Separability and the Effect of Valence
An Experimental Study on Thick Concepts

Pascale Willemsen & Kevin Reuter
Separability and the Effect of Valence
An Empirical Study of Thick Concepts

Pascale Willemsen (Pascale.Willemsen@unh.ch)  Kevin Reuter (Kevin.Reuter@unh.ch)
Institute of Philosophy, University of Zurich, Zürich, Switzerland

Abstract
Thick concepts like courage and resolve are at the heart of a variety of debates in linguistics, philosophy of language, and psychology. Central to these debates is the question of how the descriptive and evaluative components of thick concepts are related. In this paper, we present an empirical study designed to investigate whether the evaluative component of thick concepts is communicated by means of semantics or pragmatics. Our results are contrary to the common view: they also reveal an effect of valence, indicating that people rate different aspects of positive and negative thick terms. These findings were consistent with our hypothesis. We conclude that the effect of valence is best accounted for by differences in the social norms guiding evaluative language.

Keywords: Thick concepts, moral judgments, experimental methods

Introduction
Philosophers and linguists usually distinguish two types of evaluative concepts: "thin" and "thick" concepts (Kahneman, 2011; Vroon, 2019). Thin concepts evaluate an object as, for instance, "permissible", "right", "wrong", "good", "bad", or "thankworthy", but do not entail in what way the object is right or wrong. In contrast, evaluative thick concepts are incoherent of lying as wrong, they convey information about why they think so. The speaker might think that people have a right to be told the truth, but it means truthfulness, etc. Thick concepts do not merely evaluate, they also provide us with information on why the entity is evaluated. Typical examples are evaluative thick terms, such as "trait", "value", "consequence", or "unfairness". Being an agent concept allows them in particular for being willing to take risks - "traitors" also sacrifice willingness to take risks yet receive 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 are connected. According to one group of researchers, the evaluative component of a thick term is part of its semantic meaning. According to another view, the evaluation is part of the meaning but conveyed through pragmatic means. Arguments favoring this position hardly rely on linguistic intuitions about how thick terms can be used. Such intuitions often circle around the question of whether the evaluation of a thick term can be canceled without yielding contradiction. Whether this is possible is not only relevant for the inference debates about thick terms. Assumptions about the nature of thick terms provide the representativeness coreferences in metaphysical 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 test.

Separating the Evaluative from the Descriptive
Thick terms, such as "honest", "nasty", "cruel", or "deceitful", do not only evaluate an entity; they further describe in some way what entity is evaluated as positive or negative. This descriptive richness is what distinguishes thick terms from thin concepts. And it is the evaluative component that distinguishes thick from thin descriptive concepts. Some philosophers explain this descriptive richness by assuming that thick terms are a combination of thin evaluations plus some description. According to this position, thick terms express the same basic attitude, a speaker can have. Add some description to this attitude and the result is a thick term. According to this view, thick terms can be, at least in principle, divided into two distinct components, namely the evaluative and the descriptive. Hence, the thick term can be, for instance, by pragmatic means, for instance, by pragmatic means. Conventional implicatures are part of the semantic meaning and need to be inferred beyond what is literally said (Grice, 1989). By saying that an agent is nice, we convey some descriptive properties, and one further communicates the implicature that the agent is bad in terms of having these properties. Moreover, as other conventional implicatures, the negative evaluation can be canceled with a different attitude. Therefore, a speaker who states "What Tom did was wrong", but by that I am not saying something negative about Tom", makes a different statement.

The contrary position, the Separability view, denies that thick terms can be further divided into distinct components and holds that thick terms are causal (Williams, 1985; Parsons, 2002; Fox, 2010; Roberts, 2011). Those who support this independence typically argue that thick terms cannot be further reduced because their evaluation is part of the
Agenda

1. What are thick concepts?
2. Why are they philosophically interesting?
3. An empirical study on cancellability
4. Three potential explanations of a surprising result
5. Lots of open questions
Thick and Thin Terms & Concepts

- **Descriptive**: round, green, short-sighted
- **Thin**: right, wrong, good, bad
- **Thick**: discreet, cautious, industrious, chaste, just, grateful
Thick Terms & Concepts

- **Epistemology**: open/close-minded, gullible, critical, reflective, rational, dumb
- **Aesthetics**: banal, gracious, creative, elegant, dainty
- **Ethics**: friendly, helpful, generous, cold-hearted, cruel
Note: Things are not so simple

- If it is true that "ought" implies "can", it seems that even ought is thick in some sense
Thick vs. Thin

- **Allan Gibbard:** thick terms and concepts praise or condemn an action as having certain properties
- **Jonathan Dancy:** thick terms and concepts have an evaluative dimension which is inseparable from their descriptive dimension
- **Bernard Williams:** thick terms and concepts are both 1) world-guided and 2) action-guiding
What’s the Connection?

Inseparabitists

- Thick terms and concepts are inherently evaluative
- The evaluation is part of the semantic meaning
- It is impossible to separate the evaluative from the descriptive


Separabilists

- Thick terms and concepts are descriptive, yet they presuppose, implicate, etc to convey evaluative content
- The evaluation is not part of the meaning
- It is possible to separate the evaluative from the descriptive

A Thick Job in Moral Philosophy

- Thick concepts have been used to argue against
  - Non-Cognitivism
  - The (ontological) fact/value distinction
  - Ethical reductionism
- And to argue for
  - Moral Objectivism
  - The nature of reasons for actions (Internalism vs. Externalism)
- Are essential for Virtue Ethics and other character-focused ethics
A Thick Job in Moral Philosophy

- Thick concepts have been used to argue against
  - Non-Cognitivism
  - The (ontological) fact/value distinction
  - Ethical reductionism
- And to argue for
  - Moral Objectivism
  - The nature of reasons for actions (Internalism vs. Externalism)
- Are essential for Virtue Ethics and other character-focused ethics
Cognitivism vs. Non-Cognitivism

- **Cognitivism:**
  - Moral concepts are genuine concepts
  - They pick out a real feature of the world
  - Values are in the world
  - SELFISH picks our selfishness, HONEST picks out honesty, etc.
  - Moral sentences can be true or false

*18.02.2020*
Cognitivism vs. Non-Cognitivism

- **Non-Cognitivism**
  - Deny that moral concepts are genuine concepts
  - Moral concepts do not represent the world
  - Yet, they are a convenient tool for communication
  - Values are not in the world
  - Moral sentences cannot be true or false
Cognitivism vs. Non-Cognitivism

![Diagram showing a decision tree for the distinction between Cognitivism and Non-Cognitivism]

- **Cognitivism** (Yes/Yes)
  - Do indicative moral sentences represent reality in the way that ordinary sentences do?
  - Yes
- **Error Theory**
  - What do these sentences predicate of their objects?
  - False or presuppose falsity.
- **Minimal Realism**
  - Are many of our claims predicating moral properties of objects roughly true?
  - Yes.
- **Fictionalism**
  - What do these sentences predicate of their objects?
  - True.
  - Which properties do moral terms predicate?
    - Sui generis moral properties.
  - Which natural properties?
    - Natural Properties.
  - Supernaturalism
    - Being commanded by or approved of by God.
  - Moral Functionalism
    - Properties that play certain roles.
  - Cornell Realism
    - Properties that we have epistemic access to via empirical methods.
  - Subjectivism & ideal observer theories
    - Relations to the subjective responses of agents and observers.

18.02.2020
Theories in metaethics presuppose the correctness of the semantic or the pragmatic understanding of thick concepts.

The use of language seems like an obvious field of empirical research.

Testing whether this is indeed correct will have significant impact on those debates.
Pre-Registered Experiment
Methods

- Pre-pregistration: osf.io/9pbq2
- 200 participants recruited on Mturk, 50 subjects per condition
- \(4 \times 1\) between-subject design with
  - IV: Relation
    - Semantic Entailment
    - Conversational Implicature
    - Thick Positive Concepts
    - Thick Negative Concepts
  - DV: Contradiction
Thick Positive

honest,
generous,
courageous,
friendly

Please imagine that Sally said the following sentence:

"Kathy’s behavior last week was friendly, but by that I am not saying something positive about Kathy’s behavior that day."

Does Sally contradict herself?

1 - definitely not
2
3
4
5
6
7
8
9 - definitely yes

Please imagine that Sally said the following sentence:

"This is a couch, but by that I am not saying that this is a piece of furniture."

Does Sally contradict herself?

1 - definitely not
2
3
4
5
6
7
8
9 - definitely yes

Thick Negative

intolerant,
rude,
cruel,
egoistic

Please imagine that Sally said the following sentence:

"Amy’s behavior last week was egoistic, but by that I am not saying something negative about Amy’s behavior that day."

Does Sally contradict herself?

1 - definitely not
2
3
4
5
6
7
8
9 - definitely yes

Please imagine that Sally said the following sentence:

"I am hungry, but by that I am not saying that we should get something to eat."

Does Sally contradict herself?

1 - definitely not
2
3
4
5
6
7
8
9 - definitely yes

Semantic Entailment

Run,
Widow,
Couch,
Lake

Conversational Implicature

Hungry,
Dark,
Door,
Chocolate
Predictions

- There is a significant difference between SE and CI, such that Contradiction ratings are significantly higher for SE than for CI.

If the Semantic View is correct,
- There is no significant difference between SE and TC

If the Pragmatic View is correct,
- There is no significant difference between CI and TC
Results
## Results

<table>
<thead>
<tr>
<th></th>
<th>SE</th>
<th>Positive Thick</th>
<th>Negative Thick</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run</td>
<td>6.88</td>
<td>5.40 ± 1.0</td>
<td>6.98 ± 1.0</td>
<td>4.43</td>
</tr>
<tr>
<td>Widow</td>
<td>7.56</td>
<td>5.94 ± 1.0</td>
<td>7.67 ± 1.0</td>
<td>3.67</td>
</tr>
<tr>
<td>Couch</td>
<td>7.38</td>
<td>5.75 ± 1.0</td>
<td>7.82 ± 1.0</td>
<td>3.43</td>
</tr>
<tr>
<td>Lake</td>
<td>6.84</td>
<td>6.15 ± 1.0</td>
<td></td>
<td>3.43</td>
</tr>
<tr>
<td>Honest</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generous</td>
<td>5.94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Courageous</td>
<td>5.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friendly</td>
<td>6.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intolerant</td>
<td>6.98</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rude</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cruel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eggistoc</td>
<td>7.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hungry</td>
<td>4.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dark</td>
<td>3.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chocolate</td>
<td>3.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Door</td>
<td>3.43</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results

- SE: 7.2
- TPT: 5.8
- TNT: 7.2
- CI: 3.7

*p = .062

***
Predictions

- There is a significant difference between SE and CI, such that Contradiction ratings are significantly higher for SE than for CI.

If the Semantic View is correct,
- There is no significant difference between SE and TC

If the Pragmatic View is correct,
- There is no significant difference between CI and TC
Discussion

We must not overestimate the data!

- We did only test for Conversational Implicatures, not Conventional Implicatures
- We only looked at a small collection of terms
- Do laypeople’s contradiction ratings track semantic contradiction?
- Did laypeople make additional assumptions about the speaker’s intentions?
Discussion

Why do Positive and Negative Thick Terms behave differently?
Results

<table>
<thead>
<tr>
<th></th>
<th>SE</th>
<th>TPT</th>
<th>TNT</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7.2</td>
<td>5.8</td>
<td>7.2</td>
<td>3.7</td>
</tr>
</tbody>
</table>

Significance: ***
Discussion

Why do Positive and Negative Thick Terms behave differently?

1. Differences in moral evaluation
2. Availability of counterexamples
3. Differences in the social norms guiding evaluative language
Explanation 1: Differences in moral evaluation

- The terms that we tested do not only differ in the valence (positive or negative)
- They might also differ in how strongly they evaluate
- The stronger the evaluation, they more contradictory does it seem to cancel it

Prediction:
*Negative terms should be rated more negatively than positive terms are rated positively.*
Explanation 1: Results

- 400 participants recruited on Mturk, 25 subjects per condition
- 2 × 2 between-subject design
  - Question (Behavior, Sentence)
  - Valence (Positive, Negative)
- All native speakers of English
Explanation 1: Results

- 400 participants recruited on Mturk, 25 subjects per condition
- 2 × 2 between-subject design
  - Question (Behavior, Sentence)
  - Valence (Positive, Negative)
- All native speakers of English
Explanation 1: Results

Prediction:
Negative terms should be rated more negatively than positive terms are rated positively.

- 400 participants recruited on Mturk, 25 subjects per condition
- 2 × 2 between-subject design
  - Question (Behavior, Sentence)
  - Valence (Positive, Negative)
- All native speakers of English
Explanation 2: Availability of Counterexamples

- Maybe the results are an artefact of the experimental design
- In one condition, reasoning about counterexamples is triggered

Please imagine that Sally said the following sentence:

"Kathy’s behavior last week was friendly, but by that I am not saying something positive about Kathy’s behavior that day."

Does Sally contradict herself?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>definitely not</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

Please suppose that Sally said the following sentence about Tom’s behavior:

“What Tom did was courageous.”

To what extent do you disagree or agree with the following claim:

Sally committed herself to a positive evaluation of Tom’s behavior.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>definitely yes</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
Explanation 2: Availability of Counterexamples

- You are good, but for the wrong reasons
  - Courageous: show-off, want to make others look bad
  - Friendly: manipulative, devious, not trustworthy

- You are overdoing it
  - Honest: mean, offensive, indiscreet
  - Courageous: reckless, foolhardy, putting others at risk

- The consequences render your action blameworthy
Explanation 2 – Positive Thick Concepts

- While courageous acts are normally associated with positive things, it's possible to do something negative that is also courageous.
- The person could have faked it (friendly).
- Mostly, 'friendly' means nice and good things. But if you were friendly to, say, a Nazi...is that a good thing? Mostly not, probably.
- Context may present the courage displayed as unwise.
- Friendliness is not necessarily positive if the person exhibiting it has nefarious intent.
- Honesty isn't always good. He could have been honest and saying something like, "I find your face to be really ugly." It could be honest, but it wouldn't be a good thing to say, and people wouldn't think he was a good person for his honesty.
- Courageous behavior might mean they are being a bit more forward when they should really hold back more.
- Because you can be honest in a negative way. (sometimes the truth hurts, etc..)
- Jim could have been honest about something that should not have been said to be polite. In this case, it is not really positive that he was honest.
Explanation 2 – Negative Thick Concepts

- You should not be cruel to anyone in any circumstances. That is inherently negative.
- Rudeness is always negative. There's no positive way to be rude.
- being cruel at any time is a negative thing
- being rude is generally considered negative so this would be a contradiction
- In my world, being rude is not a positive behavior and I could think of no world in which it would be. I may lack sufficient imagination.
- Intolerant is a negative adjective.
- I'm not sure how it's possible to be intolerant and have that not be a negative thing
- Well rude and negative is basically the same thing.
- Intolerance is a negative thing, so you can't say that someone is intolerant and then claim you weren't saying something negative about them.
Explanation 3: Availability of Counterexamples

- If the effect arises because of a specific experimental design, using a different design should make the effect disappear.
- To test this, we used an even simpler design.
- We contrasted a thick term with its evaluation, using “but“.

Prediction:
Acceptability ratings for thick terms should be equally low for thick positive and negative terms.
Explanation 2: Results

Prediction: Acceptability ratings for thick terms should be equally low for thick positive and negative terms.

Please suppose that Sally said the following sentence about Tom's behavior:

“What Tom did was honest but good.”

To what extent does Sally’s statement sound odd or natural to you?
The norms as to when we may say something evaluative differ

- There is no harm in communicating something positive one doesn’t want to commit to
- There is a lot of harm in communicating something negative one doesn’t want to commit to

Two Social Rules:

1. **Positive thick terms may be used** even if the speaker does not mean to evaluate by doing so
2. **Negative thick terms may not be used** if the speaker does not mean to evaluate by doing so
Explanation 3: Social Norms and Moral Language

- If we use a positive thick term, we can cancel the evaluation without contradicting ourselves.
- If we use a negative thick term, we cannot cancel the evaluation without contradicting ourselves.

Prediction:
People give higher “you should not” ratings when negative thick concepts are used non-evaluatively, compared to positive thick concepts.
Explanation 3: Results

Results

- $p = 0.035$
- $p = 0.0014**$

Please imagine that Sally said the following sentence:

"Tom’s behavior last week was rude, but by that I am not saying something negative about Tom’s behavior that day."

How much do you disagree or agree with the following sentence?

If Sally doesn’t mean to say something negative, she should not use the word "rude" in the first place.
Predictions

Results

Discussion

Why do Positive and Negative Thick Terms behave differently?

1. Differences in moral evaluation
2. Availability of counterexamples
3. Differences in the social norms guiding evaluative language
Let’s do more of this!

- This results we got are surprising, given that philosophers discuss TC as a unique class of concepts.
- We should determine whether research on TC needs to treat positive and negative TC separately.
- We should determine what features can make the evaluative component change to neutral.
Get in touch!
PascaleWillemsen.com
Pascale.Willemsen@philos.uzh.ch

Special thanks to:
Bianca Cepollaro
Matti Eklund
Markus Kneer
Iza Skoczen
Alex Wiegmann
Pekka Värynen