MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„The red-anger congruence effect:
Metaphor congruency or polarity correspondence?"

verfasst von / submitted by

Marie-Luise Augsten BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2021 / Vienna 2021

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Psychologie UG2002
Betreut von / Supervisor:	Univ.-Prof. Dr. Ulrich Ansorge

Content

1. Introduction

Associations between colors and emotions or emotional valence are highly relevant for areas of daily life such as traffic, stock markets, or advertisement, to name but a few. Color-emotion associations can vary across cultures (Jonauskaite et al., 2019), though they seem to be largely universal as well, with the degree of similarity depending on linguistical and geographical proximity of countries (Jonauskaite et al., 2020). A lot of research has focused on the color red as it is assumed to be highly salient in evolutionary domains (Elliot, 2015). Connotations with red have been examined in various contexts and seem to be multifaceted and often even contradictory (Elliot & Maier, 2012; Schietecat et al., 2018b). Depending on the context, negative associations such as danger, anger, and failure as well as positive associations like love and romance are possible (Kaya & Epps, 2004; Moller et al., 2009; Pravossoudovitch et al., 2014; Winskel et al., 2021). Specifically, many studies investigated the association between red and anger (Fetterman et al., 2011; Fetterman et al., 2012; Fugate & Franco, 2019; Peromaa & Olkkonen, 2019; Sutton & Altarriba, 2016; Young et al., 2013) which is argued to arise from physiological states such as facial flushing (Drummond, 1997) and metaphoric links to heat, fire, and blood (Gibbs, 1994), and also finds expressions in language such as “seeing red” or to be “red with rage” (Lakoff, 1987).

Often, binary classification tasks are applied to investigate congruency effects between emotional stimuli and colors. This kind of tasks (e.g., Experiment 2 in Fetterman et al., 2012; Winskel et al., 2021) consists of a to-be-classified, task-relevant dimension (e.g., emotion) and a task-irrelevant dimension (e.g., color). In their second experiment, Fetterman et al. (2012), for example, presented anger- and sadness-related target words either in red or blue font color and found that anger words presented in red font color were categorized faster than anger words presented in blue font color, and sadness words presented in blue font color were categorized faster than sadness words presented in red font color. These kinds of congruency effects, which are not restricted to color-emotion associations but also occur for other conceptual-perceptual overlaps such as, for example, morality and vertical space, are often interpreted as Stroop-like response interference due to implicit, automatic associations between conceptual (e.g., emotion) and perceptual (e.g., color) stimuli, resulting from a metaphoric understanding of the world (e.g., Fetterman et al., 2012; Meier & Robinson, 2004; Meier et al., 2004; Moller et al., 2009; Wu et al., 2018).

According to metaphor representation theory (Lakoff & Johnson, 1999), conceptual thought is nearly always built on physical metaphor, because people compare their feelings to concrete perceptual experiences to better understand them. Therefore, fixed and automatic associations are built between, for example, emotions or emotional valence and metaphoric content such as the color red. This then results in facilitated processing of target alternatives which are, according to the underlying metaphor, conceptually matched to the corresponding task-irrelevant dimension alternative (e.g., anger and red). Importantly, a conceptual, that is, meaningful overlap between both alternatives is necessary to facilitate processing.

Other researchers argue that these congruency effects, examined by binary classification tasks, emerge from polarity correspondence (Kawai et al., 2020; Lakens, 2012; Lakens et al., 2012; Schietecat et al., 2018a, 2018b). The polarity correspondence principle (Proctor & Cho, 2006) is based on polarity differences, that is, asymmetries in the processing of dimensions with polar oppositions (Batistella, 1990; Greenberg, 1963; Lakoff, 1987). Asymmetries arise from the dimension's usage in language (but are not restricted to language) such that the category which names the dimension is the default mode, for instance, happy (happiness) or moral (morality) and is therefore assigned to the + pole, whereas the opposing alternative is assigned to the − pole (Lakens, 2012). Plus polar alternatives are used more often in language (e.g., Batistella, 1990; Greenberg, 1963; Zajonc, 1968), and have a processing benefit compared to − polar endpoints (Clark, 1969).

Specifically, this means that, in binary classification tasks, opposing poles are assigned to bipolar concepts on a dimension (e.g., moral [+] vs. immoral [−] on the dimension of morality) such that one dimension alternative (here, moral) is the preferred alternative and is therefore advantaged in processing whereas the other alternative (here, immoral) is not advantaged in processing. Opposing poles are also assigned to the alternatives of an irrelevant task dimension (e.g., the presentation of words up [+] or down [−] in space). Many metaphors map conceptual and perceptual dimensions in a way that polarities overlap which makes interpretation of observed congruency effects difficult (Lakens, 2012). Also, in the morality example, both metaphor congruency and polarity correspondence predict the same congruency effects such that moral words presented up and immoral words presented down on a screen elicit shorter response times than moral words presented down and immoral words presented up on the screen (e.g., Lakens, 2012; Proctor & Cho, 2006). But, crucially, according to polarity correspondence principle, in binary classification tasks, no conceptual overlap between the two dimensions is necessary to elicit congruency effects. A structural overlap, that is, both the task-relevant (e.g., moral) as well as the task-irrelevant (e.g., up) dimension alternatives are assigned to the same pole (e.g., +), is sufficient to lead to facilitated responses (Proctor & Cho, 2006).

In his meta-analysis of metaphor congruency effects, Lakens (2012) argues that often authors explain effects as being due to metaphor congruency though the response time pattern does not match the predicted crossover interference pattern described above. Rather, response time results often show an asymmetrical pattern which in fact corresponds to the summed response time benefit pattern resulting from the influences of polarity differences in the conceptual, perceptual, and response dimensions, and to polarity correspondence. Sticking to the above example, such a pattern predicts fastest responses for moral (+ polar) words presented up (+ polar), followed by moral words presented down (− polar), and equally slow categorization times for immoral (− polar) words presented either up or down (see Lakens, 2012, for a detailed description). From the polarity correspondence principle follows that conceptual-perceptual associations can be reversed and are therefore not implicit and automatic but rather context-dependent (e.g., Lakens, 2012).

In our study, we refer to the first experiment by Fetterman et al. (2012) who presented anger, fear, and neutral words in red or gray font color and found that anger-related words presented in red font color elicited facilitated responses as opposed to anger-related words presented in gray font color whereas fear-related words were categorized slightly faster when presented in gray font color compared to red font color. The authors interpret these results in favor of automatic metaphor congruency between conceptual and perceptual stimuli and argue that the differences in categorization speed of fear words presented either in red or gray font color reinforce the specificity of red-anger associations.

Critically, explicit rating studies, using more than one negative concept, for instance, anger- and fear-related terms, superficially seem to be immune against polarity correspondence because + and − poles cannot be assigned to the alternatives in an obvious manner as it is the case for moral and unmoral terms, for instance. Furthermore, when three concepts (e.g., anger, fear, and neutral words) are used, like in the study by Fetterman et al. (2012), there is no longer only one clear mapping of + and − poles to the dichotomous alternatives possible. Hence, this kind of studies seem to demonstrate the metaphoric origin of a congruency effect beyond doubt. Generally, we would agree. However, we would like to point out that there are important exceptions to this rule. Take the example of the terms above. Certainly, anger and fear are both negative and differ in their fundamental emotional concepts. Yet, it is conceivable that the corresponding terms still differ on a single dimension of negativity, for example, if fear is simply perceived as more negative than anger (or anger is perceived as more negative than fear), whereby the neutral words in Fetterman et al.'s (2012) study might have constituted the midpoint on that single dimension. Such judgments might vary inter-individually but, importantly, explicit ratings of two negative concepts would still be open to influences of polarity correspondence. Negativity is but one dimension that could be used for + and − pole assignments. Similarly, polar assignments might also be based on conceptual dimensions such as a time-related perspective (toward the future vs. the past), for instance (see Arnold, 1960), where anger would occupy the "past" pole whereas fear would occupy the "future" pole. If effects in Fetterman et al.'s (2012) study were based on polarity correspondence, the underlying processing dimension would probably be time-relation rather than negativity as the former is more likely to produce consistent polar assignments. Furthermore, despite the lack of metaphoric associations between fear and the color gray, Fetterman et al. (2012) found a slight processing benefit for fear targets presented in gray font color as opposed to fear targets presented in red font color and conclude that this effect supports the robustness of implicit red-anger associations. Instead, one might argue that this pattern rather corresponds to the polarity correspondence account such that the − poles were jointly assigned to the remaining categories (fear and gray) and their correspondence led to advantaged processing.

In our study, we aimed at disentangling the two possible mechanisms underlying congruency effects between anger and the color red: metaphor congruency and polarity correspondence. For that purpose, we implemented a between-within design. In the experimental condition (AFS condition), we employed three to-be-classified "discrete" emotion concepts, that is, anger, fear, and sadness words, that offered coding along different dimensions (from past to future or from high to low activation; see Arnold, 1960) and could not be logically broken down into only a single bipolar category. We explain this in detail below. In the control condition (AFN condition) we presented three one-dimensional concepts, that is, anger, fear, and neutral words, all located along a single dimension of time-relatedness (from past to future) as in Fetterman et al.'s (2012) study. Stimuli were presented in red or green font color. We assume that a bipolar mapping between the three discrete negative emotion categories (anger, fear, sadness) in the AFS condition is less effective than in the AFN condition containing neutral terms (anger, fear, neutral). If this is true and if polarity correspondence accounts for the effects, we should see a stronger or even selective red-anger effect in the AFN condition (due to stronger polarity correspondence effects) than in the AFS condition (due to a less effective mapping on a single dichotomous dimension).

Furthermore, in a supplementary analysis, we investigated if participants in the AFS condition classified targets according to a preferred dimension. As stated above, we suggest that there are several possible dimensions by which targets (anger, fear, and sadness words) can be classified: either time-relation dimension (past vs. future, whereby anger and sadness occupy the "past" pole and fear occupies the "future" pole) or activation dimension (high vs. low activation, whereby anger and fear occupy the "high activation" pole whereas sadness occupies the "low activation" pole). If targets are processed according to one of these dimensions a match between conceptual and perceptual dimensions (emotion category and color) and therefore a response facilitation would not only occur for anger words but also for the other emotional concept with which anger shares the pole, that is, sadness in case of time-relation dimension and fear in case of activation-dimension. We investigated this issue by analyzing the no-response patterns (see below for a more detailed explanation).

We implemented a trial-by-trial yes-no decision task. In each trial, a category word, such as "fear", was presented to the participants followed by a target word, such as "panic", and participants decided if the word belonged to the category or not by pressing one key for yes and another key for no. The category and target words changed from trial to trial and category and target were from the same category in one third of the trials and from a different category in two thirds of the trials. With offering dichotomous choice alternatives, this corresponds to standard procedures investigating polarity correspondence. Furthermore, this procedure allows to investigate underlying dimensions on which targets were classified by analyzing no-response patterns. As mentioned above, it is possible that participants in the AFS condition classify targets along the activation dimension or along the time-relation dimension. In the first case, we hypothesized participants to focus on the activation elicited by the emotion (i.e., high activation in cases of anger as well as fear vs. lower activation in cases of sadness). In this case, fear (anger) targets would be harder to reject (reflected in higher response times/lower hit rates) as belonging to the anger (fear) category than to the sadness category, for example, since fear and anger both share the “high activation” pole and sadness occupies the “low activation” pole. In the second case, we hypothesized, participants would focus on the time relation (i.e., past in the case of anger and sadness vs. future in the case of fear). In this case, sadness (anger) targets would be harder to reject as belonging to the anger (sadness) category than to the fear category, for example, since sadness and anger both share the “past” pole.

2. Method

2.1. Participants

We calculated the required sample size based on a conventional effect size of f = .25, a power of .95 and assumed α of .05, using G*Power (Faul et al., 2009). This resulted in a required total sample size of 45 participants. To counteract participant loss, we aimed at testing 85 participants. We chose a true random assignment to the two groups which unfortunately resulted in quite unequally distributed group sample sizes (49 participants in the AFS group, and 29 participants in the AFN group, after exclusions). Therefore, we recruited another 25 participants for the AFN group. Altogether, we tested a sample size of 110 (46 female) German or Austrian nationals (German as native language). Participants were tested online via Prolific. Mean age was 30.8 years (*SD* = 9.7 years). All participants had intact red-green vision. In exchange for valid participation, they received GBP 2.70. Informed consent was obtained from all participants.

2.2. Materials

Forty words (10 words each for anger, fear, sadness, and neutral categories, see Table 1) were taken from the Berlin Affective Word List Reloaded (BAWL-R; Võ et al., 2009). The numbers of nouns, adjectives, and verbs, as well as imageability, letter count, and word frequency (taken from the SUBTLEX-DE; Brysbaert et al., 2011) were kept constant across categories (Table 2). A list with all stimuli is also available on https://osf.io/g8d2f/. Target words were presented in upper case letters and either in red or green font color against white background. Instructions and response feedback were presented in black color against white background. As category words we selected “Wut” (anger), “Angst” (fear), “Kummer” (sadness), and “neutral” (neutral). These were also presented in upper case letters. The font used for all stimuli was Arial (bold) and font size was 4 vw (4% of screen width). Category words were framed. Target word colors were taken from Wilms and Oberfeld (2018) with similarly high saturation/lightness: red (L*a*b* [50, 81.29, 82.05]; HEX #F60000; RGB [245.63, 0, 0]) and green (L*a*b* [50.03, -98.59, 55.35]; HEX #009500; RGB [0, 148.83, 0]). Category words were presented in gray (L*a*b* [50, 0, 0]; HEX #777777; RGB [119, 119, 119]). Since the experiment was run online, we did not have control over the participants' monitor settings, but we asked participants to set their monitor brightness to the highest level and included a color test before the experiment (see procedure).

Table 1. Target words used in the experiment.

Anger	**Fear**	**Sadness**	**Neutral**	**Word class**
Aggression	Alptraum	Beileid	Fräulein	N
Ärgernis	Flucht	Tragik	Gigant	N
Erzfeind	Furcht	Trauer	Presse	N
Fehde	Horror	Träne	Suche	N
Streit	Panik	Wehmut	Tempo	N
Rache	Schreck	Witwer	Zeugin	N
sauer	gruselig	einsam	anonym	A
zornig	schaurig	trostlos	lenkbar	A
hassen	gefährden	bedauern	streichen	V
toben	spuken	weinen	üben	V

Note. We used 10 words each for anger, fear, sadness, and neutral categories. We chose six nouns (N), two adjectives (A), and two verbs (V). Words were taken from the Berlin Affective Word List Reloaded (BAWL-R) database.

Table 2. Number of nouns, adjectives, and verbs for anger, fear, sadness, and neutral target word categories, and mean values for valence, arousal, imageability and letters taken from the Berlin Affective Wordlist Reloaded (BAWL-R) database, and Log Word Frequencies (Subtlex Lg) taken from the SUBTLEX database.

	Valence value	Arousal value	Image value	Letters	Frequency	Nouns	Adjectives	Verbs
Anger	-1.78	3.91	3.86	6.4	2.15	6	2	2
Fear	-1.75	4.05	4.11	6.9	2.21	6	2	2
Sadness	-1.45	3.34	3.91	6.4	1.99	6	2	2
Neutral	0.00	3.01	3.54	6.2	2.21	6	2	2

2.3. Procedure

The experiment was conducted online. Participants were randomly assigned to either the AFS group (anger, fear, and sadness words) or the AFN group (anger, fear, and neutral words). After presenting informed consent and asking for demographic information, three Ishihara color plates (digital versions) were presented. If participants passed this test, intact red-green color vision could be assumed. The test also served as a monitor setting control (participants could distinguish red from green on their monitor). If participants did not pass the test, the experiment was aborted, and they were banned from further participation. Next, participants received instructions on the experimental task: In each trial, they were presented with a category word (anger [“Wut”], fear [“Angst”], or either sadness [“Kummer”] in the AFS group or neutral [“Neutral”] in the AFN group) before the target word was presented. They had to decide whether the target word belonged to the presented category or not (pressing E for yes and I for no or reversed). They were instructed to respond as fast and accurately as possible. All experimental target-category combinations were presented to the participants before the experiment started, thereby giving them the opportunity to memorize the targets’ assignments to the categories and minimizing errors due to subjective differences in category assignment between experimenters and participants. Before the main experiment started, participants completed at least nine practice trials with another set of target words (“Feind”, ”morden”, “brutal”, “Schauder”, “panisch”, “fürchten”, and either “Waise”, “trauern”, “verweint” in the AFS condition or “Kurve”, “salzig”, “plaudern” in the AFN condition, paired with a random category word). If participants gave an incorrect or no answer more than four times, the practice round had to be repeated (with new prime-target-pairs). After a maximum of three practice rounds the main experiment started, independently of the last round’s results. Each trial sequence consisted of a fixation cross (500 ms), followed by a blank screen (between 300 and 400 ms, randomly), a category word (500 ms), another blank screen (“ISI”, 400 ms), and the target word. If response time exceeded 2 s, participants were encouraged to respond faster (“Zu langsam!” [engl.: “Too slow!”]). If the given response was incorrect, feedback (“Falsch!” [eng.: “Incorrect!”] was given and the target remained on the screen until the response had been corrected. In both cases, feedback was presented for 500 ms. A schematic visualization of a trial is illustrated in Figure 1. The experiment was split into two experimental blocks with a break of self-determined length in between. In each block, each target word was presented six times (paired twice with each of the three category words) in random order, but always in the same color. If the target word was presented in red color in the first block, then it was be presented in green color in the second block (and vice versa). In total, 180 trials (10 target words per category × 3 categories × 3 category words × 2 presentations) were presented per block. Each block included the same amount of red and green stimuli.

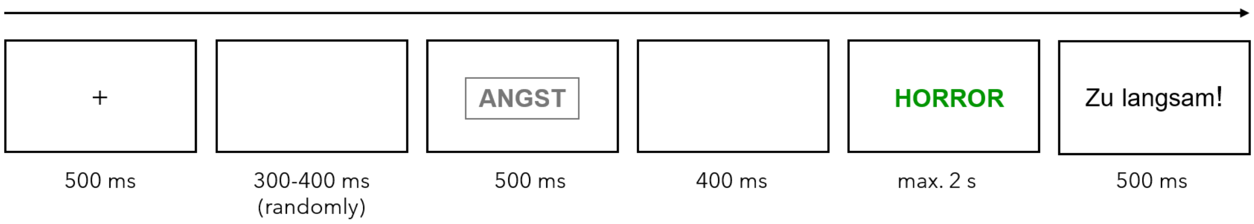

Figure 1. Example of a congruent trial that consists of a fixation cross, a blank screen, the category word, another blank screen, the target word, and (in case of false or out-of-time responses) feedback. The arrow represents the course of time. Stimulus durations are depicted below each screen.

3. Results

Nine participants were excluded from data analysis as their mean error rates (ERs) lay above 25%. We also excluded all responses with response times (RTs) out of the range of 100-2,000 ms. Furthermore, we excluded one outlier target word (“gefährden”) as it yielded an ER of 33%.

3.1. Main analysis

We calculated two 2×2×2 mixed-variable analyses of variance (ANOVAs), one for the RTs of correct yes-responses and one for the hit rates (HRs), using the within-participant variables target category (anger vs. fear), color (red vs. green), and the between-participants variable dimensionality (anger-fear-sadness [AFS] vs. anger-fear-neutral [AFN]). The main analysis had to be limited to yes responses because for no responses it is unclear which stimulus (category word or target word) is processed during response decision. We did not include responses to sadness or neutral words into analysis.

RTs are depicted in Figure 2. Results of the ANOVA of RTs revealed a main effect of target category, $F(1, 99) = 23.39$, $p < .001$, $\eta_p^2 = .191$, due to shorter RTs for anger targets ($M = 812$ ms, $SD = 17$ ms) compared to fear targets ($M = 856$ ms, $SD = 16$ ms). Also, a two-way interaction between target category and color was found, $F(1, 99) = 26.32$, $p < .001$, $\eta_p^2 = .210$, for which planned post-hoc *t*-tests revealed shorter RTs for red anger targets ($M = 801$ ms, $SD = 18$ ms) compared to green anger targets ($M = 822$ ms, $SD = 17$ ms; $t = -3.08$, $p = .003$), shorter RTs for green fear targets ($M = 841$ ms, $SD = 17$ ms) compared to red fear targets ($M = 870$ ms, $SD = 17$ ms; $t = 3.23$, $p = .001$), and shorter RTs for red anger targets compared to red fear targets ($t = -5.96$, $p < .001$). Furthermore, we found a two-way interaction between target category and dimensionality, $F(1, 99) = 16.94$, $p < .001$, $\eta_p^2 = .146$, for which planned post-hoc *t*-tests revealed longer RTs for fear targets in the AFS condition ($M = 893$ ms, $SD = 23$ ms) compared to the AFN condition ($M = 819$ ms, $SD = 23$ ms; $t = 2.28$, $p = .025$). The three-way interaction between target category, color, and dimensionality was not significant, $F(1, 99) = 2.44$, $p = .121$, $\eta_p^2 = .024$.

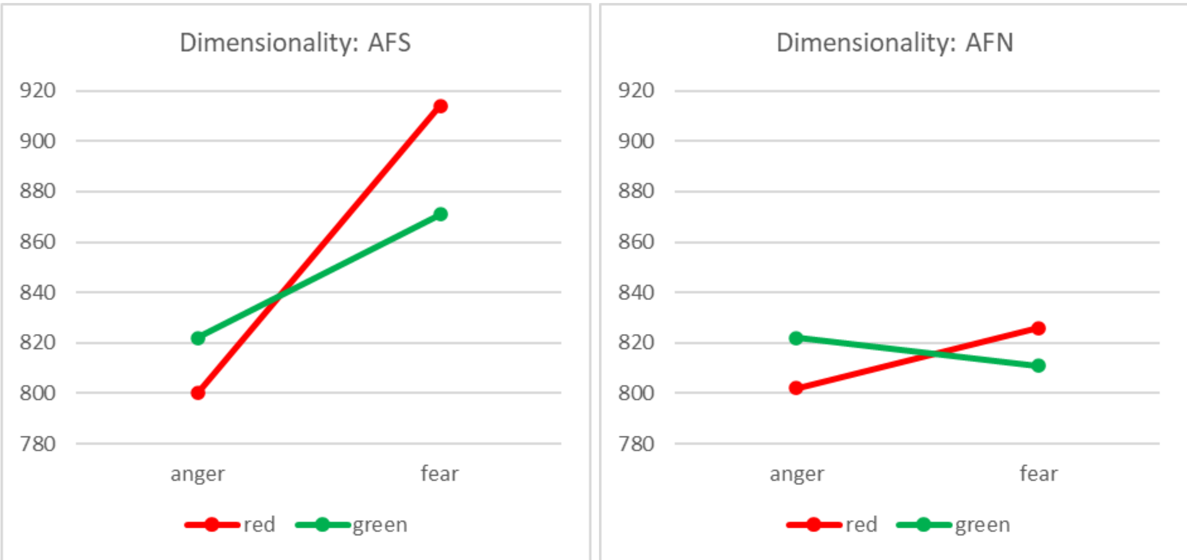

Figure 2. Response times (RTs, on the y-axis) are depicted as a function of target category (anger vs. fear, on the x-axis) and color (red vs. green, separate lines), separately for both dimensionalities (anger-fear-sadness [AFS, left] vs. anger-fear-neutral [AFN, right]). Note that for both conditions (dimensionalities) red anger targets are processed faster than green anger targets and green fear targets are processed faster than red fear targets. Also, fear targets are processed slower in the AFS condition compared to the AFN condition.

HRs are depicted in Figure 3. Results of the ANOVA of HRs revealed a two-way interaction between target category and color, $F(1, 99) = 6.16$, $p = .015$, $\eta_p^2 = .059$, for which planned post-hoc *t*-tests marginally failed to reach significance. Still, trends were quite similar to that of RTs, such that there were higher HRs for red anger targets ($M = .900$, $SD = .010$) compared to green anger targets ($M = .882$, $SD = .011$; $t = 1.64$, $p = .101$), and higher HRs for green fear targets ($M = .890$, $SD = .010$) compared to red fear targets ($M = .869$, $SD = .012$, $t = 1.91$, $p = .058$), whereby the latter occurred only in the AFS condition. The lack of significance of the post-hoc tests despite significant ANOVA interactions might be due to conservativeness of the Bonferroni correction and resulting power loss. Furthermore, we found a two-way interaction between target category and dimensionality $F(1, 99) = 13.75$, $p < .001$, $\eta_p^2 = .122$. Planned post-hoc t-tests revealed lower HRs for fear targets in the AFS condition ($M = .838$, $SD = .013$) compared to the AFN condition ($M = .921$, $SD = .013$; $t = 4.37$, $p < .001$). The three-way interaction between target category, color, and dimensionality was not significant, $F(1, 99) = 0.97$, $p = .327$, $\eta_p^2 = .010$.

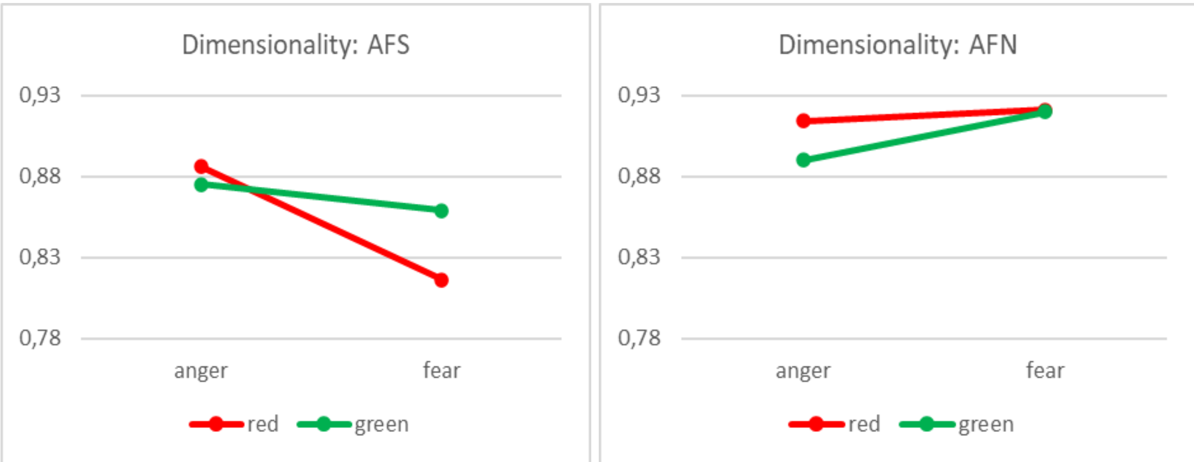

Figure 3. Hit rates (HRs, on the y-axis) are depicted as a function of target category (anger vs. fear, on the x-axis) and color (red vs. green, separate lines), separately for both dimensionalities (anger-fear-sadness [AFS, left] vs. anger-fear-neutral [AFN, right]). Note that for both conditions (dimensionalities), HRs for red anger targets are slightly higher than those for green anger targets but HRs for green fear targets are higher than those for red fear targets only in the AFS condition. Also, fear targets yielded lower HRs than anger targets only in the AFS condition.

Importantly, homoscedasticity and normality as requirements for an ANOVA were not given in the HR data. ANOVAs are robust against violations (e.g., Wilcox, 2012), yet we fitted linear mixed-effects models (LMMs) for the HR as well as for the RT data to confirm the results of the ANOVAs. As fixed factors, we entered target category (anger vs. fear), color (red vs. green), and dimensionality (AFS vs. AFN) into the models. We fitted the models with participants, target words, and a participants × target words interaction as random factors, such that we gradually added the intercepts and slopes of the random factors to the fixed factors model and compared the model fit of the different models. Usually, as it has the least assumptions on the data the unstructured covariance matrix is used if the underlying covariance matrix is unclear. Still, this matrix has many parameters to fit and often leads to convergence problems for small sample sizes or multiple repetitions (Gosho et al., 2021). This was the case here wherefore we applied the most feasible procedure for choosing the covariance matrix, that is, fitting the models with different covariance structures and choosing the one yielding the best information criteria (Kincaid, 2005). The resulting covariance structure was compound symmetry which is the one assumed for ANOVAs.

The best fitting LMM for RTs included the random intercepts of participants and target words. Similar to the ANOVA, results revealed a main effect of target category, $F(1, 18) = 4.59$, $p = .046$, a two-way interaction between target category and color, $F(1, 3599) = 20.11$, $p < .001$, and a two-way interaction between target category and dimensionality, $F(1, 3599) = 40.31$, $p < .001$, with directions of effects similar to the ANOVA.

The best fitting LMM for HRs included the random intercepts of participants, target words, and the participants × target words interaction. Results revealed a main effect of dimensionality, $F(1, 101) = 10.38$, $p = .002$, due to lower HRs in the AFS condition ($M = .859$, $SE = .016$) compared to the AFN condition ($M = .912$, $SE = .016$), a two-way interaction between target category and color, $F(1, 1919) = 6.79$, $p = .009$, and a two-way interaction between target category and dimensionality, $F(1, 1800) = 17.70$, $p < .001$, with directions of the interaction effects similar to the ANOVA.

The LMM results for RTs fully correspond to the results of the RT ANOVA, and LMM results for HRs partly correspond to the results of the HR ANOVA, as there was an additional main effect of dimensionality in the LMM only. Still, since the main issues of interest were the occurrence of the target category × color interaction and the lack of a three-way interaction, which was similar in the LMM and ANOVA results, we conclude that the ANOVA of RTs as well as that of the HRs can be interpreted at face value.

3.2. Supplementary analysis

To clarify which dimension (time or activation) was salient for target processing in the AFS condition, we split participants into two sub-groups: those who showed a numerical red-anger congruence effect (RACE group, RACE > 0) and those who did not (No-RACE group, RACE ≤ 0). We calculated this by subtracting correct mean RTs for red anger targets from correct mean RTs for green anger targets. In a second step, we calculated paired-sample *t*-tests over all no-responses in the RACE group. In case of the activation dimension, we hypothesized that, for example, fear targets would be harder to reject (reflected in higher RT/lower HR) as belonging to the anger category than to the sadness category, and anger targets would be harder to reject as belonging to the fear category than to the sadness category, since fear and anger both share the “high activation” pole and sadness occupies the “low activation” pole (cf. Arnold, 1960). In case of the time dimension, we hypothesized that, for example, sadness targets would be harder to reject as belonging to the anger category than to the fear category and anger targets would be harder to reject as belonging to the sadness category than to the fear category, since sadness and anger both share the “past” pole and fear occupies the “future” pole (cf. Arnold, 1960). See Table 3 for all hypotheses concerning the RT relations based on the two dimensions.

Table 3. Supplementary hypotheses concerning response time (RT) relations of category target combinations under two possible processing dimensions (time-relation dimension vs. activation activation).

Time-relation dimension					Activation dimension				
Category	Target	RT Relation	Category	Target	Category	Target	RT Relation	Category	Target
Fear	Anger	<	Sadness	Anger	Fear	Anger	>	Sadness	Anger
Anger	Fear	=	Sadness	Fear	Anger	Fear	>	Sadness	Fear
Anger	Sadness	>	Fear	Sadness	Anger	Sadness	=	Fear	Sadness
Anger	Fear	<	Anger	Sadness	Anger	Fear	>	Anger	Sadness
Fear	Anger	=	Fear	Sadness	Fear	Anger	>	Fear	Sadness
Sadness	Fear	<	Sadness	Anger	Sadness	Fear	=	Sadness	Anger

Note. Mathematical operators (</=/>) refer to lower/same/higher RTs, respectively.

Thirty-four participants in the AFS condition showed a numerical red-anger congruence effect and were therefore included in the RACE group analysis. Results of the no-responses showed that in the RACE group there was no difference between fear category/anger target ($M = 834$ ms, $SD = 185$ ms) and sadness category/anger target combinations ($M = 820$ ms, $SD = 192$ ms; $t = 0.96$, $p = .344$), and no difference between anger category/fear target combinations ($M = 865$ ms, $SD = 196$ ms) and sadness category/fear target combinations ($M = 872$ ms, $SD = 177$ ms; $t = -0.61$, $p = .549$). Differences between all other combination pairs were found: Anger category/sadness target combinations ($M = 829$ ms, $SD = 203$ ms) yielded lower RTs than fear category/sadness target combinations ($M = 941$ ms, $SD = 189$ ms; $t = -5.25$, $p < .001$), anger category/fear targets combinations yielded higher RTs than anger category/sadness target combinations ($t = 3.18$, $p = .003$), fear category/anger targets combinations yielded lower RTs than fear category/sadness target combinations ($t = -5.50$, $p < .001$), and sadness category/fear target combinations yielded higher RTs than sadness category/anger target combinations ($t = 3.70$, $p = .001$). This pattern does not correspond to any of the two hypothesized patterns concerning the time or activation dimensions, respectively.

4. Discussion

In the present study, we investigated whether facilitated classification of anger words presented in red font color as opposed to anger words presented in green font color is due to metaphor congruency or to polarity correspondence. One group classified anger, fear, and sadness words (AFS condition), and the other group classified anger, fear, and neutral words (AFN condition). We hypothesized that, if red-anger facilitation were due to polarity correspondence, the facilitation effect would be stronger in the group with neutral stimuli (AFN condition). In this condition the design allowed classification of stimuli only along one dimension (time-relation, from past to future), with neutral words being in the middle between past and future. Therefore, a bipolar mapping of relevant (emotion category) and irrelevant (color) stimulus categories might still be possible, leading to the assignment and correspondence of opposing poles. In contrast, in the group with sadness stimuli (AFS condition), coding along different dimensions (activation, from low to high activation, and time-relation) was possible which was assumed to weaken the possibility of each of the alternative polarity correspondences and, thus, the congruence effect, as sometimes activation and sometimes time-relation would be used as a dimension for classifications, with different results for which concepts lie at the + pole and which lie at the − pole (e.g., anger being located at the − pole of the time-relation dimension but at the + pole of the activation dimension). Thus, we hypothesized that if polarity correspondence were the underlying mechanism of the red-anger congruency effect, this effect should vanish or at least be much smaller in the AFS compared to the AFN condition. For RTs as well as for HRs, results revealed facilitated processing (reflected in lower RTs and higher HRs) of anger targets presented in red font color as opposed to anger targets presented in green font color in both conditions and no differences in facilitation between conditions. As the red-anger processing benefit was equally high in both conditions (AFS and AFN), our results highlight the robustness of the red-anger association which cannot be due to polarity correspondence effects in our design. Furthermore, processing of green fear targets was facilitated as opposed to red fear targets in both conditions regarding RTs, but regarding HRs, processing of green fear targets was only facilitated in the AFS condition. According to Laken's (2012) hypothesized summed polarity benefits, responses to green fear targets would not have been facilitated since green as well as fear represent the – poles on their dimensions, respectively, and therefore would have been processed slower than ++ or +− combinations. This means that responses to red as well as green fear targets would have been equally slow. As this was not the case and due to similarity of red-anger facilitation across conditions, our findings do not support the polarity correspondence principle. Still, it is not clear why there is such a strong green-fear association. It might arise from associations between green and apprehensions concerning the destruction of nature (cf. Lee et al., 2017). On the other hand, it might to some extent be explained by the polarity correspondence principle which assumes that if the prioritized (i.e., +) pole is occupied by one color (here, red), then the alternative color (here, green) is attributed to the remaining (i.e., −)

pole. Still, the fact that red-anger associations arose independently of context (i.e., underlying processing dimensions) supports the metaphor congruency hypothesis.

The differences in RTs to fear targets between both conditions (AFS and AFN) might reflect the difficulty to discriminate fear and sadness concepts compared to fear and neutral concepts. This assumption is also supported by the supplementary *t*-tests which show larger RTs for category word/target word combinations composed of fear and sadness words than for other combinations. Furthermore, the supplementary *t*-tests did not reveal the dominance of either activation or time-relation dimension in processing the target words which supports the strength of red-anger associations as they are not confounded by the possibility of different processing dimensions. Participants might as well have perceived targets along another, not hypothesized, dimension, that is power/dominance (or sometimes also referred to as potency-control) with power on the one pole and weakness on the other (Fontaine et al., 2007; Osgood et al., 1957). Power thereby refers to emotions like anger, hate, or contempt, and weakness refers to emotions like sadness, shame, or fear (Fontaine et al., 2007). Accordingly, in our experiment, anger would have occupied the + pole on that dimension whereas fear and sadness would have occupied the − pole. According to the dimension-specificity hypothesis (Schietecat et al., 2018a), the saliency of an underlying affective dimension is context-dependent and depends upon the relative conceptual distances between bipolar concept alternatives. This means in our experiment, participants in the AFS condition might have perceived the three emotion concepts anger, fear, and sadness on one dimension of power/dominance because anger vs. sadness and fear are most distant concepts on this dimension, whereas in the AFN condition anger and fear were most distant concepts on the dimension of time-relation. This might have resulted in difficulties in discriminating fear from sadness targets, being reflected in high response times for fear targets in the AFS condition compared to the AFN condition. To address this point, we hypothesized the RT relations of category target combinations according to this dimension (see section 9.1 in the Appendix) and compared them to the supplementary *t*-test results presented in Section 3.2. Four of the six *t*-tests correspond to the expectations. Still, although results rather correspond to the power/dominance dimension than to either time-relation or activation dimensions, they do not fully reflect the assumptions under that dimension wherefore further research is necessary to clarify that point.

One limitation of our study is the choice of covariance matrix for the linear mixed-effects model analysis. As we had convergence problems with the unstructured covariance matrix we chose the most feasible solution, that is, choosing the matrix producing the best information criteria. Still, this might not reflect the true underlying structure (Keselman et al., 1998). Gosho et al. (2021) suggest the use of a bias-corrected covariance estimator which might be implemented in further analyses. Here, we stick to the above-described analysis as the linear mixed-effects model is applied only to verify results of the ANOVAs and not as a main analysis.

5. Conclusion

Our findings support the metaphor congruency hypothesis as the red-anger facilitation was equally strong in both conditions. Also, a green-fear facilitation was found in both conditions. Large differences between conditions were revealed in processing fear targets such that the condition with anger, fear, and sadness words produced larger response times compared to the conditions with anger, fear, and neutral words which might be due to difficulties in discriminating sadness and fear emotion categories.

6. References

Arnold, M. B. (1960). *Emotion and personality.* Columbia University Press.

Batistella, E. L. (1990). *Markedness: The evaluative superstructure of language*. State University of New York Press.

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, *58*(5), 412–424. https://doi.org/10.1027/1618-3169/a000123

Clark, H. H. (1969). Linguistic processes in deductive reasoning. *Psychological Review, 76(*4), 387–404. https://doi.org/10.1037/h0027578

Drummond, P. D. (1997). Correlates of facial flushing and pallor in anger-provoking situations. *Personality and Individual Differences, 23*(4), 575–582. https://doi.org/10.1016/S0191-8869(97)00077-9

Elliot, A. J. (2015). Color and psychological functioning: A review of theoretical and empirical work. *Frontiers in Psychology, 6*, Article 368. https://doi.org/10.3389/fpsyg.2015.00368

Elliot, A. J., & Maier, M. A. (2012). Color-in-Context Theory. *Advances in Experimental Social Psychology, 45*, 61–125. https://doi.org/10.1016/B978-0-12-394286-9.00002-0

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using g*power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41(*4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fetterman, A. K., Robinson, M. D., Gordon, R. D., & Elliot, A. J. (2011). Anger as seeing red: Perceptual sources of evidence. *Social Psychological and Personality Science, 2*(3), 311–316. https://doi.org/10.1177/1948550610390051

Fetterman, A. K., Robinson, M. D., & Meier, B. P. (2012). Anger as "seeing red": Evidence for a perceptual association. *Cognition & Emotion*, *26*(8), 1445–1458. https://doi.org/10.1080/02699931.2012.673477

Fontaine, J. R., Scherer, K. R., Roesch, E. B., Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological Science, 18*(12), 1050-1057. https://doi.org/10.1111/j.1467-9280.2007.02024.x

Fugate, J. M. B., & Franco, C. L. (2019). What color is your anger? Assessing color-emotion pairings in English speakers. *Frontiers in Psychology, 10*, Article 206. https://doi.org/10.3389/fpsyg.2019.00206

Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding.* Cambridge Univ. Press.

Gosho, M., Noma, H., & Maruo, K. (2021). Practical review and comparison of modified covariance estimators for linear mixed models in small-sample longitudinal studies with missing data. *International Statistical Review*. Advance online publication. https://doi.org/10.1111/insr.12447

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of grammar* (pp. 73–113). MIT Press.

Jonauskaite, D., Abu-Akel, A., Dael, N., Oberfeld, D., Abdel-Khalek, A. M., Al-Rasheed, A. S., Antonietti, J.-P., Bogushevskaya, V., Chamseddine, A., Chkonia, E., Corona, V., Fonseca-Pedrero, E., Griber, Y. A., Grimshaw, G., Hasan, A. A., Havelka, J., Hirnstein, M., Karlsson, B. S. A., Laurent, E., … Mohr, C. (2020). Universal patterns in color-emotion associations are further shaped by linguistic and geographic proximity. *Psychological Science, 31*(10), 1245-1260. https://doi.org/10.1177/09567976209488

Jonauskaite, D., Wicker, J., Mohr, C., Dael, N., Havelka, J., Papadatou-Pastou, M., Zhang, M., & Oberfeld, D. (2019). A machine learning approach to quantify the specificity of colour-emotion associations and their cultural differences. *Royal Society Open Science, 6*(9), 190741. https://doi.org/10.1098/rsos.190741

Kawai, C., Lukács, G., & Ansorge, U. (2020). Polarities influence implicit associations between colour and emotion. *Acta Psychologica*, *209*, Article 103143. https://doi.org/10.1016/j.actpsy.2020.103143

Kaya, N., & Epps, H. H. (2004). Relationship between color and emotion: A study of college students. *College Student Journal, 38*(3), 396–405.

Keselman, H. J., Algina, J., Kowalchuk, R. K., & Wolfinger, R. D. (1998). A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements. *Communications in Statistics: Simulation and Computation, 27*(3), 591-604. https://doi.org/10.1080/03610919808813497

Kincaid, C. (2005). Guidelines for collecting the covariance structure in mixed model analysis. In SAS Users Group International (Eds.), *SUGI 30 proceedings: SAS Users Group International conference, April 10 - 13, 2005, Pennsylvania Convention Center, Philadelphia, Pennsylvania* (Paper 198-30). SAS Institute. http://www2.sas.com/proceedings/sugi30/198-30.pdf

Lakens, D. (2012). Polarity correspondence in metaphor congruency effects: Structural overlap predicts categorization times for bipolar concepts presented in vertical space. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 726–736. https://doi.org/10.1037/a0024955

Lakens, D., Semin, G. R., & Foroni, F. (2012). But for the bad, there would not be good: Grounding valence in brightness through shared relational structures. *Journal of Experimental Psychology: General*, *141*(3), 584–594. https://doi.org/10.1037/a0026468

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought* [Reprinted]. Basic Books.

Lee, Y. K., Chang, C. T., & Chen, P. C. (2017). What sells better in green communications: Fear or hope?: It depends on whether the issue is global or local. *Journal of Advertising Research, 57*(4), 379-396. https://doi.org/10.2501/JAR-2017-048

Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Association between affect and vertical position. *Psychological Science*, *15*(4), 243–247. https://doi.org/10.1111/j.0956-7976.2004.00659.x

Meier, B. P., Robinson, M. D., & Clore, G. L. (2004). Why good guys wear white. *Psychological Science*, *15*(2), 82–87. https://doi.org/10.1111/j.0963-7214.2004.01502002.x

Moller, A. C., Elliot, A. J., & Maier, M. A. (2009). Basic hue-meaning associations. *Emotion*, *9*(6), 898–902. https://doi.org/10.1037/a0017811

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. University of Illinois Press.

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin*, *132*(3), 416–442. https://doi.org/10.1037/0033-2909.132.3.416

Peromaa, T., & Olkkonen, M. (2019). Red color facilitates the detection of facial anger - but how much? *PloS One, 14*(4), Article e0215610. https://doi.org/10.1371/journal.pone.0215610

Pravossoudovitch, K., Cury, F., Young, S. G., & Elliot, A. J. (2014). Is red the colour of danger? Testing an implicit red-danger association. *Ergonomics, 57*(4), 503–510. https://doi.org/10.1080/00140139.2014.889220

Schietecat, A. C., Lakens, D., IJsselsteijn, W. A., & Kort, Y. A. W. de (2018a). Predicting context-dependent cross-modal associations with dimension-specific polarity attributions. Part 1: Brightness and aggression. *Collabra: Psychology*, *4*(1), 14. https://doi.org/10.1525/collabra.110

Schietecat, A. C., Lakens, D., IJsselsteijn, W. A., & Kort, Y. A. W. de (2018b). Predicting context-dependent cross-modal associations with dimension-specific polarity attributions. Part 2: Red and valence. *Collabra: Psychology*, *4*(1), 21. https://doi.org/10.1525/collabra.126

Sutton, T. M., & Altarriba, J. (2016). Color associations to emotion and emotion-laden words: A collection of norms for stimulus construction and selection. *Behavior Research Methods, 48*(2), 686–728. https://doi.org/10.3758/s13428-015-0598-8

Võ, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, *41*(2), 534–538. https://doi.org/10.3758/BRM.41.2.534

Wilcox, R. (2012). *Introduction to robust estimation and hypothesis testing* (3rd ed.). Academic Press.

Wilms, L., & Oberfeld, D. (2018). Color and emotion: Effects of hue, saturation, and brightness. *Psychological Research*, *82*(5), 896–914. https://doi.org/10.1007/s00426-017-0880-8

Winskel, H., Forrester, D., Hong, M., & O'Connor, K. (2021). Seeing red as anger or romance: An emotion categorisation task. *Journal of Cognitive Psychology*. Advance online publication. https://doi.org/10.1080/20445911.2021.1936538

Wu, Y., Lu, J., van Dijk, E., Li, H., & Schnall, S. (2018). The color red is implicitly associated with social status in the United Kingdom and China. *Frontiers in Psychology*, *9*, Article 1902. https://doi.org/10.3389/fpsyg.2018.01902

Young, S. G., Elliot, A. J., Feltman, R., & Ambady, N. (2013). Red enhances the processing of facial expressions of anger. *Emotion, 13*(3), 380–384. https://doi.org/10.1037/a0032471

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*(2, Pt.2), 1–27. https://doi.org/10.1037/h0025848

7. List of figures

8. List of tables

9. Appendix

9.1. Supplementary hypotheses under the power/dominance dimension

Table 4. Supplementary hypotheses concerning response time (RT) relations of category target combinations under the power/dominance dimension.

Dominance/power dimension				
Category	Target	RT Relation	Category	Target
Fear	Anger	=	Sadness	Anger
Anger	Fear	<	Sadness	Fear
Anger	Sadness	<	Fear	Sadness
Anger	Fear	=	Anger	Sadness
Fear	Anger	<	Fear	Sadness
Sadness	Fear	>	Sadness	Anger

Note. Mathematical operators (</=/>) refer to lower/same/higher RTs, respectively.

9.2. Abstract

Associations between emotional concepts and the color red have been investigated multiple times and are multifaceted and often even contradictory. Mostly, congruency effects between conceptual (e.g., emotion) and perceptual (e.g., color) domains are investigated by applying binary classification tasks. Despite the robustness of associations with anger-related words and red in these tasks there is uncertainty about the underlying processing mechanisms. Some researchers explain congruency effects between conceptual and perceptual dimensions in general and anger words and red in particular by implicit and automatic metaphor associations between the dimension alternatives. Others argue that polarity correspondence might be the reason for response time benefits of matching dimension alternatives, such that + and − poles are assigned to the binary response categories in a categorization task, and

that facilitation occurs only when the assigned poles of the conceptual and perceptual dimensions match. We investigated this issue in a mixed design where participants (*N* = 101) classified anger-, fear- and sadness-related words in one group (AFS condition) and anger, fear and neutral words in the other group (AFN condition), with all words presented in red and green font color. We assumed that a bipolar mapping between the three discrete emotion categories in the AFS condition was less effective than in the AFN condition and, thus, hypothesized that, if polarity correspondence accounts for the effects, we should see a stronger red-anger effect in the AFN condition than in the AFS condition. Results revealed no difference in red-anger facilitation between the two conditions which supports the metaphor congruency account.

9.3. Zusammenfassung

Assoziationen zwischen Emotionskonzepten und der Farbe Rot wurden oft untersucht und sind vielseitig und oft sogar gegensätzlich. Meistens werden Kongruenzeffekte zwischen konzeptuellen (z.B. Emotions-) und perzeptuellen (z.B. Farb-)Dimensionen anhand von binären Klassifizierungsaufgaben untersucht. Trotz der in dieser Art von Aufgaben auftretenden Robustheit der Verknüpfung von Wut-assoziierten Wörtern und Rot herrscht Unsicherheit über die zugrundeliegenden Verarbeitungsmechanismen. Manche Forscher*innen erklären Kongruenzeffekte zwischen konzeptuellen und perzeptuellen Dimensionen im Allgemeinen und Wut-assoziierten Wörtern und Rot im Speziellen durch automatisierte metaphorische Assoziationen zwischen den Dimensionsalternativen. Andere argumentieren, dass Polaritätskorrespondenz die Ursache für Verarbeitungsvorteile von übereinstimmenden Dimensionsalternativen sei. In diesem Fall werden den binären Antwortalternativen in einer Kategorisierungsaufgabe + und − Pole zugeordnet, wobei ein Verarbeitungsvorteil nur auftritt, wenn die zugeordneten Pole der konzeptuellen und perzeptuellen Dimensionen übereinstimmen. Dieser Sachverhalt wurde in der vorliegenden Studie in einem Inner- und Zwischensubjektdesign (*N* = 101) untersucht. Versuchspersonen der einen Gruppe (AFS-Bedingung) kategorisierten Wut-, Angst- und Kummer-assoziierte Wörter und Versuchspersonen der anderen Gruppe (AFN-Bedingung) kategorisierten Wut-assoziierte, Angst-assoziierte und neutrale Wörter, wobei alle Wörter in roter und grüner Schriftfarbe dargestellt wurden. Es wurde angenommen, dass eine bipolare Zuordnung zwischen den drei diskreten Emotionskategorien in der AFS-Bedingung weniger effektiv ist als in der AFN-Bedingung und somit die Hypothese aufgestellt, dass, falls Polaritätskorrespondenz für die Effekte verantwortlich ist, in der AFN-Bedingung ein stärkerer Rot-Wut-Effekt auftreten sollte als in der AFS-Bedingung. Die Ergebnisse zeigten keinen Unterschied im Rot-Wut-Verarbeitungsvorteil zwischen den beiden Bedingungen. Dies unterstützt die Metapherkongruenzannahme.