Title
On the formation of context-based person impressions

Authors
Huang, Lisa M
Sacchi, Dario LM
Sherman, Jeffrey W

Publication Date
2017

DOI
10.1016/j.jesp.2016.06.011

Peer reviewed

Journal of Experimental Social Psychology 68 (2017) 146–156

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

On the formation of context-based person impressions

Lisa M. Huang [a,*,1], Dario L.M. Sacchi [b,1], Jeffrey W. Sherman [a]

[a] *Department of Psychology, University of California, Davis, United States*
[b] *Department of Psychology, California State University, Sacramento, United States*

HIGHLIGHTS

- We relied on a model of category learning to study impression formation in context.
- Perceivers form multiple impressions of the same social target depending on context.
- Impressions from common contexts are learned before impressions from rare contexts.
- In rare contexts, attention shifts and focuses on differentiating attributes.
- Because of this learning process, impressions formed in rare contexts are stronger.

ARTICLE INFO

Article history:
Received 18 February 2016
Revised 29 June 2016
Accepted 29 June 2016
Available online 01 July 2016

Keywords:
Impression formation
Trait impressions
Context effects
Attention theory
Learning order
Person perception

ABSTRACT

The current research examined how people form context-based trait impressions and why some of these impressions are stronger than others. This research drew from principles of attention theory (Kruschke, 1996, 2001) in order to account for the processes underlying impression formation in context. According to attention theory, the traits expressed by an individual target person in a rarely occurring context should be more strongly associated with that context than the traits expressed in a commonly occurring context are associated with the common context. That is, people form stronger impressions of others' behavior in rare compared to common contexts. Four experiments provide support for these predictions. The current study is one of the first to examine the cognitive mechanisms by which perceivers form trait impressions of individuals across different contexts and to explain why some of these impressions are stronger than others. Implications of the nature of these impression formation processes are discussed.

Imagine that you meet a new colleague at work. Your colleague (let us call her Mary) proves herself to be a serious worker. When meeting with her colleagues or students, Mary rarely laughs, but instead engages thoughtfully and somberly in discussion. Now imagine that Mary invites you and a few other work colleagues over to her house to have dinner. To your surprise, she seems more easy-going and relaxed with her family, frequently cracking jokes and teasing her children. Although Mary is serious at work, she is relaxed with her family. You may know many people like Mary who behave one way in one context and a different way in another context. As an observer, how do you form impressions of these people and their behaviors?

As reflected in this scenario, the purpose of the current research is to examine how impressions of a person's traits are learned based on the context in which that person is encountered and how this learning process leads to different impression strengths across contexts. For example, you may form an impression of Mary as being somewhat serious at work, but you may form an impression of her as being *very* relaxed with her family. The current studies draw from the principles of attention theory of category learning (Kruschke, 1996, 2001) in order to examine how these impression formation processes occur. According to attention theory, because impressions in rarely occurring contexts are learned after impressions in commonly occurring contexts, rare context impressions are held more strongly than common context impressions. The sections that follow elaborate on these ideas.

1. Impression formation and context

A long history of impression formation research demonstrates the ways in which perceivers form trait impressions of other people. In some cases, perceivers routinely infer trait impressions from others'

* Corresponding author at: Department of Psychology, University of California, 135 Young Hall, One Shields Avenue, Davis, CA 95616, United States.
E-mail addresses: lmhuang@ucdavis.edu (L.M. Huang), dlsacchi@csus.edu (D.L.M. Sacchi), jsherman@ucdavis.edu (J.W. Sherman).
[1] Lisa M. Huang and Dario L. M. Sacchi contributed equally to this work. Order of co-first authorship was determined alphabetically.

http://dx.doi.org/10.1016/j.jesp.2016.06.011

behaviors, regardless of whether they explicitly intend to form those impressions (Srull & Wyer, 1989) or not (Uleman, Newman, & Moskowitz, 1996; Uleman, 1999). Other work suggests that perceivers form impressions in a more dynamic manner by simultaneously integrating high-level information, such as categories and stereotypes, with low-level information such as facial and vocal cues (Freeman & Ambady, 2011). In addition, much research suggests that perceivers will process behavioral information about an individual in a manner that maintains consistency in their impression of that individual (Hamilton & Sherman, 1996; Srull & Wyer, 1989). Yet, very little research has examined the mechanisms by which perceivers form impressions of people whose behavior differs across contexts. Attention theory provides one mechanism by which perceivers learn context-based impressions and how this learning process affects the relative strengths of trait impressions across contexts.

2. Attention theory and the inverse base rate effect

Attention theory was originally developed, in part, to account for the *inverse base rate effect* in human learning. In the original demonstration of the effect (Medin & Edelson, 1988), participants were asked to diagnose different diseases from patterns of symptoms. On each trial of the learning sequence, a list of symptoms was presented and participants were asked to diagnose the hypothetical patient as having one of several possible fictitious diseases. After each trial, participants were told the correct diagnosis. The basic design involved a pair of diseases, designated C (for common) and R (for rare), which occurred in a 3:1 ratio. During training, every instance of disease C had two symptoms, labeled I (for imperfect predictor; e.g., headache) and PC (for perfect predictor of the common disease; e.g., fever). Similarly, every instance of disease R had two symptoms, labeled I (headache again) and PR (for perfect predictor of the rare disease; e.g., stomachache). Thus, symptom PC always predicted disease C and never R, whereas symptom PR always predicted disease R and never C. Symptom I was an imperfect predictor of the two diseases, in that all cases of both C and R were associated with that symptom. Following training, participants were tested with combinations of symptoms not shown during training. When tested with the imperfect predictor I (headache) alone, people tended to choose the common disease, consistent with the base rates (i.e., during training, disease C occurred 75% of the time). However, when presented with the conflicting symptoms PC + PR (fever + stomach ache), participants tended to choose the rare disease, contrary (or inverse) to base rates.

Attention theory explains the effect as follows: During training, people first learn that symptoms I and PC are typical of disease C because that case occurs with high frequency. Subsequently, when learning about rare disease R, they realize that the shared symptom I is a misleading predictor because it already is associated with disease C. As a result, attention shifts away from I and is focused on the distinct symptom of R, PR, which has greater predictive diagnosticity than I. As a result, when learning about disease R, attention is focused primarily on a single, uniquely predictive symptom (PR), whereas, when learning about disease C, attention is divided between the symptoms PC and I (Fig. 1). More generally, the theory suggests that, as categories develop, greater attention is devoted to features that distinguish new categories from old ones than is devoted to features that define old ones. For these reasons, PR becomes more strongly associated with disease R than PC is associated with disease C. The relative strengths of these associations are evident from the inverse base-rate effect; when the perfect predictor symptoms are placed in direct competition with each other, the rare symptom carries more weight than the common symptom, indicating that the association and the rare disease and its symptom is stronger than the association between the common disease and its symptom (Kruschke, 1996).

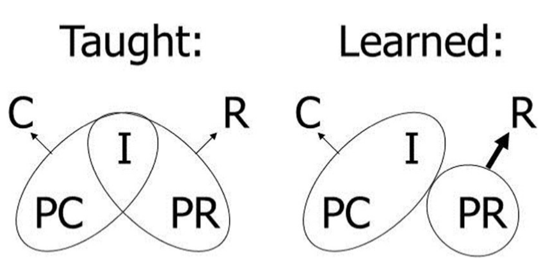

Fig. 1. Left: The core design of the inverse base-rate effect. C and R represent the common and rare disease, respectively. The symptom PC is a perfect predictor of C, the symptom PR is a perfect predictor of R, and the symptom I is an imperfect predictor of both C and R. Right: Depiction of what is learned according to attention theory.

To illustrate the concept using the above example, people first learn that a headache and fever predict disease C because disease C occurs with high frequency. They subsequently learn that a headache and stomachache predict disease R. However, because headache has already become associated with disease C, they focus their attention instead on the stomachache, which uniquely distinguishes disease R from disease C. For this reason, the stomachache becomes more strongly associated with disease R than the headache is associated with disease C. Therefore, when people are presented with the combination of fever and stomachache, they are more likely to choose disease R, contrary to base rates.

3. Attention theory and stereotype formation

Recent research has applied attention theory to understand the formation of group stereotypes (Sherman et al., 2009). In the same way that people learn to distinguish unique symptoms of diseases, they may also learn how to distinguish unique characteristics of social groups whose members are encountered with different frequency. Sherman et al. (2009) tested these predictions in a series of studies that used the inverse base-rate and illusory correlation (Hamilton & Gifford, 1976) paradigms. Their findings were consistent with attention theory predictions: when participants were asked to categorize individuals who possessed both perfect predictor traits (PC + PR), they were more likely to categorize the individual as a minority group member than as a majority group member. This finding indicates that the association between the minority group and its predictor trait is stronger than the association between the majority group and its perfect predictor trait. Further research also showed that similar processes may occur when perceivers classify individuals of mixed-race ancestry as minority group members rather than as majority group members (Halberstadt, Sherman, & Sherman, 2011).

4. The present research: Attention theory and context-based trait impressions

Theoretically, the same attentional processes that perceivers use to form associations between diseases and their symptoms and between social groups and their traits may also be used to form associations between particular contexts and the behaviors of single individuals. If the same processes occur, then the association that perceivers form between an individual's rare behavior and the context of that rare behavior should be stronger than the association they form between the same individual's common behavior and the context of that common behavior. Thus, perceivers should develop a stronger rare context than common context impression, and the rare impression should predominate when perceivers encounter the individual in a combined context. For example, consider again your colleague who behaves very differently when she is at work versus when she is with her family. If you frequently encounter this person in the work context and only later encounter her in the family context, attention theory would suggest that you will pay particularly close attention to whatever behaviors distinguish her family behavior from her work behavior. Subsequently, your impression of her in the family context will be held more strongly than your impression of her in the work context. Consider now the case in which you encounter your colleague when she is with both her work colleagues and her family members. Because your impression of your colleague with her family is stronger than your impression of her at work, your impressions of her family behavior should predominate when you encounter her with both work colleagues and family.

5. Experiment 1

5.1. Overview and predictions

The current research extends the logic of attention theory to the study of context-based impression formation. Experiment 1 attempted to replicate the inverse base-rate effect, but using associations between traits and contexts rather than between symptoms and diseases or between traits and social groups. Participants learned about a target person named Dave and how his traits varied according to context. Just as in the standard inverse base-rate paradigm, there was a common context (C; occurring 75% of the time) and a rare context (R; occurring 25% of the time). The contexts in the current experiment were represented by different colored rooms. There was a perfect predictor trait for the common room (PC), a perfect predictor trait for the rare room (PR), and an imperfect predictor trait (I) that was present for all descriptions of the target. During learning, participants were presented with different trait descriptions of Dave and had to guess which colored room he was in. In each trial, Dave always possessed a combination of two traits – a perfect predictor trait plus the imperfect predictor trait (PC + I or PR + I). After accurately learning the context-based trait impressions in the learning phase, participants completed a test phase in which they were presented with different configurations of the traits and were once again asked to guess which room Dave was in.

Our key prediction was that participants would form stronger impressions of Dave in the rare context than in the common context. In this case, context-based impressions refer to the associations between a person's traits and the contexts in which the traits are expressed. The strengths of those impressions are assessed via the categorization judgments participants make when the traits from the two contexts are placed into direct competition with one another. If rare context impressions are stronger than common context impressions, then rare traits should be weighted more heavily than common traits when they are placed in direct competition with each other. Therefore, participants should select the rare context more often than the common context when the target person possesses both perfect predictor traits (PC + PR).

5.2. Method

5.2.1. Participants

One hundred thirty-five undergraduates (88 females, 46 males, 1 unspecified) with a mean age of 19.29 years ($SD = 1.68$) from the University of California, Davis participated in exchange for partial course credit. Given an estimated effect size of $w = .19$ (taken from the inverse base-rate effect obtained in Sherman et al., 2009), an observation size of 1024, and an alpha of .05, the power to observe the predicted inverse base-rate effect is > .99. The observation size was determined by multiplying the number of participants (128 after exclusions; see below) by the number of inverse base-rate trials performed by each participant (8 trials).

5.2.2. Learning phase task

In this task, participants learned about a person named Dave. Participants were told that they would read descriptions about Dave and the traits that he possessed and that they would identify which room Dave was in for each description of him given. In each trial, participants read a description about Dave and two personality traits that he currently possessed (e.g., "Dave is Intelligent and Honest"). After reading each description, participants guessed the colored room by selecting the key corresponding to each of the four colored room options (Blue, Brown, Yellow, or Green). The Blue and Yellow rooms shared an imperfect predictor trait, as did the Brown and Green rooms. Thus, each participant learned two replications of the inverse base-rate design with two pairs of rooms. In addition, one of the room pairs always consisted of the traits Intelligent, Creative, and Honest and the other room pair always consisted of the traits Friendly, Loyal, and Reliable. Within each room pair, the status of the three traits was counterbalanced across participants, with each trait sometimes serving as the perfect predictor for the common room (PC), sometimes serving as the perfect predictor for the rare room (PR), and sometimes serving as the imperfect predictor (I). Also, the trait grouping that was associated with each room pair was counterbalanced across participants – for half of the participants, the traits Intelligent, Creative, and Reliable were grouped with the Blue and Yellow rooms and the traits Friendly, Loyal, and Reliable were grouped with the Brown and Green rooms. These groupings were reversed for the other half of participants. Following each trial, participants received feedback regarding the correct colored room (e.g., "Dave was in the Blue room"). The feedback was presented on the computer screen for 2000 ms before proceeding to the next trial.

The learning phase consisted of 12 blocks of trials, with eight trials per block. Within each block, the two traits that were perfect predictors of the common rooms (PC) occurred three times each (75% of trials) and the two traits that were perfect predictors of the rare rooms (PR) occurred one time each (25% of trials). The imperfect predictor traits (I) always occurred with PC and with PR. Thus, in each trial, Dave always possessed two traits, either PC + I or PR + I.

5.2.3. Test phase

After completing the learning phase, participants completed the test phase. In this phase, they were once again presented with descriptions of Dave and were asked to guess which colored room he was in. In these trials, however, Dave possessed novel combinations of the traits. For each colored room pair (Blue + Yellow, Brown + Green), there were two trials in which Dave possessed only the perfect predictor for the common room (PC), two trials in which he possessed only the perfect predictor for the rare room (PR), two trials in which he possessed only the imperfect predictor (I), and four trials in which he possessed both perfect predictors (PC + PR). Overall, there were a total of 20 trials in the test phase. Unlike in the learning phase, participants did not receive feedback after each trial regarding the correct colored room.

5.3. Results

5.3.1. Analytic method

First, accuracy rates from the learning phase were analyzed in order to determine learning order. Because common impressions occur more frequently than rare impressions, they should be learned prior to rare impressions. We reasoned that, if common impressions are learned first, then participants should learn the common impressions in fewer learning blocks than the rare impression (i.e., they should achieve greater accuracy on the common trials faster than on the rare trials). For each block, we calculated the proportion of correct responses for the common context trials and the proportion of correct responses for the rare context trials. Then, within each block, we subtracted the rare context accuracy from the common context accuracy. These difference values were regressed onto block number (1–12). If common impressions are learned before rare impressions, then the relationship between accuracy difference and block number should be negative; the accuracy of common context trials should be initially higher than the accuracy of rare context trials, but accuracy difference should decrease as learning of the rare context impression improves across successive blocks.

Second, we analyzed the test phase data in order to measure participants' selection of the context that Dave was in, given each novel combination of traits that he possessed. For each combination of traits (PC, PR, I, PC + PR), counts for each response option were summed across all participants. Then, context preference was analyzed using the chi-square goodness-of-fit test. Because there were four response options, with one pair of responses belonging to one iteration of the inverse base-rate design and another pair belonging to the second iteration of the design, the observed and expected values that were entered into the analyses were based only on the total number of responses from the correct response pair.

5.3.2. Learning phase

Six participants were excluded from analysis: three participants for incomplete data due to technical issues and another three participants for poor performance during the last block of the learning phase (<75% accuracy), indicating a failure to learn the associations between the traits and contexts. This left a total of 129 participants for analysis of the learning phase.

There was a significant negative relationship between accuracy difference and block number, $\beta = -.93$, $t(10) = -7.87$, $p < .001$, Adjusted $R^2 = .85$. Participants were more accurate on the common context trials than on the rare context trials in the earlier learning blocks, but the accuracy differences decreased as learning of the rare contexts improved across blocks. This result indicates that participants learned the common impressions prior to learning the rare impressions.

5.3.3. Test phase

One additional participant was excluded from the test phase analyses due to incomplete data, leaving a total of 128 participants for analyses.

First, results showed that participants successfully learned the perfect predictors associated with each context. When given the perfect predictors of the common contexts (PC), participants correctly selected the common contexts significantly more often (93%) than the rare contexts (4%), $\chi^2(1, N = 128) = 424.76$, $p < .001$, $w = .93$. Similarly, when given the perfect predictors of the rare contexts (PR), participants correctly selected the rare contexts significantly more often (92%) than the common contexts (4%), $\chi^2(1, N = 128) = 416.80$, $p < .001$, $w = .92$. When participants were given the imperfect predictor traits (I), they were more likely to select the common contexts (54%) than the rare contexts (34%), $\chi^2(1, N = 128) = 23.58$, $p < .001$, $w = .23$. This indicates that participants associated the imperfect predictor more with the common context than with the rare context. This shows that participants attended to base rate information; they learned appropriately that the common contexts occurred with greater frequency than the rare contexts.

Importantly, participants demonstrated the inverse base rate effect, as predicted. When given both perfect predictors (PC + PR), participants were more likely to select the rare context (53%) than the common context (43%), $\chi^2(1, N = 128) = 11.20$, $p < .001$, $w = .11$. This pattern of results is contrary to base rates, which alone suggest that participants should select the common contexts more often than the rare contexts because they occur with greater frequency.[2]

5.4. Discussion

We predicted that rare context impressions would be held more strongly than common context impressions. Results were consistent with these predictions. Participants demonstrated the inverse base-rate effect: they were more likely to assign the target to the rare context than to the common context when the target person possessed both perfect predictor traits (PC + PR). This result supports the idea that rare traits and contexts become more strongly associated with each other than do common traits and contexts.

6. Experiment 2

6.1. Overview and predictions

In Experiment 2, we directly replicated the Experiment 1 procedures and paradigm, except that we also investigated whether the second-learned impression would be stronger than the first-learned impression when the valence of the target's traits differed across the contexts.

Experiment 2 procedures were exactly the same as Experiment 1 procedures, except that participants learned about a target person who possessed both positive and negative traits, in contrast to Experiment 1, in which targets possessed only positive traits. For half of the participants, the perfect predictor of the common context was positive and the perfect predictor of the rare context was negative; thus, positive traits occurred with greater frequency than negative traits. For the other half of participants, the reverse was true – the perfect predictor of the common context was negative, whereas the perfect predictor of the rare context was positive; thus negative traits occurred with greater frequency than positive traits.

6.2. Method

6.2.1. Participants

Fifty-two undergraduates (30 females) with a mean age of 19.18 years ($SD = 1.42$) from the University of California, Davis participated in exchange for partial course credit. Given an estimated effect size of $w = .11$ (taken from the inverse base-rate effect obtained in Experiment 1), an observation size of 384, and an alpha of .05, the power to observe the predicted inverse base-rate effect is .58.[3] The observation size was determined by multiplying the number of participants (48 after exclusions; see below) by the number of inverse base-rate trials performed by each participant (8 trials).

6.2.2. Learning phase task

The learning phase in Experiment 2 was identical to the learning phase in Experiment 1, except that trait valence was manipulated between participants. For half of the participants, the common traits were positive and the rare traits were negative (CommonPos valence condition) and for the other half of participants, the common traits were negative and the rare traits were positive (CommonNeg valence condition). For all participants, the shared traits were positive. The positive traits used were the same as those used in Experiment 1 (Intelligent, Creative, Honest, Friendly, Loyal, and Reliable). These six traits were counterbalanced to represent either a perfect predictor trait (PC in the CommonPos condition and PR in the CommonNeg condition) or a shared trait (I). The negative traits used were Rude, Obnoxious, Cruel, Boring, Lazy, and Greedy. The negative traits were also counterbalanced.

[2] Across all four experiments, we included supplementary scale-based measures to assess the strength of trait impressions across contexts. In these measures, participants were asked to rate on a 1–8 scale the extent to which a target possessed a given trait depending on the context he was in. In all experiments, participants provided higher ratings on the perfect predictor traits (PC or PR) when the target was in the corresponding context. Additionally, in two out of the four experiments, participants provided higher ratings on the imperfect predictor trait (I) when the target was in the common compared to the rare context. Ratings on the PR trait when the target was in the rare context were generally equivalent to ratings on the PC trait when the target was in the common context. According to attention theory, we expected the former to be consistently higher than the latter. We suggest that these null findings could be due to a ceiling effect in ratings for the perfect predictor traits. Indeed, when trait ratings for the perfect predictors were not at ceiling (Negative condition of Experiment 4) we found evidence for stronger rare context impressions than common context impressions.

[3] The power in Experiment 2 was substantially lower than in Experiments 1, 3, and 4 (see below). The sample includes all the participants we were able to obtain before the end of the academic quarter. The experiments in this paper were conducted prior to the field's increased focus on power and sample size. Regardless, there is high power in all experiments except for Experiment 2.

6.2.3. Test phase

After completing the learning phase, participants completed the test phase. The test phase procedure in Experiment 2 was identical to the test phase procedure in Experiment 1.

6.3. Results

6.3.1. Analytic strategy

The data were analyzed using the same methods as in Experiment 1, with the addition of analyses for examining the possible moderation of trait valence. In the learning phase data, the moderation of trait valence was analyzed by regressing the difference in accuracy rates between the common and rare trials onto a two-way interaction between block number and valence condition. In the test phase data, the moderation of trait valence was analyzed using the chi-square test of independence or Fisher's exact test when the expected value of at least one of the cells was less than five.

6.3.2. Learning phase

Four participants were excluded from analysis: one participant for forgetting instructions and three participants for poor performance during the last block of the learning phase (<75% accuracy), indicating a failure to learn the associations between the traits and contexts. Excluded participants were evenly distributed between the two trait valence conditions. A total of 48 participants remained for data analyses.

As in Experiment 1, there was a significant negative correlation between accuracy difference and block number, $\beta = -.49$, $t(20) = -2.86$, $p = .010$, Adjusted $R^2 = .33$, indicating that participants learned the common impressions before the rare impressions. There was also a marginal main effect of valence condition, $\beta = -.73$, $t(20) = -2.00$, $p = .059$, indicating that the overall magnitude of difference between the common trait accuracy and the rare trait accuracy was larger in the CommonNeg condition than in the CommonPos condition. However, the block x valence interaction was not significant, $\beta = .41$, $t(20) = 1.13$, $p = .273$, indicating that the common traits were still learned prior to the rare traits, regardless of valence condition.

6.3.3. Test phase

Just as in the Experiment 1 test phase, participants in Experiment 2 also successfully learned the perfect predictor traits associated with each context. When given the perfect predictor traits of the common contexts (PC), participants correctly selected the common contexts significantly more often (94%) than the rare contexts (0%), $\chi^2(1, N = 48) = 179.00$, $p < .001$, $w = .99$. This finding was not moderated by the valence condition ($p > .999$). Similarly, when given the perfect predictor traits of the rare contexts (PR), participants correctly selected the correct rare contexts significantly more often (92%) than the common contexts (4%), $\chi^2(1, N = 48) = 151.58$, $p < .001$, $w = .91$. Although there was some evidence for moderation by the valence condition ($p = .068$), participants still selected the rare contexts significantly more often than the common contexts in both the CommonPos condition, $\chi^2(1, N = 23) = 83.05$, $p < .001$, $w = .98$, and the CommonNeg condition, $\chi^2(1, N = 25) = 71.02$, $p < .001$, $w = .86$.

When participants were given the imperfect predictor traits (I), they were more likely to select the common context (59%) than the rare context (31%), $\chi^2(1, N = 48) = 16.86$, $p < .001$, $w = .31$. Again, this indicates that participants were appropriately attending to base rate information during learning. This effect was moderated by trait valence, $\chi^2(1, N = 48) = 17.36$, $p < .001$, $w = .32$. In the CommonNeg condition (common traits were negative, rare traits were positive) there was no significant difference between the selection of the common context (49%) and the selection of the rare context (45%), $\chi^2(1, N = 25) = 0.17$, $p = .680$, $w = .04$. In the CommonPos condition (common traits were positive, rare traits were negative), however, participants selected the common context significantly more often (65%) than they selected the rare context (14%), $\chi^2(1, N = 23) = 32.92$, $p < .001$, $w = .65$.

Importantly, the inverse base rate effect shown in Experiment 1 was replicated in Experiment 2. When given both perfect predictors (PC + PR), participants were more likely to select the rare context (54%) than the common context (37%), $\chi^2(1, N = 48) = 11.34$, $p < .001$, $w = .18$. This basic inverse base rate effect was moderated by the valence condition, $\chi^2(1, N = 48) = 18.26$, $p < .001$, $w = .23$. In the CommonNeg condition, there was no significant difference between the selection of the common context (49%) and the selection of the rare context (45%), $\chi^2(1, N = 25) = 0.08$, $p = .773$, $w = .02$. In the CommonPos condition, however, participants selected the rare context significantly more often (57%) than they selected the common context (23%), $\chi^2(1, N = 23) = 29.27$, $p < .001$, $w = .43$.

6.4. Discussion

The results again suggest that second-learned, context-based trait impressions are stronger than first-learned, context-based trait impressions. Once again, participants demonstrated the inverse base-rate effect – when the target possessed both perfect predictor traits, participants were more likely to classify the target as being in the rare context than in the common context, contrary to base rates.

Interestingly, valence moderated the strength of the inverse base rate effect. The inverse base rate effect was observed when the common context was positive and the rare context was negative, but not when the reverse was true. Prior research on the negativity bias suggests that negative traits are weighted more heavily than are positive traits in impressions (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Fiske, 1980; Hamilton & Zanna, 1972; Pratto & John, 1991; Skowronski & Carlston, 1989). Because of biased attention toward negative traits, the inverse base rate effect may have been augmented when the rare traits were negative and the common traits were positive. The combined influence of numerical distinctiveness (i.e. stronger rare impressions than common impressions) and the stronger weight of negative traits may have led participants to be especially likely to categorize the target into the rare context when the target possessed both perfect predictor traits. However, when the rare traits were positive but the common traits were negative, the inverse base-rate effect may have been attenuated due to the opposing influences of numerical distinctiveness and negativity bias.

The valence condition also moderated how participants assigned the target to the common and rare contexts when the target possessed the imperfect predictor trait. In considering this finding, it is important to remember that the imperfect predictor traits were always positive and, as such, these traits may be expected to become more strongly associated with the evaluatively positive context (i.e. the context that is associated with the positive perfect predictor trait) than with the evaluatively negative context (i.e. the context that is associated with the negative perfect predictor trait), regardless of the frequency of trait context. In the CommonNeg condition, in which the positive traits were rare, participants may have been torn between assigning the imperfect predictor traits to the common context, which is consistent with base rates, and the rare context, which is evaluatively positive due to its association with a positive perfect predictor trait. In the CommonPos condition, however, participants are not faced with this conflict because the common context and the evaluatively positive context are the same.

7. Experiment 3

7.1. Overview and predictions

Attention theory argues that the learning order of context-based trait impressions is responsible for the attention shifting mechanism that causes stronger associations to be formed between rare traits and contexts than between common traits and contexts. Specifically, common impressions are learned prior to rare impressions. It is this learning order that leads to the rare impression (i.e. second-learned impression) being stronger than the common impression (i.e. first-learned impression).

The goals of Experiment 3 were twofold. First, we sought a more direct test of the role of learning order in the formation of stronger rare context impressions than common context impressions. Although Experiments 1 and 2 provided evidence that the common impression was learned prior to the rare impression, the learning of the two impressions was interwoven, providing a relatively weak test of the role of learning order. Therefore, in Experiment 3, we directly manipulated the learning order of context-based trait impressions.

The second and related goal of Experiment 3 was to rule out an alternative explanation for the inverse base-rate effect found in Experiments 1 and 2. The eliminative inference model, the major competing model of the inverse base-rate effect, argues that the inverse base-rate effect can be explained by a strategic response mechanism rather than an attentional mechanism (Juslin, Wennerholm, & Winman, 2001; Winman, Wennerholm, Juslin, & Shanks, 2005). According to this model, participants acquire a more firmly established knowledge structure of common events than of rare events because common events occur with greater frequency. When presented with the competing predictors PC + PR, participants judge the set to be sufficiently dissimilar to the firmly established rules associated with the common event and, therefore, eliminate the common event as a choice. Thus, they are more likely to select the rare event because there is less securely established knowledge about it. This model would imply that rare context impressions are not stronger than common context impressions. In fact, based on this model, common context impressions would be expected to be stronger because they have been more firmly established.

The logic of the eliminative inference model implies that one event (i.e., the common event) must occur more frequently than the other event (i.e., the rare event) in order for the inverse base-rate effect to be observed. That is the basis for elimination of the common event in the inverse base-rate effect. In contrast, attention theory does not require that one event occur more frequently than the other event in order to observe an inverse base-rate effect. According to attention theory, all that matters is that one event (or impression) is learned *before* the other, regardless of the ultimate frequency of the events. Regardless of ultimate frequency, attention will shift to distinguish the second event from the first one, leading to a stronger impression of the second event.

In Experiment 3, the contextualized impressions were presented with equal frequency. However, one impression was learned first. This permits a test of the competing models. In this case, participants should acquire equally established knowledge structures of the first-learned and second-learned impressions because they appear with equal frequency. If this is true, then according to the eliminative inference model, participants should select the first-learned and second-learned contexts equally often when given the competing predictors PC + PR. However, if participants show a preference for the second-learned context when given the competing predictors, then the data would support the attention theory model.

In Experiment 3, participants formed contextualized impressions in sequential order. Impressions occurred with equal frequency. In the first half of the learning phase, participants learned about the contexts that four different target people appeared in and the traits that they possessed when in those contexts. Participants learned about four targets rather than one target, as in Experiments 1 and 2, because the sequential learning task would otherwise be insufficiently challenging. As in the standard inverse base-rate paradigm, each target always possessed two traits – a perfect predictor and an imperfect predictor. In the second half of the learning phase, participants learned about the same targets but in new contexts. This time, the targets possessed new perfect predictor traits plus the same imperfect predictor traits that they learned about in the first half of the learning phase. As in Experiments 1 and 2, participants were told the traits that each target possessed and were then asked to guess the context that the target was in. Participants then completed a test phase that was similar to the previous experiments.

In accordance with the attention theory model, we predicted that participants would form stronger impressions of the targets in the second-learned contexts than in the first-learned contexts. This differential strength of impressions would be made evident by the inverse base-rate effect – when the targets possess both perfect predictor traits (i.e. the first-learned trait plus the second-learned trait), we expected that participants would be more likely to classify the targets as being in the second-learned context than in the first-learned context.[4]

Also, in order to test whether trait valence effects from Experiment 2 would replicate, participants in Experiment 3 learned that the targets possessed both positive and negative traits across contexts. Half of the participants learned that the targets possessed positive traits in the first context and negative traits in the second context and the other half learned that the targets possessed negative traits in the first context and positive traits in the second context.

7.2. Method

7.2.1. Participants

One hundred twenty-six undergraduates (97 females) with a mean age of 19.06 years ($SD = 1.42$) from the University of California, Davis participated in exchange for partial course credit. Given an estimated effect size of $w = .11$ (taken from the inverse base-rate effect obtained in Experiment 1), an observation size of 928, and an alpha of .05, the power to observe the predicted inverse base-rate effect is .92. The observation size was determined by multiplying the number of participants (116 after exclusions; see below) by the number of inverse base-rate trials performed by each participant (8 trials).

7.2.2. Learning phase task

The learning phase in Experiment 3 employed a variation of the procedures from Experiments 1 and 2. Rather than learning about contexts and traits that occurred with greater or lesser frequency, participants learned about traits and contexts in sequential order. In addition, participants learned about four different target people (Steve, Bill, Chuck, and Dave) rather than a single target person. In the first half of the learning phase, participants were told that they would read descriptions of the traits that each of the targets possessed and that they would identify which room a target was in for each description of him given. As in Experiments 1 and 2, the contexts were represented by colored rooms. As in the standard inverse base-rate paradigm, the targets always possessed two traits – one unique predictor trait ("Trait1," akin to a perfect predictor trait of a common context) and one trait that the targets possessed in all trials for both the first and second halves of the learning phase ("Shared Trait," akin to the imperfect predictor trait). In addition, each of the targets was in a different colored room. The second half of the learning phase was identical to the first half, except that the perfect predictor traits in the first half were replaced with new perfect predictor traits ("Trait2," akin to a perfect predictor trait of a rare context). The targets were also in different contexts than those in the first half of the learning phase. To summarize, the participants learned about one set of traits and contexts in the first half of the learning phase (Trait1 + Shared Trait in Context1) and they learned about a different set of traits and contexts in the second half of the learning phase (Trait2 + Shared Trait in Context2).

The trait valence also was manipulated between subjects. As in Experiment 2, one perfect predictor trait for each target was always positive and the other perfect predictor trait was always negative. For half of the participants, the first-learned traits (Trait1) were positive and the

[4] This classification pattern is not technically an inverse base-rate effect because participants learn both impressions with equal frequency, but we will continue to use the term here for the sake of consistency.

second-learned traits (Trait2) were negative (POSfirst condition). For the other half of participants, the first-learned traits were negative and the second-learned traits were positive (NEGfirst condition). In each half of the learning phase, each of the four traits corresponded uniquely to each of the four targets. The positive traits used were: Friendly (Steve), Loyal (Bill), Intelligent (Chuck), and Honest (Dave). The negative traits used were: Greedy (Steve), Lazy (Bill), Cruel (Chuck), and Rude (Dave). For all participants, the Shared Trait was always neutral in valence. Each of the four Shared Traits also corresponded uniquely to each of the four targets. The Shared neutral traits were: Practical (Steve), Tidy (Bill), Careful (Chuck), and Curious (Dave).

In the first half of the learning phase, participants learned about Steve, Bill, Chuck, and Dave. Participants were told that they would read descriptions of the traits that each target possessed and that they would identify which room a target was in for each description of that target given. In each trial, participants read a description of a target and the two personality traits that he currently possessed (e.g. "Steve is Friendly and Practical"). After reading each description, participants guessed the colored room by selecting the key corresponding to each of four colored room options (Light Blue, Red, Purple, or Brown). Each of the four colored room options corresponded uniquely to each of the four targets; Steve was always in the Light Blue room, Bill was always in the Purple room, Chuck was always in the Red room, and Dave was always in the Brown room. Following each trial, participants received feedback regarding the correct colored room (e.g., "Steve was in the Light Blue room"). The feedback was presented on the computer screen for 2000 ms before proceeding to the next trial. The first half of the learning phase consisted of five blocks of trials, with eight trials per block. Each of the four target people were presented two times per block.

Participants completed the second half of the learning phase immediately after finishing the first half. The second half was identical to the first half, except that the first-learned traits were replaced with the second learned traits. Also, the first set of colored room contexts was replaced with the second set of colored room contexts, each of which also corresponded uniquely to each of the four targets: Orange (Steve), Yellow (Bill), Green (Chuck), and Dark Blue (Dave).

7.2.3. Test phase

After completing the learning phase, participants completed two blocks of the test phase. In the first test phase block, participants were told that they would once again be presented with descriptions of two of the target people, Steve and Bill, and would be asked to guess which colored room they were in. Each trial presentation consisted of either a description of Steve and a combination of his associated traits (i.e., Friendly, Greedy, and Practical), or a description of Bill and a combination of his associated traits (i.e., Loyal, Lazy, and Tidy). As in Experiments 1 and 2, each target person possessed a novel combination of traits. For each target person, there was one trial in which he possessed only the perfect predictor of the first context (Trait1), one trial in which he possessed only the perfect predictor of the second context (Trait2), one trial in which he possessed only the imperfect predictor (Shared Trait), and two trials in which he possessed both of the perfect predictor traits (Trait1 + Trait2). Overall, there were a total of 10 trials in the first test phase block. The four colored room response options consisted of the two contexts associated with Steve during the learning phase (Light Blue and Orange) plus the two contexts associated with Bill during the learning phase (Purple and Yellow). Unlike in the learning phase, participants did not receive feedback after each trial regarding the correct colored room.

The second test phase block was the same as the first test phase block, except that participants completed trials pertaining to the other two target people, Chuck and Dave. Also, the traits and contexts used in the second test phase block were those that were associated with Chuck and Dave during the learning phase.

7.3. Results

7.3.1. Analytic strategy

Learning accuracy was calculated as the percentage of correct responses in the last block of each half of the learning phase combined. Also, the average accuracy rate was separated for each trait valence condition (POSfirst and NEGfirst). A *t*-test was performed in order to check that learning was not higher in one condition than another.

The test phase data were analyzed using the same methods as in Experiment 1, with two exceptions. The counts for each response option were summed across all four of the target people that the participants learned about. Also, as in Experiment 2, the moderation of trait valence was analyzed using the chi-square test of independence.

7.3.2. Learning phase

Ten participants were excluded from analysis: two for incomplete data and eight for poor performance in the learning phase (<75% average accuracy in the last block of each half of the learning phase combined), indicating a failure to learn the associations between the traits and the contexts. Excluded participants were evenly distributed between the two trait valence conditions. A total of 116 participants remained for data analyses.

Accuracy during learning was high in both trait valence conditions (POSfirst – $M = .95$, $SD = .07$; NEGfirst – $M = .95$, $SD = .08$). There was no significant effect of valence on accuracy, $t(114) = 0.24$, $p = .814$, $d = 0.04$.

7.3.3. Test phase

Participants successfully learned the trait-context associations. When given Trait1, participants correctly selected the correct Context1 that corresponded to that trait significantly more often (54%) than they selected the corresponding Context2 (15%), $\chi^2(1, N = 116) = 110.36$, $p < .001$, $w = .49$. Furthermore, this finding was not moderated by valence condition $\chi^2(1, N = 116) = 0.01$, $p = .920$, $w < .01$. When given Trait2, participants correctly selected the correct Context2 that corresponded to that trait significantly more often (59%) than they selected the corresponding Context1 (18%), $\chi^2(1, N = 116) = 104.54$, $p < .001$, $w = .47$. This finding also was not moderated by valence condition, $\chi^2(1, N = 116) = 0.66$, $p = .417$, $w = .04$.

When participants were given the shared trait, participants selected that trait's corresponding Context1 (31%) and corresponding Context2 (36%) about equally, $\chi^2(1, N = 116) = 1.20$, $p = .273$, $w = .05$. Contrary to the results of Experiment 2, this finding was not moderated by valence condition, $\chi^2(1, N = 116) = 0$, $p > .999$, $w < .01$.

When participants were given both perfect predictor traits (Trait1 + Trait2), they selected Context2 (40%) more frequently than Context1 (35%). Although this difference is in the expected direction, it was not reliable, $\chi^2(1, N = 116) = 2.48$, $p = .115$, $w = .05$. Contrary to the results of Experiment 2, this finding was not moderated by trait valence, $\chi^2(1, N = 116) = 0.72$, $p = .396$, $w = .03$. Overall, these results only weakly support the prediction that the second-learned impression is stronger than the first-learned impression.

7.4. Discussion

In Experiment 3, we directly manipulated the learning order of context-based trait impressions in order to directly test whether learning order does, in fact, lead to different impression strengths. Rather than learning about traits and contexts that occurred with differing frequency, participants learned context-based trait impressions in sequential order that occurred with equal frequency. Results only weakly support the causal influence of learning order. When a target possessed both perfect predictor traits (Trait1 + Trait2), participants were not significantly more likely to assign the target to the second-learned context (Context2) than to the first-learned context (Context1), though results were trending in the expected direction. In contrast to Experiment 2, valence did not moderate the classification pattern.

In this study, we also attempted to rule out an alternative explanation for the data based on the eliminative inference model. This model suggests that participants should have developed equally established knowledge structures for the first-learned and second-learned context impressions because they occurred with equal frequency. If this were the case, then participants would have selected Context1 and Context2 equally when given both perfect predictor traits (Trait1 + Trait2). Although the trend in the data is consistent with attention theory, the eliminative inference model cannot be entirely ruled out. Another possibility is that attention shifting does occur when perceivers learn rare impressions, but that it is not caused by learning order.

Attention theory also states that learning order leads to the imperfect predictor trait being more strongly associated with the first-learned context than the second-learned context. We would expect, then, that participants would be more likely to classify the target as being in Context1 than in Context2 when the target possesses the Shared trait. Contrary to these predictions, however, participants selected Context1 and Context2 about equally when a target possessed the Shared trait. However, these results do show that participants were classifying the target in accordance with the base rate frequencies (i.e. 50:50 base-rate appearance of first-learned and second-learned trait impressions).

8. Experiment 4

8.1. Overview and predictions

Results from Experiments 1and 2 show that participants formed stronger impressions of targets in rare contexts than in common contexts. In these two experiments, participants were presented with a target person possessing given traits and they then were asked to guess which context the target was in. In real life situations, however, people are probably more likely to infer another person's traits based on the context in which that person is encountered than to infer the context based on the traits. For example, people are more likely to infer Mary's behavior depending on whether she is at work or at home than to infer where she is depending on her behavior. Therefore, in Experiment 4, we examined whether the same pattern of impression strengths would hold if we reversed the classification order. We presented participants with the same frequency learning paradigm as in Experiments 1 and 2, except that we switched which information about the target was given and which classification needed to be learned. Specifically, participants were presented with Dave in different contexts and they then had to guess which trait he possessed in that context. It is important to note, then, that the *traits* (rather than the contexts) now represent the common (C) and rare (R) events and the *contexts* (rather than the traits) represent the perfect (PC, PR) and imperfect (I) predictors of those events.

A few other changes are worth noting. In the current experiment, the contexts were represented by the people that Dave was with, rather than by colored rooms. Using a different type of context allowed us to test the generalizability of our findings. We also tested whether trait valence would influence the strength of the context-based trait impressions and the classifications made during the test phase. However, whereas the targets in Experiments 2 and 3 had both positive and negative traits, the target in the current experiment possessed either all positive traits or all negative traits. That is, he was evaluatively consistent across both the common and rare contexts.

8.2. Method

8.2.1. Participants

One hundred eight undergraduates (63 females) with a mean age of 19.41 years ($SD = 1.53$) from the University of California, Davis participated in exchange for partial course credit. Given an estimated effect size of $w = .11$ (taken from the inverse base-rate effect obtained in Experiment 1), an observation size of 792, and an alpha of .05, the power to observe the predicted inverse base-rate effect is .87. The observation size was determined by multiplying the number of participants (99 after exclusions; see below) by the number of inverse base-rate trials performed by each participant (8 trials).

8.2.2. Learning phase task

The learning paradigm in Experiment 4 was similar to the learning paradigms in Experiments 1 and 2. However, there were several differences. First, the context in Experiment 4 was represented by the people that Dave was with, rather than the colored rooms that he was in. Second, we switched which information about Dave was given and which classification needed to be learned. In Experiments 1, 2, and 3, participants were given information about Dave's traits and then had to guess which context he was in. In Experiment 4, participants instead were given the contexts and then asked to guess which trait Dave possessed. Just as how Dave always possessed two traits at a time in Experiments 1, 2, and 3, so too was he always with two people at a time in Experiment 4. Third, we included a between-groups valence condition. However, whereas participants in both valence conditions in Experiments 2 and 3 learned that Dave had both positive and negative traits, participants in Experiment 4 learned that Dave had either all positive traits (Positive condition) or all negative traits (Negative condition). The positive traits used were: Creative, Friendly, Honest, and Intelligent. The negative traits used were: Cruel, Greedy, Lazy, and Rude.

In the learning phase of Experiment 4, participants were told that they would read descriptions of Dave and the people he was with, and that they would identify which trait Dave possessed for each description of him given. In each trial, participants read a description about Dave and the two people he was with (e.g., “Dave is with Bob and Chris”). After reading each description, participants guessed Dave's trait by selecting the key corresponding to each of four trait options (e.g. Creative, Friendly, Honest, Intelligent). Two of the traits (e.g. Creative and Honest) always shared an imperfect predictor context, as did the other two traits (e.g. Friendly and Intelligent). Thus, each participant learned two replications of the inverse base-rate design with two pairs of traits. In addition, one of the trait pairs always consisted of the context people Bob, John, and Chris, and the other trait pair always consisted of the context people Luke, Tom, and Nick. Within each trait pair, the status of the three context people was counterbalanced across participants, with each person sometimes serving as the perfect predictor for the common trait (PC), sometimes serving as the perfect predictor for the rare trait (PR), and sometimes serving as the imperfect predictor (I). Also, the context person group that was associated with each trait pair was counterbalanced across participants – for half of the participants, Bob, John, and Chris were grouped with the traits Creative and Honest and Luke, Tom, and Nick were grouped with the traits Friendly and Intelligent. These groupings were switched for the other half of participants. Following each trial, participants received feedback regarding the correct trait (e.g., “Dave was Creative”). The feedback was presented on the computer screen for 2000 ms before proceeding to the next trial.

The learning phase consisted of 12 blocks of trials, with eight trials per block. Within each block, the two context people that were perfect predictors of the common traits (PC) occurred three times each (75% of trials), and the two context people that were perfect predictors of the rare traits (PR) occurred one time each (25% of trials). The imperfect context person predictors (I) always occurred with PC and with PR. Thus, in each trial, Dave was always with two people, either PC + I or PR + I.

8.2.3. Test phase

After completing the learning phase, participants completed the test phase. The test phase in Experiment 4 was similar to the test phases in Experiments 1–3. However, participants were presented with the contexts and then asked to guess the trait, rather than being presented with the traits and then asked to guess the context. In this task, participants were once again presented with descriptions of Dave and the people he was with, and they then were asked to guess which trait he possessed. In these trials, however, Dave was with novel combinations of people. For each trait pair (e.g., Creative + Honest; Friendly + Intelligent), there were two trials in which Dave was with only the perfect predictor context person for the common trait (PC), two trials in which he was with only the perfect predictor person for the rare trait (PR), two trials in which he was with only the imperfect predictor person (I), and four trials in which he was with both of the perfect predictor people (PC + PR). Overall, there were a total of 20 trials in the test phase. Unlike in the learning phase, participants did not receive feedback after each trial regarding the correct trait.

8.3. Results

8.3.1. Analytic strategy

The data were analyzed using the same methods as in Experiment 2.

8.3.2. Learning phase

Nine participants were excluded from analyses – three due to a programming error and six for poor performance during the last block of the learning phase (<75% accuracy), indicating a failure to learn the associations between the contexts and the traits. Three of the excluded participants were from the positive valence condition, and six were from the negative valence condition. A total of 99 participants remained for data analyses.

The results in the current experiment replicated those from Experiments 1 and 2. There was a significant negative relationship between accuracy difference and the block number, $\beta = -.70$, $t(20) = -4.66$, $p < .001$, Adjusted $R^2 = .48$, indicating that participants learned the common impressions prior to the rare impressions. There was no main effect of the valence condition, $\beta = -.11$, $t(20) = -0.34$, $p = .737$, nor was there a significant Block Number x Valence Condition interaction, $\beta = -.15$, $t(20) = -0.47$, $p = .645$. These data indicate that common contexts were learned before the rare contexts, regardless of trait valence.

8.3.3. Test phase

First, results showed that participants learned the correct associations between the contexts and the traits. When given the perfect predictor context people of the common traits (PC), participants correctly selected the common traits significantly more often (89%) than the rare traits (2%), $\chi^2(1, N = 99) = 329.71$, $p < .001$, $w = .96$. This finding was not moderated by the valence condition ($p > .999$). Similarly, when given the perfect predictors of the rare traits (PR), participants correctly selected the rare traits significantly more often (88%) than the common traits (6%), $\chi^2(1, N = 99) = 289.20$, $p < .001$, $w = .88$. This finding also was not moderated by the valence condition, $\chi^2(1, N = 99) = 0.35$, $p = .552$, $w = .03$.

When participants were given the imperfect predictor people (I), they were more likely to select the frequent trait (63%) than the rare trait (22%), $\chi^2(1, N = 99) = 77.57$, $p < .001$, $w = .48$. Again, these results indicate that participants were appropriately attending to base rate information during learning. This effect was not moderated by trait valence, $\chi^2(1, N = 99) < 0.01$, $p = .955$, $w < .01$.

Importantly, the inverse base-rate effect was replicated. When given both of the perfect predictor context people (PC + PR), participants were more likely to select the rare trait (49%) than the common trait (43%), $\chi^2(1, N = 99) = 4.00$, $p = .046$, $w = .07$. In addition, this effect was moderated by the valence condition, $\chi^2(1, N = 99) = 5.37$, $p = .021$, $w = .09$. In the Positive condition, there was no significant difference between selection of the common trait (46%) and selection of the rare trait (45%), $\chi^2(1, N = 52) = 0.02$, $p = .878$, $w = .01$. In the Negative condition, however, participants were significantly more likely to select the rare trait (54%) than the common trait (39%), $\chi^2(1, N = 47) = 9.31$, $p = .002$, $w = .16$.

8.4. Discussion

Overall, the results of Experiment 4 replicated the inverse base-rate effect, which suggests that second-learned, context-based trait impressions are stronger than first-learned, context-based trait impressions. Participants were more likely to select the rare trait than the common trait when the target was with both perfect predictor people (PC + PR), indicating stronger rare context impressions than common context impressions. In addition, participants were more likely to select the common trait than the rare trait when the target was with the imperfect predictor person, indicating a stronger association of the imperfect predictor context with the common trait than with the rare trait. The same pattern of results emerged in Experiment 4 as in Experiments 1 and 2, despite switching the classification order of traits and contexts and changing the type of context learned.

In addition, the inclusion of a trait valence manipulation resulted in the moderation of the inverse base-rate effect. The inverse base-rate effect was observed when the traits were negative, but not when they were positive. It is not clear why this might be the case. Unlike in Experiment 2, the results in the current experiment cannot be explained by negativity bias. In Experiment 2, participants learned that the target possessed both positive and negative traits. Therefore, they may have attended more to the negative traits and weighted those traits more heavily in their impressions of the target. However, in the current experiment, participants formed evaluatively consistent impressions of the target. Depending on the valence condition, the target possessed either all positive traits or all negative traits, but not a combination of both. Thus, there is no reason to expect that participants would attend to one trait more than any other within the same valence condition.

The current experiment, however, differed from Experiments 1–3 in two important ways. First, the classification order of traits and contexts was reversed. In Experiments 1–3, participants were given the traits that the target possessed and then were asked to guess the context he was in. In the current experiment, participants were given the contexts and then asked to guess the traits. Second, the contexts in the current experiment were represented by people rather than by room colors. These changes may have influenced the effect of valence on trait impressions. Participants may have perceived the positive traits as being less conflicting with one another than they perceived the negative traits, and thus selected the common and rare positive traits with equal frequency when told that the target was with both perfect predictor context people. This explanation, however, is merely speculative.

9. General discussion

The purpose of the current research was to examine how impressions of a person's traits are learned based on the context in which that person is encountered and how this learning process leads to different impression strengths across contexts. We drew from the principles of attention theory (Kruschke, 1996, 2001) in order to examine these processes. Attention theory suggests that rare impressions are learned after common impressions. Furthermore, this learning order leads to stronger rare context impressions than to common context impressions because perceivers shift attention toward attributes that are unique to the second context rather than attend to attributes that are shared by both contexts.

Experiments 1 and 2, which used the inverse base-rate paradigm, showed that participants learned the traits in the common contexts prior to the traits in the rare contexts. Also, participants were more likely to classify the target as being in the rare context when he possessed both perfect predictor traits (PC + PR), contrary to what would be expected from base rate occurrences alone. This inverse base-rate effect demonstrates that impressions in the rare context were stronger than impressions in the common context. When in direct competition, the rare context traits were weighted more heavily than the common context traits. Experiment 3 directly tested the causal effect of learning order by manipulating the order in which context-based trait impressions were learned. The results of Experiment 3 only weakly support this causal role, failing to reach statistical significance. Participants did not classify a target as being in the second-learned context significantly more often than in the first-learned context when he possessed the perfect predictor traits of both contexts (PC + PR), though the results were trending the expected direction. Thus, the second-learned impressions may be stronger than the first-learned impressions, though we cannot make a definitive claim that this is the case. Although the attention theory mechanism remains a viable explanation for the results, the eliminative inference model, the major competing account of the inverse base-rate effect (e.g., Juslin et al., 2001; Winman et al., 2005), cannot be entirely ruled out. Lastly, Experiment 4 switched the classification order of traits and contexts in order to mimic a situation that is more likely to occur in real life. Rather than being given the traits and then asked to guess the contexts, participants were given the contexts in which a target person was encountered and then asked to guess the traits that the target possessed. Participants once again demonstrated the inverse base-rate effect, suggesting that impressions of the target in the rarely encountered contexts were stronger than impressions in the frequently encountered contexts.

9.1. Limitations

One limitation of the current studies is that we did not directly test the attentional mechanism that leads to the differential strength of context-based impressions. The attentional mechanism was assumed from the inverse base-rate effect, but attention was never directly measured during learning. However, findings from existing research provide evidence that people direct attention in a way that strengthens rare associations. The most direct evidence comes from Kruschke, Kappenman, and Hetrick (2005), who used eye tracking to measure participants' attention while they learned associations from an inverse base-rate design. They found that attentional preference for the perfect predictors of rare outcomes predicted stronger inverse base-rate effects. Another inverse base-rate study also demonstrated an attentional preference for rare associations as measured via ERP correlates of selective attention (Wills, Lavric, Hemmings, & Surrey, 2014). Another study that was very similar to the present ones, but showing the inverse base-rate effect in the formation of group impressions (i.e., stereotypes), showed that participants paid more attention to rare rather than common behaviors when learning about minority group members (Sherman et al., 2009). Although these studies support the attentional mechanism, the current research would benefit from additional studies that directly measure attention when perceivers form context-based impressions.

9.2. Context-based impressions and evaluations

The specific processes that underlie impression formation also may have implications for the associative links that are developed among trait impressions, contexts, and individual targets. In the current research, participants formed impressions via direct association training among a target's traits and the given contexts; participants guessed a context depending on the trait the target person possessed or they guessed a trait depending on the context that the person was in. However, it is not clear whether the trait-context association is contingent upon the target person in question. For instance, do trait-context associations hold only for the target person who was said to possess the trait or do the trait-context associations also extend to previously unknown target people? That is, if perceivers learn that Dave is intelligent when he is in the blue room, would they also infer that a new person named Steve also is intelligent in the blue room? Additional data from our lab suggests this may be the case. However, these ideas must be further tested in order to understand why trait impressions may extend to previously unknown targets and under what conditions this effect will occur.

One process for which there *is* evidence of a target contingency is spontaneous trait inference. Spontaneous trait inference is the process by which people spontaneously form trait impressions based on observed behaviors (Uleman et al., 1996). Previous research shows that spontaneous trait inferences are bound directly to the actor who performed the behavior and not to other people who may be present at the same time (Todorov & Uleman, 2004).

In their studies of contextual modulation of attitudes, Gawronski and colleagues also found that attitudinal responses were elicited specifically in response to the target person who performed the evaluation-eliciting behaviors and not in response to previously unknown people (Gawronski, Rydell, Vervliet, & De Houwer, 2010; Rydell & Gawronski, 2009). These findings seem to suggest that the association between a particular attitude and a context is contingent on the target person who performed the evaluation-eliciting behaviors. Indeed, the attitude formation process in these studies (i.e., forming attitudes about a target person based on positive or negative behaviors performed by the target) is not unlike the formation of spontaneous evaluative impressions (Schneid, Carlston, & Skowronski, 2015). Together, these findings suggest that context-based impressions may be bound only to the target person in question and not to other previously unknown people when the impression of the target was formed by a spontaneous interpretation of the target's behaviors.

9.3. Future directions: context and group impressions

The current findings may have important implications for understanding how stereotypic and prejudicial impressions of group members are maintained. For example, pre-existing impressions of a social group may influence the strength of impressions that perceivers form of an individual group member who is encountered in different contexts. Because perceivers may attempt to explain away stereotype-inconsistent behaviors, they may attend more to the context in which those behaviors occur in order to attribute the cause of the behaviors to the context rather than to the target's disposition. Consequently, they may form especially strong associations between the inconsistent behaviors and the context in which they occur. The strength of the inconsistent impression would be evident if the inconsistent impression predominates when the target is in a combined context (e.g., when the target is simultaneously with a context person who is associated with inconsistent behaviors and a context person who is associated with consistent behaviors). The implication of having strong context-based impressions of inconsistent behaviors is that perceivers may “subtype” the inconsistent impression into the context in which it was learned, thereby reducing the chances that the impression will generalize to encounters with the target in new contexts. In other words, forming especially strong context-based impressions of a target's stereotype-inconsistent behaviors may be a means of maintaining pre-existing stereotypic expectations.

Knowledge of impression formation processes may also allow us to examine how undesirable stereotypic and prejudicial impressions may be changed. If perceivers pay particularly close attention to expectancy-inconsistent behaviors when they appear in a rare context (Forster, Higgins, & Strack, 2000; Forster, Higgins, & Werth, 2004; Gawronski, Ye, Rydell, & De Houwer, 2014; Sherman & Frost, 2000; Sherman, Lee, Bessenoff, & Frost, 1998), then they may form stronger impressions of those behaviors, which may then be utilized in order to weaken pre-existing stereotypic and evaluative impressions. If the expectancy-inconsistent impression from the rare context can be introduced in more common contexts, it may undermine or even overwhelm the common impression. This would be an important and counter-intuitive finding, indeed. The current research offers one framework from which to examine such possibilities.

References

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370.

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology*, *38*(6), 889–906.

Forster, J., Higgins, E. T., & Strack, F. (2000). When stereotype disconfirmation is a personal threat: How prejudice and prevention focus moderate incongruency effects. *Social Cognition*, *18*(2), 178–197.

Forster, J., Higgins, E. T., & Werth, L. (2004). How threat from stereotype disconfirmation triggers self-defense. *Social Cognition*, *22*(1), 54–74.

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118*, 247–279.

Gawronski, B., Rydell, R. J., Vervliet, B., & De Houwer, J. (2010). Generalization versus contextualization in automatic evaluation. *Journal of Experimental Psychology: General*, *139*(4), 683–701.

Gawronski, B., Ye, Y., Rydell, R. J., & De Houwer, J. (2014). Formation, representation, and activation of contextualized attitudes. *Journal of Experimental Social Psychology*, *54*, 188–203.

Halberstadt, J., Sherman, S. J., & Sherman, J. W. (2011). Why Barack Obama is Black: A cognitive account of hypodescent. *Psychological Science*, *22*(1), 29–33.

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

Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review*, *103*, 336–355.

Hamilton, D. L., & Zanna, M. P. (1972). Differential weighting of favorable and unfavorable attributes in impressions of personality. *Journal of Experimental Research in Personality*, *6*, 204–212.

Juslin, P., Wennerholm, P., & Winman, A. (2001). High-level reasoning and base-rate use: Do we need cue-competition to explain the inverse base-rate effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 849–871.

Kruschke, J. K. (1996). Base rates in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 3–26.

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

Kruschke, J. K., Kappenman, E. S., & Hetrick, W. P. (2005). Eye gaze and individual differences consistent with learned attention in associative blocking and highlighting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 830–845.

Medin, D. L., & Edelson, S. M. (1988). Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General*, *117*, 68–85.

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology*, *61*(3), 380–391.

Rydell, R. J., & Gawronski, B. (2009). I like you, I like you not: Understanding the formation of context-dependent automatic attitudes. *Cognition and Emotion*, *23*(6), 1118–1152.

Schneid, E. D., Carlston, D. E., & Skowronski, J. J. (2015). Spontaneous evaluative inferences and their relationship to spontaneous trait inferences. *Journal of Personality and Social Psychology*, *108*(5), 681–696.

Sherman, J. W., & Frost, L. A. (2000). On the encoding of stereotype-relevant information under cognitive load. *Personality and Social Psychology Bulletin*, *26*(1), 26–34.

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

Sherman, J. W., Lee, A. Y., Bessenoff, G. R., & Frost, L. A. (1998). Stereotype efficiency reconsidered: Encoding flexibility under cognitive load. *Journal of Personality and Social Psychology*, *75*(3), 589–606.

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, *105*(1), 131–142.

Srull, T. K., & Wyer, R. S. (1989). Person memory and judgment. *Psychological Review*, *96*, 58–83.

Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology*, *87*(4), 482–493.

Uleman, J. S. (1999). Spontaneous versus intentional inferences in impression formation. In S. Chaiken, & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 141–160). New York: Guilford.

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. In M. P. Zanna (Ed.), *Advances in experimental social psychology*. *Vol. 28*. (pp. 211–279). San Diego, CA: Academic Press.

Wills, A. J., Lavric, A., Hemmings, Y., & Surrey, E. (2014). Attention, predictive learning, and the inverse base-rate effect: Evidence from event-related potentials. *NeuroImage*, *87*, 6171.

Winman, A., Wennerholm, P., Juslin, P., & Shanks, D. R. (2005). Evidence for rule-based processes in the inverse base-rate effect. *The Quarterly Journal of Experimental Psychology*, *58A*(5), 789–815.