1 Ambiguous Words

2 Nicholas R. Harp[1], Catherine C. Brown[1], & Maital Neta[1]

3 [1] University of Nebraska-Lincoln

4 Author Note

5 Nicholas R. Harp, Department of Psychology, Center for Brain, Biology, and Behavior,

6 University of Nebraska-Lincoln Catherine C. Brown, Department of Psychology, Center for

7 Brain, Biology, and Behavior, University of Nebraska-Lincoln Maital Neta, Department of

8 Psychology, Center for Brain, Biology, and Behavior, University of Nebraska-Lincoln

9 Correspondence concerning this article should be addressed to Nicholas R. Harp,

10 Postal address. E-mail: nharp@huskers.unl.edu

11   Abstract

12   We found some ambiugous words.

13   Two or three sentences explaining what the **main result** reveals in direct comparison

14   to what was thought to be the case previously, or how the main result adds to previous

15   knowledge.

16   One or two sentences to put the results into a more **general context**.

17   Two or three sentences to provide a **broader perspective**, readily comprehensible to

18   a scientist in any discipline.

19   *Keywords:* ambiguity

20   Word count: X

²¹ <div style="text-align:center">Ambiguous Words</div>

²² <div style="text-align:center">**Introduction**</div>

²³ We wanted to identify ambiguous words. Mention in this section why we also
²⁴ generated clearly positive and negative words.

²⁵ <div style="text-align:center">**Study 1: Pilot**</div>

²⁶ **Methods**

²⁷ **Participants.** Workers on Amazon's Mchanical Turk (MTurk) were invted to
²⁸ participate in an eligibility screener with the option to earn a bonus if they met the
²⁹ requirements and completed the entire study. The Workers clicked a hyperlink that directed
³⁰ them to the study. The screener task included demographic questions and one block of word
³¹ ratings that included 5 instances of the word "negative" and 5 instances of the word
³² "positive" (see Procedure below for full details). Workers were invited to complete the entire
³³ study if they indicated that they were over 18 years old, had English as their native
³⁴ language, had no history of psychological or neurological disorder, and correctly rated the
³⁵ words "positive" and "negative" as positive or negative with at least 80% accuracy. Of the
³⁶ 145 Workers who completed the screener, 119 met the eligibility requirements, and 103
³⁷ (54.37% female, 45.63% male) chose to complete the entire study. The final sample was
³⁸ 3.88% Asian, 5.83% Black, 2.91% Hispanic or Latino, 85.44% White, and 0% Other, with a
³⁹ mean(sd) age of 37.16(10.60).

⁴⁰ **Material.**

⁴¹ *Stimuli.*

⁴² We compiled an initial set of 59 words that we believed had two distinct definitions,
⁴³ one clearly positive definition and one clearly negative definition. To create lists of clearly
⁴⁴ positive and clearly negative words, we first created a master list of words that were included
⁴⁵ in both the study by Warriner, Kuperman, and Brysbaert (2013), for valence and arousal

ratings, and the English Lexicon Project online word query (Balota et al., 2007), for lexical characterisic measurements. We then elimiated any words with a mean arousal rating that was greater than 1 standard deviation away from the mean arousal of the list of 59 ambiguous words. We classified "positive" words as those with a mean valence > 7 on the 1-9 scale used by Warriner et al. (2013); "negative" words had mean valence < 3. To ensure that all words shared similar lexical characteristics, we eliminated any words from the master list whose lexical characteristics did not fall within the minimum and maximum values of the 59 ambiguous words' lexical characteristics. The following were used for the cutoffs: length, the frequency of a word as reported by the Hyperspace Analogue to Language (HAL) study (Lund & Burgess, 1996), the log of HAL frequency, number of phonemes, number of syllables, number of morphemes, lexical decision reaction time and accuracy, and naming reaction time and accuracy. The final list of pilot words included 59 ambiguous, 267 positive, and 304 negative words.

All of the calculations described in this section were scripted using R version 3.6.2 and are available in the **Supplementary Information**.

Software.

All tasks were created and presented using Gorilla Experiment Builder (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2019). The study was only accessible to participants using a computer (not a phone or tablet) within the United States.

Screener and word rating task. After giving informed consent, participants first answered demographic questions about their gender, age, race, native language, and whether they had ever been diagnosed with a psychological or neurological disorder. They were thenshown a brief self-guided instructional walkthrough of the task before completing the screener.

Using a random seed, we selected 20 positive and 20 negative words from the final pilot list for use in the screener task. These 40 words, along with 5 instances of the word

72 "positive" and 5 instances of the word "negative" were presented randomly, one at a time,

73 each following a 250 ms fixation cross. Each word remained on screen until the participant

74 indicated that they thought it was positive or negative by pressing A or L on their keyboard

75 (key pairing randomized across participants). If no response was made after 2000ms, a

76 reminder appeared on screen, "Please respond as quickly as you can! A = POSITIVE. L =

77 NEGATIVE." Participants who rated the words "positive" and "negative" with less than

78 80% accuracy were compensated for their time but were not invited to complete the rest of

79 the study. Participants were also excluded at this point if they indicated that they were

80 younger than 18, that English was not their native language, or that they had been

81 diagnosed with a psychological or neurological disorder.

82   The remaining 590 words from the final pilot list were randomly presented across 10

83 blocks of 59 words using the same button-press procedure as the screener block.

Results

85   Trials with a response time faster than 250ms were removed from the data prior to

86 analysis, as well as trials with a reaction time greater than 3 SDs above the mean reaction

87 time averaged across all trials.

88   We assessed average reaction time to identify the ambiguous words within the range of

89 35%-65% average negative rating, suggesting low response consensus. Previous work has

90 shown that ambiguous faces and images are associated with longer reaction times in a

91 forced-choice valence classification task (CITE). **Figure 1a** shows that 29 amibugous, 5

92 negative, and 6 positive words surpassed a reaction time threshold of 875ms (Why did we

93 use 875? Just visual inspection?). These 40 words were considered for inclusion in a final list

94 of ambiguous words. We removed 7 words that did not have both a clearly positive and

95 clearly negative definition ("recession", "faceless", "headstone", "inherit", "abundant",

96 "cosmic", "receive"), as well as 1 word that was redundant to another ambiguous word that

97   we included ("courtroom"), resulting in a final list of 32 ambiguous words.

98   As shown in **Figure 1b**, visual inspection of the average valence ratings revealed two

99   distinct groups of words with high response consensus: one with a clearly negative meaning

100   (n = 18, mean valence rating > 75% negative) and one with a clearly positive meaning (n =

101   20, mean valence rating < 10% negative). We removed the words "positive" and "negative"

102   from each list (explain). Because the valence bias task requires an equal number of

103   ambiguous (50%) and clearly-valenced (25% positive, 25% negative) stimuli, we included the

104   16 words with the fastest reaction time for the positive and negative word lists, respectively.

Study 1 Discussion

106   Study 1 generated a list of 32 ambigous words, as well as 16 positive and 16 negative

107   words, for use in determining whether valence bias generalizes to verbal ambiguity. Study 2

108   aimed to test this by comparing ratings of word to ratings of well-validated stimuli sets

109   consisting of faces and scenes.

Study 2: Comparison of words with valence bias and IPANAT

Methods

Participants.

Material.

Stimuli.

Valence Bias with Words.

Valence Bias with Faces.

Valence Bias with IAPS.

IPANAT.

Software.

Procedure.

Valence Bias Tasks.

122 ***IPANAT.***

123 **Data analysis.**

124 ***Valence Bias Tasks.***

125 ***IPANAT.***

126 **Results**

127 **Subjective ratings.**

128 *Valence Bias with Words.*

129 *Valence Bias with Faces.*

130 *Valence Bias with IAPS.*

131 *IPANAT.*

132 **Reaction times.**

133 *Valence Bias with Words.*

134 *Valence Bias with Faces.*

135 *Valence Bias with IAPS.*

136 *IPANAT.*

137 **Relationships between the measures.**

138 **Discussion**

139 We did this study.

References

140

141 Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2019).

142   Gorilla in our midst: An online behavioral experiment builder. *Behav Res.*

143   https://doi.org/10.3758/s13428-019-01237-x

144 Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., . . .

145   Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3),

146   445–459. https://doi.org/10.3758/BF03193014

147 Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical

148   co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2),

149   203–208. https://doi.org/10.3758/BF03204766

150 Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and

151   dominance for 13,915 English lemmas. *Behav Res*, *45*(4), 1191–1207.

152   https://doi.org/10.3758/s13428-012-0314-x

Supplementary Information

Study 1 Stimuli

Insert link to repository for "pick_words.R"