendix 1”. Figure 3 presents only the means and 95% credible intervals of accurate CARs, and Fig. 4 presents false alarms for distractors. The grand means of accurate CARs, compared between large font size items ( $M=0.278, SD=0.358$ ) and small font size items ( $M=0.256, SD=0.313$ ), pooling over configuration types and test conditions, were not significantly different,  $t(153)=0.82$ . However, when the grand means of accurate CARs were compared between Colour 1 items ( $M=0.226, SD=0.312$ ) and Colour 2 items ( $M=0.309, SD=0.354$ ), a significant difference was found,  $t(153)=2.81$ , Cohen’s  $d=0.23, p=0.006$ , indicating that participants attributed the less frequently presented colour more accurately.

### Within-subjects effects of the base-rates

If memory is informed by base-rates, the accurate mean CAR should be higher for the most frequent configuration (Colour 1 and large font size) than the least frequent configuration (Colour 2 and large font size), therefore, the one-sided alternative hypothesis is  $CAR_{C1L} > CAR_{C2L}$ . However, in the neutral condition, the Bayesian  $t$  test yielded strong evidence for a null hypothesis,  $BF_{+0} = 0.078$ , and, in the reinstated condition, the evidence for a null hypothesis was moderate,  $BF_{+0} = 0.207$ . This suggests that memory was not informed by base-rates.

A second comparison was conducted between equally frequent configurations of Colour 1 and Colour 2 with small font size. This time, if memory is informed by base-rates, we should find evidence for the null hypothesis. The alternative two-sided hypothesis is that  $CAR_{C1S} \neq CAR_{C2S}$ . In the neutral condition, we found moderate evidence for an alternative hypothesis,  $BF_{10} = 3.112$ . In the reinstated condition, the evidence was indeterminate,  $BF_{10} = 1.096$ . Therefore, in the case of equally frequent colour/size configurations, the evidence is against the use of base-rates in the neutral condition, and it is inconclusive in the reinstated condition. If the participants ignore the base-rates of the colour/size configurations but they use the general proportion of colours, the alternative hypothesis should be one-sided,  $CAR_{C1S} > CAR_{C2S}$ , since there were, in general, more items in Colour 1. However, in the neutral condition, we found strong evidence for the null hypothesis,  $BF_{+0} = 0.052$ , which means that the data is about 19 times more likely under the null hypothesis.

In the case of false alarms (FA), that is, “yes” responses to distractors (see Fig. 4), if the response bias is informed by base-rates, more “yes” responses should be found for distractors when the participants are asked about the more frequent Colour 1. In the neutral condition, the Wilcoxon signed-rank test yielded Bayes factor,  $BF_{-0} = 387.375$ , indicating very strong evidence for the one-sided alternative hypothesis,  $FA_{C2?} < FA_{C1?}$ . In the reinstated condition, the alternative hypotheses differ depending on the font size in which the distractors were presented. In comparison with the neutral condition, this time the participants need to consider not only which colour was more frequent but also what were the proportions of colours depending on the font size. We found weak evidence for the one-sided alternative hypothesis,  $FA_{LargeC2?} < FA_{LargeC1?}$ , when the distractors were presented in large font size,  $BF_{-0} = 1.959$ . However, when the distractors were presented in small font, for which colours were equally distributed at study, the evidence was moderate for the null hypothesis,  $FA_{SmallC2?} = FA_{SmallC1?}$ ,  $BF_{10} = 0.186$ . These results suggest that the response bias is informed by base-rates in both the neutral and reinstated test conditions.

### Between-subjects effects of context reinstatement

We can hypothesise that the participants in the reinstated condition should be more informed by base-rates than the participants in the neutral condition. Therefore, for the most frequent configuration (Colour 1 and large font size), the mean CAR should be lower in the neutral than in the reinstated condition,  $CAR_{C1Lneutral} < CAR_{C1Lreinstated}$ . However, we found moderate evidence for the null hypothesis,  $BF_{-0} = 0.280$ , from the independent one-sided  $t$  test. In contrast, the mean CAR should be higher in the neutral than in the reinstated condition for the least frequent configuration (Colour 2 and large font size),  $CAR_{C2Lneutral} > CAR_{C2Lreinstated}$ . We found weak evidence for the null hypothesis,  $BF_{+0} = 0.556$ . Therefore, we cannot conclude that in the reinstated condition the base-rates inform memory more than in the neutral condition, we have moderate or weak evidence for the opposite hypothesis.

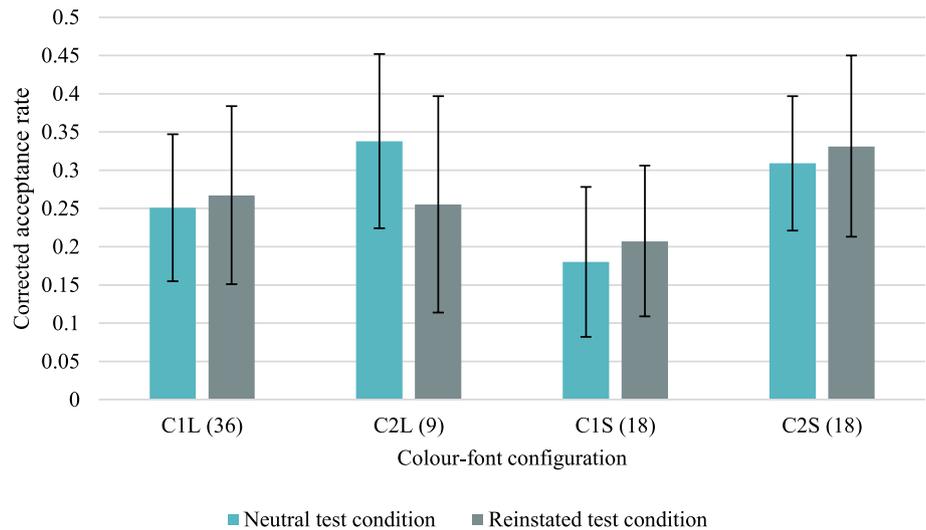
### Results based on process measures

Hierarchical analyses were conducted using the latent-trait approach (Klauer 2010) implemented in the TreeBUGS software (Heck et al. 2018). Model fit was assessed with the  $T1$  (the distance between the observed and the expected mean frequencies) and  $T2$  (the summed distance between the observed and the expected covariance statistics) (Klauer 2010; see Heck et al. 2018). Good model fit was indicated by nonsignificant test results in the neutral condition ( $T_1$ :  $p = 0.519$ ,  $T_2$ :  $p = 0.366$ ) and in the reinstated condition ( $T_1$ :  $p = 0.446$ ,  $T_2$ :  $p = 0.424$ ). Group-level dual-recollection multinomial model parameter estimates and their 95% Bayesian Credible Intervals (BCI) are presented in Table 1.

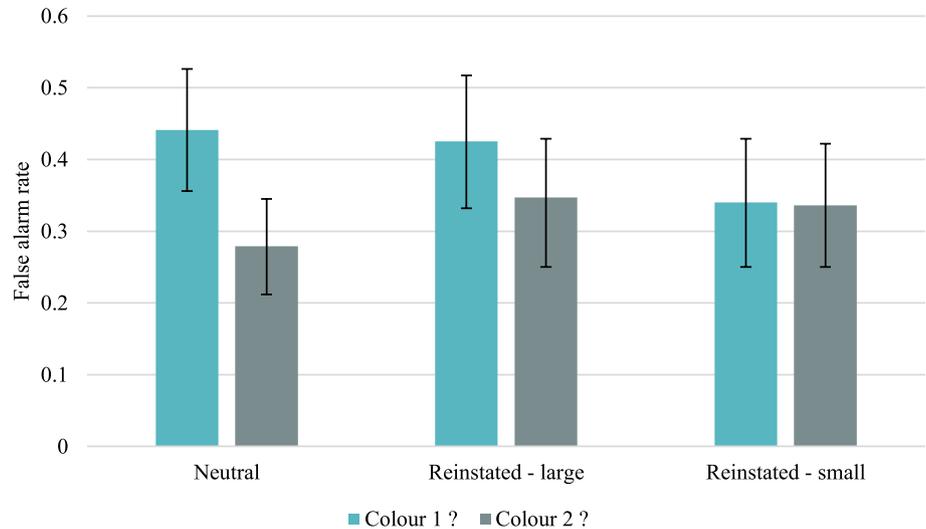
### Within-subjects effects of base-rates on context recollection and response bias

For each memory parameter, the posterior samples obtained for one configuration of features were subtracted from those obtained for another configuration. Parameters for which the 95% CI of the difference estimates overlapped with 0 do not meaningfully differ between conditions (Smith and Batchelder 2010). In the neutral condition, we compared the context recollection parameter and the response bias depending on the base-rates. The difference in the mean context recollection parameters  $\Delta RC$  between frequent versus infrequent configuration ( $RC_{C1L} - RC_{C2L}$ ) was  $M = -0.014$ , with the credibility interval of the difference  $[-0.191, 0.189]$  indicating no substantial effect. Similarly, the difference in the mean parameters between the equally frequent configurations ( $RC_{C1S} - RC_{C2S}$ ) indicated no substantial effect,  $M = -0.043$  with 95% CI  $[-0.215, 0.129]$ . However, in the case of the difference of the response bias  $\Delta b$  depending on the probe question about the frequent versus less frequent

**Fig. 3** Mean corrected-for-guessing acceptance rates for accurate colour recognition for configurations of contextual features in Experiment 1. Error bars represent 95% credible intervals. Numbers in parentheses indicate the number of items in the condition



**Fig. 4** Mean false alarm rates in Experiment 1. Error bars represent 95% credible intervals



colour ( $b_{C1} - b_{C2}$ ), a substantial effect was indicated by the credibility interval not overlapping with 0,  $M=0.149$ , 95% CI [0.056, 0.238].

In the reinstated condition, the results for the context recollection parameters indicated no substantial effect. In detail, the difference in the mean parameters  $\Delta RC$  were  $M=0.153$ , 95% CI [-0.092, 0.396] for the frequent versus infrequent configuration ( $RC_{C1L} - RC_{C2L}$ ), and  $M = -0.074$ , 95% CI [-0.272, 0.110] for equally frequent configurations ( $RC_{C1S} - RC_{C2S}$ ). For the response biases, the 95% CI of the difference estimates overlapped with 0, both when the distractors were presented in large font  $\Delta b_L$  ( $b_{C1L} - b_{C2L}$ ),  $M=0.097$ , 95% CI: [-0.045, 0.215] and when distractors were presented in small font  $\Delta b_S$  ( $b_{C1S} - b_{C2S}$ ),  $M = -0.03$ , 95% CI [-0.186, 0.117]. This suggests that the response bias did not meaningfully differ between the probes in the reinstated condition.

### Between-subjects effects of context reinstatement

In this section, we present the results for all the memory parameters of the dual-recollection model since source memory literature suggests that context reinstatement can influence verbatim trace retrieval (Nieznański and Tkaczyk 2017), which has been represented in the model by the target recollection parameter. For each memory parameter, the posterior samples obtained in the reinstated condition were subtracted from those obtained in the neutral condition. As Table 2 shows, no substantial effect of the test condition was indicated for differences in memory parameters.

### Bayesian analyses of hypotheses about context recollection and response bias

Although hierarchical models specify the parameters both for the group and the individual participants level (Heck,

**Table 1** Group-level parameter estimates (standard deviations) and 95% Bayesian Credible Intervals of the dual-recollection multinomial model obtained in Experiment 1

	Large font		Small font	
	Colour 1 (n=36)	Colour 2 (n=9)	Colour 1 (n=18)	Colour 2 (n=18)
<b>Neutral condition</b>				
<i>RT</i>	0.304 (0.041) [0.222, 0.383]	0.324 (0.073) [0.169, 0.457]	0.230 (0.047) [0.131, 0.318]	0.277 (0.048) [0.177, 0.366]
<i>RC</i>	0.185 (0.054) [0.077, 0.289]	0.219 (0.068) [0.078, 0.346]	0.111 (0.061) [0.012, 0.239]	0.143 (0.063) [0.030, 0.272]
<i>F</i>	0.127 (0.077) [0.008, 0.289]	0.352 (0.161) [0.042, 0.647]	0.297 (0.130) [0.039, 0.530]	0.121 (0.082) [0.005, 0.300]
<i>b<sub>C1</sub></i>	0.407 (0.045) [0.318, 0.493]			
<i>b<sub>C2</sub></i>	0.250 (0.034) [0.185, 0.319]			
<i>b<sub>C1or2</sub></i>	0.224 (0.027) [0.171, 0.279]			
<b>Reinstated condition</b>				
<i>RT</i>	0.287 (0.056) [0.169, 0.390]	0.295 (0.092) [0.090, 0.457]	0.275 (0.048) [0.174, 0.363]	0.340 (0.083) [0.169, 0.498]
<i>RC</i>	0.167 (0.062) [0.051, 0.290]	0.097 (0.068) [0.005, 0.253]	0.046 (0.036) [0.002, 0.133]	0.108 (0.062) [0.011, 0.241]
<i>F</i>	0.142 (0.102) [0.007, 0.375]	0.396 (0.183) [0.047, 0.740]	0.186 (0.110) [0.014, 0.418]	0.068 (0.064) [0.001, 0.230]
<i>b<sub>C1</sub></i>	0.398 (0.047) [0.304, 0.488]		0.309 (0.052) [0.208, 0.410]	
<i>b<sub>C2</sub></i>	0.287 (0.052) [0.187, 0.387]		0.322 (0.049) [0.224, 0.415]	
<i>b<sub>C1or2</sub></i>	0.228 (0.048) [0.138, 0.323]		0.260 (0.045) [0.173, 0.351]	

The italicized symbols are the parameters of the dual-recollection multinomial model: *RT*=target recollection, *RC*=context recollection, *F*=familiarity, and *b<sub>C</sub>*=response bias for Colour1?, Colour2? or Color1 or Colour2? probe questions

**Table 2** Differences in the mean dual-recollection multinomial model parameters and 95% BCIs between the neutral and reinstated conditions

Parameter (neutral–reinstated)	Large font		Small font	
	Colour 1 (n=36)	Colour 2 (n=9)	Colour 1 (n=18)	Colour 2 (n=18)
$\Delta RT$	-0.008 [-0.182, 0.138]	-0.001 [-0.262, 0.261]	-0.056 [-0.200, 0.094]	-0.117 [-0.356, 0.078]
$\Delta RC$	-0.042 [-0.302, 0.152]	0.090 [-0.171, 0.282]	0.039 [-0.142, 0.191]	-0.010 [-0.279, 0.188]
$\Delta F$	-0.059 [-0.497, 0.208]	-0.061 [-0.562, 0.422]	0.068 [-0.371, 0.406]	0.020 [-0.353, 0.249]

$\Delta RT$ =the difference in the target recollection parameter estimates,  $\Delta RC$ =the difference in the context recollection parameter estimates,  $\Delta F$ =the difference in the familiarity parameter estimates

et al. 2018), entering the individual estimates as input into a Bayesian *t* test is problematic.<sup>1</sup> This is because parameter estimates for individual participants are informed by the group means, especially in the case of less reliable estimates (a property called shrinkage); in result, the estimation error is artificially decreased (Boehm et al. 2018). Therefore, we computed the independent estimates of the parameters for each participant in a conventional way using the maximum likelihood fitting method (e.g., Riefer and Batchelder 1988) implemented in the *multiTree* software (Moshagen 2010). It must be noted, however, that estimates based on relatively few observations per participant do not allow the parameters

to be estimated precisely, which increases the error variance. The full results of the maximum likelihood analyses are presented in "Appendix 2".

In the neutral condition, a one-sided Bayesian Wilcoxon signed-rank test yielded moderate evidence for the null hypothesis when we compared the context recollection parameter between the frequent versus infrequent configuration of sources,  $RC_{CIL}$  ( $M=0.225$ ,  $SD=0.213$ ) versus  $RC_{C2L}$  ( $M=0.238$ ,  $SD=0.283$ ),  $BF_{+0}=0.145$ ,  $W=318.00$ . Similarly, a two-sided test provided moderate evidence for the null hypothesis when this parameter was compared between equally frequent configurations,  $RC_{CIS}$  ( $M=0.176$ ,  $SD=0.234$ ) versus  $RC_{C2S}$  ( $M=0.213$ ,  $SD=0.253$ ),  $BF_{10}=0.262$ ,  $W=255.00$ . However, extreme evidence was provided for the alternative hypothesis that the response

<sup>1</sup> We would like to thank Prof. Daniel W. Heck for pointing this problem out.

bias for the *Colour1?* question,  $b_{C1}$ , was higher ( $M=0.415$ ,  $SD=0.232$ ) than for the *Colour2?* question,  $b_{C2}$  ( $M=0.270$ ,  $SD=0.184$ ),  $BF_{+0}=400.365$ ,  $W=649.00$ .

In the reinstated condition, we also found moderate evidence for the null hypothesis that the context recollection parameter is equal between the frequent configuration,  $RC_{CIL}$  ( $M=0.216$ ,  $SD=0.235$ ) versus the infrequent configuration,  $RC_{C2L}$  ( $M=0.226$ ,  $SD=0.290$ ),  $BF_{+0}=0.149$ ,  $W=178.00$ . In the case of the equally frequent configurations, the two-sided test yielded weak evidence for the null hypothesis,  $RC_{CIS}$  ( $M=0.120$ ,  $SD=0.194$ ) versus  $RC_{C2S}$  ( $M=0.177$ ,  $SD=0.239$ ),  $BF_{10}=0.651$ ,  $W=96.00$ .

In the reinstated condition, though the colours were unequally distributed for the words printed in large font, we found only weak evidence for the alternative hypothesis that the response bias parameter  $b_{CIL}$  ( $M=0.406$ ,  $SD=0.242$ ) is larger than the  $b_{C2L}$  parameter ( $M=0.332$ ,  $SD=0.251$ ),  $BF_{+0}=1.225$ ,  $W=370.00$ . Moderate evidence was provided for the null hypothesis that the response biases are equal for equally probable colours presented in small font,  $b_{CIS}$  ( $M=0.333$ ,  $SD=0.239$ ) versus  $b_{C2S}$  ( $M=0.344$ ,  $SD=0.247$ ),  $BF_{10}=0.181$ ,  $W=287.00$ .

In the case of comparisons between the neutral and reinstated conditions, all Bayes factors provided weak or moderate support for the null hypotheses ( $0.244 < BF_{10} < 0.507$ , Bayesian Mann–Whitney  $U$  tests).

## Discussion

The results of Experiment 1 provided support for the base-rate neglect hypothesis over the base-rate dependency hypothesis. Analyses on the descriptive measures indicated that the base-rates experienced during the study do not influence accurate memory for context. Similarly, no differences between context recollection parameters between frequent and infrequent configurations were observed in multinomial analyses. However, evidence supported the hypothesis that the participants guess the colour in accordance with the base-rate. Evidence for probability-matching of guessing strategy was stronger in the neutral condition than in the reinstated condition, probably because the task of matching to Colour 1 / Colour 2 ratio was an easier task than matching strategy to a more complex colour and font size configuration. No support was found in favour of the hypothesis that reinstating one of the correlated features cues the retrieval of the other feature facilitating base-rate dependency. Initial analyses comparing accurate CARs between Colour 1 and Colour 2 items revealed that the less frequent colour is more accurately attributed to the test item than the more frequent colour. This suggests a kind of fan effect (Anderson 1974), in which it is easier to retrieve information when it has fewer associations. Item-to-context associations may be stronger for Colour 1 because it has a smaller fan of connections with

individual items than Colour 2. However, such a fan effect should not refer to context-to-context associations, since these connections are learned through repeated exposure to the same features (colour and size) that do not form unique pairs.

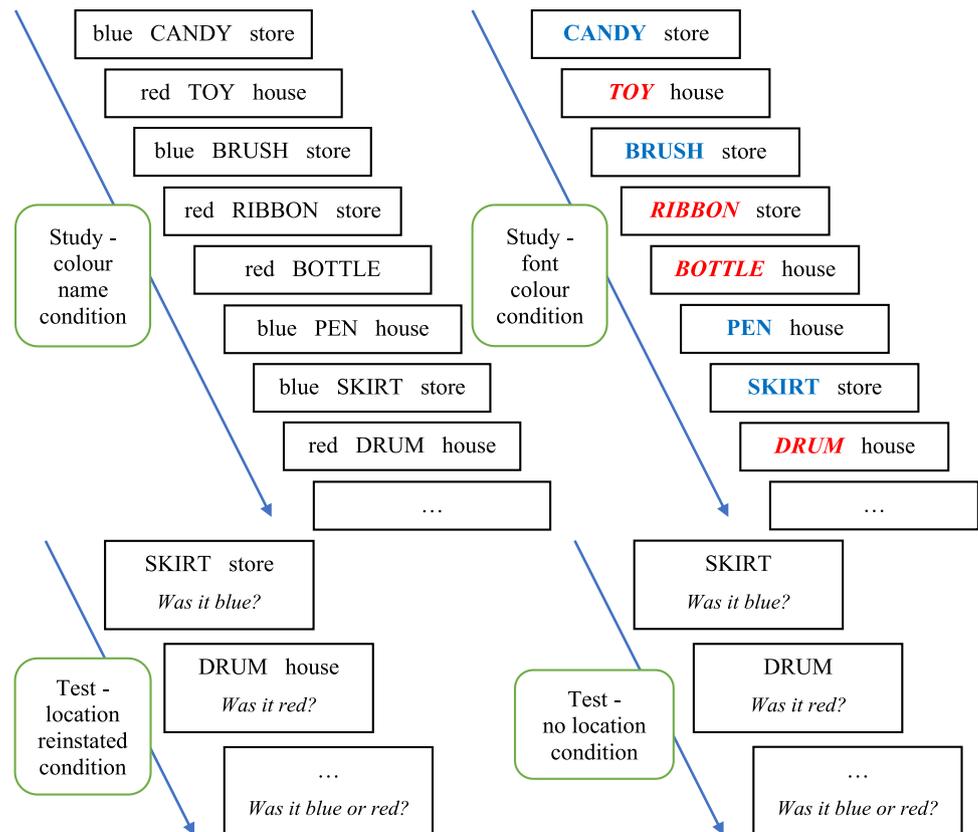
The conclusion that our memory ignores base-rate information seems premature from the above results. Although most of these results supported the base-rate neglect hypothesis, when context recollection parameters were compared between equally frequent configurations, weak to moderate evidence for the null hypothesis indicated that base-rates are not completely disregarded. In the remainder of this article, we present two more experiments with a different procedure intended to enhance the encoding of an integrated episodic trace. Binding contextual features with each other and with item information to create an episodic trace is crucial for the reconstructive processes during a memory test. It is possible, however, that binding size with colour information was not effective in Experiment 1, since it is difficult to find any semantic association between these features. In Experiment 2, we purposely used elements that are easier to bind and we explicitly asked the participants to try to create an integrated image of an object having particular features. Such an instruction can facilitate context reinstatement effects (e.g., Hanczakowski et al. 2014; Hockley 2008).

## Experiment 2: Memory for unequally distributed extrinsic versus intrinsic features

In the second experiment, we examined how retrieving one feature associated with an object depends on the base-rate of objects' feature configurations. The feature of interest (colour) was presented either as an intrinsic property of the object or as a distinct element. In the latter case, it can be argued that the participants encoded the configurations of items, similar to an associative recognition task but with repeating pairings.

The participants were required to memorise *colour-object-location* triplets (cf. Horner and Burgess 2013). In one condition, the colour information was represented by a separate word—the name of the colour, and, in another condition, it was represented as a font colour. Particular colours were frequently paired with one of two locations and infrequently with the other. Therefore, in this experiment, we examined whether the base-rate of the colour/location configuration can influence the retrieval of the colour of an object. We assumed that if such a dependency exists it should be more pronounced when the object-location pairs are reinstated at test in comparison with a neutral condition when only the objects are provided as test probes. Therefore, in comparison with Experiment 1, we extended our investigation on the base-rate dependency versus the base-rate

**Fig. 5** Procedure of Experiment 2. During study, the information about the colour, object, and place were provided, and colour information was represented as a word or a font. During the conjoint recognition test, the object or object-and-place were used as the cues, and the participants were asked about the colour information



neglect hypotheses to the features that are extrinsic to the item, and we made contextual and item information binding easier and more effective.

## Participants

In this experiment, 72 undergraduates took part in exchange for course credit. Their mean age was 21.97 years ( $SD = 4.12$ ), 8 were men. Each participant was assigned to two experimental tasks among four possible (i.e., font colour and location reinstated, colour name and location reinstated, font colour and neutral, and colour name and neutral). The two tasks assigned to a participant were prepared in two versions, differing in materials (*Blue/red and House/store* versus *Green/yellow and Forest/garden*) in order to minimize the interference between the tasks. Hence, we planned to gather 144 data sets, with the conditions being manipulated between subjects. However, due to a mistake in task assignment, a part of the participants (30) received tasks from conditions that differed in one but not the other factor (either font colour vs. name or reinstated vs. neutral condition). In order to fully preserve between-subjects design for these conditions, we had to exclude one of the two sets obtained from each of these 30

participants. Two other data sets were not recorded due to technical problems. Finally, we analysed 112 datasets: 27 in the font colour and location reinstated condition, 28 in the colour name and location reinstated condition, 28 in the font colour and neutral condition, and 29 in the colour name and neutral condition.

## Stimuli

Among Polish nouns of medium valence and frequency, medium or low arousal level, and medium or high imaginability according to the Imbir (2016) dataset, we selected 140 words. Half of them referred to objects that can occur in the blue or red colours and which can be found in a house or a store (exemplary English equivalents: *candy*, *toy*, *brush*, and *ribbon*), the other half referred to objects that can occur in green or yellow colours and which can be found in a forest or a garden (e.g., *apple*, *balloon*, *tent*, and *butterfly*). The selected 140 nouns had  $M = 6$  number of letters (range: 4–8),  $M = 1.92$  concreteness (range: 1.44–3.2, on a scale from 1 to 9, where 1 means high concreteness and 9 high abstractedness),  $M = 7.87$  imageability (range: 7.32–8.42), and  $M = 380$  frequency of appearance in the language (range: 25–1459) (Mandera et al. 2014).

## Procedure and design

The experiment was built in *OpenSesame* (Mathôt et al. 2011), and conducted online on the Jatos platform. A schematic description of the procedure is presented in Fig. 5. At study, we manipulated the way these features were represented—as a font colour in which the target word referring to an object (e.g., *apple*) was printed, or by the name of the colour paired with the target word (e.g., *green*). At test, we manipulated the reinstatement versus the absence of the word representing the location in an environment of the object that was present at study (e.g., *green-apple-garden*). The full crossing of these two between-subjects variables resulted in four experimental conditions (font colour and location reinstated; colour name and location reinstated; font colour and neutral; and colour name and neutral). In all the conditions, the frequency of a configuration (30 vs. 18 colour-location pairs) was manipulated within-subjects.

In the font colour condition, the participants were presented with 52 pairs of nouns (48 of them were targets, 2 were added as buffers at the beginning, and another 2 at the end of the study list). The first noun in each pair referred to one of the 52 objects, and the second referred to one of two places—a house or store (forest or garden in the second version) in which that object can be found. Each noun referring to an object was printed in uppercase blue or red font (green or yellow font in the second version); the words referring to places were always printed in lowercase white font.

In the colour name condition, the participants were presented with 52 triads, the second and third words in each triad referred to the object and place in the same way as in the font colour condition. However, all the words were printed in white font, and the first word in each triad was the name of one of two colours—blue or red (green or yellow in the second version). In both types of study conditions, the frequencies of colour/location configurations were not equal; among 48 targets, 30 were presented in ‘frequent’ configurations (e.g., 15 in red and house plus 15 in blue and store) and 18 in the ‘infrequent’ configurations (e.g., 9 in blue and house plus 9 in red and store).

The participants were instructed to try to remember each object with its colour and location, they were encouraged to create an image of an object in a particular colour and placed in its location. The slides were presented at a 6 s rate, with a 200 ms interstimulus interval. The stimuli were presented in *Mono* font, 38 px size; the background screen was black.

At test, 66 nouns referring to objects were presented: 48 targets and 18 distractors. In the neutral condition, a single noun was presented, that is, without the noun referring to the location. In the reinstated test condition, the targets were paired with the same location words as during the study phase. Half of the distractors were presented with

one location word, the other half with the second location word. The participants were informed that their task was to answer “yes” or “no” to the question that will be presented on a particular slide. There were three types of probe questions: (a) Was this word presented in Colour 1 (with the word *Colour 1*)?; (b) Was this word presented in Colour 2 (with the word *Colour 2*)?; and (c) Was this word presented in either Colour 1 or Colour 2 (with either the word *Colour 1* or *Colour 2*)? The slides were presented in a random order, at a self-paced rate.

The software used to conduct the experiment randomly assigned colours to objects and particular (frequent or infrequent) configurations of colour and location to the participants. The probe question types assigned to the test items were counterbalanced across the test items. The sequence of task versions (*Blue/red and House/store* vs. *Green/yellow and Forest/garden*) and test condition (reinstated vs. neutral) were assigned to approximately an equal number of participants by one of the authors who distributed the links to the online experiment to the participants by email.

## Results and discussion

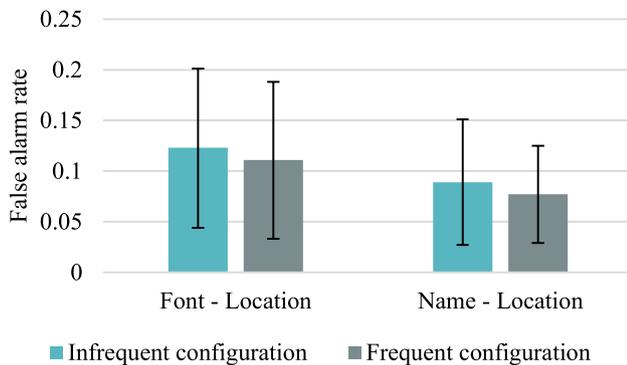
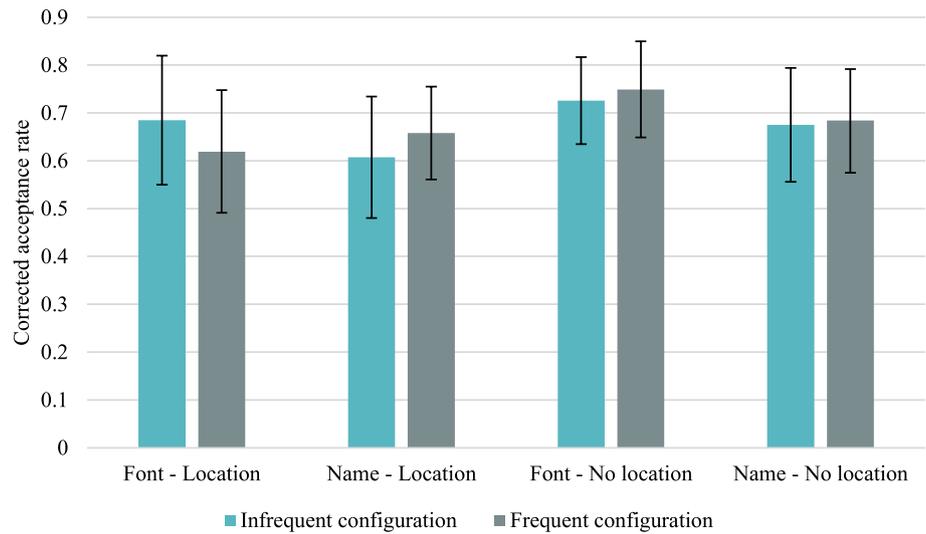
### Results based on descriptive measures

Descriptive statistics concerning the mean acceptance rates and the mean corrected acceptance rates (CAR) for frequent versus infrequent colour/location configurations in different study and test conditions are presented in Tables 18 and 19 in “Appendix 3”. The mean CARs for accurate responses are shown in Fig. 6, and false alarms for distractors are presented in Fig. 7.

### Within-subjects effects of base-rates

If memory is informed by the base-rates, the accurate mean CAR should be higher for the frequent configuration than the infrequent configuration, since more learning episodes should result in a stronger contextual trace or stronger context-context bindings, therefore, the one-sided alternative hypothesis is  $CAR_{\text{frequent}} > CAR_{\text{infrequent}}$ . However, Bayesian Wilcoxon signed-rank test yielded strong evidence for the null hypothesis in the font colour location reinstated condition,  $BF_{+0} = 0.084$ , and moderate evidence in the colour name neutral condition,  $BF_{+0} = 0.206$ . Weak evidence for the null hypothesis was obtained in the colour name location reinstated condition,  $BF_{+0} = 0.442$ , and in the font colour neutral condition,  $BF_{+0} = 0.369$ .

**Fig. 6** Mean corrected-for-guessing acceptance rates for accurate colour recognition for frequent versus infrequent configurations of features in Experiment 2. Error bars represent 95% credible intervals



**Fig. 7** False alarm rates in Experiment 2. Error bars represent 95% credible intervals

In the case of false alarms, if the response bias was informed by base-rates, the participants in the reinstated condition should respond “yes” more often to distractors presented with frequent rather than infrequent location for a particular colour. Both for font colour and colour name conditions, we found moderate evidence in favour of the null hypothesis,  $BF_{+0} = 0.167$  and  $BF_{+0} = 0.146$ , respectively. Therefore, it seems that the response bias was not informed by the base-rates.

**Between-subjects effects of the reinstated versus the neutral test condition and of the form of colour information representation**

We explored the role of representing the colour of an object as a font colour versus as a word. Table 3 presents the Bayes factors, all of which provide weak or moderate support for the null hypothesis that the form of colour representation does not affect memory performance.

As in Experiment 1, we assumed that the reinstated condition in comparison with the neutral condition should result in acceptance rates that are closer to the base-rates. Therefore, for the frequent configuration, the mean CAR should be higher in the reinstated than in the neutral condition, but for the infrequent configuration, it should be lower in the reinstated than in the neutral condition. However, Table 4 shows the Bayes factors indicating evidence for the hypothesis that the neutral and the reinstated conditions do not differ in the mean CARs.

**Results based on process measures**

Table 5 presents the results of hierarchical multinomial processing tree modelling for all the experimental conditions.

**Table 3** Results of the Bayesian Mann–Whitney *U* test for corrected acceptance rates between the font colour versus colour name study conditions

Font colour versus colour name	Frequent configuration Neutral	Infrequent configuration Neutral	Frequent configuration Reinstated	Infrequent configuration Reinstated
Bayes factor $BF_{10}$	0.343	0.296	0.283	0.444
Decision	Weak support for $H_0$	Moderate support for $H_0$	Moderate support for $H_0$	Weak support for $H_0$
Mann–Whitney test statistics	$W = 348.50$	$W = 386.00$	$W = 370.50$	$W = 445.50$

**Table 4** Results of the Bayesian Mann–Whitney *U* test for the corrected acceptance rates between the neutral versus the reinstated test conditions

Neutral versus reinstated	Frequent configuration Font colour	Infrequent configuration Font colour	Frequent configuration Colour name	Infrequent configuration Colour name
Bayes factor	BF <sub>-0</sub> =0.119	BF <sub>+0</sub> =0.326	BF <sub>-0</sub> =0.176	BF <sub>+0</sub> =0.611
Decision	Moderate support for H <sub>0</sub>	Moderate support for H <sub>0</sub>	Moderate support for H <sub>0</sub>	Weak support for H <sub>0</sub>
Mann–Whitney test statistics	W=481.00	W=382.00	W=361.00	W=348.50

**Table 5** Group-level parameter estimates (standard deviations) and 95% BCIs of the dual-recollection multinomial model obtained in Experiment 2

Parameter	Frequent configuration Font colour	Infrequent configuration Font colour	Frequent configuration Colour name	Infrequent configuration Colour name
<b>Neutral condition</b>				
<i>RT</i>	0.575 (0.065) [0.444, 0.701]	0.572 (0.061) [0.454, 0.692]	0.406 (0.056) [0.299, 0.520]	0.420 (0.064) [0.298, 0.551]
<i>RC</i>	0.585 (0.053) [0.475, 0.684]	0.482 (0.063) [0.351, 0.599]	0.599 (0.063) [0.467, 0.714]	0.548 (0.084) [0.367, 0.702]
<i>F</i>	0.176 (0.147) [0.005, 0.541]	0.243 (0.181) [0.009, 0.669]	0.227 (0.142) [0.011, 0.533]	0.271 (0.151) [0.018, 0.564]
<i>b<sub>C1</sub></i> = <i>b<sub>C2</sub></i>	0.052 (0.021) [0.017, 0.098]		0.042 (0.022) [0.009, 0.095]	
<i>b<sub>C1or2</sub></i>	0.042 (0.023) [0.007, 0.096]		0.043 (0.020) [0.011, 0.088]	
<b>Reinstated condition</b>				
<i>RT</i>	0.364 (0.087) [0.184, 0.534]	0.507 (0.099) [0.313, 0.707]	0.390 (0.051) [0.290, 0.492]	0.433 (0.069) [0.289, 0.564]
<i>RC</i>	0.457 (0.090) [0.272, 0.624]	0.461 (0.113) [0.223, 0.668]	0.492 (0.069) [0.343, 0.616]	0.339 (0.096) [0.141, 0.519]
<i>F</i>	0.268 (0.178) [0.014, 0.683]	0.193 (0.167) [0.005, 0.638]	0.177 (0.117) [0.008, 0.430]	0.255 (0.140) [0.018, 0.530]
<i>b<sub>Cf</sub></i>	0.065 (0.034) [0.013, 0.141]		0.050 (0.026) [0.010, 0.111]	
<i>b<sub>Ci</sub></i>	0.065 (0.037) [0.011, 0.151]		0.051 (0.027) [0.009, 0.113]	
<i>b<sub>Cf or i</sub></i>	0.053 (0.038) [0.005, 0.144]		0.054 (0.032) [0.008, 0.130]	

The italicized symbols are the parameters of the dual-recollection multinomial model: *RT*=target recollection, *RC*=context recollection, *F*=familiarity, and *b*=response bias depending on the probe question

**Table 6** Differences in the mean dual-recollection multinomial model parameters and 95% BCIs between the frequent and the infrequent configurations

Parameter (Frequent–Infrequent)	Neutral Font colour	Reinstated Font colour	Neutral Colour name	Reinstated Colour name
$\Delta RT$	0.002 [-0.161, 0.166]	-0.143 [-0.396, 0.106]	-0.012 [-0.172, 0.150]	-0.041 [-0.196, 0.125]
$\Delta RC$	0.102 [-0.036, 0.244]	0.003 [-0.216, 0.242]	0.046 [-0.105, 0.211]	0.153 [-0.041, 0.361]
$\Delta F$	-0.069 [-0.544, 0.383]	0.090 [-0.397, 0.562]	-0.042 [-0.419, 0.356]	-0.081 [-0.423, 0.272]
$\Delta b$	-	-0.001 [-0.09, 0.082]	-	-0.001 [-0.073, 0.068]

$\Delta RT$ =the difference in the target recollection parameter estimates,  $\Delta RC$ =the difference in the context recollection parameter estimates,  $\Delta F$ =the difference in the familiarity parameter estimates, and  $\Delta b$ =the difference in the response bias parameter estimates

A good model fit was indicated by the nonsignificant test results in the font colour reinstated location condition (T1:  $p=0.506$ , T2:  $p=0.452$ ), in the colour name reinstated location condition (T1:  $p=0.532$ , T2:  $p=0.584$ ), in the font colour neutral condition (T1:  $p=0.524$ , T2:  $p=0.195$ ), and

in the colour name neutral condition (T1:  $p=0.536$ , T2:  $p=0.362$ ).

The effects of within-subjects manipulation of the frequency of the colour/location configuration are shown in Table 6. The credible intervals included 0 for all the

**Table 7** Differences in the mean dual-recollection multinomial model parameters and 95% BCIs between the neutral and the reinstated location conditions

Parameter	Frequent configuration	Infrequent configuration	Frequent configuration	Infrequent configuration
Neutral – Reinstated	Font colour	Font colour	Colour name	Colour name
$\Delta RT$	<b>0.219</b> <b>[0.007, 0.441]</b>	0.072 [−0.159, 0.307]	0.021 [−0.124, 0.171]	−0.006 [−0.183, 0.183]
$\Delta RC$	0.135 [−0.071, 0.356]	0.038 [−0.216, 0.311]	0.124 [−0.060, 0.309]	0.230 [−0.025, 0.477]
$\Delta F$	−0.073 [−0.536, 0.401]	0.071 [−0.438, 0.577]	0.059 [−0.299, 0.438]	0.020 [−0.379, 0.424]

$\Delta RT$ =the difference in the target recollection parameter estimates,  $\Delta RC$ =the difference in the context recollection parameter estimates, and  $\Delta F$ =the difference in the familiarity parameter estimates

The difference when the credibility interval does not overlap with 0 is printed in bold font

**Table 8** Differences in the mean dual-recollection multinomial model parameters and 95% BCIs between the colour information presented as a font colour versus as a colour name

Parameter	Frequent configuration	Infrequent configuration	Frequent configuration	Infrequent configuration
Font Colour – Colour name	Neutral	Neutral	Reinstated	Reinstated
$\Delta RT$	<b>0.174</b> <b>[0.001, 0.344]</b>	0.150 [−0.028, 0.322]	−0.035 [−0.238, 0.157]	0.066 [−0.174, 0.308]
$\Delta RC$	−0.014 [−0.182, 0.159]	−0.063 [−0.272, 0.162]	−0.043 [−0.273, 0.177]	0.110 [−0.193, 0.399]
$\Delta F$	−0.051 [−0.441, 0.390]	−0.025 [−0.441, 0.481]	0.084 [−0.294, 0.529]	−0.074 [−0.441, 0.408]

$\Delta RT$ =the difference in the target recollection parameter estimates,  $\Delta RC$ =the difference in the context recollection parameter estimates, and  $\Delta F$ =the difference in the familiarity parameter estimates

The difference when the credibility interval does not overlap with 0 is printed in bold font

differences of the parameter estimates, indicating no substantial effect.

Between-subjects manipulation of the reinstated versus the neutral test condition demonstrated one substantial but unexpected effect—the target recollection parameter was higher for the neutral than the reinstated test condition in the case of the frequent configuration in the font colour condition (see Table 7).

The results of between-subjects manipulation of the colour presentation form are shown in Table 8. The only credibility interval not overlapping with 0 indicated that the target recollection parameter was substantially higher for the font colour than for the colour name condition in the case of the frequent configuration neutral condition.

### Bayesian analyses of the hypotheses about context recollection and response bias

For the purposes of Bayes factor analyses, we calculated the independent estimates of parameters for each

participant using the maximum likelihood fitting method implemented in the *multiTree* software (Moshagen 2010). The results of analyses conducted using this method and based on aggregated data are presented in "Appendix 4". Here, we focus on the hypotheses concerning the context recollection parameter.

**Within-subjects effects of base-rates on the context recollection and the response bias parameters** If memory is informed by the base-rates, the context recollection parameter should be higher for the frequent than for the infrequent configurations, therefore, the one-sided alternative hypothesis is  $RC_{\text{frequent}} > RC_{\text{infrequent}}$ . As Table 9 shows, moderate support for the alternative hypothesis was found in the font colour neutral condition, and moderate support in favour of the null hypothesis was obtained in the font colour reinstated condition. In the case of the colour name conditions, the Bayes factors were inconclusive. When it came to the response bias parameter, the Bayes factors indicated moderate evidence for the null hypothesis, both in the font colour reinstated condition,  $BF_{+0} = 0.112$ ,  $W = 38.00$ , and in the

**Table 9** Results of the Bayesian Wilcoxon Signed-Rank test comparisons of the context recollection parameter for the infrequent versus the frequent configurations

	Infrequent versus frequent	Neutral	Reinstated	Neutral	Reinstated
	Font colour	Font colour	Font colour	Colour name	Colour name
Bayes factor $BF_{-0}$	3.918	0.233	0.680	0.680	1.302
Decision	Moderate support for $H_1$	Moderate support for $H_0$	Moderate support for $H_0$	Weak support for $H_0$	Weak support for $H_1$
Wilcoxon test statistics	$W=118.50$	$W=145.00$	$W=164.50$	$W=164.50$	$W=138.00$

**Table 10** Results of the Bayesian Mann–Whitney  $U$  test comparisons of the context recollection parameter for the neutral versus the reinstated test conditions

Neutral versus reinstated	Frequent configuration	Infrequent configuration	Frequent configuration	Infrequent configuration
	Font colour	Font colour	Colour name	Colour name
Bayes factor	$BF_{-0}=0.203$	$BF_{+0}=0.205$	$BF_{-0}=0.156$	$BF_{+0}=1.117$
Decision	Moderate support for $H_0$	Moderate support for $H_0$	Moderate support for $H_0$	Weak support for $H_+$
Mann–Whitney test statistics	$W=421.50$	$W=356.00$	$W=332.50$	$W=316.00$

**Table 11** Results of the Bayesian Mann–Whitney  $U$  test comparisons of the context recollection parameter for the font colour versus the colour name study conditions

Font colour versus colour name	Frequent configuration	Infrequent configuration	Frequent configuration	Infrequent configuration
	Neutral	Neutral	Reinstated	Reinstated
Bayes factor $BF_{10}$	0.326	0.419	0.276	0.367
Decision	Moderate support for $H_0$	Weak support for $H_0$	Moderate support for $H_0$	Weak support for $H_0$
Mann–Whitney test statistics	$W=427.50$	$W=464.00$	$W=379.00$	$W=427.00$

colour name reinstated condition,  $BF_{+0}=0.243$ ,  $W=27.00$ . Therefore, the response bias seems to be not informed by the base-rates.

**Between-subjects effects of the reinstated versus the neutral test condition and of the form of colour information representation** As in previous analyses, the direction of the alternative hypothesis about the effects of reinstatement manipulation depended on the frequency of the colour/location configuration. As shown in Table 10, in most conditions, the data favoured the null hypothesis about the lack of any differences between the test conditions; Bayes factor was inconclusive only in the infrequent configuration colour name condition.

Finally, the context recollection parameter was compared between the font colour and the colour name conditions. As demonstrated in Table 11, the Bayes factors indicated weak or moderate support for the null hypothesis.

### Experiment 3: The role of the magnitude of base-rate difference: A follow-up study

The decision-making literature suggests that under some conditions people are sensitive to base-rates in their probability judgments. One of these conditions is when base-rates are made more extreme (Koehler 1996). In the final experiment, we manipulated the magnitude of the disproportion in base-rates between feature configurations to see if we would observe an effect of high disproportion on memory and response biases. In doing so, we also hoped that the experiment would help clarify the inconsistency of the results of Experiments 1 and 2 regarding the sensitivity of response bias to the base-rate.<sup>2</sup>

In Experiment 1 we obtained strong support for the hypothesis that the participants match their responses with the experienced frequency of contexts (e.g., Bayen and Kuhlmann 2011; Kuhlmann et al. 2012; Wulff et al. 2021). However, in Experiment 2, no difference in the response

<sup>2</sup> We would like to thank one of the reviewers for suggesting conducting such an experiment.

bias was found when the distractors were presented with frequent versus infrequent locations, and the Bayes factor indicated moderate evidence for the null hypothesis. This discrepancy may be due to a more salient difference in the proportions of feature configurations in Experiment 1 than in Experiment 2. In Experiment 1, Colour 1 was four times more probable than Colour 2 for a large font (36:9), but in Experiment 2, however, Colour 1 was only about 1.7 times more probable than Colour 2 (30:18); hence, it was more difficult for the participants to notice this difference in the base-rate. An alternative explanation is that participants in Experiment 2 could be aware of the base-rates, but they did not rely on them in guessing strategy due to their metamemory judgment of good memory (indeed their memory was much better in Experiment 2 than in Experiment 1), which did not require notable adjustment.

In Experiment 3, we used generally the same materials and procedure as in the font colour location reinstated condition of Experiment 2, but we varied the base-rates and increased the number of stimuli. Higher number of items in Experiment 3 (60 targets and 30 distractors) in comparison with Experiment 2 (48 targets and 18 distractors) should decrease participants' confidence in memory judgments giving more room for guessing. Concerning the base-rates manipulation, we compared a more salient disproportion with a less salient disproportion of features in particular locations, that is, in the *high disproportion condition*, one colour was four times more probable than the other in a given location (48:12), and in the *low disproportion condition*, one colour was 1.5 times more probable than the other (36:24). These differences in base-rates are similar to those in Experiment 1 and Experiment 2, where the response bias effect was and was not observed, respectively.

Moreover, to minimize instances in which participants automatically respond to probe questions, mistakenly responding to a different question than the one actually asked, we distinguished the type of question with colour font corresponding to the content of the question (e.g., “Was this word presented in GREEN?” presented in green font colour). This procedural change can improve context reinstatement when the probe question matches the target context (Symeonidou and Kuhlmann 2021).

## Participants

In this experiment, 66 first year part-time psychology students took part in exchange for course credit. Their mean age was 23.00 years ( $SD = 5.91$ ), 12 were men. The number of participants per condition was similar as in Experiment 2. Each participant was randomly assigned to the high disproportion condition ( $n = 33$ ) or the low disproportion condition ( $n = 33$ ). Two participants who failed to follow the instructions were excluded, one from each group.

## Stimuli

From the same dataset (Imbir 2016) as in Experiments 1 and 2 we selected 94 nouns. They referred to objects that can occur in green or yellow colours and which can be found in a forest or a garden. The selected nouns had  $M = 5.9$  number of letters (range: 4–8),  $M = 1.9$  concreteness (range: 1.48–3.2),  $M = 7.92$  imageability (range: 7.54–8.42), and  $M = 354$  frequency of appearance in the language (range: 36–1357).

## Procedure and design

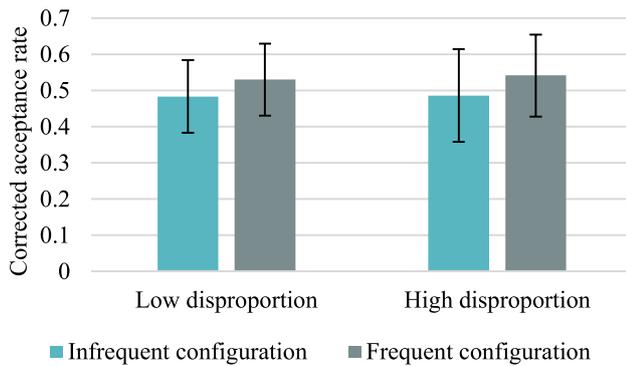
The participants were examined at individual workstations in the University Lab. The presentation of the stimuli and the response recording were conducted using the E-Prime 2.0 program.

At study, the participants were presented with 64 pairs of nouns (60 of them were targets, 2 were added as buffers at the beginning, and another 2 at the end of the study list). The first noun in each pair referred to one of the 64 objects, and the second referred to one of two places—a forest or garden. Each noun referring to an object was printed in the uppercase green or yellow font; the words referring to places were always printed in lowercase white font. Buffers at the beginning and at the end of the study list were presented in the more frequent colour-place configuration. As in Experiment 2, the participants were instructed to try to remember each object with its font colour and location and to create an image to help in remembering. The slides were displayed at a 6 s rate, with a 200 ms interstimulus interval. The stimuli were presented in *Arial* font, 48 pts size; the background screen was black.

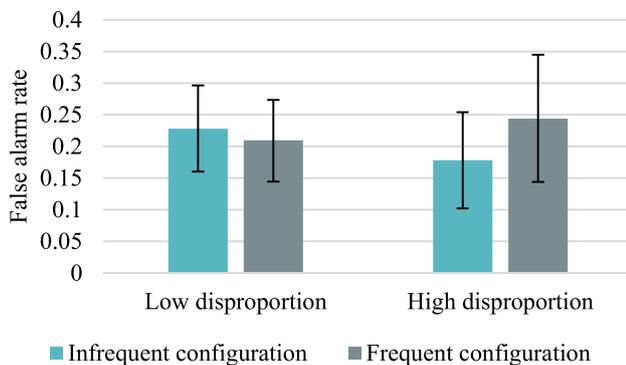
The frequency of configurations (frequent vs. infrequent colour-location pairs) was manipulated within subjects, but the magnitude of disproportion in base-rates was manipulated between subjects. In the high disproportion condition it was: 12 infrequent colour-location pairs (e.g., 6 green and garden plus 6 yellow and forest) and 48 frequent pairs (e.g., 24 yellow and garden plus 24 green and forest), and in the low disproportion condition it was: 24 infrequent pairs (e.g., 12 green and garden plus 12 yellow and forest) and 36 frequent pairs (e.g., 18 yellow and garden plus 18 green and forest).<sup>3</sup>

At test, 90 nouns referring to objects were presented: 60 targets and 30 distractors. As in the reinstated condition of Experiment 2, the targets were paired with the same location words as during the study phase. Half of the distractors were presented with one location word, the other half with

<sup>3</sup> Taking into account 4 buffer items presented only at study, these differences in base-rates are a bit higher; 12:54 in the high disproportion condition and 24:40 in the low disproportion condition.



**Fig. 8** Mean corrected-for-guessing acceptance rates for accurate colour recognition for frequent versus infrequent configurations of features in Experiment 3. Error bars represent 95% credible intervals



**Fig. 9** False alarm rates in Experiment 3. Error bars represent 95% credible intervals

the second location word. The three types of probe questions were: (a) Was this word presented in COLOUR 1?; (b) Was this word presented in COLOUR 2?; and (c) Was this word presented in either COLOUR 1 or COLOUR 2? The questions were presented in a font corresponding to the content of the question: green, yellow or white. The slides were presented in a random order, at a self-paced rate. The assignment of colours (green, yellow) to frequent or infrequent configurations and probe question types to the test items were counterbalanced across participants.

## Results and discussion

### Results based on descriptive measures

Descriptive statistics concerning the mean acceptance rates and the mean CARs for frequent versus infrequent colour/location configurations in high versus low disproportion conditions are presented in Tables 22 and 23 in "Appendix 5". The mean CARs for accurate responses are shown in Fig. 8, and false alarms for distractors are presented in Fig. 9.

### Within-subjects effects of base-rates

As in previous experiments, the one-sided alternative hypothesis is  $CAR_{\text{frequent}} > CAR_{\text{infrequent}}$ . In the low disproportion condition, the Bayesian Wilcoxon signed-rank test yielded weak evidence for the null hypothesis,  $BF_{+0} = 0.662$ . Similarly, in the high disproportion condition, the Bayesian  $t$  test yielded weak support for the null hypothesis,  $BF_{+0} = 0.480$ . Therefore, on the descriptive level of analysis, participants' memory seems to be not dependent on the base-rates, even when there is a high level of disproportion between frequent and infrequent configuration of features. This result replicates observations from Experiments 1 and 2.

In the case of false alarms, the Bayesian Wilcoxon signed-rank test indicated weak evidence in favour of the alternative hypothesis,  $BF_{+0} = 2.080$ , in the high disproportion condition, but moderate in favour of the null hypothesis,  $BF_{+0} = 0.117$ , in the low disproportion condition. Therefore, as in Experiment 2, the response bias was not informed by base-rates in the low disproportion condition, but in the high disproportion condition it is 2 times more probable that response bias is affected by the base-rates than that it is not affected.

### Between-subjects effects of the magnitude of base-rate difference

Table 12 shows the Bayes factors, all of which provide weak or moderate support for the null hypothesis that the

**Table 12** Results of the Bayesian Mann–Whitney  $U$  test for corrected acceptance rates between the high versus low disproportion conditions

High versus low disproportion	Frequent configuration Hits	Infrequent configuration Hits	Frequent configuration False alarms	Infrequent configuration False alarms
Bayes factor $BF_{10}$	0.261	0.272	0.268	0.396
Decision	Moderate support for $H_0$	Moderate support for $H_0$	Moderate support for $H_0$	Weak support for $H_0$
Mann–Whitney test statistics	$W = 526.00$	$W = 523.00$	$W = 500.50$	$W = 420.50$

**Table 13** Group-level parameter estimates (standard deviations) and 95% BCIs of the dual-recollection multinomial model obtained in Experiment 3

Parameter	Frequent configuration	Infrequent configuration	Frequent configuration	Infrequent configuration
	High disproportion	High disproportion	Low disproportion	Low disproportion
<i>RT</i>	0.460 (0.055) [0.345, 0.562]	0.507 (0.070) [0.358, 0.639]	0.442 (0.060) [0.321, 0.556]	0.383 (0.059) [0.262, 0.497]
<i>RC</i>	0.436 (0.058) [0.314, 0.543]	0.202 (0.070) [0.053, 0.331]	0.315 (0.066) [0.179, 0.438]	0.291 (0.068) [0.150, 0.418]
<i>F</i>	0.174 (0.098) [0.012, 0.371]	0.432 (0.225) [0.038, 0.872]	0.107 (0.082) [0.004, 0.300]	0.142 (0.111) [0.005, 0.409]
<i>b<sub>Cf</sub></i>	0.163 (0.057) [0.068, 0.289]		0.190 (0.036) [0.122, 0.262]	
<i>b<sub>Ci</sub></i>	0.120 (0.042) [0.048, 0.214]		0.210 (0.037) [0.139, 0.286]	
<i>b<sub>Cfori</sub></i>	0.057 (0.027) [0.014, 0.119]		0.140 (0.035) [0.077, 0.215]	

The italicized symbols are the parameters of the dual-recollection multinomial model: *RT*=target recollection, *RC*=context recollection, *F*=familiarity, and *b*=response bias depending on the probe question

magnitude of the disproportion does not affect accurate target recognition or false alarms to distractors.

### Results based on process measures

Table 13 presents the results of hierarchical multinomial processing tree modelling for both experimental conditions. An acceptable model fit was indicated by the nonsignificant test results in the high disproportion condition (T1:  $p = 0.494$ , T2:  $p = 0.342$ ), and in the low disproportion condition (T1:  $p = 0.385$ , T2:  $p = 0.055$ ).

In the high disproportion condition, the mean difference in context recollection parameters  $\Delta RC$  between frequent versus infrequent configurations ( $RC_{Cf} - RC_{Ci}$ ) was  $M = 0.240$ , with the credibility interval of the difference [0.069, 0.415] not overlapping with 0, indicating a substantial effect. In the low disproportion condition, no such effect was found for  $\Delta RC$ ,  $M = 0.024$  with 95% CI [-0.150, 0.197]. This result indicates that the context recollection becomes sensitive to the base rate when the disproportion between configurations is made very salient. It is worth noting that the target recollection parameter was numerically higher for the infrequent configuration than for the frequent configuration, indicating a dissociation between the recollection processes.

For the difference between response bias parameters ( $b_{Cf} - b_{Ci}$ ) no substantial effects were detected both in the high disproportion condition,  $M = 0.043$  with 95% CI [-0.052, 0.146], and the low disproportion condition,  $M = -0.02$  with 95% CI [-0.105, 0.066].

Between-subjects manipulation of the magnitude of disproportion in the base-rate yielded no substantial effect for the context recollection parameter, for the frequent configuration of features ( $RC_{Cf\text{high}} - RC_{Cf\text{low}}$ ),  $M = 0.119$  with 95% CI [-0.054, 0.290], and for the infrequent configuration of features ( $RC_{Ci\text{high}} - RC_{Ci\text{low}}$ ),  $M = -0.091$  with 95% CI [-0.285, 0.100].

### Bayesian analyses of the hypotheses about context recollection and response bias

For the purposes of Bayes factor analyses, we calculated the independent estimates of parameters for each participant using the maximum likelihood fitting method. The results of the analyses using this method, based on the aggregated data, are presented in Table 25 in "Appendix 6".

**Within-subjects effects of base-rates on the context recollection and the response bias** For the context recollection parameter, the Bayesian Wilcoxon signed-rank test indicated strong support for the alternative hypothesis in the high disproportion condition,  $BF_{+0} = 11.263$ ,  $W = 362.00$ , but moderate support in favour of the null hypothesis in the low disproportion condition,  $BF_{+0} = 0.196$ ,  $W = 254.00$ .

In similar vein, when it came to the response bias parameter, the Bayes factors indicated moderate evidence for the alternative hypothesis in the high disproportion condition,  $BF_{+0} = 5.624$ ,  $W = 206.00$ , but moderate evidence for the null hypothesis in the low disproportion condition,  $BF_{+0} = 0.148$ ,  $W = 178.00$ . Although the difference between the response bias parameters were not significant, Bayesian analyses supported the prediction that more salient differences in base rates influence participants' response strategy.

### Between-subjects effects of the magnitude of disproportion on the context recollection and the response bias parameters

The Bayesian Mann–Whitney *U* test favoured the null hypothesis about the lack of any differences between the high versus low disproportion conditions. For the context recollection parameter, Bayes factors weakly or moderately supported the null hypothesis, for frequent configurations,  $BF_{10} = 0.407$ ,  $W = 601.00$ , and infrequent configurations of features,  $BF_{10} = 0.314$ ,  $W = 452.00$ . For the response bias parameters, the Bayes factors also indicated weak or moderate support for the null hypothesis, in the case of guessing "yes" to frequent configuration,  $BF_{10} = 0.271$ ,  $W = 505.00$ ,

and in the case of guessing “yes” to infrequent configuration,  $BF_{10}=0.375$ ,  $W=419.00$ .

In sum, in Experiment 3 we found a significant difference in the context recollection parameter between frequent and infrequent configurations when differences in the base rates were high, but not when they were less salient. Bayes factor indicated strong support for base-rate dependency under high disproportion condition. Some support, although less unequivocal, was also provided to the hypothesis of the role of salient difference in base-rates on response bias.

## General discussion

In three experiments, using the Bayesian inference, we weighted the evidence in favour of the base-rate dependency versus the base-rate neglect hypotheses in memory for the correlated features. The first two studies and one condition of the third experiment, supported the base-rate neglect hypothesis. In Experiment 1, the memory for the font colour did not depend on whether this feature was equally or unequally distributed among the words printed in large versus small font. We found no support for such a dependency even when the correlated feature was reinstated at retrieval. This result was somewhat surprising, taking into account the stochastic dependence reported in many studies on multidimensional source memory (e.g., Arnold et al. 2019; Boywitt and Meiser 2012; Meiser 2014; Meiser and Bröder 2002) that was interpreted as indicating context-context binding in episodic memory. However, our results are consistent with the suggestions in literature that a stochastic dependence stems from item-context rather than from context-context binding (e.g., Hicks and Starns 2016; Starns and Hicks 2005; Vogt and Bröder 2007). Therefore, providing one contextual feature (font size) was probably incapable of assisting the retrieval of another contextual feature (font colour), since these context dimensions are not bound directly to one another (Hicks and Starns 2016). Another possibility is that context-context bindings do exist, but the effects of mutual cueing were undetectable since the contextual feature probed in memory test was very poorly encoded. In Experiment 1, the non-overlapping with 0 credibility intervals of all context recollection parameters for colours indicated that participants did remember colours, however, the *RC* values were relatively low, ranging from 0.046 to 0.219. Nevertheless, this interpretation is no longer convincing when we consider the results of Experiment 2, in which the *RC* values were much higher, ranging from 0.339 to 0.599, and of the low disproportion condition in Experiment 3, with the *RC* values from 0.291 to 0.315, where support for the base-rate dependency was not found either.

Better memory for context in Experiments 2 and 3 was probably achieved because we asked the participants to

create integrated images of the studied elements. In these experiments, we also tested a different form of representing features of the to-be-remembered episodes. In some of the experimental conditions, we used triplets of distinct elements representing the colour, object, and location. Following the studies on pattern completion, which suggested that the event elements are integrated into the holistic representations (e.g., Horner and Burgess 2013; Horner et al. 2015), we expected the more frequent pairings to be more effectively encoded into an episodic trace than the less frequent pairings, and that providing one of the elements at test will effectively cue another element—the stronger the binding between elements, the more effective the cueing will be. This was expected for contextual details represented by font colour, and even more so for features represented by separate items. However, in Experiment 2, we found that a manipulation of the colour/location configuration frequency did not affect memory for colour information, both when colour was represented as an intrinsic feature (font colour) and as a separate element (colour name). Instead, the Bayesian analyses supported the null hypothesis, indicating the lack of base-rate dependency. These results may be interpreted as suggesting that events are not encoded holistically (cf. Trinkler et al. 2006), at least when they do not comprise of unique triplets of elements, but two elements (colour and location) of these triplets are shared by multiple objects. Binding effects would probably be present for unique events, that is, such that are represented by specific elements.

Experiment 3 was designed with the goal of examining the role of extreme disproportion in base-rates on context memory and response bias. When there were 4 times as many targets in frequent than infrequent feature configurations, and one of the features was reinstated at test, we did find significantly higher context recollection for frequently presented than infrequently presented targets. It is noteworthy that this effect was not detected at the behavioural level, probably because context recollection and target recollection processes operated in an incompatible way. When the disproportion between base-rates was not so salient, and all other elements of the procedure and materials were the same, we again obtained support for the base-rate neglect hypothesis. The result observed in the high disproportion condition shows that under certain conditions, context recollection is sensitive to the base-rate. Similar boundary conditions for the base-rate neglect phenomenon have been found in the research on human probability judgements (Koehler 1996).

In our research, following the recent dual-recollection interpretation by Brainerd et al. (2022a), we assumed that contextual features are encoded into a contextual trace and conscious reinstatement of this trace proceeds as a phenomenon of context recollection. In one of the experiments validating the dual-recollection model, Brainerd et al. (2015, Experiment 2) presented smaller versus larger groups of

semantically related words during a study phase. A manipulation of the number of related items influenced the estimates of the context recollection contribution to memory performance but did not affect the target recollection or familiarity contribution. This result prompted us to assume that context recollection is sensitive to the number of occurrences of particular contextual features and that frequent feature-feature pairings will result in stronger contextual representation than infrequent pairings—reflecting the base-rates of features. However, in our experiments, standard comparisons of the context recollection parameter estimates between the conditions differing in the base-rates of the features indicated substantial effects only when extreme differences in base-rates were used. Similarly, the Bayesian analyses, with the exception of the high disproportion condition in Experiment 3, did not support the base-rate sensitivity hypothesis for the context recollection parameter. In most conditions, we found weak or moderate evidence in favour of the null hypothesis. It seems that the effects of the number of items sharing a contextual feature on context recollection are confined to the high disproportions in base-rates and to the semantic features and are harder to obtain for low or moderate disproportions and perceptual features such as colour.

Concerning the response bias sensitivity to the base-rates, we obtained strong (Experiment 1) or moderate (the high disproportion condition of Experiment 3) support for the hypothesis that the participants match their responses with the experienced frequency of contexts (e.g., Bayen and Kuhlmann 2011; Kuhlmann et al. 2012; Wulff et al. 2021). This effect was limited to conditions with a salient difference in the base-rates and disappeared when this disproportion was not so striking (Experiment 2 and the low disproportion condition of Experiment 3).

In our research, we did not assume that the base-rate dependence requires an intentional or explicit learning of the base-rates; as a matter of fact, there is no reason to assume that base-rate sensitivity cannot be acquired from implicit learning (cf. Wismer and Bohil 2017). Moreover, at least in Experiment 1, the participants were instructed at study that one colour font is more frequent in a particular font size than in another, and they used this knowledge to inform their guessing strategy, in spite of the fact that they neglected the base-rates in their memory for context.

To conclude this article, our findings supported the hypothesis that context memory is not sufficiently informed by the frequency of the feature pairings and is prone to base-rate neglect, at least when differences in the base-rates are not extreme. The conditional probability of one feature given another feature is not sufficiently reflected in the strength of the contextual memory trace. This memory insensitivity to the structure of real-world events is at odds with the observations of Anderson and Schooler (1991), but it resembles deep distortions; a family of memory biases

(overdistribution, super-overdistribution, non-additivity, and impossible conjunctions) that violate some axioms and rules of classical probability (Brainerd 2021, 2022). In the case of base-rate neglect in memory for context, the essential standard of the Bayes' theorem seems to be violated (Lu and Nieznański 2020). Further research is needed to define boundary conditions for the base-rate neglect in memory for contextual features. Our experiments indicated that such conditions can be created by using salient disproportion in base-rates simultaneously with facilitating feature binding. It would be worth investigating whether base-rate dependency would be present for 'integral' contextual features, in which changes in one dimension cannot be ignored when attention is paid to another dimension<sup>4</sup> (Garner 1974). Previous research has also suggested that base-rate dependency for context recollection can be observed when contexts are defined by groups of semantically related words (Brainerd et al. 2015).

## Appendix 1: Results of descriptive measures used in Experiment 1

See Tables 14 and 15.

**Table 14** Mean acceptance probabilities depending on the probe type in Experiment 1

	C1? <i>M</i> ( <i>SD</i> )	C2? <i>M</i> ( <i>SD</i> )	C1or 2? <i>M</i> ( <i>SD</i> )
Neutral condition			
C1L (n=36)	<b>0.692 (0.173)</b>	0.394 (0.185)	<b>0.654 (0.191)</b>
C1S (n=18)	<b>0.621 (0.233)</b>	0.388 (0.231)	<b>0.688 (0.211)</b>
C2L (n=9)	0.475 (0.310)	<b>0.617 (0.298)</b>	<b>0.775 (0.255)</b>
C2S (n=18)	0.471 (0.256)	<b>0.588 (0.207)</b>	<b>0.637 (0.220)</b>
New (n=42)	0.441 (0.265)	0.279 (0.208)	0.230 (0.140)
Reinstated condition			
C1L (n=36)	<b>0.691 (0.158)</b>	0.437 (0.209)	<b>0.698 (0.221)</b>
C1S (n=18)	<b>0.554 (0.252)</b>	0.518 (0.254)	<b>0.667 (0.242)</b>
C2L (n=9)	0.550 (0.344)	<b>0.595 (0.295)</b>	<b>0.784 (0.251)</b>
C2S (n=18)	0.482 (0.225)	<b>0.667 (0.272)</b>	<b>0.635 (0.269)</b>
NewL (n=21)	0.425 (0.277)	0.340 (0.268)	0.266 (0.241)
NewS (n=21)	0.347 (0.260)	0.336 (0.259)	0.301 (0.235)

Results printed in boldface indicate correct responses. C1, C2 denote Colour 1, Colour 2, respectively; L, S denote large font, small font, respectively

<sup>4</sup> We would like to thank one of the reviewers for pointing this interesting area of investigation.

**Table 15** Mean corrected acceptance probabilities for colour recognition in Experiment 1

	C1?	C2?	C1or2?
	<i>M</i> ( <i>SD</i> )	<i>M</i> ( <i>SD</i> )	<i>M</i> ( <i>SD</i> )
Neutral condition			
C1L (n = 36)	<b>0.251 (0.300)</b>	0.115 (0.203)	<b>0.424 (0.197)</b>
C1S (n = 18)	<b>0.180 (0.306)</b>	0.109 (0.278)	<b>0.457 (0.253)</b>
C2L (n = 9)	0.034 (0.414)	<b>0.338 (0.357)</b>	<b>0.545 (0.252)</b>
C2S (n = 18)	0.230 (0.321)	<b>0.309 (0.275)</b>	<b>0.407 (0.232)</b>
Reinstated condition			
C1L (n = 36)	<b>0.267 (0.350)</b>	0.097 (0.248)	<b>0.432 (0.303)</b>
C1S (n = 18)	<b>0.207 (0.296)</b>	0.182 (0.291)	<b>0.365 (0.289)</b>
C2L (n = 9)	0.125 (0.349)	<b>0.255 (0.424)</b>	<b>0.517 (0.321)</b>
C2S (n = 18)	0.135 (0.273)	<b>0.331 (0.356)</b>	<b>0.334 (0.365)</b>

Results printed in boldface indicate correct responses

## Appendix 2: Results of multinomial modelling analyses using maximum likelihood methods for Experiment 1

The probabilities of elementary cognitive states, which are represented by the model parameters, were estimated by applying maximum likelihood procedures. Traditionally, multinomial processing tree models are fitted using data aggregated across participants and items (Table 16). Such analyses are conducted under the homogeneity assumption that the individual frequencies are identically distributed (Smith and Batchelder 2008). However, this assumption is usually violated, therefore, in the main text of this paper, results based on more sophisticated Bayesian hierarchical

analyses that explicitly account for the heterogeneity of participants are provided. The hypotheses were tested with the log-likelihood ratio statistic ( $G^2$ ), which is distributed asymptotically as a  $\chi^2$  distribution. At  $\alpha$  level of 0.05,  $G^2(1) = 3.84$  indicates a critical value. The computations were carried out with the multiTree computer program (Moshagen 2010). In Experiment 1, sensitivity power analyses ensured a high test power ( $1 - \beta = 0.95$ ) for the parameter comparisons across the conditions. With  $\alpha = 0.05$  and the number of responses across participants ranging from 4551 in the reinstated condition to 4920 in the neutral condition, small effect sizes ranging from  $w = 0.051$  to  $0.053$  were detectable (analyses computed with G\*Power 3; Faul et al. 2007). The parameter estimates based on the aggregated data are presented in Table 17.

In the neutral condition, the only significant difference was found between the guessing parameters, in particular, guessing “yes” was significantly higher when the participants were asked about Colour 1 than when they were asked about Colour 2,  $\Delta G^2(1) = 32.31$ ,  $p < 0.001$ .

In the reinstated condition, the familiarity parameter ( $F$ ) was higher for the configuration Colour1 and small font than for the equally frequent configuration Colour 2 and small font,  $\Delta G^2(1) = 4.45$ ,  $p = 0.03$ . Moreover, the  $b$  guessing parameter was significantly higher for the distractors presented in large font when participants were asked about the more frequent Colour 1 than the less frequent Colour 2,  $\Delta G^2(1) = 3.96$ ,  $p < 0.05$ . This result supports the conclusion that the response bias is affected by the base-rates. The between-group comparisons of memory parameters revealed no significant difference.

**Table 16** Aggregated frequencies of responses in Experiment 1

Question probe	C1?		C2?		C1or2?	
	Yes	No	Yes	No	Yes	No
Neutral condition ( $N=40$ )						
C1L (n = 36)	332	148	189	291	314	166
C1S (n = 18)	149	91	93	147	165	75
C2L (n = 9)	57	63	74	46	93	27
C2S (n = 18)	113	127	141	99	153	87
New (n = 42)	247	313	156	404	129	431
Reinstated condition ( $N=37$ )						
C1L (n = 36)	307	137	194	250	310	134
C1S (n = 18)	123	99	115	107	148	74
C2L (n = 9)	61	50	66	45	87	24
C2S (n = 18)	107	115	148	74	141	81
NewL (n = 21)	110	149	88	171	69	190
NewS (n = 21)	90	169	87	172	78	181

**Table 17** Dual-recollection multinomial model parameter estimates (standard errors) for aggregated data in Experiment 1

	C1L (n=36)	C2L (n=9)	C1S (n=18)	C2S (n=18)
<b>Neutral condition</b>				
<i>RT</i>	0.303 (0.035)	0.312 (0.065)	0.226 (0.043)	0.277 (0.047)
<i>RC</i>	0.208 (0.040)	0.228 (0.059)	0.123 (0.056)	0.210 (0.043)
<i>F</i>	0.185 (0.084)	0.450 (0.114)	0.401 (0.080)	0.176 (0.099)
<i>b<sub>C1</sub></i>	0.441 (0.021)			
<i>b<sub>C2</sub></i>	0.279 (0.019)			
<i>b<sub>C1or2</sub></i>	0.230 (0.018)			
<b>Reinstated condition</b>				
<i>RT</i>	0.322 (0.041)	0.320 (0.064)	0.295 (0.043)	0.372 (0.047)
<i>RC</i>	0.209 (0.042)	0.097 (0.067)	0.030 (0.056)	0.184 (0.050)
<i>F</i>	0.233 (0.092)	0.520 (0.106)	0.304 (0.094)	0.000 (0.140)
<i>b<sub>C1</sub></i>	0.425 (0.031)		0.347 (0.030)	
<i>b<sub>C2</sub></i>	0.340 (0.029)		0.338 (0.029)	
<i>b<sub>C1or2</sub></i>	0.266 (0.027)		0.299 (0.028)	

The italicized symbols are the parameters of the dual-recollection multinomial model: *RT*=target recollection, *RC*=context recollection, *F*=familiarity, *b*=response bias for Colour1?, Colour2?, and Color1 or Colour2? probe questions

### Appendix 3: Results of descriptive measures used in Experiment 2

See Tables 18 and 19.

**Table 18** Mean acceptance probabilities depending on the probe type in Experiment 2

Experimental conditions	Ci? <i>M (SD)</i>	Cf? <i>M (SD)</i>	Ciorf? <i>M (SD)</i>
<b>Font colour – neutral</b>			
Frequent (n=30)	0.246 (0.137)	<b>0.832 (0.172)</b>	<b>0.846 (0.206)</b>
Infrequent (n=18)	<b>0.809 (0.174)</b>	0.315 (0.205)	<b>0.839 (0.195)</b>
New (n=18)	0.083 (0.124)		0.077 (0.147)
<b>Colour name – neutral</b>			
Frequent (n=30)	0.190 (0.132)	<b>0.779 (.193)</b>	<b>0.828 (0.171)</b>
Infrequent (n=18)	<b>0.770 (0.246)</b>	0.218 (0.184)	<b>0.833 (0.161)</b>
New (n=18)	0.095 (0.168)		0.052 (0.090)
<b>Font colour – reinstated</b>			
Frequent (n=30)	0.248 (0.228)	<b>0.730 (0.192)</b>	<b>0.833 (0.171)</b>
Infrequent (n=18)	<b>0.809 (0.225)</b>	0.278 (0.249)	<b>0.827 (0.204)</b>
New (n=18)	0.123 (0.199)	0.111 (0.196)	0.142 (0.280)
<b>Colour name – reinstated</b>			
Frequent (n=30)	0.225 (0.165)	<b>0.736 (0.195)</b>	<b>0.800 (0.170)</b>
Infrequent (n=18)	<b>0.696 (0.253)</b>	0.304 (0.261)	<b>0.804 (0.182)</b>
New (n=18)	0.089 (0.160)	0.077 (0.124)	0.119 (0.212)

Results printed in boldface indicate correct responses. Ci? denotes probe question referring to an infrequently presented colour, Cf? denotes probe question referring to a frequently presented colour, Ciorf? denotes probe question about any colour (infrequent or frequent)

**Table 19** Mean corrected acceptance probabilities for colour recognition in Experiment 2

	Ci? <i>M (SD)</i>	Cf? <i>M (SD)</i>	Ciorf? <i>M (SD)</i>
<b>Font colour – neutral</b>			
Frequent (n=30)	0.163 (0.115)	<b>0.749 (0.259)</b>	<b>0.769 (0.261)</b>
Infrequent (n=18)	<b>0.726 (0.234)</b>	0.232 (0.178)	<b>0.762 (0.254)</b>
<b>Colour name – neutral</b>			
Frequent (n=30)	0.095 (0.122)	<b>0.684 (0.285)</b>	<b>0.776 (0.207)</b>
Infrequent (n=18)	<b>0.675 (0.313)</b>	0.124 (0.137)	<b>0.782 (0.209)</b>
<b>Font colour – reinstated</b>			
Frequent (n=30)	0.125 (0.206)	<b>0.619 (0.324)</b>	<b>0.691 (0.361)</b>
Infrequent (n=18)	<b>0.685 (0.341)</b>	0.167 (0.217)	<b>0.685 (0.379)</b>
<b>Colour name – reinstated</b>			
Frequent (n=30)	0.136 (0.163)	<b>0.658 (0.250)</b>	<b>0.681 (0.298)</b>
Infrequent (n=18)	<b>0.607 (0.328)</b>	0.226 (0.245)	<b>0.684 (0.266)</b>

Results printed in boldface indicate correct responses

### Appendix 4: Results of multinomial modelling analyses using maximum-likelihood methods for Experiment 2

In Experiment 2, the sensitivity power analyses ensured a high test power ( $1 - \beta = 0.95$ ) for the parameter comparisons across the conditions. With  $\alpha = 0.05$  and the number of responses across participants ranging from 1782 in the font colour reinstated condition to 1914 in the colour name neutral condition, small effect sizes ranging from  $w = 0.082$  to  $0.085$  were detectable (*G\*Power 3*; Faul et al. 2007). Parameter estimates were based on the aggregated data presented in Table 20.

The parameter estimates of the model are presented in Table 21. When the context recollection parameter was compared between the frequent versus the infrequent configurations, only one difference on a trend level was found—in the colour name reinstated condition, the context recollection parameter was higher for the frequent than the infrequent configuration  $\Delta G^2(1) = 3.82, p = 0.05$ . In the remaining three conditions, the differences were nonsignificant,  $\Delta G^2s(1) < 2.55$ . In the reinstated location conditions, the response bias parameter was not different depending on the probe question referring to the colour frequently versus infrequently paired with the reinstated location,  $\Delta G^2s(1) < 0.16$ .

**Table 20** Aggregated frequencies of the responses in Experiment 2

	Ci?		Cf?		Ciorf?	
	Yes	No	Yes	No	Yes	No
Font colour – neutral condition ( $N=28$ )						
Frequent (n = 30)	69	211	233	47	237	43
Infrequent (n = 18)	136	32	53	115	141	27
New (n = 18)	28	308			13	155
Colour name – neutral condition ( $N=29$ )						
Frequent (n = 30)	55	235	226	64	240	50
Infrequent (n = 18)	134	40	38	136	145	29
New (n = 18)	33	315			9	165
Font colour – reinstated condition ( $N=27$ )						
Frequent (n = 30)	67	203	197	73	225	45
Infrequent (n = 18)	131	31	45	117	134	28
New (n = 18)	18	144	20	142	23	139
Colour name – reinstated condition ( $N=28$ )						
Frequent (n = 30)	63	217	206	74	224	56
Infrequent (n = 18)	117	51	51	117	135	33
New (n = 18)	13	155	15	153	20	148

**Table 21** Dual-recollection multinomial model parameter estimates (standard errors) for aggregated data in Experiment 2

Parameter	Frequent configuration Font colour	Infrequent configuration Font colour	Frequent configuration Colour name	Infrequent configuration Colour name
Neutral condition				
<i>RT</i>	0.558 (0.045)	0.589 (0.050)	0.406 (0.046)	0.433 (0.056)
<i>RC</i>	0.586 (0.034)	0.494 (0.047)	0.590 (0.033)	0.552 (0.045)
<i>F</i>	0.091 (0.177)	0.162 (0.200)	0.254 (0.128)	0.308 (0.153)
<i>b<sub>Ci</sub></i> = <i>b<sub>C2</sub></i>	0.083 (0.015)		0.095 (0.016)	
<i>b<sub>Cior2</sub></i>	0.077 (0.021)		0.052 (0.017)	
Reinstated condition				
<i>RT</i>	0.409 (0.043)	0.537 (0.057)	0.410 (0.042)	0.455 (0.046)
<i>RC</i>	0.485 (0.039)	0.528 (0.048)	0.514 (0.037)	0.389 (0.051)
<i>F</i>	0.361 (0.111)	0.077 (0.221)	0.207 (0.127)	0.331 (0.133)
<i>b<sub>Cf</sub></i>	0.111 (0.025)		0.077 (0.021)	
<i>b<sub>Ci</sub></i>	0.123 (0.026)		0.089 (0.022)	
<i>b<sub>Cfori</sub></i>	0.142 (0.027)		0.119 (0.025)	

The italicized symbols are the parameters of the dual-recollection multinomial model: *RT*=target recollection, *RC*=context recollection, *F*=familiarity, and *b*=response bias depending on the probe question

### Appendix 5: Results of descriptive measures used in Experiment 3

See Tables 22 and 23.

**Table 22** Mean acceptance probabilities depending on the probe type in Experiment 3

Experimental conditions	Ci?	Cf?	Ciorf?
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
High disproportion			
Frequent (n = 48)	0.313 (0.167)	<b>0.785 (0.149)</b>	<b>0.795 (0.159)</b>
Infrequent (n = 12)	<b>0.664 (0.281)</b>	0.492 (0.280)	<b>0.836 (0.207)</b>
New (n = 30)	0.178 (0.211)	0.244 (0.278)	0.106 (0.162)
Low disproportion			
Frequent (n = 36)	0.372 (0.212)	<b>0.740 (0.202)</b>	<b>0.740 (0.202)</b>
Infrequent (n = 24)	<b>0.711 (0.194)</b>	0.355 (0.242)	<b>0.723 (0.241)</b>
New (n = 30)	0.228 (0.189)	0.209 (0.178)	0.172 (0.175)

Results printed in boldface indicate correct responses

**Table 23** Mean corrected acceptance probabilities for colour recognition in Experiment 3

	Ci?	Cf?	Ciorf?
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
High disproportion			
Frequent (n = 48)	0.134 (0.211)	<b>0.541 (0.315)</b>	<b>0.689 (0.227)</b>
Infrequent (n = 12)	<b>0.486 (0.355)</b>	0.248 (0.377)	<b>0.730 (0.246)</b>
Low disproportion			
Frequent (n = 36)	0.144 (0.245)	<b>0.530 (0.276)</b>	<b>0.568 (0.235)</b>
Infrequent (n = 24)	<b>0.483 (0.278)</b>	0.146 (0.278)	<b>0.551 (0.272)</b>

Results printed in boldface indicate correct responses

**Table 24** Aggregated frequencies of the responses in Experiment 3

	Ci?		Cf?		Ciorf?	
	Yes	No	Yes	No	Yes	No
High disproportion condition ( <i>N</i> = 32)						
Frequent (n = 48)	160	352	402	110	407	105
Infrequent (n = 12)	85	43	63	65	107	21
New (n = 30)	57	263	78	242	34	286
Low disproportion condition ( <i>N</i> = 32)						
Frequent (n = 36)	143	241	284	100	284	100
Infrequent (n = 24)	182	74	91	165	185	71
New (n = 30)	73	247	67	253	55	265

### Appendix 6: Results of multinomial modelling analyses using maximum-likelihood methods for Experiment 3

In Experiment 3, the sensitivity power analyses ensured a high test power ( $1 - \beta = 0.95$ ) for the parameter comparisons. With the total number of 2880 responses across participants in the high or the low disproportion condition, small effect size ( $w$ ) of 0.067 was detectable (Faul et al. 2007). Parameter estimates were based on the aggregated data presented in Table 24.

The parameter estimates of the model are presented in Table 25. When the context recollection parameter was compared between the frequent versus the infrequent configurations, a significant difference was found in the high disproportion condition,  $\Delta G^2(1) = 14.501, p < 0.001$ , but not in the low disproportion condition,  $\Delta G^2(1) = 0.186$ . Moreover, in the high disproportion condition guessing parameter was significantly higher for the probe question about frequent than infrequent configuration,  $\Delta G^2(1) = 4.153, p = 0.04$ . When the context recollection parameter was compared between the high versus the low disproportion conditions, a significant difference was found only for the infrequent configuration of features,  $\Delta G^2(1) = 4.127, p = 0.04$ .

**Table 25** Dual-recollection multinomial model parameter estimates (standard errors) for aggregated data in Experiment 3

Parameter	Frequent configuration	Infrequent configuration	Frequent configuration	Infrequent configuration
	High disproportion	High disproportion	Low disproportion	Low disproportion
<i>RT</i>	0.480 (0.034)	0.490 (0.049)	0.474 (0.036)	0.425 (0.042)
<i>RC</i>	0.454 (0.030)	0.199 (0.060)	0.373 (0.035)	0.348 (0.043)
<i>F</i>	0.192 (0.102)	0.551 (0.106)	0.045 (0.122)	0.474 (0.036)
<i>b<sub>Cf</sub></i>	0.244 (0.024)		0.209 (0.023)	
<i>b<sub>ci</sub></i>	0.178 (0.021)		0.228 (0.023)	
<i>b<sub>Cfori</sub></i>	0.106 (0.017)		0.172 (0.021)	

The italicized symbols are the parameters of the dual-recollection multinomial model: *RT*=target recollection, *RC*=context recollection, *F*=familiarity, and *b*=response bias depending on the probe question

**Acknowledgements** We thank Yong Lu for the inspiration to investigate the base-rate neglect phenomenon in the domain of memory research. We would also like to thank anonymous reviewers for helpful comments and suggestions on a previous draft of this paper.

**Funding** This work was supported by the National Science Centre, Poland under Grant No. 2018/31/B/HS6/00511.

**Data availability** Raw data can be accessed online at: <https://osf.io/sz4vp/>.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethics approval** The article does not contain clinical studies or patient data. The experiments presented in this article do not require mandatory ethics review. The national and institutional guidelines have been followed.

**Informed consent** Informed consent was obtained from all individual participants included in the studies.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Anderson JR (1974) Retrieval of propositional information from long-term memory. *Cogn Psychol* 6:451–474
- Anderson JR, Schooler LJ (1991) Reflections of the environment in memory. *Psychol Sci* 2:396–408
- Arnold NR, Heck DW, Bröder A, Meiser T, Boywitt CD (2019) Testing hypotheses about binding in context memory with a

- hierarchical multinomial modeling approach. *Exp Psychol* 66:239–251
- Bayen UJ, Kuhlmann BG (2011) Influences of source–item contingency and schematic knowledge on source monitoring: tests of the probability-matching account. *J Mem Lang* 64:1–17
- Bayen UJ, Murnane K, Erdfelder E (1996) Source discrimination, item detection, and multinomial models of source monitoring. *J Exp Psychol Learn Mem Cogn* 22:197–215
- Bayen UJ, Nakamura GV, Dupuis SE, Yang C-L (2000) The use of schematic knowledge about sources in source monitoring. *Mem Cogn* 28:480–500
- Bell R, Mieth L, Buchner A (2020) Source attributions for detected new items: persistent evidence for schematic guessing. *Q J Exp Psychol* 73:1407–1422
- Boehm U, Marsman M, Matzke D, Wagenmakers E-J (2018) On the importance of avoiding shortcuts in applying cognitive models to hierarchical data. *Behav Res Methods* 50:1614–1631
- Boywitt CD, Meiser T (2012) Bound context features are integrated at encoding. *Q J Exp Psychol* 65:1484–1501
- Brainerd CJ (2021) Deep memory distortions. *Cogn Psychol* 126:101386
- Brainerd CJ (2022) Deep distortion. *Memory* 30:5–9
- Brainerd CJ, Reyna VF (1990) Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Dev Rev* 10:3–47
- Brainerd CJ, Reyna VF (2002) Fuzzy-trace theory and false memory. *Curr Dir Psychol Sci* 11:164–169
- Brainerd CJ, Reyna VF (2004) Fuzzy-trace theory and memory development. *Dev Rev* 24:396–439
- Brainerd CJ, Gomes CFA, Moran R (2014a) The two recollections. *Psychol Rev* 121:563–599
- Brainerd CJ, Holliday RE, Nakamura K, Reyna VF (2014b) Conjunction illusions and conjunction fallacies in episodic memory. *J Exp Psychol Learn Mem Cogn* 40:1610–1623
- Brainerd CJ, Gomes CFA, Nakamura K (2015) Dual recollection in episodic memory. *J Exp Psychol Gen* 144:816–843
- Brainerd CJ, Nakamura K, Reyna VF, Holliday RE (2017) Overdistribution illusions: categorical judgments produce them, confidence ratings reduce them. *J Exp Psychol Gen* 146:20–40
- Brainerd CJ, Bialer DM, Chang M (2022a) Fuzzy-trace theory and false memory: meta-analysis of conjoint recognition. *J Exp Psychol Learn Mem Cogn* 48(11):1680–1697
- Brainerd CJ, Bialer DM, Upadhyay P (2022b) A fundamental asymmetry in human memory: old ≠ not-new and new ≠ not-old. *J Exp Psychol Learn Mem Cogn* 48:1850–1867
- Ernst HM, Kuhlmann BG, Vogel T (2019) The origin of illusory correlations. *Exp Psychol* 66:195–206

- Faul F, Erdfelder E, Lang A-G, Buchner A (2007) G\*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 39:175–191
- Garner WR (1974) The processing of information and structure. Lawrence Erlbaum Associates, Mahwah
- Hanczakowski M, Zawadzka K, Coote L (2014) Context reinstatement in recognition: memory and beyond. *J Mem Lang* 72:85–97
- Heck DW, Arnold NR, Arnold D (2018) TreeBUGS: an R package for hierarchical multinomial-processing-tree modeling. *Behav Res Methods* 50:264–284
- Hicks JL, Starns JJ (2016) Successful cuing of gender source memory does not improve location source memory. *Mem Cognit* 44:650–659
- Hockley WE (2008) The effect of environmental context on recognition memory and claims of remembering. *J Exp Psychol Learn Mem Cogn* 34:1412–1429
- Horner AJ, Burgess N (2013) The associative structure of memory for multi-element events. *J Exp Psychol Gen* 142:1370–1383
- Horner AJ, Bisby JA, Bush D, Lin W-J, Burgess N (2015) Evidence for holistic episodic recollection via hippocampal pattern completion. *Nat Commun* 6(1):7462
- Imbir K (2016) Affective norms for 4900 Polish words reload (ANPW\_R): assessments for valence, arousal, dominance, origin, significance, concreteness, imageability and age of acquisition. *Front Psychol* 7:1081
- James E, Ong G, Henderson LM, Horner AJ (2020) Make or break it: boundary conditions for integrating multiple elements in episodic memory. *R Soc Open Sci* 7:200431
- JASP Team (2019) JASP (Version 0.16) [Computer software]. <https://jasp-stats.org/>
- Johnson MK, Hashtroudi S, Lindsay DS (1993) Source monitoring. *Psychol Bull* 114:3–28
- Kahneman D, Tversky A (1973) On the psychology of prediction. *Psychol Rev* 80:237–251
- Klauer KC (2010) Hierarchical multinomial processing tree models: a latent-trait approach. *Psychometrika* 75:70–98
- Koehler JJ (1996) The base rate fallacy reconsidered: descriptive, normative, and methodological challenges. *Behav Brain Sci* 19:1–17
- Kuhlmann BG, Vaterrodt B, Bayen UJ (2012) Schema bias in source monitoring varies with encoding conditions: support for a probability-matching account. *J Exp Psychol Learn Mem Cogn* 38:1365–1376
- Lu Y, Nieznański M (2020) The base rate neglect in episodic memory. *Memory* 28:270–277
- Mandera P, Keuleers E, Wodniecka Z, Brysbaert M (2014) Subtlex-pl: subtitle-based word frequency estimates for Polish. *Behav Res Methods* 47:471–483
- Mathôt S, Schreij D, Theeuwes J (2011) OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behav Res Methods* 44:314–324
- Meiser T (2014) Analyzing stochastic dependence of cognitive processes in multidimensional source recognition. *Exp Psychol* 61:402–415
- Meiser T, Bröder A (2002) Memory for multidimensional source information. *J Exp Psychol Learn Mem Cogn* 28:116–137
- Moshagen M (2010) multiTree: a computer program for the analysis of multinomial processing tree models. *Behav Res Methods* 42:42–54
- Nakamura K, Brainerd CJ (2017) Disjunction and conjunction fallacies in episodic memory. *Memory* 25:1009–1025
- Nieznański M (2013) Effects of resource demanding processing on context memory for context-related versus context-unrelated items. *J Cogn Psychol* 25:745–758
- Nieznański M, Tkaczyk D (2017) Effects of pictorial context reinstatement on correct and false recognition memory: insights from the simplified conjoint recognition paradigm. *J Cogn Psychol* 29:866–881
- Nieznański M, Obidziński M, Niedziałkowska D, Zyskowska E (2019) False memory for orthographically related words: research in the simplified conjoint recognition paradigm. *Am J Psychol* 132:57–69
- Reyna VF (2012) A new intuitionism: meaning, memory, and development in Fuzzy-Trace Theory. *Judgem Decis Mak* 7:332–359
- Riefer DM, Batchelder WH (1988) Multinomial modeling and the measurement of cognitive processes. *Psychol Rev* 95:318–339
- Smith JB, Batchelder WH (2008) Assessing individual differences in categorical data. *Psychon Bull Rev* 15:713–731
- Smith JB, Batchelder WH (2010) Beta-MPT: Multinomial processing tree models for addressing individual differences. *J Math Psychol* 54:167–183
- Spaniol J, Bayen UJ (2002) When is schematic knowledge used in source monitoring? *J Exp Psychol Learn Mem Cogn* 28:631–651
- Starns JJ, Hicks JL (2005) Source dimensions are retrieved independently in multidimensional monitoring tasks. *J Exp Psychol Learn Mem Cogn* 31:1213–1220
- Starns JL, Hicks JL (2013) Internal reinstatement hides cuing effects in source memory tasks. *Mem Cognit* 41:953–966
- Symeonidou N, Kuhlmann BG (2021) A novel paradigm to assess storage of sources in memory: the source recognition test with reinstatement. *Memory* 29:507–523
- Trinkler I, King J, Spiers H, Burgess N (2006) Part or parcel? Contextual binding of events in episodic memory. In: Zimmer HD, Mecklinger A, Lindenberger U (eds) *Handbook of binding and memory: perspectives from cognitive neuroscience*. Oxford University Press, Oxford, pp 53–83
- Tversky A, Kahneman D (1983) Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychol Rev* 90:293–315
- Tversky A, Koehler DJ (1994) Support theory: a nonextensional representation of subjective probability. *Psychol Rev* 101:547–567
- Van Doorn J, van den Bergh D, Böhm U, Dablander F, Derks K, Draws T et al (2020) The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychon Bull Rev* 28:813–826
- Vogt V, Bröder A (2007) Independent retrieval of source dimensions: an extension of results by Starns and Hicks (2005) and a comment on the ACSIM measure. *J Exp Psychol Learn Mem Cogn* 33:443–450
- Wisner AJ, Bohil CJ (2017) Base-rate sensitivity through implicit learning. *PLoS ONE* 12:e0179256
- Wulff L, Bell R, Mieth L, Kuhlmann BG (2021) Guess what? Different source-guessing strategies for old versus new information. *Memory* 29:416–426
- Yonelinas AP (2002) The nature of recollection and familiarity: a review of 30 years of research. *J Mem Lang* 46:441–517

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.