Journal of Experimental Social Psychology 97 (2021) 104193

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

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

Is less of an unhealthy ingredient healthy or unhealthy? Effects of mere co-occurrence and quantitative relations on attribute judgments☆

Bertram Gawronski *, Dillon M. Luke, Nyx L. Ng

University of Texas at Austin, United States

ARTICLE INFO

This paper has been recommended for acceptance by Vanessa Bohns.

Keywords:
Associative learning
Dual-process theory
Impression formation
Multinomial modeling
Propositional learning

ABSTRACT

Research suggests that evaluative responses to an object can be jointly influenced by the mere co-occurrence of the object with a pleasant or unpleasant stimulus (e.g., mere co-occurrence of object A with unpleasant event B) and the qualitative relation of the object to that stimulus (e.g., object A starts vs. stops unpleasant event B). Expanding on these findings, the current research investigated effects of mere co-occurrence and quantitative relations (e.g., product A includes more vs. less of unhealthy ingredient B) on attribute judgments. Seven experiments obtained strong effects of quantitative relations and rather weak evidence for mere co-occurrence effects. Although processing conditions during encoding and judgment moderated effects of quantitative relations in a manner consistent with the predictions of extant theories, the evidence for predicted moderators of mere co-occurrence effects was mixed. The results are explained via a combination of propositional inferences during learning and selective retrieval during judgment.

Imagine an advertisement for a specific food product stating that the product contains less sodium. Will this advertisement make people think of the product as being healthy or unhealthy? Although the former outcome may seem more plausible, research suggests that either one could happen (Gawronski, Brannon, & Luke, 2021). On the one hand, people may think of the product as being healthy based on the information that it contains less of an unhealthy ingredient. On the other hand, the mere pairing of the product with an unhealthy ingredient in the advertisement may lead people to think of the product as being unhealthy. Whereas the former outcome would reflect an effect of relational information, the latter outcome would reflect an effect of mere co-occurrence.

In the current research, we used a multinomial modeling approach (see Hütter & Klauer, 2016) to investigate effects of mere co-occurrence and information about quantitative relations on judgments about object attributes. The work was inspired by evidence suggesting that evaluative responses to an object can be influenced by (1) the object's mere co-occurrence with a pleasant or unpleasant stimulus (e.g., mere co-occurrence of object A and negative stimulus B) and (2) the object's qualitative relation to the co-occurring stimulus (e.g., object A starts vs. stops negative stimulus B). Expanding on prior research studying effects of mere co-occurrence and relational information on *evaluative responses* (e.g., Heycke & Gawronski, 2020; Hu, Gawronski, & Balas, 2017; Hughes, Ye, Van Dessel, & De Houwer, 2019; Kukken, Hütter, & Holland, 2020; Moran & Bar-Anan, 2013), the current research investigated effects on *attribute judgments* (i.e., judgments of whether a product is healthy or unhealthy; see Högden & Unkelbach, 2021). Moreover, different from the dominant focus on *qualitative relations* in prior studies (e.g., A causes vs. prevents B; A starts vs. stops B; A likes vs. dislikes B; A is similar to vs. different from B; see Kurdi & Dunham, 2020), the current research investigated effects of messages involving *quantitative relations* (i.e., product A has less vs. more of ingredient B). Addressing whether earlier findings regarding effects of mere co-occurrence and qualitative relations on evaluative judgments generalize to effects of mere co-occurrence and quantitative relations on attribute judgments, our main questions were: (1) Does mere co-occurrence influence judgments regarding specific attributes irrespective of information about quantitative relations? (2) How do processing conditions during encoding and judgment moderate effects of mere co-occurrence and quantitative relations?

1. Mere co-occurrence and relational information

Early evidence for joint effects of mere co-occurrence and relational

☆ This paper has been recommended for acceptance by Dr Vanessa Bohns.

* Corresponding author.
E-mail address: gawronski@utexas.edu (B. Gawronski).

https://doi.org/10.1016/j.jesp.2021.104193
Received 1 January 2021; Received in revised form 5 July 2021; Accepted 8 July 2021
Available online 15 July 2021

information on evaluative responses came from several studies using a task-dissociation approach (e.g., Hu et al., 2017; Moran & Bar-Anan, 2013). The main finding of these studies was that evaluative responses on explicit measures reflected effects of relational information, whereas evaluative responses on implicit measures reflected effects of mere co-occurrence (for an overview of implicit measures, see Gawronski & De Houwer, 2014). For example, when participants were presented with information that a pharmaceutical product prevents a negative health condition, they showed a positive response to the product on explicit measures, reflecting its causal relation to the negative health condition. Yet, participants showed a negative response to the product on implicit measures, reflecting its mere co-occurrence with the negative health condition (Hu et al., 2017, Experiments 1 and 2).

Although some studies support the idea that evaluative responses on implicit and explicit measures differ in their sensitivity to effects of mere co-occurrence and relational information, the available evidence is rather mixed and inconclusive, in that several studies found a dominant effect of relational information on both explicit and implicit measures without obtaining any evidence for mere co-occurrence effects (e.g., Gawronski, Walther, & Blank, 2005, Experiment 1; Hu et al., 2017, Experiment 3; for a review, see Kurdi & Dunham, 2020). Another factor undermining strong conclusions is that implicit and explicit measures differ in numerous ways, which renders the meaning of dissociative effects on the two kinds of measures theoretically ambiguous (see Bading, Stahl, & Rothermund, 2020; Calanchini, 2020; Corneille & Hütter, 2020; Payne, Burkley, & Stokes, 2008).

To overcome these limitations, recent work has adopted a multinomial modeling approach (see Hütter & Klauer, 2016) to quantify effects of mere co-occurrence and relational information (e.g., Gawronski & Brannon, 2021; Heycke & Gawronski, 2020; Kukken et al., 2020). Different from the comparison of responses across measures in the task-dissociation approach, a major advantage of the multinomial modeling approach is that it allows researchers to quantify the contributions of mere co-occurrence and relational information to responses on a single task (see Corneille & Hütter, 2020; Sherman, Krieglmeyer, & Calanchini, 2014). The two kinds of effects are captured by separate parameters quantifying the probabilities that (1) responses reflect the object's mere co-occurrence with a positive or negative stimulus and (2) responses reflect the object's relation to the co-occurring stimulus. Although studies using a multinomial modeling approach have identified several contextual factors that moderate the impact of mere co-occurrence and relational information (Gawronski & Brannon, 2021; Heycke & Gawronski, 2020; Kukken et al., 2020), the results obtained in these studies support the idea that mere co-occurrence and relational information jointly influence evaluative responses.

2. Theoretical explanations

A common explanation for the effects of mere co-occurrence and relational information is that they are the products of two functionally distinct learning mechanisms. For example, the associative-propositional evaluation (APE) model (Gawronski & Bodenhausen, 2006, 2011, 2018) suggests that mere co-occurrence effects are the product of an associative learning mechanism involving the automatic formation of mental associations between co-occurring stimuli. In contrast, effects of relational information are claimed to be the product of a propositional learning mechanism involving the non-automatic generation and truth assessment of mental propositions about the relation between co-occurring stimuli. Based on the hypothesis that effects of mere co-occurrence and relational information arise from two learning mechanisms with distinct functional properties, we refer to such explanations as *dual-learning accounts*.

An alternative explanation is offered by theories that interpret all learning effects as outcomes of a single propositional mechanism involving the non-automatic generation and truth assessment of mental propositions about stimulus relations (e.g., De Houwer, 2009, 2018; De Houwer, Van Dessel, & Moran, 2020). According to these theories, distinct effects of mere co-occurrence and relational information result from processes during the retrieval of stored propositional information rather than two functionally distinct learning mechanisms. For example, based on the assumptions of the Integrated Propositional Model (IPM; De Houwer, 2018), mere co-occurrence effects can be expected to occur when the retrieval of stored propositions about stimulus relations is incomplete (e.g., retrieval of *A is related to B* rather than *A stops B*; see Van Dessel, Gawronski, & De Houwer, 2019). Based on the hypothesis that effects of mere co-occurrence and relational information arise from the incomplete retrieval of stored propositional information about stimulus relations, we refer to such explanations as *selective-retrieval accounts*.

3. Quantitative relations and attribute judgments

In the current research, we used a multinomial modeling approach to investigate whether effects of mere co-occurrence and relational information on evaluative responses generalize to judgments regarding specific attributes (i.e., judgments of whether a product is healthy or unhealthy; see Högden & Unkelbach, 2021). Moreover, going beyond the dominant focus on qualitative relations in prior research (e.g., A causes vs. prevents B; A starts vs. stops B; A likes vs. dislikes B; A is similar vs. dissimilar to B; see Kurdi & Dunham, 2020), we investigated effects of mere co-occurrence and relational information in messages involving quantitative relations (i.e., A has less vs. more of ingredient B).

From a dual-learning view, people may form a mental association between two co-occurring stimuli regardless of whether the relation between the two stimuli is qualitative or quantitative. For example, an advertisement stating that a food product contains less sodium may create a mental association between the food product and sodium. To the extent that sodium is mentally associated with the attribute *unhealthy*, spread of activation along these associations may lead people to judge the product as unhealthy (i.e., effect of mere co-occurrence). At the same time, the information in the advertisement may lead people to draw the propositional inference that having less of an unhealthy ingredient is healthy, leading to judgments of the product as healthy (i.e., effect of relational information). Because the learning process of associative link formation is assumed to operate independent of processing goals (Gawronski & Bodenhausen, 2014), observed co-occurrences should shape mental representations regardless of the operation and outcomes of propositional inferences during encoding. While associative link formation based on mere co-occurrences should primarily depend on the frequency of the observed co-occurrences, effects of quantitative relations resulting from propositional learning should depend on any factor that facilitates or interferes with propositional inferences during encoding.

From a selective-retrieval view, people may encode and store episodic information about specific stimulus relations regardless of whether these relations are qualitative or quantitative, and the resulting episodic representations may lead to mere co-occurrence effects when the retrieval of these representations is incomplete (De Houwer, 2018). For example, an advertisement stating that a food product contains less sodium may create an episodic representation of the specific relation between the food product and sodium, and judgments of the product's attributes may depend on whether retrieval of this representation is complete or incomplete. When retrieval is complete (i.e., *the product contains less sodium*), people may draw the propositional inference that having less of an unhealthy ingredient is healthy, leading to judgments of the product as healthy (i.e., effect of relational information). Yet, when retrieval is incomplete in the sense that people fail to retrieve the quantitative qualifier (i.e., *the product contains sodium*), they may draw the propositional inference that having an unhealthy ingredient is unhealthy, leading to judgments of the product as unhealthy (i.e., effect of mere co-occurrence). From this perspective, any factor that influences the likelihood of complete versus incomplete retrieval should have compensatory effects on the impact of mere co-occurrence and quantitative relations. That is, greater likelihood of complete retrieval should be associated with stronger effects of quantitative relations and weaker effects of mere co-occurrence. Conversely, greater likelihood of incomplete retrieval should be associated with weaker effects of quantitative relations and stronger effects of mere co-occurrence.

4. The current research

In the current research, we used a multinomial modeling approach to investigate effects of mere co-occurrence and information about quantitative relations on judgments about object attributes. The two main questions guiding this work were: (1) Does mere co-occurrence influence judgments regarding specific attributes irrespective of information about quantitative relations? (2) How do processing conditions during encoding and judgment moderate effects of mere co-occurrence and quantitative relations?

Toward this end, participants were presented with health-related information about ingredients of food products. The information varied as a function of whether a given product was said to have more or less of a healthy or an unhealthy ingredient. Some products were said to have more of a healthy ingredient; some products were said to have less of a healthy ingredient; some products were said to have more of an unhealthy ingredient; and some products were said to have less of an unhealthy ingredient. Participants were asked to form an impression of the products in terms of whether they are healthy or unhealthy. Afterwards, participants were asked to indicate for each product if it is healthy or unhealthy. Responses were analyzed using a modified version of Heycke and Gawronski's (2020) RCB model (see Fig. 1), which provides numerical estimates for (1) the probability that information about quantitative relations drives judgments (captured by the model's *R* parameter); (2) the probability that mere co-occurrence drives judgments if information about quantitative relations does not drive judgments (captured by the model's *C* parameter); and (3) the probability that a general positivity or negativity bias drives judgments if neither information about quantitative relations nor mere co-occurrence drive judgments (captured by the model's *B* parameter).[1]

To investigate effects of mere co-occurrence and quantitative relations on attribute judgments, we conducted seven experiments. Experiment 1 investigated whether attribute judgments are influenced by both mere co-occurrence and information about quantitative relations. Expanding on the findings of Experiment 1, Experiments 2–5 investigated the impact of various contextual conditions during encoding and judgment on the effects of mere co-occurrence and quantitative relations. Experiment 2 investigated the influence of time for encoding; Experiment 3 investigated the influence of information repetition during encoding; Experiment 4 investigated the influence of time during judgment; and Experiment 5 investigated the influence of temporal delay between encoding and judgment. Expanding on the findings of Experiments 1–5, Experiment 6 investigated whether the weak evidence for mere co-occurrence effects in these studies was due to a strong emphasis on relational information in the learning instructions. Finally, Experiment 7 aimed to provide more compelling evidence for the assumption that the obtained findings reflect effects of attribute rather than evaluative learning. The data for each study were collected in one shot without intermittent statistical analyses. We report all measures, all conditions, and all data exclusions. The materials, raw data, and analysis files for all studies are publicly available at https://osf.io/3p5y2/.

Following Heycke and Gawronski (2020), we aimed to recruit 100 participants for the one study without additional manipulations (Experiment 1), 400 participants for studies using a between-subjects manipulation with two conditions (Experiments 2, 4, 6, 7), and 200 participants for studies using a within-subjects manipulation with two conditions (Experiments 3, 5). For the four studies using a between-subjects manipulation with two conditions, a sample of 400 participants provides a power of 80% in detecting a small effect of $d = 0.28$ in a traditional *t*-test for independent means (two-tailed). For the two studies using a within-subjects manipulation with two conditions, a sample of 200 participants provides a power of 80% in detecting a small effect of $d = 0.20$ in a traditional *t*-test for dependent means (two-tailed).[2] By default, we excluded all participants who (1) started the study but did not complete it until the end, (2) disclosed that they did not pay attention to the stimuli or did not take their responses seriously (see Aust, Diedenhofen, Ullrich, & Musch, 2013), (3) failed to pass an instructional attention check (see Oppenheimer, Meyvis, & Davidenko, 2009), (4) failed to pass a materials comprehension check, or (5) responded to less than 50% of all trials within the 1000 millisecond response window on our main dependent measure. Cases with the same subject code were treated as duplicate submissions from the same participant. In such cases, we kept the first submission and excluded all following submissions. A flow chart depicting the sequence of a priori exclusion decisions is depicted in Fig. 2.

5. Experiment 1

The main goal of Experiment 1 was to investigate whether attribute judgments are influenced by both mere co-occurrence and information about quantitative relations. Toward this end, participants were presented with images of hypothetical food products and information about whether a given product includes more or less of a healthy or an unhealthy ingredient. Participants were asked to form an impression of the products in terms of whether they are healthy or unhealthy. Afterwards, participants were asked to indicate for each product if it is healthy or unhealthy. Responses were analyzed using Heycke and Gawronski's (2020) RCB model to quantify the extent to which participants' judgments of the products were influenced by (1) their mere co-occurrence with a healthy or unhealthy ingredient and (2) the quantitative relation specified in the message.

5.1. Method

5.1.1. Participants

We aimed to recruit 100 participants from Amazon's Mechanical Turk (MTurk). The data collection was completed in January 2019. Eligibility for participation was restricted to MTurk workers from the United States who had successfully completed at least one previous assignment, had an approval rating of at least 95% on past assignments, and had not completed an earlier assignment from our lab using similar materials. Of the 108 participants who started the assessment (112 submissions), 100 participants completed the assessment in full. Of these participants, 1 participant was excluded because they disclosed that they were inattentive or did not take their responses seriously; 4 participants were excluded for failing the attention check; 9 participants were excluded for failing the materials comprehension check; and 2 participants were excluded for failing to respond to at least 50% of all

[1] Following Heycke and Gawronski (2020), we use *R* for the parameter capturing effects of relational information, *C* for the parameter capturing effects of mere co-occurrence, and *B* for the parameter capturing general response biases. In a multinomial model that is structurally equivalent to the one depicted in Figure 1, Kukken et al. (2020) used *m* instead of *R* (referring to *meaning*), *p* instead of *C* (referring to *pairing*), and *g* instead of *B* (referring to *guessing*).

[2] Because power analyses within multinomial modeling require simulations with expected population values for the three parameters and any specific expectations in this regard would be arbitrary, we made our a priori sample-size decision in a heuristic fashion based on the sample sizes in Heycke and Gawronski's (2020) research and simple comparisons of mean values using *t*-tests.

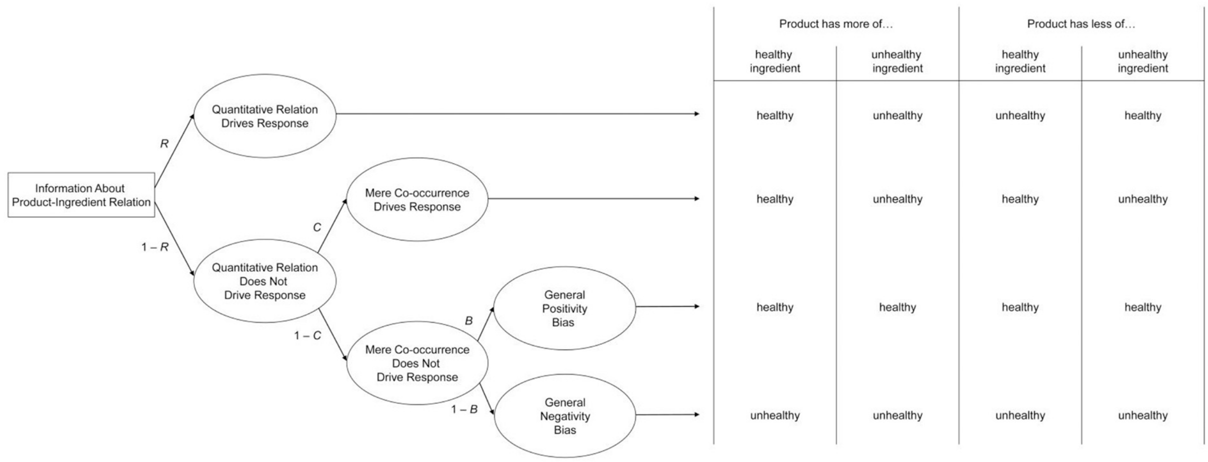

Fig. 1. Multinomial processing tree depicting effects of stimulus relation, stimulus co-occurrence, and general response biases on health judgments (healthy vs. unhealthy) as a function of relational information (more vs. less) and focal ingredient (healthy vs. unhealthy).

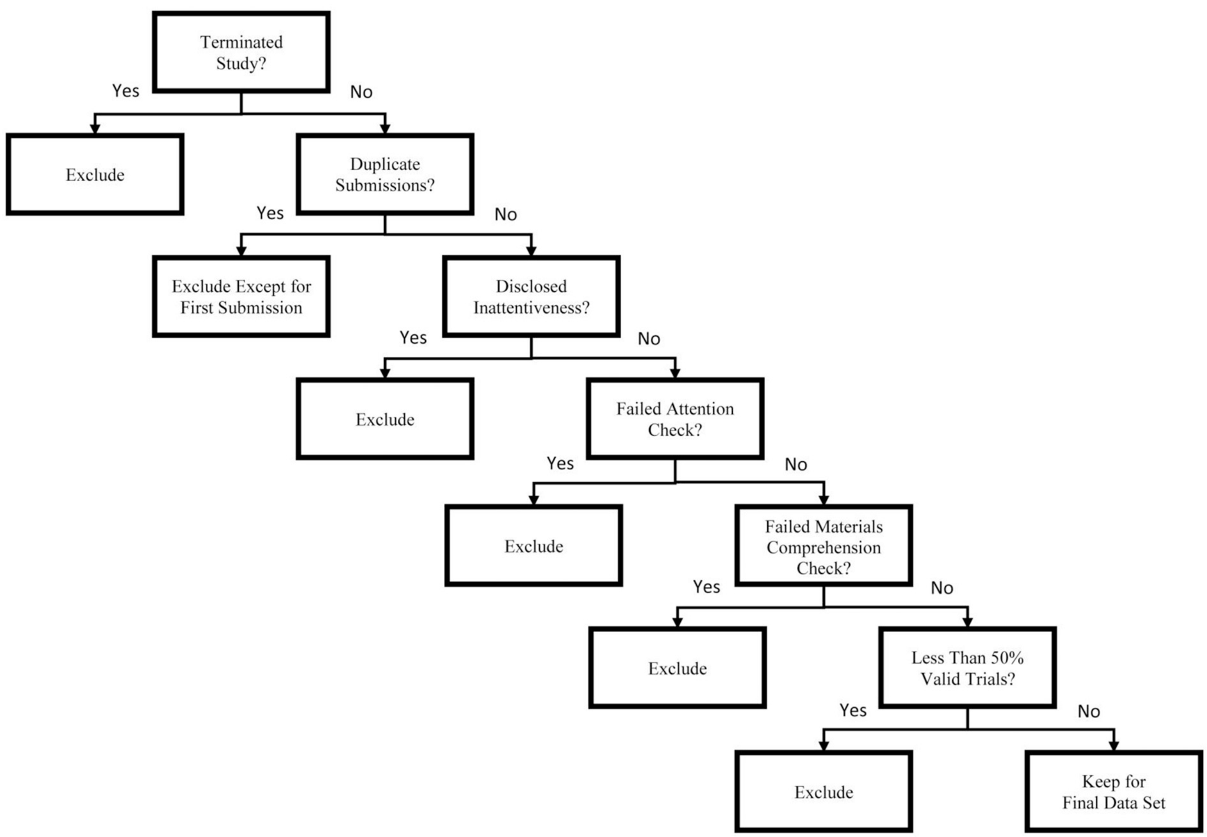

Fig. 2. Flow chart depicting the sequence of exclusion decisions in Experiments 1–7.

trials within the 1000 millisecond response window on our main dependent measure, resulting in a final sample of 84 participants (44.05% female, 55.95% male; $M_{\text{age}} = 35.57$, $SD_{\text{age}} = 10.47$). Participants were compensated $2.00 for their time.

5.1.2. Materials

For the target stimuli, we created 16 images of hypothetical food product brands. To avoid potential influences from prior impressions, brand names were generated such that they (1) were not already in use for an existing brand and (2) did not make reference to a specific food item (e.g., burgers). Brand names were displayed in the center of each image on a solid or two-toned background. The font, font color(s), and background color(s) of each image varied across brands. For the ingredients, we selected eight types of nutrition information, with four ingredients as presumed healthy ingredients (calcium, iron, protein, vitamin D) and four ingredients as presumed unhealthy ingredients (fat, sodium, sugar, calories). Nutrition information was displayed via images, with the ingredient being shown in the center of each image in white font on a solid black background.

5.1.3. Learning task

For the learning task, participants were informed that the main goal of the study was to investigate how people form impressions of food products. The specific instructions for the learning task were as follows:

The main goal of the present study is to investigate how people form impressions of food products. For this purpose, please imagine that you are trying to evaluate if a product is healthy or unhealthy. To make this judgment, you will be presented with nutrition facts about each product. The nutrition facts will indicate whether a given food product is healthy or unhealthy. Specifically, the nutrition facts will indicate whether a given food product has more or less of an ingredient. Some of these ingredients may be healthy and some of these ingredients may be unhealthy. We ask that you use this information to evaluate whether a product is healthy or unhealthy. For example, if a product is said to have more of a healthy ingredient, then this product is healthy. Conversely, if a product is said to have more of an unhealthy ingredient, then this product is unhealthy. Similarly, if a product is said to have less of a healthy ingredient, then this product is unhealthy. Conversely, if a product is said to have less of an unhealthy ingredient, then this product is healthy. Please form an impression of the products based on the presented nutrition information.

On each learning trial, participants were presented with an image of a food product on the left side of the screen, an image of a piece of nutritional information on the right side of the screen, and relational information in the center of the screen. The relational information indicated whether the product contained *more* or *less* of the displayed nutritional ingredient, which was further qualified by a specific percentage which could take on the value of 30%, 40%, 50%, or 60%.[3] The images of the food products and the nutritional information were equal in size.

Four of the 16 food products were presented with information indicating more of a healthy ingredient, four were presented with information indicating less of a healthy ingredient, four were presented with information indicating more of an unhealthy ingredient, and four were presented with information indicating less of an unhealthy ingredient. For each participant, the same product (e.g., *Summer Sammy's*) was always displayed with the same relational information (e.g., *30% more*) and the same ingredient (e.g., *fat*). The pairing of a given product with a given type of relation (more of a healthy ingredient, less of a healthy ingredient, more of an unhealthy ingredient, less of an unhealthy ingredient) was counterbalanced by means of a Latin Square.

The learning task was divided into three blocks, with each block presenting each product-ingredient pairing twice. Within each block, the presentation of product-ingredient pairings was randomized with the constraint that every pairing be displayed once before repeating. Following Heycke and Gawronski (2020), each pairing was displayed for 3000 milliseconds, with an inter-trial interval of 1000 milliseconds. After each block, participants were provided feedback on their progress through the learning task and were prompted to start the next block when ready. With the total number of 16 unique pairings, each block comprised 32 trials, summing up to a total of 96 trials with each pairing being displayed a total of six times across the three blocks.

5.1.4. Judgment task

After completion of the learning task, participants completed a timed judgment task, which asked them to indicate whether they considered a given food product to be healthy or unhealthy. The specific instructions for the judgment task were as follows:

In the following part, we ask you to evaluate how healthy each product is. For this purpose, we will show you pictures of different food products, and we ask you to indicate whether that product is healthy or unhealthy. Please indicate quickly for each product whether it is healthy or unhealthy by pressing the corresponding key on your keyboard: Unhealthy = 'A' Healthy = 'K' You will have only 1 s to make your decision. You do not need to justify your decision: just go with your first impression. Please place your two index fingers on these keys. In short: Indicate quickly for each product whether it is healthy or unhealthy: Unhealthy = 'A'; Healthy = 'K' You will have only 1 s to make your decision. Please place your two index fingers on these keys.

On each trial, the image of a food product was displayed in the center of the screen, presented directly above the question *Is this product unhealthy or healthy?* Instructional reminders regarding the response options were displayed at the bottom left (*Unhealthy* = *A*) and bottom right side (*Healthy* = *K*) of the screen. The judgment task was divided into three blocks, with each block displaying each product once (in random order), summing up to a total of 48 trials displaying each product three times. Following the procedure in Heycke and Gawronski's (2020) studies, each product was presented for 1000 milliseconds. If participants did not respond within this timeframe by pressing a valid response key, the trial timed out and the error message *Too slow* was displayed in red font in the center of the screen for 750 milliseconds. Each trial began with a blank screen for 100 milliseconds, followed by a fixation cross for 900 milliseconds in the center of the screen.

5.1.5. Additional measures

After the judgment task, participants completed a materials comprehension check, a set of demographic questions (i.e., gender, age, ethnicity), and an instructional attention check. The materials comprehension check was included to confirm that participants considered the ingredients to be healthy and unhealthy in a manner consistent with their actual health status. If that is not the case (e.g., participants do not have accurate knowledge of an ingredient's health status), a central premise for our manipulation of health status would not be met, which should lead to theoretically trivial null effects of both mere co-occurrence and relational information. Thus, participants were asked for each ingredient to indicate whether they considered it healthy or unhealthy using a binary answer choice. By default, we excluded participants who judged the health status of more than two of the eight ingredients counter to their actual health status.

Next, participants were presented with a reading-intensive attention check, which read as follows:

Most modern theories of decision-making recognize the fact that decisions do not take place in a vacuum. Individual preferences and knowledge, along with situational variables can greatly impact the decision process. In order to facilitate our research on decision-making we are interested in knowing certain factors about you, the decision maker. Specifically, we are interested in whether you actually take the time to read the directions; if not, then some of our manipulations that rely on changes in the instructions will be ineffective. So, in order to demonstrate that you have read the instructions, please ignore the sports items below. Instead, simply continue on to the next page after the options. Thank you very much. Which of these activities do you engage in regularly? (check all that apply).

Answer choices to the attention check included *football*, *soccer*, *dancing*, *watersports*, *triathlon*, *running*, *volleyball*, and *I engage in other activities*. Given that the attention check instructs participants to ignore the answer choices to demonstrate their attentiveness, participants who selected any of the answer choices were excluded from analyses (see Oppenheimer et al., 2009). As a final quality check, participants were asked whether they (1) paid attention to the images presented through the entire task and (2) took the requested response seriously. Participants were informed that their answers to either of these questions

[3] The primary question of the current research concerned the quantitative relations of having *more* vs. *less* of a particular ingredient. The specific percentages were included to create a more diverse stimulus set, but the percentages are irrelevant for the main question of whether attribute judgments are influenced by both mere co-occurrence and quantitative relations. While having more or less of a particular ingredient are quantitative relations, the percentages merely specify the relative strength of the described relation.

would not affect their payment. Participants who reported that they either did not pay attention to the images presented through the entire task or did not take the requested responses seriously were excluded from further analyses (see Aust et al., 2013). Finally, participants were debriefed, thanked for their participation, and given a code for compensation.

5.2. Results

Attribute judgments were aggregated by calculating the sums of *healthy* and *unhealthy* judgments for each of the four product categories (i.e., more of a healthy ingredient, less of a healthy ingredient, more of an unhealthy ingredient, less of an unhealthy ingredient). Means and 95% confidence intervals of the relative proportion of *healthy* (vs. *unhealthy*) judgments as a function of product information are presented in Table 1. RCB model analyses were conducted following the procedure by Heycke and Gawronski (2020). The procedural details are explained in Appendix A. Overall, the RCB model fit the data well with three free parameters, $G^2(1) = 2.04$, $p = .153$, $w = 0.023$. Parameter estimates obtained with the baseline model are presented in Table 2. The *R* parameter was significantly greater than zero, $\Delta G^2(1) = 604.42$, $p < .001$, $w = 0.397$, indicating that attribute judgments were influenced in a manner consistent with the specified quantitative relations. The *C* parameter was marginally greater than zero, $\Delta G^2(1) = 3.36$, $p = .067$, $w = 0.030$, indicating that mere co-occurrence tended to influence judgments despite the described quantitative relations. Finally, the *B* parameter was significantly greater than its reference point of 0.5, $\Delta G^2(1) = 4.31$, $p = .038$, $w = 0.036$, indicating that participants showed a general response tendency to judge the products as healthy.

5.3. Discussion

The results of Experiment 1 indicate that health judgments are shaped by information about quantitative relations, as reflected in a significant effect on the *R* parameter. Evidence for a mere co-occurrence effect was comparatively weaker, in that the *C* parameter was only marginally different from zero. Indeed, whereas the effect size for the impact of relational information qualifies as a medium in terms of Cohen's (1988) conventions, the effects size for the impact of mere co-occurrence falls below the benchmark of a small effect.[4] However, both dual-learning and selective-retrieval accounts suggest that mere co-occurrence effects may vary in size depending on contextual conditions during encoding and judgment. In the following experiments, we tested predictions derived from the two accounts regarding the impact of time for encoding (Experiment 2), information repetition (Experiment 3), time during judgment (Experiment 4), and temporal delay (Experiment 5).

6. Experiment 2

Experiment 2 investigated the impact of time for encoding on the effects of mere co-occurrence and quantitative relations. According to selective-retrieval accounts, contextual factors influencing the retrieval of information from memory should have compensatory effects on the impact of mere co-occurrence and quantitative relations, in that any factor that impairs the complete retrieval of stored propositions about stimulus relations should increase effects of mere co-occurrence and decrease effects of quantitative relations. Conversely, any factor that supports the complete retrieval of stored propositions about stimulus relations should decrease effects of mere co-occurrence and increase effects of quantitative relations. Because more time for encoding supports the storage of information in memory, and thereby the subsequent retrieval of this information (Craik & Lockhart, 1972), more time for encoding should increase the impact of quantitative relations and reduce the impact of mere co-occurrence. Conversely, less time for encoding should reduce the impact of quantitative relations and increase the impact of mere co-occurrence. Together, these assumptions imply that more (vs. less) time for encoding should increase scores on the *R* parameter and decrease scores on the *C* parameter.

In contrast, dual-learning accounts suggest that mere co-occurrence effects are the product of an associative learning mechanism involving the automatic formation of mental associations between co-occurring stimuli, whereas effects of relational information are claimed to be the product of a propositional learning mechanism involving the non-automatic generation and truth assessment of mental propositions about the relation between co-occurring stimuli. Based on these assumptions, more time for encoding should increase the impact of quantitative relations, whereas less time for encoding should decrease the impact of quantitative relations. In contrast, mere co-occurrence effects should be unaffected by time for encoding, given that mere co-occurrence effects are assumed to be driven by a resource-independent process involving the automatic formation of mental associations between co-occurring stimuli. Together, these assumptions imply that more (vs. less) time for encoding should increase scores on the *R* parameter without affecting scores on the *C* parameter. The main goal of Experiment 2 was to test these competing predictions.

6.1. Method

6.1.1. Participants

We aimed to recruit 400 participants from Amazon's MTurk. The data collection was completed in November 2019. The same eligibility criteria for participation in Experiment 1 were used in Experiment 2. Of the 445 participants who started the assessment (469 submissions), 410 participants completed the assessment in full. Two participants had more than one complete submission, in which case only the first submission of each participant was retained. Of the 410 participants in the data set, 25 participants were excluded because they reported that they were inattentive or did not take their responses seriously, 46 participants were excluded for failing the attention check, 30 participants were excluded for failing the materials comprehension check, and 1 participant was excluded for failing to provide valid responses to at least 50% of the judgment trials, resulting in a final sample of 308 participants (44.48% female, 55.19% male, 0.32% prefer not to answer; $M_{age} = 36.53$, $SD_{age} = 10.34$). Participants were compensated \$2.00 for their time.

6.1.2. Procedure

The materials, learning task, judgment task, and additional measures were identical to Experiment 1 with one exception. In Experiment 1, each product-ingredient pairing was presented for 3000 milliseconds in the learning task. In Experiment 2, participants were randomly assigned to one of two conditions in which each product-ingredient pairing in the learning task was presented for either 1000 milliseconds (short duration condition) or 5000 milliseconds (long duration condition). The presentation times of 1000 and 5000 milliseconds were adopted from Heycke and Gawronski (2020, Experiment 2b).

6.2. Results

Attribute judgments were aggregated in line with the procedures in Experiment 1. Means and 95% confidence intervals of the relative proportion of *healthy* (vs. *unhealthy*) judgments as a function of product information and time for encoding are presented in Table 1. The RCB model showed suboptimal fit when the model was fit to the data with six free parameters (i.e., three per condition), $G^2(2) = 5.80$, $p = .055$, $w = 0.020$. Because large sample sizes increase the likelihood of significant

[4] For the effect size *w*, the conventional benchmark for a small effect is 0.1, the benchmark for a medium effect is 0.3, and the benchmark for a large effect is 0.5 (see Cohen, 1988).

Table 1
Mean proportions and 95% confidence intervals of *healthy* (vs. *unhealthy*) judgments of food products that include more or less of a healthy or unhealthy ingredient. Higher scores reflect higher proportions of *healthy* (vs. *unhealthy*) judgments in all cases except for the tastiness-judgment condition in Experiment 7, where higher scores reflect higher proportions of *tasty* (vs. *bland*) judgments.

	Product has more of…				Product has less of			
	Healthy ingredient		Unhealthy ingredient		Healthy ingredient		Unhealthy ingredient	
	M	95% CI	*M*	95% CI	*M*	95% CI	*M*	95% CI
Experiment 1								
baseline	0.71	[0.67, 0.76]	0.30	[0.25, 0.35]	0.35	[0.29, 0.41]	0.71	[0.66, 0.75]
Experiment 2								
1000 ms (encoding)	0.57	[0.54, 0.61]	0.41	[0.38, 0.45]	0.46	[0.43, 0.50]	0.57	[0.54, 0.61]
5000 ms (encoding)	0.63	[0.59, 0.67]	0.39	[0.36, 0.43]	0.37	[0.32, 0.41]	0.64	[0.60, 0.68]
Experiment 3								
4 repetitions	0.65	[0.60, 0.71]	0.36	[0.30, 0.42]	0.40	[0.34, 0.46]	0.68	[0.63, 0.73]
24 repetitions	0.74	[0.69, 0.79]	0.30	[0.25, 0.35]	0.34	[0.29, 0.39]	0.73	[0.68, 0.78]
Experiment 4								
750 ms (judgment)	0.60	[0.57, 0.64]	0.38	[0.34, 0.42]	0.41	[0.37, 0.45]	0.64	[0.60, 0.67]
2500 ms (judgment)	0.67	[0.63, 0.71]	0.32	[0.28, 0.36]	0.35	[0.31, 0.40]	0.67	[0.63, 0.71]
Experiment 5								
immediate	0.70	[0.66, 0.74]	0.32	[0.27, 0.36]	0.36	[0.32, 0.41]	0.69	[0.65, 0.73]
2-day delay	0.62	[0.58, 0.66]	0.36	[0.32, 0.40]	0.40	[0.36, 0.44]	0.63	[0.59, 0.67]
Experiment 6								
relational instructions	0.62	[0.58, 0.67]	0.41	[0.36, 0.45]	0.43	[0.38, 0.47]	0.60	[0.55, 0.64]
minimal instructions	0.60	[0.57, 0.64]	0.43	[0.39, 0.47]	0.46	[0.41, 0.50]	0.61	[0.58, 0.65]
Experiment 7								
healthiness judgment	0.64	[0.61, 0.68]	0.41	[0.37, 0.45]	0.41	[0.37, 0.46]	0.66	[0.62, 0.69]
tastiness judgment	0.57	[0.53, 0.61]	0.54	[0.50, 0.58]	0.53	[0.50, 0.57]	0.55	[0.52, 0.59]

Table 2
Parameter estimates without model restrictions, Experiment 1.

Parameter	Estimate (*SE*)	95% CI	Difference to reference point
R	0.39 (0.01)	[0.36, 0.42]	$\Delta G^2(1) = 604.42$, $p < .001$, $w = 0.397$
C	0.04 (0.02)	[−0.00, 0.09]	$\Delta G^2(1) = 3.36$, $p = .067$, $w = 0.030$
B	0.53 (0.01)	[0.50, 0.55]	$\Delta G^2(1) = 4.31$, $p = .038$, $w = 0.036$

Note. The *R* parameter captures effects of relational information; the *C* parameter captures effects of co-occurrence; the *B* parameter captures general response biases. The neutral reference point for *R* and *C* is 0; the neutral reference point for *B* is 0.5, with scores higher than 0.5 reflecting a general bias toward positive responses and scores lower than 0.5 reflecting a general bias toward negative responses.

discrepancies between actual and predicted response probabilities, and the effect size of the observed discrepancies fell far below Cohen's (1988) benchmark for a small effect (see Footnote 4), we nevertheless tested whether the obtained estimates for the three parameters were significantly different across conditions.

Parameter estimates obtained with the baseline model are presented in Table 3. The *R* parameter was significantly smaller in the 1000-milliseconds condition compared to the 5000-milliseconds condition, $\Delta G^2(1) = 48.94$, $p < .001$, $w = 0.059$, indicating that relational information had a greater impact when time for encoding was long than when it was short. Conversely, the *C* parameter was significantly greater in the 1000-milliseconds condition compared to the 5000-milliseconds condition, $\Delta G^2(1) = 4.18$, $p = .041$, $w = 0.017$, indicating that mere co-occurrence had a greater impact when time for encoding was short than when it was long. There was no significant effect of encoding time on the *B* parameter, $\Delta G^2(1) < 0.01$, $p = .986$, $w < 0.001$.

6.3. Discussion

Experiment 2 obtained evidence for a compensatory impact of time for encoding. Whereas effects of quantitative relations were greater when participants had more time for encoding than when they had less time for encoding, effects of mere co-occurrence were greater when participants had less time for encoding than when they had more time for encoding. Although both selective-retrieval and dual-learning accounts predict the obtained effect of time for encoding on the impact of quantitative relations, the obtained effect on the impact of mere co-occurrence is consistent with the prediction derived from selective-retrieval accounts, but it is inconsistent with the prediction derived from dual-learning accounts.

Table 3
Parameter estimates without model restrictions as a function of time for encoding (1000 ms vs. 5000 ms), Experiment 2.

Parameter	Estimate (*SE*)	95% CI	Difference to reference point
R			
1000 ms	.14 (.01)	[.12, .16]	$\Delta G^2(1) = 145.19$, $p < .001$, $w = .140$
5000 ms	.26 (.01)	[.23, .28]	$\Delta G^2(1) = 426.25$, $p < .001$, $w = .257$
C			
1000 ms	.03 (.01)	[.00, .05]	$\Delta G^2(1) = 4.69$, $p = .030$, $w = .025$
5000 ms	.00 (.02)	[-.03, .03]	$\Delta G^2(1) = 0.00$, $p = 1.00$, $w < .001$
B			
1000 ms	.51 (.01)	[.49, .52]	$\Delta G^2(1) = 1.01$, $p = .316$, $w = .012$
5000 ms	.51 (.01)	[.49, .52]	$\Delta G^2(1) = 0.76$, $p = .383$, $w = .011$

Note. The *R* parameter captures effects of relational information; the *C* parameter captures effects of co-occurrence; the *B* parameter captures general response biases. The neutral reference point for *R* and *C* is 0; the neutral reference point for *B* is 0.5, with scores higher than 0.5 reflecting a general bias toward positive responses and scores lower than 0.5 reflecting a general bias toward negative responses.

Table 4
Parameter estimates without model restrictions as a function of information repetition during encoding (4 repetitions vs. 24 repetitions), Experiment 3.

Parameter	Estimate (*SE*)	95% CI	Difference to reference point
R			
4 Repetitions	0.30 (0.02)	[0.27, 0.34]	$\Delta G^2(1) = 222.25$, $p < .001$, $w = 0.306$
24 Repetitions	0.42 (0.02)	[0.39, 0.46]	$\Delta G^2(1) = 448.95$, $p < .001$, $w = 0.432$
C			
4 Repetitions	0.01 (0.03)	[−0.04, 0.07]	$\Delta G^2(1) = 0.25$, $p = .616$, $w = 0.010$
24 Repetitions	0.05 (0.03)	[−0.02, 0.11]	$\Delta G^2(1) = 2.18$, $p = .140$, $w = 0.030$
B			
4 Repetitions	0.53 (0.01)	[0.50, 0.56]	$\Delta G^2(1) = 5.09$, $p = .024$, $w = 0.046$
24 Repetitions	0.55 (0.02)	[0.51, 0.58]	$\Delta G^2(1) = 7.90$, $p = .005$, $w = 0.057$

Note. The *R* parameter captures effects of relational information; the *C* parameter captures effects of co-occurrence; the *B* parameter captures general response biases. The neutral reference point for *R* and *C* is 0; the neutral reference point for *B* is 0.5, with scores higher than 0.5 reflecting a general bias toward positive responses and scores lower than 0.5 reflecting a general bias toward negative responses.

7. Experiment 3

Experiment 3 investigated the impact of information repetition on the effects of mere co-occurrence and quantitative relations. To the extent that repetition supports the storage of new information in memory, and thereby the subsequent retrieval of this information, selective-retrieval accounts suggest that repetition should increase the impact of quantitative relations and reduce the impact of mere co-occurrence. In contrast, dual-learning accounts suggest that repetition should have corresponding effects on the impact of mere co-occurrence and quantitative relations. On the one hand, repetition should increase the impact of quantitative relations by supporting the storage of information about quantitative relations in memory. On the other hand, repetition should increase the impact of mere co-occurrence by strengthening newly formed associations between co-occurring stimuli (Smith & DeCoster, 2000). Thus, while selective-retrieval accounts predict that more frequent repetition should increase scores on *R* parameter and decrease scores on the *C* parameter, dual-learning accounts predict that more frequent repetition should increase scores on both the *R* and the *C* parameter. The main goal of Experiment 3 was to test these competing predictions.

7.1. Method

7.1.1. Participants

We aimed to recruit 200 participants from Amazon's MTurk. The data collection was completed in February 2020. The same eligibility criteria for participation in Experiments 1 and 2 were used in Experiment 3. Of the 230 participants who started the assessment (241 submissions), 203 participants completed the assessment in full. Two participants had more than one complete submission, in which case only the first submission of each participant was retained. Of the 203 participants in the data set, 6 participants were excluded because they reported that they were inattentive or did not take their responses seriously, 13 participants were excluded for failing the attention check, 21 participants were excluded for failing the materials comprehension check, and 1 participant was excluded for failing to provide valid responses to at least 50% of the judgment trials, resulting in a final sample of 162 participants (45.06% female, 54.94% male; $M_{age} = 37.43$, $SD_{age} = 10.94$). Participants were compensated $2.00 for their time.

7.1.2. Procedure

The materials, learning task, judgment task, and additional measures of Experiment 3 were identical to Experiment 1 with a few exceptions. Changes in the learning task included: (1) the presentation of product-ingredient pairings for one of two sets of 8 food product brands (rather than the full 16 food product brands), (2) the manipulation of repetition of product-ingredient pairings within-subjects, such that each type of product-ingredient relation was repeated either one time (low repetition condition) or six times (high repetition condition) per block, and (3) the inclusion of an additional fourth block. These changes resulted in eight within-subjects conditions for each participant, reflecting the manipulations of health status (healthy vs. unhealthy), quantitative relation (more vs. less), and repetition (low vs. high). Each learning block included 28 trials. Thus, given the additional fourth block, there were 112 learning trials in total, with each product-ingredient pairing displayed either 4 (low repetition condition) or 24 (high repetition condition) times in total. To avoid confounding percentage information (e.g., 30%, 40%, 50%, 60%) with different within-subject conditions in the learning task, percentage information was held constant at 50% for all pairings. The use of specific products and ingredients for each within-subjects condition in the learning task was counterbalanced by means of a Latin Square.

Changes in the judgment task included an additional fourth block of trials. Given that product-ingredient pairings were displayed for only 8 food products (rather than the full set of 16 products), each block consisted of 8 judgment trials per block. With 4 blocks, there were 32 judgment trials across blocks presenting each product four times in total.

7.2. Results

Attribute judgments were aggregated in line with the procedures in Experiment 1. Means and 95% confidence intervals of the relative proportion of *healthy* (vs. *unhealthy*) judgments as a function of product information and repetition are presented in Table 1. The RCB model fit the data well with six free parameters (i.e., three per condition), $G^2(2) = 3.08$, $p = .215$, $w = 0.025$. Parameter estimates obtained with the baseline model are presented in Table 4. The *R* parameter was significantly greater in the high-repetition condition compared to the low-repetition condition, $\Delta G^2(1) = 20.21$, $p < .001$, $w = 0.065$, indicating that relational information had a greater impact on attribute judgments when it was presented more frequently than when it was presented less frequently. The *C* parameter did not significantly differ across repetition conditions, $\Delta G^2(1) = 0.61$, $p = .436$, $w = 0.011$. There was also no significant effect of repetition on the *B* parameter, $\Delta G^2(1) = 0.47$, $p = .493$, $w = 0.010$.

7.3. Discussion

Consistent with the shared prediction of selective-retrieval and dual-learning accounts, repetition increased the impact of quantitative relations. However, counter to the unique predictions of the two accounts, repetition had no significant effect on the impact of mere co-occurrence. Whereas selective-retrieval accounts suggest that repetition should decrease the impact of mere co-occurrence, dual-learning accounts suggest that repetition should increase the impact of mere co-occurrence. Although the pattern of means obtained for the *C* parameter was directionally consistent with the predictions derived from dual-learning accounts, the difference between conditions was not statistically significant.[5]

8. Experiment 4

Experiment 4 investigated the impact of time during judgment on the effects of mere co-occurrence and quantitative relations. According to selective-retrieval accounts, more time during judgment should support the complete retrieval of stored information about stimulus relations, which should increase the impact of quantitative relations and decrease the impact of mere co-occurrence. Similarly, dual-learning accounts such as the APE model (Gawronski & Bodenhausen, 2006, 2011, 2018) suggest that effects of activated associations on judgments and behavior should be reduced when deliberate propositional reasoning leads to a rejection of automatically activated associations during judgment. Thus, in line with the predictions of selective-retrieval accounts, dual-learning accounts suggest that more time during judgment should increase the impact of quantitative relations and decrease the impact of mere co-occurrence. Applied to the RCB model, these assumptions imply that more (vs. less) time during judgment should increase scores on the *R* parameter and decrease scores on the *C* parameter. The main goal of Experiment 4 was to test these predictions.

8.1. Method

8.1.1. Participants

We aimed to recruit 400 participants from Amazon's MTurk. The data collection was completed in December 2019. The same eligibility criteria for participation in Experiments 1–3 were used in Experiment 4. Of the 438 participants who started the assessment (453 submissions), 401 participants completed the assessment in full. One participant had more than one complete submission, in which case only the first submission was retained. Of the 401 participants in the data set, 21 participants were excluded because they reported that they were inattentive or did not take their responses seriously, 28 participants were excluded for failing the attention check, 39 participants were excluded for failing the materials comprehension check, and 4 participants were excluded for failing to provide valid responses to at least 50% of the judgment trials, resulting in a final sample of 309 participants (43.37% female, 56.63% male; M_{age} = 36.20, SD_{age} = 10.77). Participants were compensated $2.00 for their time.

8.1.2. Procedure

The materials, learning task, judgment task, and additional measures in Experiment 4 were identical to Experiment 1 with one exception. In Experiment 1, participants were given 1000 milliseconds to indicate whether a given food product brand is healthy or unhealthy. In Experiment 4, participants were randomly assigned to one of two conditions in which they had either 750 milliseconds (short duration condition) or 2500 milliseconds (long duration condition) to indicate whether a given product is healthy or unhealthy. The response deadlines of 750 and 2500 milliseconds were adopted from Heycke and Gawronski (2020, Experiment 4). Modifications were also made to the instructions for the judgment task reflecting this change.

Table 5
Parameter estimates without model restrictions as a function of time during judgment (750 ms vs. 2500 ms), Experiment 4.

Parameter	Estimate (*SE*)	95% CI	Difference to reference point
R			
750 ms	0.23 (0.01)	[0.21, 0.25]	$\Delta G^2(1) = 366.09, p < .001, w = 0.230$
2500 ms	0.33 (0.01)	[0.31, 0.36]	$\Delta G^2(1) = 802.37, p < .001, w = 0.338$
C			
1000 ms	0.00 (0.02)	[−0.03, 0.03]	$\Delta G^2(1) = 0.02, p = .891, w = 0.002$
5000 ms	0.03 (0.02)	[−0.01, 0.06]	$\Delta G^2(1) = 2.47, p = .116, w = 0.019$
B			
1000 ms	0.51 (0.01)	[0.50, 0.53]	$\Delta G^2(1) = 1.82, p = .177, w = 0.016$
5000 ms	0.51 (0.01)	[0.49, 0.52]	$\Delta G^2(1) = 0.36, p = .550, w = 0.007$

Note. The *R* parameter captures effects of relational information; the *C* parameter captures effects of co-occurrence; the *B* parameter captures general response biases. The neutral reference point for *R* and *C* is 0; the neutral reference point for *B* is 0.5, with scores higher than 0.5 reflecting a general bias toward positive responses and scores lower than 0.5 reflecting a general bias toward negative responses.

8.2. Results

Attribute judgments were aggregated in line with the procedures in Experiment 1. Means and 95% confidence intervals of the relative proportion of *healthy* (vs. *unhealthy*) judgments as a function of product information and time during judgment are presented in Table 1. The RCB model showed poor fit when the model was fit to the data with six free parameters (i.e., three per condition), $G^2(2) = 8.15$, $p = .017$, $w = 0.024$. Because large sample sizes increase the likelihood of significant discrepancies between actual and predicted response probabilities, and the effect size of the observed discrepancies fell far below Cohen's (1988) benchmark for a small effect (see Footnote 4), we nevertheless tested whether the obtained estimates for the three parameters were significantly different across conditions.

Parameter estimates obtained with the baseline model are presented in Table 5. The *R* parameter was significantly greater in the 2500-milliseconds condition compared to the 750-milliseconds condition, $\Delta G^2(1) = 41.95$, $p < .001$, $w = 0.055$, indicating that relational information had a greater impact on attribute judgments when time during judgment was long than when it was short. The *C* parameter did not significantly differ across judgment time conditions, $\Delta G^2(1) = 1.16$, $p = .282$, $w = 0.001$. There was also no significant effect of judgment time on the *B* parameter, $\Delta G^2(1) = 0.20$, $p = .658$, $w = 0.004$.

8.3. Discussion

Consistent with the shared prediction of selective-retrieval and dual-learning accounts, more time during judgment increased the impact of quantitative relations. However, counter to the shared prediction of the two accounts, time during judgment had no significant effect on the impact of mere co-occurrence. Both selective-retrieval and dual-learning accounts suggest that more time during judgment should increase the impact of quantitative relations and decrease the impact of mere co-occurrence. If anything, the pattern of means obtained for the *C* parameter suggests an influence in the opposite direction, in that more time during judgment increased rather than decreased mere co-occurrence effects (for similar findings, see Heycke & Gawronski, 2020). However, the difference between conditions was not statistically significant.

9. Experiment 5

Experiment 5 investigated the impact of temporal delay between encoding and judgment on the effects of mere co-occurrence and quantitative relations. According to selective-retrieval accounts, longer delays between encoding and judgment should impair the complete

[5] To investigate whether the lack of a significant effect on the *C* parameter is due to insufficient statistical power, we conducted a follow-up study with a sample twice as large as the one in Experiment 3 (N = 323). Replicating the results of Experiment 3, the follow-up study revealed a significant effect of repetition on the *R* parameter, $\Delta G^2(1) = 33.39$, $p < .001$, and no significant effect on the *C* parameter, $\Delta G^2(1) = 0.37$, $p = .542$. The details of the follow-up study are presented in the Supplemental Materials.

Table 6
Parameter estimates without model restrictions as a function of measurement delay between encoding and judgment (immediate vs. 2-day delay), Experiment 5.

Parameter	Estimate (*SE*)	95% CI	Difference to reference point
R			
Immediate	0.36 (0.01)	[0.34, 0.38]	$\Delta G^2(1) = 844.91$, $p < .001$, $w = 0.365$
2-day delay	0.24 (0.01)	[0.22, 0.27]	$\Delta G^2(1) = 381.28$, $p < .001$, $w = 0.244$
C			
Immediate	0.04 (02)	[0.01, 0.08]	$\Delta G^2(1) = 5.74$, $p = .017$, $w = 0.030$
2-day delay	0.01 (0.02)	[−0.02, 0.05]	$\Delta G^2(1) = 0.86$, $p = .354$, $w = 0.012$
B			
Immediate	0.53 (0.01)	[0.51, 0.55]	$\Delta G^2(1) = 8.96$, $p = .003$, $w = 0.038$
2-day delay	0.50 (0.01)	[0.49, 0.52]	$\Delta G^2(1) = 0.25$, $p = .618$, $w = 0.006$

Note. The *R* parameter captures effects of relational information; the *C* parameter captures effects of co-occurrence; the *B* parameter captures general response biases. The neutral reference point for *R* and *C* is 0; the neutral reference point for *B* is 0.5, with scores higher than 0.5 reflecting a general bias toward positive responses and scores lower than 0.5 reflecting a general bias toward negative responses.

retrieval of stored information about stimulus relations, which should decrease the impact of quantitative relations and increase the impact of mere co-occurrence. Together, these assumptions imply that long (vs. short) delays between encoding and judgment should reduce scores on the *R* parameter and increase scores on the *C* parameter. A different set of predictions can be derived from dual-learning theories suggesting that mental representations of relational information involve multiple layers within associative networks (Doumas, Hummel, & Sandhofer, 2008; McClelland, McNaughton, & O'Reilly, 1995; Smith & DeCoster, 2000). According to such multi-layer network theories, activated concepts at higher levels specify the relation between activated concepts at lower levels (Gawronski & Bodenhausen, 2018; Gawronski, Brannon, & Bodenhausen, 2017). Thus, to the extent that hierarchical representations involving multiple layers of associative links are more likely affected by memory decay compared to direct associative links between two concepts, effects of mere co-occurrence should be more stable over time compared to effects of relational information. From this perspective, longer temporal delays between encoding and judgment should reduce the impact of quantitative relations, with the impact of mere co-occurrence being less affected by temporal delays. Together, these assumptions imply that long (vs. short) delays between encoding and judgment should reduce scores on the *R* parameter without affecting scores on the *C* parameter. The main goal of Experiment 5 was to test these competing predictions by measuring attribute judgments immediately after encoding and then again after a two-day delay.

9.1. Method

9.1.1. Participants

We aimed to recruit 200 participants from Amazon's MTurk to complete assessments at two time points, approximately two days apart. Based on prior research from our lab, we expected that approximately 33% of participants who completed the assessment at Time 1 would not accept the invitation to complete the assessment at Time 2. We therefore oversampled at Time 1 by recruiting 300 participants. Data collection at Time 1 was completed over a period of approximately 24 h between March 25 and March 26, 2020. Of the 329 participants who started the assessment at Time 1 (345 submissions), 303 participants completed the assessment in full. Of these participants, 300 participants were invited back for participation at Time 2 via follow-up emails through the bonus payment system in MTurk.[6] The data collection at Time 2 began roughly 48 h after the data collection for Time 1 was finished, and was completed over a 48-h time period spanning between March 28 and March 30, 2020. Of the 209 participants who started the assessment at Time 2 (205 submissions), 202 participants completed the assessment in full. Four participants in the data set had more than one submission, in which case only the first submission was retained.

To merge the data from participants at Time 1 and Time 2, participants provided one-digit responses to five personal questions to form a unique 5-digit code (see below). Of the 202 participants completing the assessment in full at Time 2, 103 participants provided fully matching codes at the two time points.[7] To link the submissions of the remaining participants, data were merged across time points if (1) at least 3 digits of the codes provided across time points matched and (2) the demographic information (gender, age, ethnicity) provided across time points was identical, with the exception of age which we allowed to be one year greater at Time 2. If a submission at Time 2 met these criteria for multiple submissions at Time 1, then the submission could not be uniquely linked to a submission at Time 1. Using this procedure, we were able to link the submissions of 165 participants across time points. Of these participants, 6 participants were excluded because they reported that they were inattentive or did not take their responses seriously at either Time 1 or Time 2, 7 participants were excluded for failing the attention check at either Time 1 or Time 2, and 12 participants were excluded for failing the materials comprehension check, resulting in a final sample of 140 participants (41.43% female, 58.57% male; $M_{\text{age}} = 36.55$, $SD_{\text{age}} = 10.44$).

9.1.2. Procedure

The materials, learning task, judgment task, and additional measures in Experiment 5 were identical to Experiment 1 with four exceptions. First, in addition to measuring attribute judgments immediately after encoding, we measured attribute judgments a second time after a two-way delay. Second, to link participants' responses across time points, participants were asked to provide one-digit answers to a series of five personal questions (e.g., *please type in the second letter of your first name*) at the end of the assessment at each time point. Answers to these five questions were concatenated to create a unique 5-digit code for each participant. Third, to avoid revealing our central research question, the debriefing information was modified at the end of the assessment at Time 1. Finally, participants were not asked to complete the materials comprehension check at Time 2, because this information was already obtained at Time 1.

9.2. Results

Attribute judgments were aggregated in line with the procedures in Experiment 1. Means and 95% confidence intervals of the relative

[6] Of the 303 participants who completed the assessment at Time 1, three participants either did not submit a completion code or submitted an incorrect completion code. As a consequence, these participants were not sent a follow-up email via MTurk for participation at Time 2.

[7] Two participants submitted the same code at Time 1, and only one of these participants completed the assessment and submitted their code at Time 2. In this case, the submission at Time 2 was linked to one of the two submissions at Time 1 by matching demographic information between the submissions.

proportion of *healthy* (vs. *unhealthy*) judgments as a function of product information and time delay are presented in Table 1. The RCB model fit the data well with six free parameters (i.e., three per condition), $G^2(2) = 4.49$, $p = .106$, $w = 0.019$. Parameter estimates obtained with the baseline model are presented in Table 6. The *R* parameter was significantly greater in short-delay condition compared to the long-delay condition, $\Delta G^2(1) = 48.56$, $p < .001$, $w = 0.062$, indicating that relational information had a greater impact when judgments were measured immediately after encoding compared to a two-day delay. The *C* parameter did not significantly differ across time delay conditions, $\Delta G^2(1) = 1.42$, $p = .233$, $w = 0.011$. The *B* parameter showed a marginal effect of time delay, $\Delta G^2(1) = 3.84$, $p = .050$, $w = 0.017$, indicating a greater general tendency to judge the products as healthy when judgments were measured immediately after encoding compared to a two-day delay.

9.3. Discussion

Consistent with the shared prediction of selective-retrieval and dual-learning accounts, a longer temporal delay between encoding and judgment reduced the impact of quantitative relations. Yet, temporal delay had no significant effect on the impact of mere co-occurrence. The latter finding is consistent with predictions derived from dual-learning accounts, but it is inconsistent with predictions derived from selective-retrieval accounts. Whereas selective-retrieval accounts suggest that longer temporal delays between encoding and judgment should decrease the impact of quantitative relations and increase the impact of mere co-occurrence, dual-learning accounts suggest that longer temporal delays between encoding and judgment should decrease the impact of quantitative relations, with the impact of mere co-occurrence being less affected by temporal delays.

10. Experiment 6

Across Experiments 2–5, selective-retrieval and dual-learning accounts fared very well in predicting contextual influences on the effect of quantitative relations, in that their shared predictions were confirmed in every single study. However, the two accounts fared less well in predicting the functional properties of mere co-occurrence effects. An important aspect for the interpretation of these mixed results is that, while effects of quantitative relations were relatively large, mere co-occurrence effects were extremely small overall. In fact, the *C* parameter reflecting mere co-occurrence effects was significantly different from zero in only two out of nine cases, showing effect sizes that consistently fell below the conventional benchmark of a small effect (see Tables 2-6). The results were remarkably different for the effect of quantitative relations captured by the *R* parameter, which was significantly different from zero in all nine cases with an average effect size that qualifies as medium in terms of Cohen's (1988) conventions (see Tables 2-6). These findings stand in contrast to earlier research using the RCB model to investigate effects of mere co-occurrence and qualitative relations on evaluative judgments, which obtained (1) much stronger effects of mere co-occurrence and (2) much weaker effects of relational information compared to the current studies (e.g.,Gawronski & Brannon, 2021; Heycke & Gawronski, 2020).

The discrepancy in the obtained effect sizes raises the question of whether the strong emphasis on relational information in the learning instructions of the current studies enhanced effects of relational information, which might suppress the emergence of mere co-occurrence effects. Such a compensatory impact would be consistent with selective-retrieval accounts and studies suggesting that a focus on overall outcomes in the processing of relational information can reduce mere co-occurrence effects (Moran, Bar-Anan, & Nosek, 2015). Although prior research using the RCB model obtained strong evidence for mere co-occurrence effects on evaluative judgments with instructions that included a similarly strong emphasis on relational information (Gawronski & Brannon, 2021; Heycke & Gawronski, 2020), Experiment 6 aimed to investigate whether mere-occurrence effects become more pronounced when the strong emphasis on relational information is removed from the learning instructions.

10.1. Method

10.1.1. Participants

We aimed to recruit 400 participants from Amazon's MTurk. The data collection was completed in April 2021. The same eligibility criteria for participation in Experiments 1–5 were used in Experiment 6 with the exception that MTurk workers were required to have successfully completed 100 previous assignments rather than only one previous assignment. Of the 443 participants who started the assessment (472 submissions), 406 participants completed the assessment in full. One participant had more than one complete submission, in which case only the first submission was retained. Of the 406 participants in the data set, 20 participants were excluded because they reported that they were inattentive or did not take their responses seriously, 40 participants were excluded for failing the attention check, 57 participants were excluded for failing the materials comprehension check, and 2 participants were excluded for failing to provide valid responses to at least 50% of the judgment trials, resulting in a final sample of 287 participants (39.37% female, 60.28% male, 0.35% prefer not to answer; $M_{age} = 37.37$, $SD_{age} = 11.13$). Participants were compensated $2.00 for their time.

10.1.2. Procedure

The materials, learning task, judgment task, and additional measures in Experiment 6 were identical to Experiment 1 with one exception. In Experiment 1, the instructions for the learning task asked participants to form impressions of food products as healthy or unhealthy based on whether the product is said to have more or less of an ingredient, thereby emphasizing the importance of relational information in the formation of impressions. In Experiment 6, participants were randomly assigned to one of two conditions in which they either received the original instructions for the learning task (relational-instructions condition) or a revised set of instructions that did not emphasize the importance of relational information (minimal-instructions condition). The specific revised set of instructions were as follows:

> *The main goal of the present study is to investigate how people form impressions of food products. Toward this end, you will be presented with information about various food products. Please form an impression of these products based on the presented information.*

Table 7
Parameter estimates without model restrictions as a function of learning task instructions (relational vs. minimal), Experiment 6.

Parameter	Estimate (*SE*)	95% CI	Difference to reference point
R			
Relational	0.20 (0.01)	[0.17, 0.22]	$\Delta G^2(1) = 232.23$, $p < .001$, $w = 0.134$
Minimal	0.17 (0.01)	[0.15, 0.19]	$\Delta G^2(1) = 197.83$, $p < .001$, $w = 0.124$
C			
Relational	0.03 (0.02)	[−0.00, 0.06]	$\Delta G^2(1) = 2.94$, $p = .087$, $w = 0.015$
Minimal	0.01 (0.01)	[−0.02, 0.04]	$\Delta G^2(1) = 0.55$, $p = .459$, $w = 0.007$
B			
Relational	0.51 (0.01)	[0.50, 0.53]	$\Delta G^2(1) = 2.65$, $p = .103$, $w = 0.014$
Minimal	0.53 (0.01)	[0.52, 0.54]	$\Delta G^2(1) = 17.52$, $p < .001$, $w = 0.037$

Note. The *R* parameter captures effects of relational information; the *C* parameter captures effects of co-occurrence; the *B* parameter captures general response biases. The neutral reference point for *R* and *C* is 0; the neutral reference point for *B* is 0.5, with scores higher than 0.5 reflecting a general bias toward positive responses and scores lower than 0.5 reflecting a general bias toward negative responses.

10.2. Results

Attribute judgments were aggregated in line with the procedures in Experiment 1. Means and 95% confidence intervals of the relative proportion of *healthy* (vs. *unhealthy*) judgments as a function of product information and learning-task instructions are presented in Table 1. The RCB model fit the data well with six free parameters (i.e., three per condition), $G^2(2) = 1.72$, $p = .423$, $w = 0.012$. Parameter estimates obtained with the baseline model are presented in Table 7. Neither the *R* parameter, $\Delta G^2(1) = 2.16$, $p = .141$, $w = 0.013$, nor the *C* parameter, $\Delta G^2(1) = 0.58$, $p = .445$, $w = 0.007$, significantly differed across learning-task instruction conditions, indicating that the learning-task instructions did not moderate the impact of either relational or co-occurrence information on attribute judgments. There was also no significant effect of learning-task instructions on the *B* parameter, $\Delta G^2(1) = 2.54$, $p = .111$, $w = 0.014$.

10.3. Discussion

Counter to the idea that the small size of mere co-occurrence effects in Experiments 1–5 might have been due to the strong emphasis on relational information in the instructions for the learning task, scores on *C* parameter were unaffected by whether the learning instructions did or did not include a strong emphasis on relational information. If anything, scores on the C parameter became smaller (rather than larger) when the emphasis on relational information was removed from the instructions. Yet, scores on the *C* parameter did not significantly differ from zero regardless of learning instructions (see Table 7). These results rule out potential concerns that the small size of mere co-occurrence effects in Experiments 1–5 is an artifact of the employed instructions. However, together with the small effects sizes obtained in Experiments 1–5, the findings of Experiment 6 raise further questions about the extent to which mere co-occurrence influences attribute judgments in cases involving quantitative relations. We will return to this question in the General Discussion where we discuss implications of our findings.

11. Experiment 7

A central assumption that guided the current research is that the observed judgments reflect mental representations of specific attributes. However, compelling evidence for this idea is still lacking, in that the observed judgmental effects may be driven by broad evaluative representations rather representations of specific attributes. Specifically, it is possible that participants formed broad evaluative representations of the products as “good” or “bad” during the learning task, and then used these representations as a basis for their judgments of the products in terms of semantic attributes with a positive (healthy) or negative (unhealthy) connotation. Experiment 7 aimed to rule out this alternative interpretation. Toward this end, all participants completed the same basic learning task. Following the procedure in Experiments 1–6, half of the participants were then asked to judge whether the products are healthy or unhealthy (healthiness-judgment condition). The remaining half was asked to judge whether the products are tasty or bland (tastiness-judgment condition). The rationale underlying this manipulation is that effects of specific attribute representations should be limited to the focal attribute dimension, whereas effects of broad evaluative representations should lead to corresponding effects for other attribute dimensions with evaluative connotations. That is, if the obtained results are driven by semantic representations of specific attributes, effects of quantitative relations on the *R* parameter should be significantly greater in the healthiness-judgment condition compared to the tastiness-judgment condition, with scores being significantly different from zero only in the healthiness-judgment condition but not in the tastiness-judgment condition. In contrast, if the obtained results are driven by broad evaluative representations, effects of quantitative relations on the *R* parameter should not differ between the healthiness-judgment condition and the tastiness-judgment condition, with scores in both conditions being significantly different from zero.

11.1. Method

11.1.1. Participants

We aimed to recruit 400 participants from Amazon’s MTurk. The data collection was completed in May 2021. The same eligibility criteria for participation in Experiment 6 were used for the current study. Of the 462 participants who started the assessment, 409 completed the assessment in full. Three participants had more than one submission, in which case only the first submission was retained. Of the 406 participants in the data set, 22 were excluded because they reported that they were inattentive or did not take their responses seriously, 43 were excluded for failing the attention check, 43 were excluded for failing the materials comprehension check, and 9 were excluded for failing to provide valid responses to at least 50% of the judgment trials. The final sample thus comprised 289 participants (44.98% female, 54.32% male, 0.35% prefer not to answer, 0.35% other; $M_{age} = 40.39$, $SD_{age} = 12.92$). Participants were compensated $2.00 for their time.

11.1.2. Procedure

The materials, learning task, judgment task, and additional measures in Experiment 7 were identical to Experiment 1 with one exception. In Experiment 1, the instructions for the judgment task asked participants to indicate whether each product is healthy or unhealthy. To determine if responses in the previous experiments reflect broad evaluative representations (e.g., good vs. bad) rather than representations of specific attributes (i.e., healthy vs. unhealthy), participants in Experiment 7 were randomly assigned to one of two judgment conditions. In the

Table 8
Parameter estimates without model restrictions as a function of judgment type (healthiness vs. tastiness), Experiment 7.

Parameter	Estimate (*SE*)	95% CI	Difference to reference point
R			
Healthiness	0.24 (0.01)	[0.22, 0.26]	$\Delta G^2(1) = 401.09$, $p < .001$, $w = 0.176$
Tastiness	0.02 (0.01)	[−0.00, 0.05]	$\Delta G^2(1) = 3.25$, $p = .071$, $w = 0.016$
C			
Healthiness	0.00 (0.02)	[−0.03, 0.03]	$\Delta G^2(1) = 0.00$, $p = .999$, $w = 0.000$
Tastiness	0.00 (0.01)	[−0.02, 0.03]	$\Delta G^2(1) = 0.06$, $p = .812$, $w = 0.002$
B			
Healthiness	0.54 (0.01)	[0.52, 0.55]	$\Delta G^2(1) = 24.71$, $p < .001$, $w = 0.044$
Tastiness	0.55 (0.01)	[0.54, 0.56]	$\Delta G^2(1) = 62.28$, $p < .001$, $w = 0.070$

Note. The *R* parameter captures effects of relational information; the *C* parameter captures effects of co-occurrence; the *B* parameter captures general response biases. The neutral reference point for *R* and *C* is 0; the neutral reference point for *B* is 0.5, with scores higher than 0.5 reflecting a general bias toward positive responses and scores lower than 0.5 reflecting a general bias toward negative responses.

healthiness-judgment condition, participants were asked to indicate whether the products are healthy or unhealthy. In the tastiness-judgment condition, participants were asked to indicate whether the products are tasty or bland.

11.2. Results

The judgments were aggregated in line with the procedures in Experiment 1. Means and 95% confidence intervals of the relative proportions of *healthy* (vs. *unhealthy*) and *tasty* (vs. *bland*) judgments as a function of product information and judgment-task instructions are presented in Table 1. The RCB model fit the data well with six free parameters (i.e., three per condition), $G^2(2) = 1.64$, $p = .440$, $w = 0.011$. Parameter estimates obtained with the baseline model are presented in Table 8. Whereas the C parameter, $\Delta G^2(1) = 0.06$, $p = .812$, $w = 0.002$, and the B parameter, $\Delta G^2(1) = 1.66$, $p = .198$, $w = 0.011$, did not significantly differ across judgment conditions, there was a significant effect of Judgment Task on the R parameter, $\Delta G^2(1) = 158.12$, $p < .001$, $w = 0.111$, indicating that relational information had a greater impact on judgments in the healthiness-judgment condition compared to the tastiness-judgment condition. The R parameter was significantly different from zero in the healthiness-judgment condition, but not in the tastiness-judgment condition (see Table 8).

11.3. Discussion

Results of Experiment 7 support our assumption that the judgments observed in Experiments 1–6 reflect mental representations of specific attributes rather than broad evaluative representations. Consistent with this assumption, effects of health-related information about quantitative relations on the R parameter were significantly greater when participants were asked to judge the presented products in terms of their healthiness than when they were asked to judge the products in terms of their tastiness. Moreover, scores on the R parameter were significantly different from zero only in the healthiness-judgment condition but not in the tastiness-judgment condition. If the obtained results were driven by broad evaluative representations, effects of quantitative relations on the R parameter should not differ between the two judgment conditions and scores on the R parameter should be significantly different from zero in both conditions.

12. General discussion

The current research aimed to address two questions: (1) Does mere co-occurrence influence judgments regarding specific attributes irrespective of information about quantitative relations? (2) How do processing conditions during encoding and judgment moderate effects of mere co-occurrence and quantitative relations? We will first discuss the obtained evidence regarding the second question, before we return to the first question.

Overall, selective-retrieval and dual-learning accounts fared very well in predicting contextual influences on the effect of quantitative relations. Consistent with the shared predictions of the two accounts, information about quantitative relations had a greater impact on attribute judgments when time for encoding was long rather than short (Experiment 2), when the information was presented more frequently rather than less frequently (Experiment 3), when participants had more time to make a judgment than when they had less time (Experiment 4), and when participants made their judgments immediately after encoding than when they made their judgments after a two-day delay (Experiment 5).

However, different from the high accuracy in predicting contextual influences on the effect of quantitative relations, the two accounts fared less well in predicting the functional properties of mere co-occurrence effects. Selective-retrieval accounts correctly predicted the finding that mere co-occurrence effects were greater when there was less time for encoding than when there was more time for encoding (Experiment 2). However, the predictions derived from selective-retrieval accounts conflict with the obtained null effects of information repetition (Experiment 3), time during judgment (Experiment 4), and temporal delay (Experiment 5). According to selective-retrieval accounts, mere co-occurrence effects should be greater when there is less (vs. more) time for encoding, when relational information is presented less (vs. more) frequently, when there is less (vs. more) time to make a judgment, and when the delay between encoding and judgment is long (vs. short).

Dual-learning accounts correctly predicted the finding that mere co-occurrence effects were unaffected by temporal delay (Experiment 5). However, the predictions derived from dual-learning accounts conflict with the obtained null effects of information repetition (Experiment 3) and time during judgment (Experiment 4). They are also inconsistent with the finding that more time for encoding reduced mere co-occurrence effects (Experiment 2). According to dual-learning accounts, mere co-occurrence effects should be greater when relational information is presented more (vs. less) frequently and when there is less (vs. more) time to make a judgment. Moreover, mere co-occurrence effects should be unaffected by time for encoding and delays between encoding and judgment.

An important aspect for the interpretation of these mixed results is that effects of quantitative relations were relatively large overall, whereas mere co-occurrence effects were very small. In fact, the C parameter reflecting mere co-occurrence effects was significantly different from zero in only two out of 14 cases, showing effect sizes that consistently fell below the conventional benchmark of a small effect.[8] The results were remarkably different for the effect of quantitative relations captured by the R parameter, which was significantly different from zero in all 14 cases with an average effect size that is close to a medium-size effect in terms of Cohen's (1988) conventions. These findings stand in contrast to earlier research using the same multinomial modeling approach to investigate effects of mere co-occurrence and qualitative relations on evaluative judgments, which obtained (1) much stronger effects of mere co-occurrence and (2) much weaker effects of relational information compared to the current studies (e.g., Gawronski & Brannon, 2021; Heycke & Gawronski, 2020).

Considering the rather small effects of mere co-occurrence in conjunction with the obtained evidence regarding contextual influences on the effects of mere co-occurrence and quantitative relations, the current findings might be best explained via a combination of propositional processes during learning and selective retrieval during judgment. When learning about quantitative relations (e.g., *product A has less sodium*), people may infer specific attributes during encoding via propositional reasoning and store the outcome of these inferences in memory (e.g., *product A is healthy*). Because abstract representations of specific attributes do not include episodic information about co-occurring stimuli (as would be the case for *product A has less sodium*), factors that influence the storage or retrieval of abstract attribute information should moderate only the impact of quantitative relations without producing mere co-occurrence effects and without influencing the size of mere co-occurrence effects in a compensatory fashion. These post-hoc assumptions would explain why the contextual factors investigated in the current studies consistently showed the predicted effects on the impact of relational information, with mere co-occurrence effects being close to zero regardless of contextual conditions. Nevertheless, extreme time pressure during encoding (as in Experiment 2) may disrupt propositional inferences of abstract attributes, changing the content of stored information from abstract attributes (e.g., *product A is healthy*) to

[8] The 14 cases include all individual conditions of Experiments 1–7 and Experiment S1 reported in the Supplemental Materials, the only exception being the tastiness-judgment condition in Experiment 7 for which our attribute-judgment account would not predict any significant effects on the C and the R parameter.

episodic memories of specific relations (e.g., *product A has less sodium*). Because extreme time pressure during encoding should also impair the retrieval of stored information (see Craik & Lockhart, 1972), it may not only reduce the impact of quantitative relations but also produce mere co-occurrence effects. Although these assumptions are admittedly post-hoc, we believe they provide a parsimonious, yet comprehensive, explanation of the current pattern of results.

12.1. Comparison to prior findings

An interesting question is how the functional properties obtained in the current studies compare to the functional properties obtained in previous research on effects of mere co-occurrence and qualitative relations on evaluative judgments. Consistent with the pattern obtained in the current studies, Heycke and Gawronski (2020) found that information about qualitative relations had a greater impact on evaluative judgments when time for encoding was long rather than short, when the information was presented more frequently rather than less frequently, when participants had more time to make a judgment than when they had less time, and when the delay between encoding and judgment was short rather than long. In addition, Heycke and Gawronski obtained null effects of repetition and temporal delay on the impact of mere co-occurrence, consistent with the null effects obtained in the current studies. Yet, there are two discrepancies between the current and earlier findings. First, Heycke and Gawronski found that mere co-occurrence effects significantly increased as a function of time during judgment, which conflicts with the null effect obtained in the current studies. Second, Heycke and Gawronski found no significant effect of time for encoding on the impact of mere co-occurrence, which conflicts with the current finding that more time for encoding significantly reduced mere co-occurrence effects. Because the current studies showed a mean-level pattern for time during judgment that is consistent with the effect obtained by Heycke and Gawronski, it seems possible that the null effect in the current studies is a false negative due to insufficient statistical power (Maxwell, Lau, & Howard, 2015). However, the conflicting effects of time for encoding are more difficult to reconcile, given that the mean-level pattern obtained by Heycke and Gawronski is directionally opposite compared to the significant effect in the current studies. Considering the conceptual differences between the two lines of work, the different effects of time for encoding might be driven by (1) a difference between information about qualitative versus quantitative relations or (2) a difference between evaluative versus attribute judgments (or both). Future research directly comparing the different cases in studies investigating effects of time for encoding may help to identify the mechanisms underlying the discrepant outcomes.[9]

12.2. Potential objections

Although the current research was guided by the difference between dual-learning and selective-retrieval accounts, it is worth noting that mere co-occurrence effects in cases involving quantitative relations could be rooted in an alternative mechanism that is different from the ones proposed by the two accounts. This mechanism may involve a propositional inference during encoding that a given product must contain a certain amount of a given ingredient if the product is said to have less of that ingredient. For example, learning that a product has less of an unhealthy ingredient may lead people to infer that the product must have a certain amount of the unhealthy ingredient, making it unhealthy even if it has comparatively less of that ingredient. Such inferences could lead to mere co-occurrence effects in the current paradigm over and above the hypothesized mechanisms of associative link formation and selective retrieval. Yet, despite this theoretical possibility, the operation of such a mechanism seems rather unlikely in light of the finding that mere co-occurrence effects were extremely small and not statistically significant in most cases, with information about quantitative relations showing relatively large effects. If anything, inferences about default ingredients should lead to stronger (not weaker) co-occurrence effects in the current paradigm compared to previous research on mere co-occurrence effects in information about qualitative relations.

Another potential concern is that asymmetric perceptions of healthiness might influence RCB model estimates in the current paradigm. For example, although having less of an unhealthy ingredient may be perceived as healthy, having more of a healthy ingredient may be perceived as even healthier. Similarly, although having less of a healthy ingredient may be perceived unhealthy, having more of an unhealthy ingredient may be perceived as even unhealthier. Within our multinomial modeling approach, such asymmetries should negatively affect the reliability of the *R* parameter, but it has no implications for the reliability of the *C* parameter, the latter of which depends exclusively on health perceptions of the ingredients independent of the described quantitative relations (e.g., perceptions of sodium as being unhealthy). Hence, potential asymmetries between the four cases should reduce the likelihood of detecting effects on the *R* parameter, but not the *C* parameter. Yet, counter to this concern, the *R* parameter consistently showed effects that were in line with the shared predictions derived from extant theories.

A related concern is that, if the predictions of selective-retrieval accounts had been consistently confirmed, the obtained evidence for ubiquitous compensatory relations between *C* and *R* would question a major premise of multinomial modeling, which requires that model parameters can vary independently (see Hütter & Klauer, 2016). Although evidence for a ubiquitous compensatory relation would be consistent with the predictions of selective-retrieval accounts and inconsistent with the predictions of dual-learning accounts, critics may object that theoretical conclusions from such findings should be made with great caution because a basic premise for the use of the RCB model would be violated. Again, this concern is ruled out by the findings that (1) mere co-occurrence effects on the *C* parameter were almost non-existent in the current study and (2) the *R* and the *C* parameters showed compensatory effects in only one of seven studies that investigated effects of contextual factors on *R* and *C* (including Experiment S1 reported in the Supplemental Materials).

A final question is whether the small, almost non-existent co-occurrence effects on the *C* parameter could be interpreted as counterevidence against the validity of the RCB model to study effects of mere co-occurrence and quantitative relations on attribute judgments. This relatively broad argument can be interpreted in two ways. First, one might argue that the very small, oftentimes non-significant scores on the model's *C* parameter indicate that the parameter is not measuring what it is supposed to measure. In response to this claim, we would argue that lack of a statistically significant score on the *C* parameter would provide evidence against the construct validity of the *C* parameter only if one can be certain that mere co-occurrence did have a meaningful effect, suggesting that the *C* parameter was unable to capture it. However, there is no independent evidence for the latter assumption, rendering claims about lack of construct validity premature. Second, one might argue that the very small, oftentimes non-significant scores on the model's *C* parameter indicate that the parameter is not necessary for describing

[9] It is worth noting that, when deriving predictions from single-process propositional accounts, Heycke and Gawronski (2020) not only considered the possibility of compensatory effects on *R* and *C* resulting from selective retrieval, but also the possibility of parallel effects resulting from fully disrupted learning and retrieval. A major problem with a joint consideration of the two possibilities is that it makes single-process propositional accounts consistent with any potential outcome for the *C* parameter, including compensatory effects, parallel effects, and null effects. Because we deem accounts uninformative if they do not prohibit any potential outcome (see Gawronski & Bodenhausen, 2015), we focused primarily on the notion of selective retrieval, which offers testable predictions that can be subject to empirical disconfirmation.

patterns of responses in the current paradigm. In response to this claim, we agree that consistent absence of a significant effect on the *C* parameter would provide evidence that the parameter is not necessary for describing patterns of responses. However, the parameter did show significant scores in a small number of cases and it was influenced by time for encoding in a theoretically meaningful way. These findings suggest that a reduced version of the model that does not include a parameter for mere co-occurrence effects would miss a determinant of attribute judgments that seems theoretically important even if its impact is relatively small overall.

13. Conclusion

In sum, the current findings provide only weak support for the idea that mere co-occurrence influences judgments regarding specific attributes irrespective of information about quantitative relations. Although mere co-occurrence effects seem to be more pronounced when there is little time for encoding, mere co-occurrence effects were extremely small overall and statistically significant in only a small number of cases. Although selective-retrieval and dual-learning accounts face difficulties in explaining the full set of evidence regarding the functional properties of effects of mere co-occurrence and quantitative relations, the findings can be explained via a combination of propositional inferences during learning and selective retrieval during judgment.

Open practices

The materials, raw data, and analysis files for all studies are publicly available at https://osf.io/3p5y2/.

Acknowledgements

The reported research was supported by National Science Foundation Grant BCS-1649900. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Appendix A

The data analytic approach of Heycke and Gawronski's (2020) RCB model can be illustrated by means of a multinomial processing tree that specifies potential patterns of judgments as a function of whether a product is said to have either more of less of either a healthy or an unhealthy ingredient (see Fig. 1). The four paths on the left side of the figure depict the four potential cases that (1) judgments of the product reflect the quantitative relation specified in the message, (2) judgments of the product reflect its mere co-occurrence with a healthy or unhealthy ingredient, (3) judgments of the product reflect a general positivity bias to respond *healthy*, and (4) judgments of the product reflect a general negativity bias to respond *unhealthy*. The table on the right side of the figure depicts the patterns of judgments for each of the four cases as a function of relational information (i.e., more vs. less) and the nature of the ingredient (i.e., healthy vs. unhealthy).

If judgments of a given product are driven by the quantitative relation in the message, participants should judge the product as healthy when it has more of a healthy ingredient and less of an unhealthy ingredient, and participants should judge the product as unhealthy when it has less of a healthy ingredient and more of an unhealthy ingredient (first path in Fig. 1). If judgments of a given product are driven by mere co-occurrence, participants should judge the product as healthy when it co-occurred with a healthy ingredient and as unhealthy when it co-occurred with an unhealthy ingredient (second path in Fig. 1). If judgments of a given product are driven by a general positivity bias, participants should judge the product as healthy regardless of the co-occurring ingredient and the quantitative relation in the message (third path in Fig. 1). Conversely, if judgments of a given product are driven by a general negativity bias, participants should judge the product as unhealthy regardless of the co-occurring ingredient and the quantitative relation in the message (fourth path in Fig. 1).

Based on the processing tree depicted in Fig. 1, multinomial modeling provides numerical estimates for (1) the probability that information about quantitative relations drives judgments (captured by the parameter *R* in Fig. 1); (2) the probability that mere co-occurrence drives judgments if information about quantitative relations does not drive judgments (captured by the parameter *C* in Fig. 1); and (3) the probability that a general positivity or negativity bias drives judgments if neither information about quantitative relations nor mere co-occurrence drive judgments (captured by the parameter *B* in Fig. 1). Numerical scores for the three probabilities are estimated by means of four non-redundant mathematical equations derived from the processing tree (see Appendix B).[10] These equations include the three model parameters *R*, *C*, and *B* as unknowns (henceforth, RCB model; see Heycke & Gawronski, 2020) and the empirically observed probabilities of *healthy* versus *unhealthy* judgments in the four product conditions (i.e., more of healthy ingredient; less of healthy ingredient; more of unhealthy ingredient; less of unhealthy ingredient) as known values. Using maximum likelihood statistics, multinomial modeling generates numerical estimates for the three unknowns that minimize the discrepancy between the empirically observed probabilities of *healthy* versus *unhealthy* judgments in the four product conditions and the probabilities of *healthy* versus *unhealthy* judgments predicted by the model equations using the generated parameter estimates.

The adequacy of the model in describing the data can be evaluated by means of goodness-of-fit statistics, with poor model fit being reflected in a statistically significant deviation between the empirically observed probabilities in a given data set and the probabilities predicted by the model. The estimated scores for each parameter can vary between 0 and 1. For the *R* parameter, scores significantly greater than zero indicate that responses were affected by information about quantitative relations. For the *C* parameter, scores significantly greater than zero indicate that responses were affected by mere co-occurrence. Finally, for the *B* parameter, scores significantly greater than 0.5 indicate a general positivity bias and scores significantly lower than 0.5 indicate a general negativity bias.

Differences from these reference points can be tested by enforcing a specific value for a given parameter and comparing the fit of the restricted model to the fit of the unrestricted baseline model. If setting a given parameter equal to a specific reference point leads to a significant reduction in model fit, it can be inferred that the parameter estimate is significantly different from that reference point. For example, to test whether mere co-occurrence influenced judgments, the *C* parameter is set equal to zero and the resulting model fit is compared to the fit of the model that does not include any restrictions for the *C* parameter. To the extent that enforcing a parameter estimate of zero leads to a significant reduction in model fit, it can be inferred that mere co-occurrence significantly influenced participants' judgments. The same approach can be used to test the influence of information about quantitative relations captured by the *R* parameter. For the *B* parameter, comparisons to reference values are equivalent, except

[10] Because multinomial modeling is based on binary responses with p(positive response) = 1 – p(negative response), there are only four non-redundant equations in the set of eight equations listed in Appendix B.

that the reference value reflecting the absence of a general response bias is 0.5. Similar tests can be conducted to investigate whether estimates for a given parameter significantly differ across groups, which can be tested by enforcing equal estimates for that parameter across groups. If setting a given parameter equal across groups leads to a significant reduction in model fit, it can be inferred that the parameter estimates for the two groups are significantly different.

In the current research, multinomial modeling analyses were conducted using the free software multiTree v0.43 (Moshagen, 2010) and multiTree template files for RCB model analyses provided by Heycke and Gawronski (2020) at https://osf.io/7ac4d/. Following Heycke and Gawronski (2020), all of the reported studies used the same estimation algorithm with random start values, two replications, and a maximum of 90,000 iterations.

Appendix B

Model equations for the estimation of effects of stimulus relations (*R*), stimulus co-occurrence (*C*), and general response bias (*B*) on health judgments of objects that have more or less of a healthy or unhealthy ingredient.

$$p(\text{healthy} \mid \text{more of healthy ingredient}) = R + [(1–R) \times C] + [(1–R) \times (1–C) \times B]$$

$$p(\text{healthy} \mid \text{more of unhealthy ingredient}) = (1–R) \times (1–C) \times B$$

$$p(\text{healthy} \mid \text{less of healthy ingredient}) = [(1–R) \times C] + [(1–R) \times (1–C) \times B]$$

$$p(\text{healthy} \mid \text{less of unhealthy ingredient}) = R + [(1–R) \times (1–C) \times B]$$

$$p(\text{unhealthy} \mid \text{more of healthy ingredient}) = (1–R) \times (1–C) \times (1–B)$$

$$p(\text{unhealthy} \mid \text{more of unhealthy ingredient}) = R + [(1–R) \times C] + [(1–R) \times (1–C) \times (1–B)]$$

$$p(\text{unhealthy} \mid \text{less of healthy ingredient}) = R + [(1–R) \times (1–C) \times (1–B)]$$

$$p(\text{unhealthy} \mid \text{less of healthy ingredient}) = [(1–R) \times C] + [(1–R) \times (1–C) \times (1–B)]$$

Appendix C. Supplemental Materials

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2021.104193.

References

Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods, 45*, 527–535.

Bading, K., Stahl, C., & Rothermund, K. (2020). Why a standard IAT effect cannot provide evidence for association formation: The role of similarity construction. *Cognition and Emotion, 34*, 128–143.

Calanchini, J. (2020). How multinomial processing trees have advanced, and can continue to advance, research using implicit measures. *Social Cognition, 38*, s165–s186.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Routledge.

Corneille, O., & Hütter, M. (2020). Implicit? What do you mean? A comprehensive review of the delusive implicitness construct in attitude research. *Personality and Social Psychology Review, 24*, 212–232.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684.

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior, 37*, 1–20.

De Houwer, J. (2018). Propositional models of evaluative conditioning. *Social Psychological Bulletin, 13*(3), Article e28046.

De Houwer, J., Van Dessel, P., & Moran, T. (2020). Attitudes beyond associations: On the role of propositional representations in stimulus evaluation. *Advances in Experimental Social Psychology, 61*, 127–183.

Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review, 115*, 1–43.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*, 692–731.

Gawronski, B., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology, 44*, 59–127.

Gawronski, B., & Bodenhausen, G. V. (2014). The associative-propositional evaluation model: Operating principles and operating conditions of evaluation. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 188–203). New York: Guilford Press.

Gawronski, B., & Bodenhausen, G. V. (2015). Theory evaluation. In B. Gawronski, & G. V. Bodenhausen (Eds.), *Theory and explanation in social psychology* (pp. 3–23). New York: Guilford Press.

Gawronski, B., & Bodenhausen, G. V. (2018). Evaluative conditioning from the perspective of the associative-propositional evaluation model. *Social Psychological Bulletin, 13*(3), Article e28024.

Gawronski, B., & Brannon, S. M. (2021). Attitudinal effects of stimulus co-occurrence and stimulus relations: Range and limits of intentional control. *Personality and Social Psychology Bulletin* (in press).

Gawronski, B., Brannon, S. M., & Bodenhausen, G. V. (2017). The associative-propositional duality in the representation, formation, and expression of attitudes. In R. Deutsch, B. Gawronski, & W. Hofmann (Eds.), *Reflective and impulsive determinants of human behavior* (pp. 103–118). New York: Psychology Press.

Gawronski, B., Brannon, S. M., & Luke, D. M. (2021). Unintentional influences in intentional impression formation. In E. Balcetis, & G. B. Moskowitz (Eds.), *The handbook of impression formation: A social psychological approach*. New York: Routledge (in press).

Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 283–310). New York: Cambridge University Press.

Gawronski, B., Walther, E., & Blank, H. (2005). Cognitive consistency and the formation of interpersonal attitudes: Cognitive balance affects the encoding of social information. *Journal of Experimental Social Psychology, 41*, 618–626.

Heycke, T., & Gawronski, B. (2020). Co-occurrence and relational information in evaluative learning: A multinomial modeling approach. *Journal of Experimental Psychology: General, 149*, 104–124.

Högden, F., & Unkelbach, C. (2021). The role of relational qualifiers in attribute conditioning: Does disliking an athletic person make you unathletic? *Personality and Social Psychology Bulletin, 47*, 643–656.

Hu, X., Gawronski, B., & Balas, R. (2017). Propositional versus dual-process accounts of evaluative conditioning: I. the effects of co-occurrence and relational information on implicit and explicit evaluations. *Personality and Social Psychology Bulletin, 43*, 17–32.

Hughes, S., Ye, Y., Van Dessel, P., & De Houwer, J. (2019). When people co-occur with good or bad events: Graded effects of relational qualifiers on evaluative conditioning. *Personality and Social Psychology Bulletin, 45*, 186–208.

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

Kukken, N., Hütter, M., & Holland, R. (2020). Are there two independent evaluative conditioning effects in relational paradigms? Dissociating effects of CS-US pairings and their meaning. *Cognition and Emotion, 34*, 170–187.
Kurdi, B., & Dunham, Y. (2020). Propositional accounts of implicit evaluation: Taking stock and looking ahead. *Social Cognition, 38*, s42–s67.
Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *American Psychologist, 70*, 487–498.
McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*, 419–457.
Moran, T., & Bar-Anan, Y. (2013). The effect of object–valence relations on automatic evaluation. *Cognition and Emotion, 27*, 743–752.
Moran, T., Bar-Anan, Y., & Nosek, B. (2015). Processing goals moderate the effect of co-occurrence on automatic evaluation. *Journal of Experimental Social Psychology, 60*, 157–162.
Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods, 42*, 42–54.
Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*, 867–872.
Payne, B. K., Burkley, M., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology, 94*, 16–31.
Sherman, J. W., Krieglmeyer, R., & Calanchini, J. (2014). Process models require process measures. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 121–138). New York: Guilford Press.
Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*, 108–131.
Van Dessel, P., Gawronski, B., & De Houwer, J. (2019). Does explaining social behavior require multiple memory systems? *Trends in Cognitive Sciences, 23*, 368–369.