

Social Psychology

Is "Neutral" Really Neutral? Mid-point Ratings in the Affective Norms English Words (ANEW) May Mask Ambivalence

Farid Anvari¹, Jacqueline Bachmann², Jeffrey Sanchez-Burks³, Iris K. Schneider^{1a}

¹ Technical University Dresden, Dresden, Germany, ² University of Cologne, Cologne, Germany, ³ University of Michigan, Ann Arbor, MI, USA

Keywords: mixed feelings, ambivalence, ANEW, neutrality, indifference

<https://doi.org/10.1525/collabra.82204>

Collabra: Psychology

Vol. 9, Issue 1, 2023

The Affective Norms for English Words (ANEW) is a stimulus set that provides researchers with English language words that have been pre-rated on bipolar scales for valence, dominance, and arousal. Researchers rely on these pre-ratings to ensure that the words they select accurately reflect the affective responses these words elicit. Each word has a valence rating reflecting the degree to which people experience the word as positive or negative, with midpoint ratings on this scale presumably reflecting neutrality. However, neutral words tend to vary substantially in arousal, suggesting that not all neutral words are the same. Some researchers account for this by using the bipolar valence ratings in conjunction with the arousal ratings, selecting low-arousal neutral words when neutrality is what they seek. We argue that the varying levels of arousal in neutral words is due to varying levels of ambivalence. However, the idea that midpoint valence ratings for ANEW stimuli may hide varying levels of ambivalence has not yet been examined. This article provides evidence that words in the ANEW that appear neutral actually vary markedly in the levels of ambivalence they elicit and that this is related to their levels of arousal. These findings are relevant for research, past and present, that use the ANEW because ambivalence has different psychological consequences than neutrality, and therefore complicates the ability to draw clear inferences and maintain experimental control.

The Affective Norms for English Words (ANEW) (Bradley & Lang, 1999, 2010) is a stimulus set that provides normative ratings for a number of English words. The ANEW is used to study a variety of topics, including facial expressions of affect (Larsen et al., 2003), mindfulness (Chambers et al., 2008), and picture processing (Bradley et al., 2001). The ANEW has over 3500 citations (Google Scholar, February 03, 2023). As such, a large number of psychology scholars, over a wide variety of domains, rely on the stimuli in the ANEW. Each word in ANEW has a valence rating reflecting the degree to which people experience the word as positive or negative and an arousal rating reflecting how much people experience the word as activating (e.g., stimulated) or deactivating (e.g., relaxed). Although arousal ratings are generally lower for words rated near the midpoint on valence, there is substantial variability, suggesting that not all neutral words are equal. Indeed, some researchers recognize the importance of accounting for arousal in the ANEW word stimuli that have mid-point valence ratings, such as by selecting low arousal neutral words when neutrality is what they seek (e.g., Chan & Singhal, 2013). What

might account for these varying levels of arousal? One explanation might be that the bipolar midpoint rated words mask varying levels of ambivalence (Schneider et al., 2016; Thompson et al., 1995). However, this has not been empirically examined in the ANEW stimulus set.

ANEW stimuli were normed using a bipolar valence scale ranging from positive (happy, pleased, satisfied, contented, hopeful) to negative (unhappy, annoyed, unsatisfied, melancholic, despaired, bored) and a bipolar arousal scale ranging from excited (stimulated, excited, frenzied, jittery, wide-awake, or aroused) to calm (relaxed, calm, sluggish, dull, sleepy, or unaroused) (Bradley & Lang, 1994, 1999, 2010). Treating these bipolar scales as distinct from one another implies that positive and negative feelings are opposing endpoints of the same dimension and constitute a zero-sum: when feeling positive, there is no negativity and vice versa. However, this conceptualization is at odds with daily experience and prior research that shows that people can experience positivity and negativity at the same time (Trampe et al., 2015), a state referred to as ambivalence

a Corresponding author: iris.schneider@tu-dresden.de
Zellescher Weg 17, 01069, Dresden, Germany

(Kaplan, 1972; Larsen et al., 2003; Schneider & Schwarz, 2017; Thompson et al., 1995).

The notion that people can have positive and negative reactions to the same stimulus at the same time is formalized in the Evaluative Space Model (ESM) (Cacioppo & Berntson, 1994). In this model, neutrality is characterized by a lack of activation of positivity and negativity. A perfectly neutral stimulus would elicit no positive and no negative reaction at all and thus no activation of valence (Cacioppo & Berntson, 1994; Schneider & Mattes, 2021). On the other hand, ambivalence refers to the presence of positive and negative reactions simultaneously and is characterized by activation of valence (Cacioppo & Berntson, 1994; Schneider & Schwarz, 2017; Thompson et al., 1995). Ambivalence and neutrality have starkly different psychological consequences. One important consequence of ambivalence is the experience of conflict and arousal. Specifically, when people feel ambivalent, their physiological (van Harreveld et al., 2009) and self-reported arousal (Schneider et al., 2016) increases (for an overview and exceptions, see van Harreveld et al., 2015). Neutrality, on the other hand, is not associated with conflict (Schneider & Mattes, 2021) or arousal. Thus, the variance in arousal for the midpoint rated ANEW stimuli might be indicative of varying levels of ambivalence, which would not be detectable on a bipolar scale of valence.

The bipolar valence scale used to norm the ANEW cannot by itself distinguish between neutrality and ambivalence (Kaplan, 1972; Schneider et al., 2016; Thompson et al., 1995). For instance, the word “vase” (a “neutral” word in the ANEW; Bradley & Lang, 2010) might not evoke strong positive or negative feelings. People indicate this neutral response by rating the midpoint of the scale. However, the word “hospital” (also a “neutral” word in the ANEW; Bradley & Lang, 2010) may evoke both positive *and* negative feelings, or ambivalence. How should this be expressed? Potentially, also on the scale’s midpoint, because a rating to either side of the midpoint would not do the experience justice. As a consequence, midpoint ratings on a bipolar valence scale can indicate either (i) the absence of positive and negative feelings (i.e., neutrality), or (ii) the presence of positive *and* negative feelings (i.e., ambivalence).

The potential confusion between neutrality and ambivalence is essential for researchers using the ANEW stimuli because ambivalence has consequences different from neutrality. Ambivalence can cause uncertainty (van Harreveld et al., 2009), affect information processing (Rees et al., 2013), and increase illusory pattern perception (van Harreveld et al., Keskinis, 2014), (for overviews, Conner & Sparks, 2002; van Harreveld et al., 2015). Researchers aiming to induce “neutrality” might introduce these consequences into their research.

Therefore, it is crucial that researchers are able to disentangle truly neutral words in ANEW from those that are ambivalent. Because ambivalence is associated with arousal, the neutral words with high arousal ratings are likely to be more ambivalent, or at least less likely to be truly neutral, than those with low arousal ratings. That

is, arousal ratings for the ANEW stimuli may indicate ambivalence. Yet, not all researchers take this into consideration, given that they select neutral words which vary greatly on arousal or that hardly differ on arousal from the positive/negative words that the researchers selected (e.g., Legare & Souza, 2014; Minnema & Knowlton, 2008; Sereno et al., 2015; Wadlinger & Isaacowitz, 2008). On the other hand, many researchers do recognize the importance of taking arousal into consideration when selecting neutral stimuli based on bipolar scales. Indeed, when selecting neutral stimuli from ANEW some researchers either (i) control for arousal, (ii) choose neutral stimuli that have low arousal ratings, or (iii) choose neutral stimuli with lower arousal ratings than the positive/negative stimuli (e.g., Chan & Singhal, 2013; Isaac et al., 2012; Jeon et al., 2020; Levens & Phelps, 2008; Trammell & Clore, 2014; Yang et al., 2013; Zhang et al., 2019). This suggests that some researchers are aware that words rated on the midpoint of the bipolar valence scale may not be truly neutral but instead hide some other properties of the stimuli, related to arousal, such as ambivalence. However, there has not yet been an empirical examination of the extent that neutral words in the ANEW vary in ambivalence or the extent that the variance in arousal ratings for the neutral words is related to variance in ambivalence.

There is good reason to expect that ambivalent stimuli are hidden among the mid-point rated stimuli in the ANEW and that this ambivalence is related to arousal. For example, the International Affective Picture System (IAPS; Lang et al., 1999) uses the same bipolar scale as the ANEW and also contains several stimuli that have midpoint ratings, suggesting neutrality. However, Schneider et al. (2016) had participants rate these neutral images on bipolar valence and arousal scales, unipolar positivity and negativity scales (used to calculate an index known as objective ambivalence), and a subjective ambivalence scale. They found that the neutral images actually varied in ambivalence, and that this ambivalence was positively related to arousal, but negatively with bipolar valence ratings. Moreover, the unique relationship of arousal ratings was strongest with subjective ambivalence ratings, beyond the objective ambivalence measure and the bipolar valence ratings. Therefore, midpoint valence ratings for images in the IAPS hid varying levels of ambivalence, and the arousal ratings were an indicator of this.

The present research consists of 2 studies examining whether the findings of Schneider et al. (2016) generalize to word stimuli, specifically from the ANEW database. Note that our stimuli are selected from the stimulus set we obtained directly from the Center for Emotion and Attention at the University of Florida (Bradley & Lang, 2010; the PDF can be found on the OSF link provided below in the methods for Study 1).

Study 1

We examined whether the midpoint valence ratings in the ANEW mask mixed feelings (i.e., ambivalence), expecting that midpoint rated stimuli will likely encompass both relatively neutral as well as relatively ambivalent stimuli.

Furthermore, we examined, for words with mid-point ratings on valence, whether the ratings of subjective ambivalence correlate with self-reported arousal as found by Schneider et al. (2016) for IAPS images.

We used three established ways to examine ambivalence for each word. First, we used a measure of subjective ambivalence that asks participants to self-report their experience of ambivalence (Jamieson, 1993; Priester & Petty, 1996; Schneider & Schwarz, 2017). Second, we used a measure known as objective ambivalence, calculated using the formula $((P+N)/2) - |P - N|$, where P is the positive rating, and N is the negative rating (Kaplan, 1972; Schneider et al., 2016; Thompson et al., 1995). This index ranges from -3 to 9, higher scores indicating more objective ambivalence, and 1 indicating neutrality. Finally, we used the minimum statistic (or MIN stat) as an index of ambivalence (Kreibig & Gross, 2017; Schimmack, 2001). The MIN stat equals the lower of the two ratings given for the positivity and negativity of a stimulus. A higher MIN stat reflects a greater intensity of ambivalence and mixed feelings. For example, if a stimulus has a rating of 3 for positivity and 6 for negativity, the MIN stat is 3; if a stimulus has a rating of 6 for negativity and 6 for positivity, the MIN stat is 6; if a stimulus has a rating of 1 for positivity and 6 for negativity, the MIN stat is 1.

Method

Participants rated mid-point words from the ANEW on ambivalence, as well as on the ANEW norming measures of arousal and valence. Materials, data, and analysis scripts are [here: https://osf.io/c3yvt/?view_only=ba8f2e230c574b1cb08cc75766a88f29](https://osf.io/c3yvt/?view_only=ba8f2e230c574b1cb08cc75766a88f29).

Study 1 was not preregistered and we had no a priori power analyses for this study. The study was relatively long because of the number of stimuli. We had no expectations about effect sizes in any hypothesis tests and aimed to examine how people would rate the neutral words from ANEW. With this aim, we recruited 100 participants, each of whom rated all stimuli. Sensitivity analyses revealed that 100 participants would give a two-sided paired samples *t*-test, with alpha of .05, 80% to detect an effect size of $d = 0.28$ (relevant for the comparison between the 10 most ambivalent and 10 least ambivalent words). We report all manipulations, measures, and exclusions.

Participants and Design

One-hundred-eleven participants from Amazon Mechanical Turk (MTurk) participated in a 25-35 minute survey for \$3.00. Due to a programming error, we did not collect demographic details. However, in general, MTurk samples consist of slightly more females than males (about 55% vs. 60%, respectively), with a mean age of around 32 years (Berinsky et al., 2012; Buhrmester et al., 2016; Mason & Suri, 2012). For our study, we recruited only participants with a U.S. location.

Stimuli

Following Schneider et al. (2016), we selected words close to the mid-point with an average valence rating between 4.95 to 5.05 (on a 9-point scale). This resulted in 60 words (see [Table 1](#)).

Procedure

After participants signed the informed consent, we informed them that they would be rating different words for emotional responses. The survey consisted of two parts. In the first part, we measured valence and arousal using a procedure that closely followed the original norming procedure of ANEW (Bradley & Lang, 1999). We introduced participants to the Self-Assessment Manikin (SAM) scales using the instructions from the original work by Bradley and Lang (1999). After this, they rated valence and arousal for each word, in that order. The presentation order of the words was randomized. After each word, we checked for understanding by asking participants whether they knew what the word meant ("Do you know what this word means?" - yes/no). After this, participants were offered to take a short break. In the second part of the survey, all words were presented again in random order, along with the positivity and negativity ratings and subjective ambivalence measure, in that order. Participants were offered self-paced breaks after each block of 15 words. At the end of the survey, participants indicated their motivation on one item that asked, "*Sometimes some of our participants are less motivated or more tired than usual. In such a case, we would like to know so we can exclude the answers of those participants. Should we use your answers? NOTE: This does not affect your compensation! We just want to know so that our data quality remains high! (Definitely use my answers /Probably use my answers /Probably discard my answers /Definitely discard my answers)*". Participants were then given the option to leave comments before being thanked for their participation.

Measures

In line with the original ANEW norming procedure, we measured valence and arousal with the Self-Assessment Manikin (SAM) scales (Bradley & Lang, 1994). The SAM scales are a visual depiction of a 9-point scale using images of human-like figures in various states. The valence scale is anchored by negative feelings at 1 (most left figure) and positive feelings at 9 (most right figure). Arousal is anchored with no arousal at 1 (most left figure) and high arousal at 9 (most right figure) (for details, see Bradley & Lang, 1999).

We measured subjective ambivalence by asking participants whether they felt conflict (1 = *feel no conflict at all*, 11 = *feel maximum conflict*), indecision (1 = *feel no indecision at all*, 11 = *feel maximum indecision*), and mixed reactions (1 = *completely one-sided reactions*, 11 = *completely mixed reactions*)¹ (Priester & Petty, 1996). The average Cronbach's alpha for each stimulus was .95 (range: .89-.98). We averaged these three items into a single index where higher scores indicate more subjective ambivalence.

Participants separately rated positivity and negativity for each word. Specifically, we asked participants, “Think about this word. When you think about the **positive (negative)** aspects of this word, while ignoring the negative (positive) aspects, how **positive (negative)** is your evaluation of this word?”. We used the positivity and negativity ratings to calculate objective ambivalence using the equation described earlier as well as the MIN stat.

Results

No participants indicated that we discard their answers. We removed 132 responses for words that participants indicated they did not understand. The vast majority of participants understood most words: the least understood words were “fervor,” which was not understood by eleven participants, and “rebuff,” which ten participants did not understand. With the remaining responses, for each word, we computed mean scores for valence, arousal, subjective ambivalence, objective ambivalence, and the MIN stat.

We expected that “neutral” midpoint ratings on the valence scale in the ANEW might mask underlying variation in ambivalence and that this might be indicated by arousal. [Table 1](#) displays each word and its mean (and standard deviation) on the different measures. Arousal ratings ranged from 2.16 to 6.50. The neutral words thus varied substantially in arousal. Correspondingly, the neutral words also varied substantially in ambivalence. Subjective ambivalence ranged from 2.18 to 5.09, objective ambivalence scores ranged from 0.36 to 3.17, and the MIN stat ranged from 2.26 to 4.05.

[Table 1](#) shows that some “neutral” words were clearly more ambivalent than others. For example, when comparing the mean subjective ambivalence ratings of the 10 words with the highest ratings ($M = 4.48$, $SD = 0.32$) against the 10 words with the lowest ratings ($M = 2.43$, $SD = 0.13$), the difference is considerable, $t(12.13) = 18.88$, $p < .001$, Cohen’s $d = 8.45$, 95%CI [5.49, 11.40]. We observe similar results for objective ambivalence when comparing the 10 words with the highest scores ($M = 2.34$, $SD = 0.34$) against the 10 words with the lowest scores ($M = 0.71$, $SD = 0.32$), $t(17.91) = 10.96$, $p < .001$, Cohen’s $d = 4.90$, 95%CI [3.02, 6.78]. And also when comparing the 10 words with the highest MIN stat scores ($M = 3.69$, $SD = 0.20$) against the 10 words with the lowest scores ($M = 2.49$, $SD = 0.10$), $t(13.55) = 17.03$, $p < .001$, Cohen’s $d = 7.62$, 95%CI [4.92, 10.32].

[Table 2](#) displays the means and standard deviations for each measure averaged across all of the words and the correlations between the different measures. The 3 measures of ambivalence were strongly, positively correlated. Replicating previous research (Newby-Clark et al., 2002), subjective ambivalence correlated relatively strongly with objective ambivalence and very strongly with the MIN stat, and

the MIN stat correlated very strongly with objective ambivalence.

Arousal was related to ambivalence. As expected, arousal was positively correlated with subjective ambivalence, replicating previous findings (Schneider et al., 2016), as well as with the MIN stat. However, arousal was not significantly correlated with objective ambivalence. The more ambivalence the word induced, based on the subjective ambivalence ratings and the MIN stat, but not the objective ambivalence measure, the more arousal participants experienced.

Valence was statistically significantly and negatively correlated with subjective ambivalence and the MIN stat, but not with objective ambivalence. Valence was also negatively related to arousal.

To further examine the hypothesis that arousal may be an indicator of ambivalence, we complemented regression analyses with commonality analyses (see Gustavson et al., 2018; Nimon et al., 2008). Whereas regression analyses indicate how much unique variance in the outcome is explained by each predictor, commonality analyses quantify the amount of variance in the total regression effect that is explained by the commonality (or shared variance) between the predictor variables. Therefore, commonality analyses account for predictors that are highly correlated, as is the case when different indices of the same construct are used. In the present study, the subjective ambivalence measure, the objective ambivalence measure, and the MIN stat are all strongly correlated and are designed to ostensibly tap into the same construct (i.e., ambivalence). Therefore, commonality analyses can help reveal how much variance in a regression model’s effect on the outcome is explained by the shared variance between various combinations of the different measures of ambivalence. In this way, commonality analyses can help inform us about the relative importance of the predictor variables in relation to the outcome and thus which are the most appropriate predictors to include in the regression analyses and which are redundant.

We performed commonality analyses with arousal as the outcome variable and subjective ambivalence, objective ambivalence, the MIN stat, and valence as the predictors. The commonality across all predictors had a negative coefficient ($B = -.38$, $R^2 = -.48$) indicating a suppressor effect which seemed to be due to the commonality between objective ambivalence and the MIN stat, because every commonality that included both of these measures had a negative coefficient. The commonality between subjective ambivalence, the MIN stat, and valence had the highest coefficient and contributed the highest amount of explained variance in the regression effect ($B = .46$, $R^2 = .57$). The commonality between subjective ambivalence, objective ambivalence, and valence had a substantially lower coefficient and R^2 value ($B = .26$, $R^2 = .33$). Objective ambivalence may thus

1 Due to a technical error, 43 participants were presented with a 9-point scale for these items. To match the rest of the data, we converted these scores in the following manner: New score = $1 + 1.25 \times (\text{Old score} - 1)$. This maps scores in the desired manner (1 to 1; 5 to 6; 9 to 11). Excluding participants who saw a 9-point scale did not change results. For the interested reader, we have included these analyses in the Supplemental Materials.

Table 1. Means (standard deviations) for valence, subjective ambivalence, objective ambivalence, and the MIN stat for each word. The list is sorted in ascending order by subjective ambivalence.

Word	Number	Valence	Arousal	Subjective Ambivalence	Objective Ambivalence	MIN stat
society	2231	4.87 (1.73)	4.11 (2.31)	5.09 (3.06)	2.46 (3.02)	3.90 (2.04)
hospital	215	2.78 (1.51)	5.60 (2.07)	4.90 (3.34)	2.14 (3.48)	3.87 (2.30)
intent	1730	5.19 (1.02)	3.41 (2.09)	4.76 (3.02)	3.17 (2.84)	4.05 (2.11)
trick	2385	4.16 (1.55)	4.60 (1.94)	4.41 (3.00)	1.83 (3.02)	3.49 (2.02)
wolf	2463	4.56 (1.80)	5.36 (2.21)	4.36 (2.98)	2.22 (2.94)	3.70 (1.99)
storm	1000	4.63 (2.07)	5.46 (2.33)	4.34 (2.87)	1.77 (3.01)	3.58 (2.01)
flaunt	1566	4.84 (1.37)	4.30 (1.86)	4.32 (2.94)	1.97 (2.79)	3.46 (1.94)
fervor	1552	4.86 (1.54)	5.26 (2.43)	4.29 (2.80)	1.76 (2.51)	3.34 (1.70)
apology	1093	5.21 (1.80)	4.23 (1.93)	4.20 (2.97)	1.10 (3.06)	3.12 (1.88)
rattle	346	4.68 (1.30)	4.30 (2.16)	4.17 (2.92)	2.11 (2.59)	3.44 (1.85)
descent	1415	4.63 (1.22)	3.84 (2.08)	3.97 (2.68)	1.95 (2.44)	3.31 (1.69)
picket	1965	4.33 (1.21)	3.76 (2.09)	3.95 (2.70)	2.67 (2.66)	3.69 (2.01)
frenzy	1589	3.95 (1.59)	6.50 (2.25)	3.92 (2.77)	1.02 (2.72)	3.04 (1.75)
compel	1326	4.68 (1.35)	4.38 (2.27)	3.85 (2.68)	2.14 (2.83)	3.51 (1.87)
crocodile	1372	4.03 (1.82)	5.79 (2.27)	3.83 (2.83)	1.45 (2.71)	3.26 (1.69)
ingest	1722	5.06 (1.14)	3.52 (2.00)	3.76 (2.60)	2.01 (2.45)	3.31 (1.75)
intoxicated	1732	3.86 (2.39)	5.78 (2.36)	3.75 (2.98)	0.14 (2.93)	2.61 (1.81)
hitch	1690	4.92 (0.88)	2.98 (1.88)	3.74 (2.56)	2.08 (2.46)	3.19 (1.71)
rigor	2098	4.89 (1.36)	4.17 (2.08)	3.73 (2.69)	1.71 (2.73)	3.30 (1.89)
squash	2256	4.81 (1.43)	3.47 (2.22)	3.72 (2.53)	2.07 (2.76)	3.41 (1.93)
ad	1051	3.87 (1.59)	3.51 (2.15)	3.67 (2.77)	1.26 (2.93)	3.13 (1.83)
sentry	2161	4.68 (0.97)	3.54 (2.16)	3.65 (2.66)	2.01 (2.82)	3.35 (1.98)
belly	1169	4.97 (1.37)	3.18 (2.01)	3.63 (2.62)	2.13 (2.82)	3.42 (1.93)
daze	1394	4.46 (1.45)	3.47 (2.01)	3.62 (2.64)	1.15 (2.43)	2.77 (1.56)
desert	1416	5.03 (1.79)	3.98 (2.16)	3.56 (2.68)	2.14 (3.06)	3.60 (2.00)
rebuff	2054	4.09 (1.32)	3.95 (2.10)	3.55 (2.63)	0.96 (2.57)	2.78 (1.62)
scissors	974	4.97 (1.06)	3.16 (1.97)	3.48 (2.73)	2.07 (2.86)	3.29 (2.02)
noisy	904	3.42 (1.65)	5.76 (1.96)	3.44 (2.59)	0.19 (2.43)	2.60 (1.47)
pulp	2029	4.83 (1.23)	2.89 (1.83)	3.42 (2.41)	1.57 (2.58)	3.01 (1.76)
herd	1682	5.28 (1.14)	3.17 (2.21)	3.41 (2.52)	2.13 (2.58)	3.30 (1.86)
taxi	1008	4.93 (1.14)	3.64 (2.10)	3.39 (2.13)	2.08 (2.38)	3.31 (1.64)
tire	2356	5.19 (1.05)	3.03 (2.06)	3.36 (2.61)	2.18 (2.93)	2.81 (1.90)
inhabitant	813	4.77 (1.01)	2.78 (1.84)	3.36 (2.49)	1.24 (2.74)	3.39 (2.18)
subject	2287	4.80 (0.98)	2.64 (1.90)	3.32 (2.43)	1.99 (2.80)	3.19 (1.97)
steam	2265	5.34 (0.82)	3.26 (2.25)	3.27 (2.38)	2.07 (2.70)	3.30 (2.02)
impart	1712	5.20 (1.09)	2.79 (1.75)	3.23 (2.43)	1.38 (2.81)	2.99 (1.79)
adhere	1053	5.08 (1.09)	2.83 (1.78)	3.21 (2.42)	1.40 (2.58)	2.95 (1.70)
bone	1198	4.66 (0.96)	2.98 (1.78)	3.18 (2.42)	1.66 (2.61)	3.06 (1.80)
limb	1788	4.89 (1.04)	2.76 (1.84)	2.87 (2.30)	1.51 (2.70)	2.83 (1.71)
freezer	1588	5.02 (1.00)	2.93 (2.05)	2.83 (2.20)	1.23 (2.65)	2.79 (1.80)
stove	1001	5.55 (1.04)	3.06 (2.17)	2.79 (2.32)	1.55 (2.84)	2.93 (2.00)
seat	380	5.19 (0.92)	2.45 (1.86)	2.77 (2.47)	0.93 (2.89)	2.56 (1.85)
cod	1318	5.05 (1.09)	2.89 (2.00)	2.76 (2.34)	1.38 (2.50)	2.76 (1.63)
icebox	799	5.22 (1.07)	2.93 (2.15)	2.75 (2.25)	0.69 (2.54)	2.41 (1.57)
dune	1459	5.27 (1.14)	2.71 (1.99)	2.74 (2.25)	1.31 (2.65)	2.77 (1.79)
sock	2232	5.23 (1.11)	2.34 (1.69)	2.70 (2.33)	1.01 (2.83)	2.65 (1.75)
metal	874	5.27 (1.02)	2.85 (2.02)	2.69 (2.15)	1.14 (2.45)	2.72 (1.64)
foot	757	5.07 (0.85)	2.16 (1.67)	2.64 (2.33)	1.26 (2.74)	2.75 (1.87)
ink	229	5.09 (0.72)	2.70 (2.05)	2.63 (2.13)	1.27 (2.86)	2.54 (1.66)
hydrant	564	5.22 (1.07)	2.89 (2.08)	2.63 (2.41)	0.93 (2.71)	2.81 (1.97)
cabinet	675	5.07 (0.84)	2.42 (1.76)	2.61 (2.18)	1.29 (2.96)	2.72 (1.99)

Word	Number	Valence	Arousal	Subjective Ambivalence	Objective Ambivalence	MIN stat
knee	1760	4.92 (1.03)	2.45 (2.03)	2.57 (2.14)	1.46 (2.61)	2.85 (1.76)
van	2412	5.10 (1.04)	2.90 (1.93)	2.52 (1.88)	1.41 (2.58)	2.78 (1.78)
lid	1786	5.36 (1.21)	2.90 (2.07)	2.48 (2.07)	1.41 (2.76)	2.46 (1.51)
dryer	1454	5.06 (0.65)	2.26 (1.87)	2.48 (2.06)	0.72 (2.52)	2.68 (1.77)
barrel	651	5.16 (0.87)	2.62 (2.10)	2.46 (2.10)	1.27 (2.41)	2.58 (1.65)
pail	1933	4.91 (0.92)	2.64 (1.91)	2.40 (2.09)	1.14 (2.66)	2.55 (1.73)
liter	1794	4.54 (1.46)	3.33 (2.38)	2.33 (1.91)	0.95 (2.73)	2.45 (1.68)
gallon	1611	5.15 (0.79)	2.48 (2.01)	2.28 (2.07)	1.13 (2.49)	2.44 (1.62)
vase	2414	5.43 (1.04)	2.62 (1.88)	2.18 (1.84)	0.57 (2.69)	2.26 (1.63)

Note. Objective ambivalence could range from -3.5 to 10, subjective ambivalence and the MIN stat could range from 1 to 10. The number column indicates the word reference in the ANEW dataset.

Table 2. Descriptive statistics and correlations between the variables

	Valence	Arousal	Subjective	Objective	MIN	M (SD)
Valence		-.52[-.70, -.29]***	-.45[-.65, -.19]**	-.44[-.64, -.18]**	-.47[-.66, -.22]***	5.00 (0.23)
Arousal	-.71[-.82, -.56]***		.79[.66, .88]***	.49[.24, .68]***	.70[.52, .82]***	3.24 (0.80)
Subjective	-.50[-.67, -.28]***	.72[.57, .82]***		.78[.65, .87]***	.92[.86, .95]***	3.27 (0.71)
Objective	.04[-.22, .29]	.08[-.18, .32]	.59[.40, .74]***		.94[.89, .97]***	1.60 (0.53)
MIN	-.28[-.50, -.02]*	.47[.24, .64]***	.87[.78, .92]***	.89[.82, .93]***		3.04 (0.42)
M (SD)	4.82 (0.51)	3.51 (1.04)	3.37 (0.71)	1.55 (0.58)	3.05 (0.42)	

Note. The cells below the shaded diagonal display the correlations [and 95% confidence intervals] across the stimuli when all words are included. The cells above the diagonal display the correlations after excluding the 11 words that had valence ratings greater than 0.5 from the midpoint. *N* for correlations below the shaded diagonal = 60 (words). *N* for correlations above the shaded diagonal = 49 (words). The bottom row contains the means (and standard deviations) when all words are included and the column to the far right the means (and standard deviations) after excluding the 11 words that had valence ratings greater than 0.5 from the midpoint. ****p* < .001, ***p* < .01, **p* < .05.

be considered redundant for the regression analyses because of a combination of 3 factors: (i) the suppression effect between objective ambivalence and the MIN stat, (ii) the commonality explaining the largest amount of variance included the MIN stat and not objective ambivalence, and (iii) the high correlation between the MIN stat and objective ambivalence.

Therefore, for the multiple regression analyses, we used arousal as the outcome variable and valence, subjective ambivalence, and the MIN stat as predictors. We conducted regression analyses with various combinations of these predictors. The results of these analyses are in Table 3. The full model explained 71% of the variance in arousal. The unique relationship of subjective ambivalence was positive and strong, whereas the unique relationships of valence and the MIN stat were both statistically significant and negative. Removing valence from the full model reduced the percentage of variance explained to 61%, with subjective ambivalence remaining a strong, positive, and unique predictor and the MIN stat remaining a statistically significant unique negative predictor. Adding valence thus increased the explained variance by 10%. Including only valence and subjective ambivalence as predictors reduced the explained variance from 71% (the full model) to 68%. Thus, including the MIN stat explained only an additional 3% of the variance in arousal. Including only valence and the MIN stat as predictors reduced the explained variance to 59%, with the unique relationship of the MIN stat becoming positive. Therefore, including subjective ambivalence explained an additional 12% of the variance.

Taken together, the most parsimonious model that explained the most variance in arousal included only valence and subjective ambivalence. Including the MIN stat in the model explained very little additional variance and, furthermore, produced a strange and unexpected result such that the MIN stat had a negative unique relationship with arousal. This may be due to the collinearity between the MIN stat and subjective ambivalence measures, given that they were very strongly positively correlated.

In designing this study, we included only words from the ANEW with valence ratings between 4.95 and 5.05 (5 was the mid-point). This was quite a tight margin that we decided on due to the large number of word stimuli in ANEW (Bradley & Lang, 2010). However, when we examined the valence ratings for each word in our dataset we found substantial variation such that some words were rated quite far from the midpoint of valence. To test the robustness of our results, we repeated the analyses using only words that were rated close to the midpoint on valence by the participants in our study. Given that we had far fewer stimuli than the full ANEW dataset, we used a wider margin around the midpoint (i.e., ratings between 4.5 and 5.5, within 0.5 scale points from the midpoint). For 11 of the 60 words, mean valence did not fall between 4.5 and 5.5. Thus, we repeated the analyses with these “non-neutral” words removed.

As can be seen in Table 2 (above the shaded diagonal), the descriptive statistics still showed substantial variability for subjective ambivalence ratings (*range* = 2.18-5.09), objective ambivalence (*range* = 0.57-3.17), and the MIN stat (*range* = 2.26-4.05). The correlations also remained quali-

tatively unchanged for subjective ambivalence and the MIN stat (both were positively correlated with arousal and negatively with valence). However, as can be seen above the shaded diagonal in [Table 2](#), some slight differences arose for the relationships of objective ambivalence. Objective ambivalence was significantly and positively correlated with arousal, and significantly negatively correlated with valence. Therefore, when considering only words that were rated within 0.5 of the midpoint of the valence scale, as compared to when considering all words, the correlations of objective ambivalence with arousal and valence became larger and statistically significant.

When we conducted the regression analyses using only the words that had valence ratings within 0.5 of the midpoint, the results for the MIN stat were a little different compared to when we included all words. The results from these models are in [Table 4](#). The full model explained 68% of the variance, and arousal had a unique positive relationship with subjective ambivalence, and a unique negative relationship with valence, but its unique relationship with the MIN stat was not statistically significant. Removing valence from the model reduced the explained variance to 63%, and the unique relationship of the MIN stat with arousal remained nonsignificant, whereas the unique relationship of subjective ambivalence remained strong and positive. The model that excluded the MIN stat explained 66% of the variance in arousal, thus changing very little in terms of total variance explained, with the unique relationships of valence and subjective ambivalence with arousal being statistically significant. In contrast, the model without subjective ambivalence explained 53% of the variance in arousal (i.e., decreased the explained variance of the full model by 15%) with arousal having a negative unique relationship with valence and a positive unique relationship with the MIN stat, both statistically significant.

Therefore, when we examined only the words that were rated within 0.5 of the midpoint of the valence scale, the negative coefficient for the unique relationship between the MIN stat and arousal, after controlling for subjective ambivalence, was always statistically nonsignificant. Once again, the most parsimonious model was that which included only valence and subjective ambivalence as predictors of arousal, given that adding the MIN stat explained very little additional variance.

Study 2

Study 1 showed that “neutral” words in the ANEW database are not all neutral *per se* but vary in their ambivalence. Specifically, the midpoint ratings in the ANEW ratings mask that some, but of course not all, of the words that are presumably neutral may actually be better understood as more or less ambivalent. Study 1 also showed a strong positive relationship between subjective ambivalence and arousal. However, Study 1 does not show whether ambivalence is unique to the midpoint-rated words in the ANEW. It may be the case that positive and negative words are also perceived as ambivalent, and that the variation from low to high ambivalence in the neutral words is not unique to the neutral words. In addition, Study 1 presented participants

with mid-point rated words only. Participants might have given more extreme positive and/or negative ratings to the ANEW neutral words in a list with *only* neutral words compared to when the neutral words are in a list that also includes positive and negative words. That is, the variance in ambivalence for neutral words might have been inflated because participants did not see any unequivocally non-ambivalent words (i.e. positive and negative words).

Study 2 addressed the above issues by testing whether the highly ambivalent “neutral” words from Study 1 (i.e., midpoint rated in ANEW but rated as highly ambivalent) are rated as more ambivalent than words rated in ANEW as positive and words rated as negative. In addition, the goal of Study 2 was to intersperse words of different valences based on the ANEW rating system to assess whether context matters for how ambivalent the words are rated. We selected the 10 words with the highest and the 10 words with the lowest subjective ambivalence ratings from Study 1 and added positive and negative words.

In Study 2, we preregistered using the minimum statistic (or MIN stat) rather than objective ambivalence. We focused on the MIN stat because the objective ambivalence measure sometimes fails to capture ambivalence in the same way the MIN stat can. For example, a word with a rating of 2 for positivity and 4 for negativity would have the same objective ambivalence (i.e., 1) as a word with a rating of 3 for positivity and 9 for negativity, even though the latter might arguably be considered as more ambivalent given that it has higher ratings for both positivity and negativity. In contrast, the two words would be distinguished by the MIN stat which, as described earlier, is given by the lower of the two ratings for the positivity and negativity of a stimulus. To reduce the burden on participants in Study 2, we used a single item for subjective ambivalence. We hypothesized that neutral high-ambivalence words would have higher mean subjective ambivalence ratings and MIN stat than positive words (H1a & H2a), negative words (H1b & H2b), and neutral low-ambivalence words (H1c & H2c).

Method

The methods, analyses, focal hypotheses, planned sample size and exclusion criteria were pre-registered here: (https://osf.io/v65u2/?view_only=c3bade9ee776469d8e8960fb3d5948b6). There were no deviations from the pre-registered analysis plan, though there was a slight deviation in the prescreening criteria we used to recruit participants from [Prolific.co](#). We report all manipulations, measures, and exclusions.

Participants and Design

We conducted power analyses using G*Power (Erdfelder et al., 1996) for one-tailed paired-samples *t*-tests against a null hypothesis of zero, with an alpha of .05, for an expected effect size of $d = 0.29$ (the size of a manipulation of mixed feelings using video stimuli in past research; Berrios et al., 2018). To achieve 90% power, we needed a sample size of 104 participants. We therefore recruited 104 participants from the U.S. using [Prolific.co](#) (www.Prolific.co),

Table 3. The coefficients (and standard errors) from the regression analyses predicting arousal with standardized predictors (using all words).

	Full model	Model without Valence	Model without MIN	Model without Subjective
Valence	-0.42 (0.09)***	-	-0.49 (0.09)***	-0.66 (0.09)***
MIN	-0.38 (0.16)*	-0.64 (0.17)***	-	0.30 (0.09)**
Subjective	0.87 (0.18)***	1.30 (0.17)***	0.50 (0.09)***	-
Total	$F_{(3,56)} = 46.04^{***}$, $R2 = 71\%$	$F_{(2,57)} = 44.28^{***}$, $R2 = 61\%$	$F_{(2,57)} = 61.29^{***}$, $R2 = 68\%$	$F_{(2,57)} = 40.76^{***}$, $R2 = 59\%$

Note. Subjective = subjective ambivalence. $N = 60$ (words). *** $p < .001$, ** $p < .01$, * $p < .05$.

Table 4. The coefficients (and standard errors) from the regression analyses predicting arousal with standardized predictors (using only words with valence ratings within 0.5 of midpoint).

	Full model	Model without Valence	Model without MIN	Model without Subjective
Valence	-0.19 (0.08)*	-	-0.17 (0.08)*	-0.20 (0.09)*
MIN	-0.23 (0.17)	-0.16 (0.18)	-	0.46 (0.09)***
Subjective	0.76 (0.17)***	0.78 (0.18)***	0.55 (0.08)***	-
Total	$F_{(3,45)} = 31.23^{***}$, $R2 = 68\%$	$F_{(2,46)} = 39.55^{***}$, $R2 = 63\%$	$F_{(2,46)} = 45.21^{***}$, $R2 = 66\%$	$F_{(2,46)} = 26.44^{***}$, $R2 = 53\%$

Note. Subjective = subjective ambivalence. $N = 49$ (words). *** $p < .001$, ** $p < .01$, * $p < .05$.

with a mean age of 36.4 years ($SD = 15.1$) ranging from 19 to 79 years, 78 females and 26 males. Sensitivity analyses revealed that a sample of 104 participants would give one-sided paired-samples t -tests, with an alpha of .05, 80% power to detect an effect size of $d = 0.25$. Participants rated 10 high ambivalence, 10 low ambivalence, 10 positive, and 10 negative rated words for how positive, negative, and mixed, the words made them feel.

We pre-registered using [Prolific.co](https://prolific.co)'s prescreening criteria to include only participants *fluent* in English, but we actually included those *whose first language was English*. We excluded any participants with incomplete data as per the pre-registered plan.

Stimuli

We used the average valence ratings in the ANEW database (Bradley & Lang, 2010) to categorize words from ANEW into negative (average ratings of 1-3) or positive (7-9). We categorized the negative words into *strongly negative* (1-2) or *medium negative* (2.01-3) and the positive words into *strongly positive* (8-9) or *medium positive* (7-7.99). We used R to randomly select 5 medium negative and 5 strong negative words to have a list of 10 negative words. We then randomly selected 5 medium positive and 5 strong positive words to have a list of 10 positive words. For the high ambivalent mid-point rated words, we selected the 10 words with the highest subjective ambivalence ratings in Study 1 that had a valence rating between 4.5-5.5 (i.e., within 0.5 of the midpoint of the valence scale) to create the neutral high ambivalence category. We also selected the 10 words with the lowest subjective ambivalence ratings in Study 1 that had a valence rating within 0.5 of the midpoint of the valence scale to create the neutral low ambivalence category. [Table 5](#) contains the lists of words from each category.

Procedure

Participants read an information sheet and gave informed consent. Participants were told that they would be presented with 40 words and that for each word they would rate how positive they feel, how negative they feel, and to what degree they experienced mixed thoughts and feelings, based on their first and immediate reaction as they read each word. Participants were then presented with the 40 words in random order, one at a time. For each word, participants gave ratings from 1 (not at all) to 9 (extremely) on three items. For the MIN stat, participants rated each word on how positive and how negative it made them feel. For subjective ambivalence, participants rated each word for how much it made them have mixed feelings. These three items were presented with minimal wording (i.e., "positive", "negative", "mixed feelings") to be rated on the 9-point scales. Participants then reported their age and sex and answered whether English was their native language.

Results

We calculated the mean positive, mean negative, and mean subjective ambivalence ratings across the 10 words of

Table 5. Selected words for the categories negative, positive, neutral low ambivalence and neutral high ambivalence.

Negative		
Word	Number	Valence
toothache	443	1.98
depressed	107	1.83
paralysis	926	1.98
disaster	121	1.73
slave	398	1.84
nag	1880	2.9
infect	1719	2.42
dent	1412	2.93
lonely	261	2.17
slime	400	2.68
Positive		
Word	Number	Valence
beach	34	8.03
excellence	151	8.38
smile	2216	8.16
sex	384	8.05
enjoy	1492	8.17
boat	1196	7.79
spirit	406	7
politeness	320	7.18
child	70	7.08
bride	670	7.34
Neutral High Ambivalence		
Word	Number	Valence
society	2231	5.03
intent	1730	5
wolf	2463	5
storm	1000	4.95
flaunt	1566	5.04
fervor	1552	5
apology	1093	5
rattle	346	5.03
descent	1415	5
compel	1326	4.97
Neutral Low Ambivalence		
Word	Number	Valence
vase	2414	4.97
gallon	1611	5.03
liter	1794	5
pail	1933	5.04
barrel	651	5.05
dryer	1454	5.03
lid	1786	5.03
van	2412	4.97

knee	1760	5.03
cabinet	675	5.03

Note. "Number" refers to the word's reference number in the ANEW, and "Valence" is the valence rating of that word in the ANEW.

each of the four word categories for each participant. In addition, for each word for each participant, we calculated the MIN stat and then the mean MIN stat across the 10 words of each of the four word categories. For all of the analyses, we used dependent samples *t*-tests.

Manipulation Checks

As a positive control, we wanted to make sure the positive words were rated as more positive than the words from the other three word-categories and the negative words as more negative than the other three word-categories. One-sided *t*-tests showed that indeed, the 10 positive words ($M = 7.16$, $SD = 1.08$) were rated as more positive than the negative words ($M = 1.80$, $SD = 0.85$), the neutral high-ambivalence words ($M = 4.52$, $SD = 1.19$), and the neutral low-ambivalence words ($M = 4.24$, $SD = 1.79$), all $t_s > 18$, all $p_s < .001$, all $d_s > 1.87$. Likewise, the 10 negative words ($M = 7.48$, $SD = 1.05$), were rated as more negative than the positive words ($M = 2.20$, $SD = 0.83$), the neutral high-ambivalence words ($M = 4.25$, $SD = 1.06$), and the neutral low-ambivalence words ($M = 2.90$, $SD = 1.38$), all $t_s > 27$, all $p_s < .001$, all $d_s > 3.07$. We then moved on to testing the hypotheses.

Results of Hypothesis Tests

H1a-H1c were supported. The neutral high-ambivalence words had higher mean ratings for subjective ambivalence ($M = 3.98$, $SD = 1.38$) than the positive words ($M = 2.82$, $SD = 1.34$), $t_{(103)} = 9.61$, $p < .001$, $d = 0.85$ $CI_{95\%}[0.65, 1.06]$; negative words ($M = 2.50$, $SD = 1.35$), $t_{(103)} = 12.42$, $p < .001$, $d = 1.09$ $CI_{95\%}[0.87, 1.31]$; and the neutral low-ambivalence words ($M = 3.00$, $SD = 1.56$), $t_{(103)} = 7.73$, $p < .001$, $d = 0.66$ $CI_{95\%}[0.48, 0.85]$. Therefore, the neutral high-ambivalence words were rated as eliciting more subjective ambivalence than all of the other three word-categories.

H2a-H2c were also supported. The neutral high-ambivalence words had higher mean MIN stat ($M = 3.08$, $SD = 0.93$) than the: positive words ($M = 2.08$, $SD = 0.79$), $t_{(103)} = 14.21$, $p < .001$, $d = 1.14$ $CI_{95\%}[0.93, 1.34]$; negative words ($M = 1.66$, $SD = 0.68$), $t_{(103)} = 17.46$, $p < .001$, $d = 1.69$ $CI_{95\%}[1.39, 1.99]$; and the neutral low-ambivalence words ($M = 2.68$, $SD = 1.36$), $t_{(103)} = 4.01$, $p < .001$, $d = 0.32$ $CI_{95\%}[0.16, 0.48]$. Therefore, the neutral high-ambivalence words were also rated as more ambivalent than all of the other three word-categories according to the MIN stat.

Study 2 thus showed that the results of Study 1 were not driven by people giving more extreme ratings due to being presented only with neutral words. Moreover, Study 2 showed that ambivalence is unique to some of the words classified as neutral in the ANEW norming process due to the bipolar valence scale. That is, some of the words as-

sumed as neutral in ANEW may actually be better understood as ambivalent.

Exploratory analyses showed that even the neutral low-ambivalence words were rated as more ambivalent than the positive and negative words on both the subjective ambivalence item and according to the MIN stat (though the comparison with positive words on the subjective ambivalence item was not significant, $t_{(103)} = 1.38$, $p = .169$, $d = 0.12$; all other $t_s > 3.27$, all other $p_s < .002$, all other $d_s > 0.34$). Finally, we also performed robustness checks by comparing the neutral high- and low-ambivalent words with the medium-negative and medium-positive words and found the neutral high-ambivalence words to be rated as more ambivalent than both and that even the neutral low-ambivalent words were rated as more ambivalent depending on the index examined (see Supplemental Materials).

General Discussion

We examined whether midpoint valence ratings in the norms for the ANEW masked ambivalence and whether this was related to arousal. Even after selecting the most "neutrally" rated ANEW words (within +/- .05 around the 5.0 midpoint of the valence scale), there were substantial differences in the degree to which these words evoked ambivalence and, correspondingly, arousal. In addition to a positive correlation between arousal and different indices of ambivalence, arousal ratings were strongly, positively, and uniquely related to subjective ambivalence ratings even after controlling for ratings on the bipolar valence scale. Moreover, in Study 2 we found that the relatively high ambivalence ratings that we observed in Study 1 are unique to the words with ratings near the midpoint of the valence scale, since positive and negative words had substantially lower ambivalence ratings. Thus, the neutral ANEW words with midpoint valence ratings do not necessarily induce neutral responses in participants but instead evoke differing levels of ambivalence.

Some researchers seem to have intuited that higher arousal for mid-point rated stimuli may be indicative of variance in other properties, given that they account for arousal when selecting neutral stimuli (e.g., Chan & Singhal, 2013; Isaac et al., 2012; Jeon et al., 2020; Levens & Phelps, 2008; Trammell & Clore, 2014; Yang et al., 2013; Zhang et al., 2019). Past research shows that this other property may actually be varying levels of ambivalence (Schneider et al., 2016). Our studies empirically show that the word stimuli from the ANEW with neutral valence ratings actually hide varying levels of ambivalence and that this ambivalence is positively related to arousal. Our findings conceptually replicate previous research showing the same for image stimuli (Schneider et al., 2016).

When researchers seek neutral stimuli, therefore, the bipolar valence ratings must be used in conjunction with the bipolar arousal ratings in order to avoid selecting stimuli that are ambivalent rather than neutral. However, to be more certain, it would be best if norming procedures for stimuli included subjective ambivalence ratings. This is because although arousal and subjective ambivalence may be strongly correlated, as they were in Study 1 ($r_s = .72$ & $.79$),

the correlation is not perfect. Indeed, as the list of words in [Table 1](#) shows, some words with relatively low arousal ratings may actually have relatively high subjective ambivalence ratings (e.g., “intent”) or vice versa (e.g., “noisy”). And, research shows that even when 2 measures are correlated quite strongly (e.g., $r = .85$), they may produce different results (Carlson & Herdman, 2012). Nevertheless, until norming procedures include subjective ambivalence ratings, we suggest that researchers who desire neutral stimuli should choose those that are rated near the mid-point of the valence scale and which also have low arousal ratings. Researchers may even consider using the results from our Study 1 as a guide for selecting low ambivalence neutral words for truly neutral stimuli and high-ambivalence neutral words for ambivalent stimuli. If possible, researchers should also conduct pilot tests using participants from their desired population to make sure that their neutral stimuli have low subjective ambivalence ratings.

In sum, our research shows that mid-point ratings for valence in the ANEW may mask considerable differences in ambivalence, a state substantially different from neutrality. As such, “neutral” might not be so neutral after all.

Funding

Farid Anvari was supported by funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 883785.

The Article Processing Charges (APC) were funded by the joint publication funds of the TU Dresden, including Carl Gustav Carus Faculty of Medicine, and the SLUB Dresden as well as the Open Access Publication Funding of the DFG.

Competing Interests

The authors have no competing interests to declare.

Contributions

Contributed to conception and design: FA, JSB, IKS

Contributed to acquisition of data: JSB, JB, IKS

Contributed to analysis and interpretation of data: FA, JSB, IKS

Drafted and/or revised the article: FA, JSB, JB, IKS

Approved the submitted version for publication: FA, JSB, JB, IKS

Data Accessibility Statement

All materials, analyses scripts, and data can be found here:

https://osf.io/c3yvt/?view_only=ba8f2e230c574b1cb08cc75766a88f29

The pre-registration for Study 2 is here

https://osf.io/v65u2/?view_only=c3bade9ee776469d8e8960fb3d5948b6

Submitted: February 14, 2023 PDT, Accepted: April 30, 2023 PDT



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license’s legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

References

- Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. *Political Analysis*, 20(3), 351–368. <https://doi.org/10.1093/pan/mpr057>
- Berrios, R., Totterdell, P., & Kellett, S. (2018). When Feeling Mixed Can Be Meaningful: The Relation Between Mixed Emotions and Eudaimonic Well-Being. *Journal of Happiness Studies*, 19(3), 841–861. <https://doi.org/10.1007/s10902-017-9849-y>
- Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion*, 1(3), 276–298. <https://doi.org/10.1037/1528-3542.1.3.276>
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49–59. [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
- Bradley, M. M., & Lang, P. J. (1999). Affective Norms for English Words (ANEW): Affective ratings of words and instruction manual. In *Technical Report C-1: Vol. Technical Report C-2*. The Center for Research in Psychophysiology, University of Florida.
- Bradley, M. M., & Lang, P. J. (2010). Affective Norms for English Words (ANEW): Affective ratings of words and instruction manual. In *Technical Report C-2: Vol. Technical Report C-2*. University of Florida.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2016). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research (4th ed.)*. (pp. 133–139). American Psychological Association. <https://doi.org/10.1037/14805-009>
- Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, 115(3), 401–423. <https://doi.org/10.1037/0033-2909.115.3.401>
- Carlson, K. D., & Herdman, A. O. (2012). Understanding the impact of convergent validity on research results. *Organizational Research Methods*, 15(1), 17–32. <https://doi.org/10.1177/1094428110392383>
- Chambers, R., Lo, B. C. Y., & Allen, N. B. (2008). The Impact of Intensive Mindfulness Training on Attentional Control, Cognitive Style, and Affect. *Cognitive Therapy and Research*, 32(3), 303–322. <https://doi.org/10.1007/s10608-007-9119-0>
- Chan, M., & Singhal, A. (2013). The emotional side of cognitive distraction: Implications for road safety. *Accident Analysis & Prevention*, 50, 147–154. <https://doi.org/10.1016/j.aap.2012.04.004>
- Conner, M., & Sparks, P. (2002). Ambivalence and attitudes. *European Review of Social Psychology*, 12(1), 37–70. <https://doi.org/10.1080/14792772143000012>
- Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, 28(1), 1–11. <https://doi.org/10.3758/bf03203630>
- Gustavson, D. E., du Pont, A., Whisman, M. A., & Miyake, A. (2018). Evidence for Transdiagnostic Repetitive Negative Thinking and Its Association with Rumination, Worry, and Depression and Anxiety Symptoms: A Commonality Analysis. *Collabra: Psychology*, 4(1), 13. <https://doi.org/10.1525/collabra.128>
- Isaac, L., Vrijzen, J. N., Eling, P., van Oostrom, I., Speckens, A., & Becker, E. S. (2012). Verbal and facial-emotional Stroop tasks reveal specific attentional interferences in sad mood. *Brain and Behavior*, 2(1), 74–83. <https://doi.org/10.1002/brb3.38>
- Jamieson, D. W. (1993). *The attitude ambivalence construct: Validity, utility, and measurement*. Annual Meeting of the American Psychological Association, Toronto.
- Jeon, Y. A., Resnik, S. N., Feder, G. I., & Kim, K. (2020). Effects of emotion-induced self-focused attention on item and source memory. *Motivation and Emotion*, 44(5), 719–737. <https://doi.org/10.1007/s11031-020-09830-w>
- Kaplan, K. J. (1972). On the ambivalence-indifference problem in attitude theory and measurement: A suggested modification of the semantic differential technique. *Psychological Bulletin*, 77(5), 361–372. <https://doi.org/10.1037/h0032590>
- Kreibig, S. D., & Gross, J. J. (2017). Understanding mixed emotions: Paradigms and measures. *Current Opinion in Behavioral Sciences*, 15, 62–71. <https://doi.org/10.1016/j.cobeha.2017.05.016>
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). *International affective picture system (IAPS): Instruction manual and affective ratings*.
- Larsen, J. T., Norris, C. J., & Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, 40(5), 776–785. <https://doi.org/10.1111/1469-8986.00078>
- Legare, C. H., & Souza, A. L. (2014). Searching for Control: Priming Randomness Increases the Evaluation of Ritual Efficacy. *Cognitive Science*, 38(1), 152–161. <https://doi.org/10.1111/cogs.12077>
- Levens, S. M., & Phelps, E. A. (2008). Emotion processing effects on interference resolution in working memory. *Emotion*, 8(2), 267–280. <https://doi.org/10.1037/1528-3542.8.2.267>
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, 44(1), 1–23. <https://doi.org/10.3758/s13428-011-0124-6>
- Minnema, M. T., & Knowlton, B. J. (2008). Directed forgetting of emotional words. *Emotion*, 8(5), 643–652. <https://doi.org/10.1037/a0013441>

- Newby-Clark, I. R., McGregor, I., & Zanna, M. P. (2002). Thinking and caring about cognitive inconsistency: When and for whom does attitudinal ambivalence feel uncomfortable? *Journal of Personality and Social Psychology*, *82*(2), 157–166. <https://doi.org/10.1037/0022-3514.82.2.157>
- Nimon, K., Lewis, M., Kane, R., & Haynes, R. M. (2008). An R package to compute commonality coefficients in the multiple regression case: An introduction to the package and a practical example. *Behavior Research Methods*, *40*(2), 457–466. <https://doi.org/10.3758/brm.40.2.457>
- Priester, J. R., & Petty, R. E. (1996). The gradual threshold model of ambivalence: Relating the positive and negative bases of attitudes to subjective ambivalence. *Journal of Personality and Social Psychology*, *71*(3), 431–449. <https://doi.org/10.1037/0022-3514.71.3.431>
- Rees, L., Rothman, N. B., Lehavy, R., & Sanchez-Burks, J. (2013). The ambivalent mind can be a wise mind: Emotional ambivalence increases judgment accuracy. *Journal of Experimental Social Psychology*, *49*(3), 360–367. <https://doi.org/10.1016/j.jesp.2012.12.017>
- Schimmack, U. (2001). Pleasure, displeasure, and mixed feelings: Are semantic opposites mutually exclusive? *Cognition and Emotion*, *15*(1), 81–97. <https://doi.org/10.1080/0269993004200123>
- Schneider, I. K., & Mattes, A. (2021). Mix is different from nix: Mouse tracking differentiates ambivalence from neutrality. *Journal of Experimental Social Psychology*, *95*, 104106. <https://doi.org/10.1016/j.jesp.2021.104106>
- Schneider, I. K., & Schwarz, N. (2017). Mixed feelings: The case of ambivalence. *Current Opinion in Behavioral Sciences*, *15*, 39–45. <https://doi.org/10.1016/j.cobeha.2017.05.012>
- Schneider, I. K., Veenstra, L., van Harreveld, F., Schwarz, N., & Koole, S. L. (2016). Let's not be indifferent about neutrality: Neutral ratings in the International Affective Picture System (IAPS) mask mixed affective responses. *Emotion*, *16*(4), 426–430. <https://doi.org/10.1037/emo0000164>
- Sereno, S. C., Scott, G. G., Yao, B., Thaden, E. J., & O'Donnell, P. J. (2015). Emotion word processing: Does mood make a difference? *Frontiers in Psychology*, *6*. <https://doi.org/10.3389/fpsyg.2015.01191>
- Thompson, M. M., Zanna, M., & Griffin, D. W. (1995). Let's not be indifferent about (attitudinal) ambivalence. In *Attitude Strength: Antecedents and Consequences* (pp. 361–386). Psychology Press.
- Trammell, J. P., & Clore, G. L. (2014). Does stress enhance or impair memory consolidation? *Cognition and Emotion*, *28*(2), 361–374. <https://doi.org/10.1080/02699931.2013.822346>
- Trampe, D., Quoidbach, J., & Taquet, M. (2015). Emotions in Everyday Life. *PLOS ONE*, *10*(12), e0145450. <https://doi.org/10.1371/journal.pone.0145450>
- van Harreveld, F., Nohlen, H. U., & Schneider, I. K. (2015). The ABC of Ambivalence: Affective, behavioral, and cognitive consequences of attitudinal conflict. In *Advances in Experimental Social Psychology* (Vol. 52, pp. 285–324). Academic Press. <https://doi.org/10.1016/bs.aesp.2015.01.002>
- van Harreveld, F., Rutjens, B. T., Rotteveel, M., Nordgren, L. F., & van der Pligt, J. (2009). Ambivalence and decisional conflict as a cause of psychological discomfort: Feeling tense before jumping off the fence. *Journal of Experimental Social Psychology*, *45*(1), 167–173. <https://doi.org/10.1016/j.jesp.2008.08.015>
- Wadlinger, H. A., & Isaacowitz, D. M. (2008). Looking happy: The experimental manipulation of a positive visual attention bias. *Emotion*, *8*(1), 121–126. <https://doi.org/10.1037/1528-3542.8.1.121>
- Yang, H., Yang, S., & Isen, A. M. (2013). Positive affect improves working memory: Implications for controlled cognitive processing. *Cognition & Emotion*, *27*(3), 474–482. <https://doi.org/10.1080/02699931.2012.713325>
- Zhang, W., Gross, J., & Hayne, H. (2019). An age-related positivity effect in semantic true memory but not false memory. *Emotion*, *21*(3), 526–535. <https://doi.org/10.1037/emo0000715>

Supplementary Materials

Peer Review History

Download: https://collabra.scholasticahq.com/article/82204-is-neutral-really-neutral-mid-point-ratings-in-the-affective-norms-english-words-anew-may-mask-ambivalence/attachment/169089.docx?auth_token=b3PR-ncCtr3DGqoTSEHo

Supplemental Material

Download: https://collabra.scholasticahq.com/article/82204-is-neutral-really-neutral-mid-point-ratings-in-the-affective-norms-english-words-anew-may-mask-ambivalence/attachment/169090.docx?auth_token=b3PR-ncCtr3DGqoTSEHo
