# Separability and the Effect of Valence
## An Empirical Study of Thick Concepts

**Pascale Willemsen (Pascale.Willemsen@uzh.ch)** **Kevin Reuter (Kevin.Reuter@uzh.ch)**

Institute of Philosophy, University of Zurich, Zürichbergstrasse 43, 8044 Zurich, Switzerland

**Abstract**

Thick terms and concepts, such as *honesty* and *cruelty,* are at the heart of a variety of debates in linguistics, philosophy of language, and metaethics. Central to these debates is the question of how the descriptive and evaluative components of thick concepts are related and whether they can be separated from each other. So far, no empirical data on how thick terms are used in ordinary language has been collected to inform these debates. In this paper, we present the first empirical study, designed to investigate whether the evaluative component of thick concepts can be separated. Our study might be considered to falsify the view that evaluation is conversationally implicated. However, our study also reveals an effect of valence, indicating that people reason differently about positive and negative thick terms. While evaluations cannot be cancelled for negative thick terms, they can be for positive ones. Three follow-up studies were conducted to explain this effect. We conclude that the effect of valence is best accounted for by a difference in the social norms guiding evaluative language.

**Keywords:** Thick concepts; moral judgments; experimental metaethics; evaluative language

## Introduction

Philosophers and linguists usually distinguish two types of evaluative terms and concepts: "thin" and "thick" ones (Eklund, 2011, Väyrynen, 2019). Thin terms and concepts evaluate an object as, for instance, "permissible", "right", "wrong", "good", "bad", or "blameworthy", yet they do not explicate in what way the object is right or wrong. If a speaker evaluates an instance of lying as wrong, they convey no information as to why they think so. It might be easy for you to guess what reasons the speaker has in mind: The speaker might think that people have a right to be told the truth, that it ruins friendships, etc. – but note that the term "wrong" all by itself does not provide this information. Thick concepts do not merely evaluate, they also provide information on why the entity is evaluated in this way. Typical examples are ethical thick terms and concepts, such as "rude", "cruel", "courageous", or "trustworthy". Calling an agent courageous evaluates them positively for being willing to take risks – "reckless" also ascribes willingness to take risks yet assigns a negative evaluation to it.

While there is widespread consensus that thick concepts form an additional class of concepts, a heated disagreement exists over the way in which the evaluative and the descriptive component of thick terms and concepts are connected. According to one group of researchers, the evaluative component of a thick term is part of its semantic meaning; according to another, the evaluation is not part of the semantic meaning but conveyed through pragmatic means. The issue at hand is a question about the location of the evaluative – does it belong to the semantics or the pragmatics of a thick term or concept? Arguments in favor of either position heavily rely on linguistic intuitions about how thick terms expressing thick concepts can be used. Such intuitions often circle around the question of whether the evaluation of a thick term can be cancelled without yielding contradiction. A related debate exists over the question of whether the evaluation, independent of how it is connected to the descriptive content, can be separated from it. Some philosophers argue that the evaluation is inseparable, while others deny this. Whether this is possible is not only relevant for the linguistic debate about thick terms. Assumptions about the nature of thick terms provide the argumentative cornerstones in metaethical and normative-ethical debates as well. Therefore, by putting these assumptions to the test, we can provide a more solid basis for theorizing about thick concepts in various disciplines. In this paper, we present the first empirical data of this sort.

## Separating the Evaluative from the Descriptive?

Thick terms and concepts, such as "honest", "friendly", "cruel" or "rude", do not only evaluate an entity, they further describe *in virtue of what* this entity is evaluated as positive or negative. This descriptive richness is what distinguishes thick from thin terms and concepts. And it is the evaluative component that distinguishes thick from merely descriptive terms and concepts. So far, philosophers have mostly relied on their intuitions to determine whether or not a concept is thick (but see Reuter et al. (forthcoming) which includes an empirical study on whether the concepts *friend*, *colleague*, and *rival* are thick concepts; Reuter, Baumgartner, & Willemsen (ms) discuss various methods, including tools from corpus analysis, to delineate thick concepts from descriptive concepts as well as value-associated concepts.)

Some philosophers explain the descriptive richness of thick concepts by assuming that thick terms are basic and amalgams of description and evaluation (Williams, 1985, Putnam, 2002, Kirchin, 2010, Roberts, 2011). Call this the Inseparability View. Inseparabilists claim that the evaluation is part of the semantic meaning such that the descriptive meaning itself evaluates.

According to the contrary position, the Separabilist View (Väyrynen, 2019), thick terms can be, at least in principle,

divided into two distinct components, namely the evaluative and the descriptive (Hare, 1952; Blackburn, 1992; Elstein & Hurka 2009). Separabilist Views fall into two camps: First, *Pragmatic Separabilists* assume that the descriptive and the evaluative are connected by pragmatic means, for instance, by conversational implicature (Stevenson 1938, Hare 1963; Blackburn, 1992; for discussions of these positions see Eklund, 2011, Kyle, 2013, and Väyrynen, 2013, 2019). Conversational implicatures are part of the speaker meaning and need to be inferred beyond what is literally said (Grice, 1985). By saying that an agent is rude, one ascribes some descriptive properties, and one further communicates the implicature that the agent is bad in virtue of having these properties. However, as other conversational implicatures, the negative evaluation can be cancelled without creating a contradiction. Therefore, a speaker who utters "What Tom did was rude, but by that I'm not saying something negative about Tom" makes a felicitous statement. Second, *Semantic Separabilists* claim that that the evaluative and the descriptive are connected via semantic entailment. Whenever a speaker says that Tom is rude, calling him rude entails a negative evaluation – there is no way to ever call Tom rude and not evaluate him negatively. Evaluating negatively is part of what the term "rude" means, and saying "What Tom did was rude, but by that I'm not saying something negative about Tom" would be infelicitous.

The aim of this paper is to empirically test whether the evaluation of a thick term can be cancelled like a conversational implicature. If we find such evidence, this support the Pragmatic Separability View and falsify both the Inseparability and the Semantic Separability View.

## Positive and Negative Thick Concepts

Within the thick concepts debate, scholars usually speak about thick terms and concepts as if they were a homogenous group, such that whether a concept evaluates positively or negatively does not matter. Consequently, every theoretical claim that is being made about a positive concept can be equally applied to negative concepts, and vice versa. We are yet skeptical that such an assumption should be made without further empirical evidence.

Over the past 20 years, a growing body of empirical evidence suggests that moral valence has a significant effect on a series of non-moral phenomena. Whether an act is evaluated positively or negatively affects judgments about causation (Sytsma et al., 2019, for an overview see Willemsen & Kirfel, 2019), intentionality (Knobe, 2003), knowledge (Beebe & Buckwalter, 2010), just to name a few.

Thick terms seem to fall into two groups, namely those evaluating positively and those evaluating negatively. Due to this systematic difference in valence, we believe that there is a possibility that thick positive and negative evaluations have different effects on cancellability ratings. In our pre-registered study, we initially follow the philosophical and linguistic theories and formulate predictions that treat positive and negative thick terms alike. Nevertheless, we also explore the possibility that things are more complicated than theorists have assumed and that the valence of a thick term might affect how easy cancelling its evaluation is.

## Experimental Linguistics and Cancellability

Experimental linguistics provides the means to test the Pragmatic Separability View empirically, namely the cancellability test (see Zakkou, 2018). If the evaluative is part of the meaning of a thick term and semantically entailed, a person who says that Tom is rude but at the same time cancels the evaluation should be considered to contradict herself. Take another semantic entailment as an example: "Tom is a bachelor" semantically entails that Tom is unmarried. A speaker who utters "Tom is a bachelor, but by that I am not saying he is unmarried" contradicts herself. If the evaluation of a thick term is also semantically entailed, we should expect a sentence in which the evaluation is cancelled to be equally contradictory.

However, if the Pragmatic Separabilists are correct and the evaluative aspect is only conversationally implicated, cancelling the evaluation should not lead to a contradiction. For instance, the sentence "There is the door" usually not only communicates the location of a door, but further carries the conversational implicature that the addressee is asked to leave the room. However, saying "There is the door, but I am not saying you should leave" does not yield a contradiction. If the evaluation of a thick term is conversationally implicated by a thick term, cancelling the evaluation should be equally non-contradictory as other conversational implicatures. With this well-established test at our disposal, we designed and pre-registered an experiment aiming to test for the Pragmatic Separabilist View.

At this point, we would like to emphasize that in this paper, we test whether the evaluation of a thick concept is cancellable just as conversational implicatures are. However, conversational implicatures are not the only way of conveying content beyond what is literally said. In addition, a statement can *conventionally* implicate or *presuppose* content. Crucially, conventional implicatures or presuppositions are also not cancellable. Zakkou (ms) commits to the first alternative and claims that thick concepts conventionally implicate evaluation. Väyrynen (2013) argues that thick concepts presuppose their evaluative content. We decided to simplify the debate in this way, as conversational implicatures are relatively easy to test for. If we can show that the evaluation can be cancelled, we have good evidence for a Separabilist view according to which thick concepts convey evaluation by means of conversational implicature. If our results do not support this interpretation of the Separabilist View, additional research needs to be conducted on whether other Pragmatic Separabilist or Semantic Separabilist or even the Inseparabilist positions are more adequate.

## Study 1

The goal of the first study is to provide initial evidence as to whether the evaluative component of a thick term is connected to the descriptive component through semantic or pragmatic means. To this end, we presented 206 participants

with sentences in which a Conversational Implicature, a Semantic Entailment, or the Evaluation of a Thick Term was first communicated and then canceled. We asked participants to what extent the speaker contradicted herself. We predicted that for Conversational Implicatures, cancelling the implicated meaning was possible without creating a contradiction. For Semantic Entailment, cancelling should not be possible and result in high contradiction ratings. For Thick Terms, we hypothesized that if the Pragmatic Separabilists are correct, cancelling the evaluation should provide contradiction ratings similar to those for Conversational implicatures. If the evaluative is part of the semantics of a thick term, contradiction ratings will resemble those of Semantic Entailments.

The experimental design, predictions, and statistical models were pre-registered with the Open Science Framework (https://osf.io/9pbq2/).

## Methods

**Participants** Participants were recruited via MTurk and completed an online survey implemented in Qualtrics. All participants were required to be at least 18 years old, English native speakers, and to have an approval rating of previous studies on the platform of at least 95%. These conditions did also apply to all other studies presented in this paper. All 205 participants who finished the survey were included in the analysis (67.8% male, 31.7% female, 0.5% non-binary; $M_{Age}$ = 35.69)

**Design, Procedure, and Materials** We implemented a 4×1 between-subject design with the independent variable *Condition* (Thick Negative Terms [short: TNT], Thick Positive Terms [short: TPT], Semantic Entailment [short: SE], Conversational Implicature [short: SI]) and the dependent variable *Contradiction*. As stimuli, we used:

- 4 positive thick concepts: Honest, Generous, Courageous, and Friendly (condition TPT)
- 4 negative thick concepts: Intolerant, Rude, Cruel, and Egoistic (condition TNT)
- 4 Conversational Implicatures: Hungry, Dark, Door, Chocolate (condition CI)
- 4 Semantic Entailments: Run, Widow, Couch, Lake (condition SE)

All stimulus sentences, as well as the instructions we gave to the participants, can be found in the Appendix. Here are three concrete examples of the sentences we used: *Conversational Implicature*: "I am hungry, but by that I am not saying that we should get something to eat." *Semantic Entailment*: "This is a couch, but by that I am not saying that this is a piece of furniture." *Thick Negative*: "Amy's behavior last week was egoistic, but by that I am not saying something negative about Amy's behavior that day." Participants then answered the question "Does Sally contradict herself" on a scale from "1 = definitely not" to "9 = definitely yes".

Participants in TNT and TPT read the stimuli for all 4 negative or positive thick terms, participants in CI read the stimuli for all 4 conversational implicatures, participants in SE read the stimuli for all 4 semantic entailments. All stimuli were presented in randomized order. As philosophers in the debate do not predict an effect of valence for thick concepts, we collapsed TNT with TPT to TT for our pre-registered statistical analyses.

## Results

The results of Study 1 are summarized in Figures 1 and 2. We conducted a 4×1 Anova with *Condition* as a between-subject factor. The analysis revealed a significant main effect of *Condition*, $F(3, 201) = 32.15$, $p < .001$. In accordance with our preregistered hypothesis, we conducted a planned contrast for SE ($M = 7.17$) and CI ($M = 3.74$) and detected a significant difference ($F(1, 201) = 71.34$, $p < .001$). Two additional planned contrasts revealed a significant difference between TT ($M = 6.49$) and CI ($F(1, 201) = 62.49$, $p < .0001$) and no significant difference between TT and SE ($F(1, 201) = 3.53$, $p = .062$).

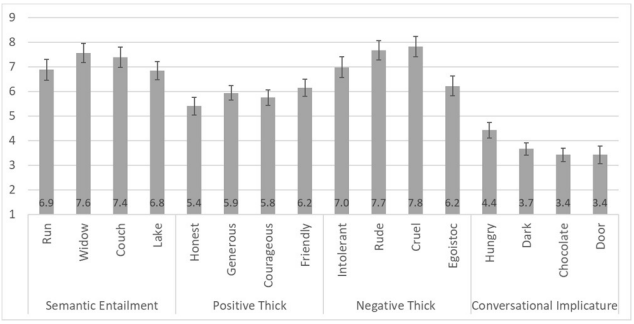


Figure 1: Participant's' mean contradiction ratings for all 16 items (1= "definitely not"; 9 = "definitely yes"). Error bars indicate the standard error around the mean.

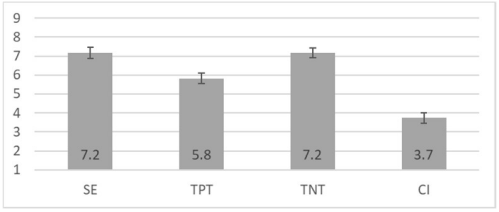


Figure 2: Participants' mean contradiction ratings as a function of Condition. Error bars indicate the standard error around the mean.

In addition, we conducted a planned contrast between TNT ($M = 7.17$) and TPT ($M = 5.81$). The analysis revealed a significant difference between both conditions ($F(1, 201) = 10.92$, $p < .01$).

## Discussion

The prediction of the Pragmatic Separabilist View was not met, since we found a significant difference between contradiction ratings for conversational implicatures and thick terms. In contrast, the results speak in favor of the Semantic View, according to which contradiction ratings for cancelling the evaluation of a thick term should resemble contradiction ratings for semantic entailments. As predicted by the Semanticists, the difference between ratings for

semantic entailments and thick concepts was insignificant. However, this result should be taken with some caution. A p-value close to .05 might have been significant, had we increased the statistical power. Our analysis also revealed an effect of valence that would not have been predicted by either group of theorists. Negative terms received significantly higher contradiction ratings than positive terms. It is this unexpected effect that sparks skepticism that the evidence provided should be taken to falsify the Pragmatic Separability View in the first place. Even though the philosophical literature treats positive and negative thick terms alike, the empirical results suggest that they are not and might better be discussed individually.

## Explaining the effect of valence

How can we explain the effect of valence? After all, this effect would not have been predicted by any theory. Instead, thick concepts are usually discussed as a homogenous group. One obvious but rather bold move would be to suggest that positive and negative thick terms simply do not work alike. Rather, it might be proposed, we need two separate accounts of thick concepts, one for positive and negative concepts. While we do not wish to dismiss this suggestion all too quickly, we believe that we should only draw this conclusion after eliminating alternative explanations of the effect. Here are three explanations that we consider plausible:

**Differences in evaluative intensity** Since we tested only a limited sample of four positive and four negative concepts, the ones we selected might differ in the extent to which they evaluate positively and negatively respectively. Thus, it might well be that the four positive terms do not evaluate as positively as the negative terms evaluate negatively. If that is the case, it should not be surprising that negative terms yield higher contradiction ratings: The stronger the evaluation, the more implausible it might be to cancel the evaluation.

**Differences in the availability of counterexamples** When thinking about honesty and courage, we might think of cases in which an agent is being too honest or too courageous. We might also think of cases in which an agent is honest or courageous, yet for the wrong reasons. In all of these cases, being honest and courageous is not such a good thing but has (at least partially) turned into something negative. For negative thick terms, however, such counterexamples do not come to mind easily. Therefore, it might be argued, attempts to cancel a usually communicated evaluation of a thick term is dependent on our reasoning about counterexamples. For positive terms, they can be easily triggered, while for negative ones, they are not.

**Differences in the social norms guiding evaluative language** Finally, one might wonder whether the effect of valence can be explained by different social norms that guide evaluative language. Uttering a positive thick term without the intention to commit to a positive evaluation seems relatively harmless. Being misunderstood in cases of negative thick terms has a potentially greater impact. If mistaken, a speaker communicates a negative evaluation they initially did not want to commit to. Since negative evaluations harm others by diminishing their social status and reputation, people are less willing to accept a cancellation of a negative evaluation. We tested all three explanations by running additional empirical tests:

## Study 2: Differences in evaluative intensity

Explanation 1 holds that the thick terms we used differ in their evaluative intensity. In Study 2 we aim to test this explanation and assign the following prediction to its advocates: If the effect of valence can be explained by differences in intensity ratings, negative terms should be rated more negatively than positive terms are rated positively. For all eight thick concepts used in Study 1, we collected intensity ratings. We used two measures for intensity, one targeting the goodness or badness of the behavior, another one targeting the valence of the statement featuring a thick term.

### Methods

**Participants** Of all 409 participants who finished the survey, 10 participants were excluded because they were not native speakers of English. 399 participants were included in the analysis (0% non-binary, 47.4% male, 52.6% female; $M_{Age} = 38.74$).

**Design, Procedure, and Materials** We implemented a 2×2 between-subject design with the independent variable *Valence* (Positive; Negative) and *Question* (Behavior; Sentence), and the dependent variable *Intensity*. In the Behavior condition, participants answered a question with the following structure on a 9-point Likert item, reaching from "1 = not bad/good at all" to "9 = very bad/good": How bad/good is it if a person's behavior is [thick term]? For instance, for Rude, the question read: How bad is it if a person's behavior is rude? In the Sentence condition, participants read "Suppose that Sally said the following thing about Tom: "What Tim did was [thick term]". Afterwards they answered the question "Is this a negative/positive statement about Tom?" on a 9-point Likert item, anchored at "1 = definitely not" and "9 = definitely yes".

### Results

The results of Study 2 are summarized in Figure 3. We conducted a 2×2 Anova for intensity ratings with the independent factors *Valence* and *Question*. There was a significant main effect of *Valence*, $F(1, 394) = 70.22$, $p < .001$, such that positive thick terms were given higher intensity ratings (Behavior: $M = 8.1$, Sentence: $M = 8.4$) than negative thick terms (Behavior: $M = 7.1$, Sentence: $M = 6.4$). There was no significant main effect of *Question*, $F(1, 394) = 1.36$, $p = .244$. The two-way interaction was significant, $F(1, 394) = 6.42$, $p = .012$, as the difference between Negative and Positive was larger for Sentence compared to Behavior.

## Discussion

Based on our results, we can reject the hypothesis that the effect of valence can be explained by differences in the intensity of the thick concepts we used. Our positive thick terms were rated even more positively than our negative terms were rated negatively. This effect occurred for both the Behavior and the Sentence condition.

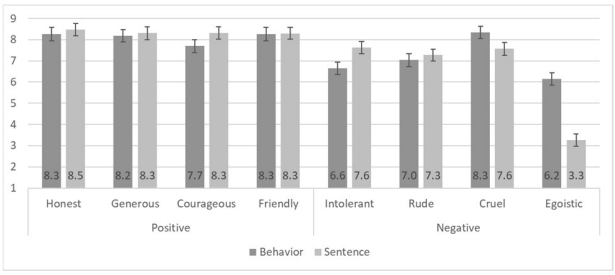


Figure 3: Participants' mean intensity ratings for all 8 items. Error bars indicate the standard error around the mean.

## Study 3: Availability of counterexamples

A second explanation of the effect of valence suggests that putative counterexamples that allow for cancellation of the evaluative component are more easily available for positive compared to negative thick terms. When a speaker calls someone courageous, but immediately cancels the typically communicated evaluation, participants are likely to start thinking of situations in which being courageous is not a good thing. This seems to be the case when a person is *too* courageous. A similar case can hardly be construed for negative thick concepts: being too rude or too cruel is not deemed positive. Consequently, a sentence with a positive thick term might be considered less contradictory, as participants can think of cases in which the sentence applies to possible situations. The effect of valence could therefore be reinterpreted as a pragmatic effect resulting from the particular experimental design we used. In order to investigate this possibility, we need to change our design such that attempts to make sense of our target sentence are less likely triggered. We opted for a relatively simple design making use of the contrastive word "but". While "but" and "and" are truth-conditionally equivalent, only "but" conventionally implicates a contrast between the conjuncts. Statements like "What Tom did was courageous but good." are presumably less likely to cause people to imagine possible counterexamples, especially because we asked people to merely state how natural the statement sounded to them, not whether that person makes a contradictory claim. Consequently, advocates of this explanation will make the following prediction: If the effect of valence can be explained by the differential availability of counterexamples, using a design that prevents thinking about counterexamples makes the effect disappear.

## Methods

**Participants** 220 participants were recruited. 8 participants were excluded because they were not native speakers of English. Of the remaining 212 participants that were included in the analysis, there were 50.5% female, 0.5% did not identify, 49.0% male with $M_{Age} = 38.25$.

**Design, Procedure, and Materials** We used a 2×1 between-subject design with the independent variable *Valence* (Positive; Negative) and the dependent variable *Naturalness*. All participants were presented with the following vignette: Please suppose that Sally said the following sentence about Tom's behavior: "What Tom did was [thick term], but good/ bad." (E.g., "What Tom did was courageous, but good", What Tom did was rude, but bad"). Participants were then asked "To what extent does Sally's statement sound odd or natural to you?" People's responses were recorded on a 9-point Likert scale with '1' labelled as "very odd", and '9' labelled as "very natural" for the same eight thick terms used before.

## Results

The results are displayed in Figure 4. A 2×1 Anova for naturalness ratings with the independent factor *Valence* revealed that positive thick terms were rated to sound significantly more natural ($M = 3.81$) than negative thick terms ($M = 2.71$), $F(1, 210) = 13.92$, $p < 0.001$.

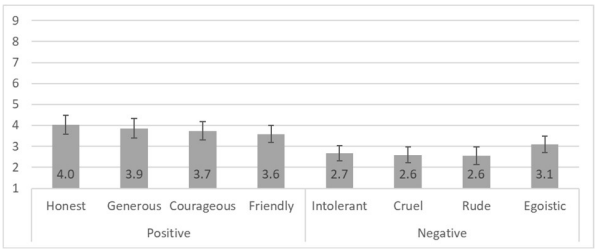


Figure 4: Participants' mean naturalness ratings for all 8 items. Error bars indicate the standard error around the mean.

## Discussion

Study 3 shows two things. First, the potentially differential availability of counterexamples for positive and negative thick terms is unlikely to account for the effect of valence we recorded in Study 1. Experiment 3 was designed to reduce the likelihood to think of possible ways in which the evaluative component of a thick term might be cancelled. Still, the same effect was found. Second, the results of Study 3 provide independent evidence that valence has an important impact on people's reasoning with thick concepts. Having used a different method to examine the (in-)separability of the descriptive and the evaluative component of thick terms, we collected additional data indicating that it is harder for people to disentangle the badness from negative thick concepts than the goodness from positive thick concepts. In future work we plan to conduct a more extensive version of this study to compare the mean ratings for thick concepts with ratings involving semantic entailment and pragmatic implicature.

## Study 4: Differences in social norms

The third explanation suggests that the effect of valence can be accounted for by differences in the norms guiding

evaluative language. Accordingly, social norms more strongly prohibit the use of negatively evaluating language (compared to positive language), unless the evaluation is absolutely intended. Given the substantial evidence on the effect of norm violations on non-normative judgments, differences in contradiction ratings could be explained in this way. We tested the norms underlying evaluative language in this experiment. We believe that advocates of the social norm explanation make the following prediction: People give higher impermissibility ratings when negative thick terms are used non-evaluatively, compared to positive thick terms.

### Methods

**Participants** 198 participants finished the survey and 8 participants were excluded because they were not native speakers of English. 190 participants were included in the analysis (48.9% female, 0% non-binary, 51.1% male; $M_{Age}$ = 38.24)

**Design, Procedure, and Materials** We implemented a 2×1 between-subject design with the independent variable *Valence* (Positive; Negative) and the dependent variable *Impermissibility*. Participants first read one of the thick term sentences from Study 1. Afterwards they answered the question "How much do you disagree or agree to the following sentence? If Sally doesn't mean to say something negative/positive, she should not have used the word "[thick term]" in the first place." on a 9-point Likert item, reaching from "1 = fully disagree" to "9 = fully agree".

### Results

The results of Study 2 are summarized in Figure 5.

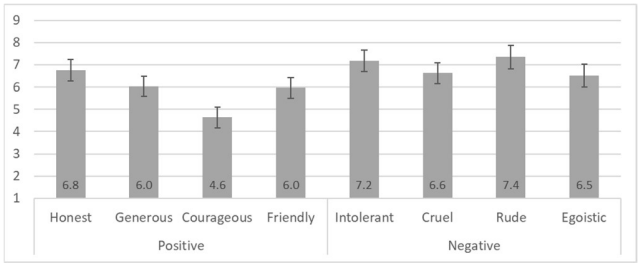


Figure 5: Participants' mean impermissibility ratings for all 8 items. Error bars indicate the standard error around the mean.

We conducted a 2×1 Anova for impermissibility ratings with the independent factor *Valence*. There was a significant main effect of *Valence*, $F(1, 189) = 9.96$, $p < .01$, such that negative thick terms were given higher impermissibility ratings ($M = 6.9$) than positive thick terms ($M = 5.8$).

### Discussion

In line with the prediction stated above, people did give higher impermissibility ratings when a speaker used a negative thick term, yet did not intend to communicate its evaluation, compared to positive thick terms. This effect suggests that the norms guiding the permissible use of thick terms differ, such that using a negative thick term non-evaluatively constitutes a norm-violation. This difference, given our initial reasoning, appears to be a promising candidate to explain differences in contradiction ratings.

## General Discussion

How do thick terms carry their evaluative force? Are these evaluative aspects conveyed by means of conventional implicature – as many Pragmatic Separabilists argue –, or is the evaluation part of the semantic meaning of thick concepts, connected via semantic entailment? In this paper, we presented the results of the first set of empirical studies on thick concepts focusing both on the relation between the evaluative and descriptive aspects of thick concepts, as well as on possible differences in the way positive and negative thick concepts work.

Study 1 demonstrated that the evaluative component of thick concepts is significantly harder to cancel compared to the conversational implicatures we tested. Additionally, contradiction ratings for thick terms were *not* significantly different from ratings for semantic entailments. These results put pressure on the Pragmatic Separabilist View. However, the effect should be taken with caution, as a higher statistical power might have resulted in a statistically significant difference. In addition, we should note, that we selected only particularized conversational implicatures, whose force depends on the given context. Had we used generalized implicatures or other kinds of implicatures that are harder to cancel, the results might have been different.

Study 1 also revealed an effect of valence on contradiction ratings. For positive thick terms, contradiction ratings were significantly lower compared to negative thick terms as well as semantic entailments. This effect of valence is hitherto unknown and has not been predicted by any of the various accounts of thick concepts. In fact, such an effect provides further hope for the pragmatist, as a more complicated picture seems to be emerging.

We then put forward three potential explanations of the observed effect of valence. Our follow-up studies suggest that the effect can neither be explained by different evaluative intensities of our stimuli, nor by differences in the availability of possible counterexamples. Instead, social norms seem to play a crucial role in the application of positive and negative thick terms. Ascribing a negative thick term without intending to communicate its evaluation is considered less acceptable compared to positive thick terms. Consequently, an agent who does not want to communicate a negative evaluation *should* not use a negative thick term in the first place, as she is perceived to violate a social norm guiding evaluative language.

Some caution is required at this stage. The evidence presented in this paper is the first of its kind. It is quite plausible that we have not explored all reasonable options or have dismissed too quickly alternative explanations. The effect of valence, however, seems to be a rather robust phenomenon: Two different studies (Study 1 and Study 3) using different methods yielded a similar outcome. If social

norms have a decisive effect on the differential applicability of thick concepts, then this will have important consequences for the philosophical and linguistic debates on thick concepts.

Given our knowledge about the effects of norm violations on a variety of non-normative concepts, it seems quite plausible to assume that norm-violations can affect contradiction ratings differently. One might even go so far in arguing that the recorded effect of people's contradiction ratings for negative thick concepts is increased due to a negativity bias. This bias will arguably be weakened when it comes to positive thick terms, suggesting that the results we collected for positive thick concepts give us a less distorted view of the relation between the descriptive and evaluative components of thick concepts. On the other hand, it is likely that the use of positive thick terms is also subject to social norms, even if less stringent. Disentangling the effects of social norms from the mere linguistic aspects will be a serious challenge yet to be overcome.

## References


Alicke, M. (1992). Culpable causation. *Journal of Personality and Social Psychology*, 63(3), 368–378.

Beebe, J., Buckwalter, W. (2010). The Epistemic Side-Effect Effect. *Mind & Language*, *25*(4), 474–498.

Blackburn, S. (1992). Through Thick and Thin. *Proc. of the Aristotelian Society*, supplementary volume 66: 284–99.

Eklund, M. (2011). What Are Thick Concepts? *Canadian Journal of Philosophy,* 41(1): 25–49.

Elstein, D., Hurka, T. (2009). From Thick to Thin: Two Moral Reduction Plans. *Canadian Journal of Philosophy.* 39(4): 515–36.

Grice, P. (1989). Logic and Conversation. In P. Grice (1989), Studies in the Way of Words (pp. 22-40). Cambridge, Mass.: Harvard University Press.

Hare, R.M. (1952). *The Language of Morals*. Oxford: Clarendon Press.

Kirchin, S. (2010). The Shapelessness Hypothesis. *Philosophers' Imprint*, 10(4), 1-28.

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, 63(279), 190–194.

Kyle, B. (2013). How Are Thick Terms Evaluative? *Philosophers' Imprint*, 13(1).

Putnam, H. (2002). *The Collapse of the Fact/Value Dichotomy and Other Essays*, Cambridge, MA: Harvard University Press.

Reuter, K., Baumgartner, L., & Willemsen, P. (ms). Tracing Thick Concepts Through Corpora.

Reuter, K., Löschke, J., Betzler, M. (forthcoming). What is a colleague? The descriptive and normative dimension of a dual character concept. *Philosophical Psychology*.

Roberts, D. (2011). Shapelessness and the Thick. *Ethics*, 121(3): 489–520.

Stevenson, C. L. (1938). Persuasive Definitions. *Mind*, 47(187), 331-350.

Sytsma, J., Bluhm, R., Willemsen, P., Reuter, K. (2019). Causal Attributions and Corpus Linguistics, In E. Fischer and M. Curtis (eds) *Methodological Advances in Experimental Philosophy*. London: Bloomsbury.

Väyrynen, P. (2019). Thick Ethical Concepts. In E. N. Zalta (Edit), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/sum2019/entrieshick-ethical-concepts/

Williams, B. (1985). *Ethics and the Limits of Philosophy*, Cambridge, MA: Harvard University Press.

Willemsen, P., Kirfel, L. (2019). Recent Empriical Work on the Relationship Between Causal Judgments and Norms. *Philosophy Compass*, 14(1).

Zakkou, J. (2018). The Cancellability Test for Conversational Implicatures. *Philosophy Compass*, 13(12).

Zakkou, J. (ms). Conventional Evaluativity.


## Appendix

### Instructions given to participants in Study 1

**Contradictions occur when a person says two things that exclude each other**. The easiest and most obvious way to contradict yourself is to say *"This thing is round and also not round"*. A thing cannot be *round* and *not round* at the same time. But sometimes contradictions are a bit less obvious. For instance, *"This thing is round and it has edges"*. You need to know that round things don't have edges to see that **the speaker contradicts himself.**

In contrast, imagine the following sentence: *"Dave is tall but he is also very thin"*. **This sentence is not a contradiction at all**, even though it describes a contrast. Think about what it means to be tall and about what it means to be very thin. It is perfectly ok to be tall and very thin at the same time.

Here is a last example: *"Joanna is 20 years old, but I don't mean to say that she is under 30"*. **This statement clearly is a contradiction**. A speaker who says that Joanna is 20 cannot say that he does not mean that she is under 30. If you know what it means to be 20, you know that "being 20" means "being under 30".

### Stimulus Sentences used in Study 1

(X indicates a first name and was varied throughout the stimulus sentences)

**Positive Thick Terms**

- X's behavior last week was [positive thick term], but by that I am not saying something positive about X's behavior that day.

**Negative Thick Terms**

- X's behavior last week was [negative thick term], but by that I am not saying something negative about X's behavior that day.

**Semantic Entailment**

- Sven is running, but by that I am not saying that he is moving.
- Ann is a widow, but by that I am not saying that the person she was married to died.
- This is a couch, but by that I am not saying that this is a piece of furniture.
- This is a lake, but by that I'm not saying that it consists of water.

**Conversational Implicature**

- I am hungry, but by that I am not saying that we should get something to eat.
- It is dark in here, but by that I am not saying we should turn on the lights.
- This chocolate is good value-for-money, but by that I am not saying that we should buy it.
- There is the door, but by that I am not saying that I want you to leave.