A First Look at Human Values-Violation in App Reviews

Humphrey O. Obie [1] Waqar Hussain [1] Xin Xia [1] John Grundy [1] Li Li [1] Burak Turhan [2] Jon Whittle [3] Mojtaba Shahin [1]

[1]Monash University, Melbourne, Australia

[2]University of Oulu, Oulu, Finland

[3]CSIRO's Data61, Melbourne, Australia

May, 2021

Outline

1. Background
2. Methodology
3. Results

Outline

1. Background
2. Methodology
3. Results

Introduction - Mobile Apps and Human Values

Introduction - Consequences of Values-violation

"AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity" - [1]

Human Values

Definition

Human values are the "guiding principles of what people consider important in life [2]. Examples include achievement, power, tradition, benevolence, security, etc.

- Determine human behaviour and attitude
- Guide for actions and a vehicle for expressing need

Example Scenario

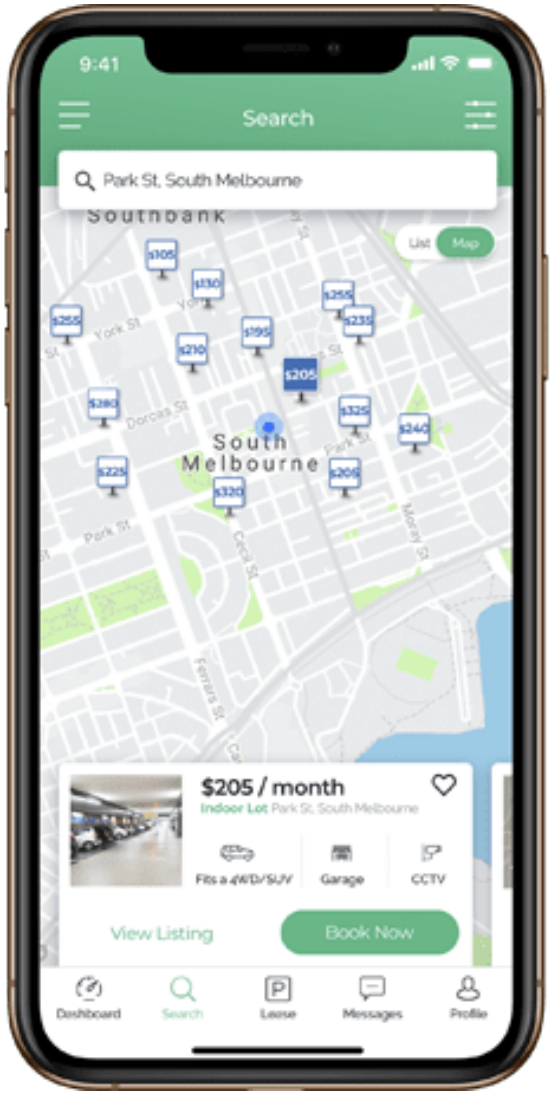

Figure: An example parking app.

Example Scenario

Privacy

There is no way I'm giving your app my location... or access to my photos.

Honesty

...It is unscrupulous about signing you up to a subscription when you are skipping past the in-app ads. It is not made clear once you have subscribed and there is no way of cancelling it through the app.

Research Aim

Aim

To help developers to understand and identify end user values violations in their apps with a view to being able to fix them and avoid them in future.

Research Questions

RQ1 What are the most common human values violations perceived by app users as documented in their mobile app reviews?

RQ2 Which reviews indicating human values violations by a mobile app are the most supported by other end users?

RQ3 Are app description features related to specific human values violations?

Outline

1 Background

2 Methodology

3 Results

Schwartz Theory of Basic Human Values

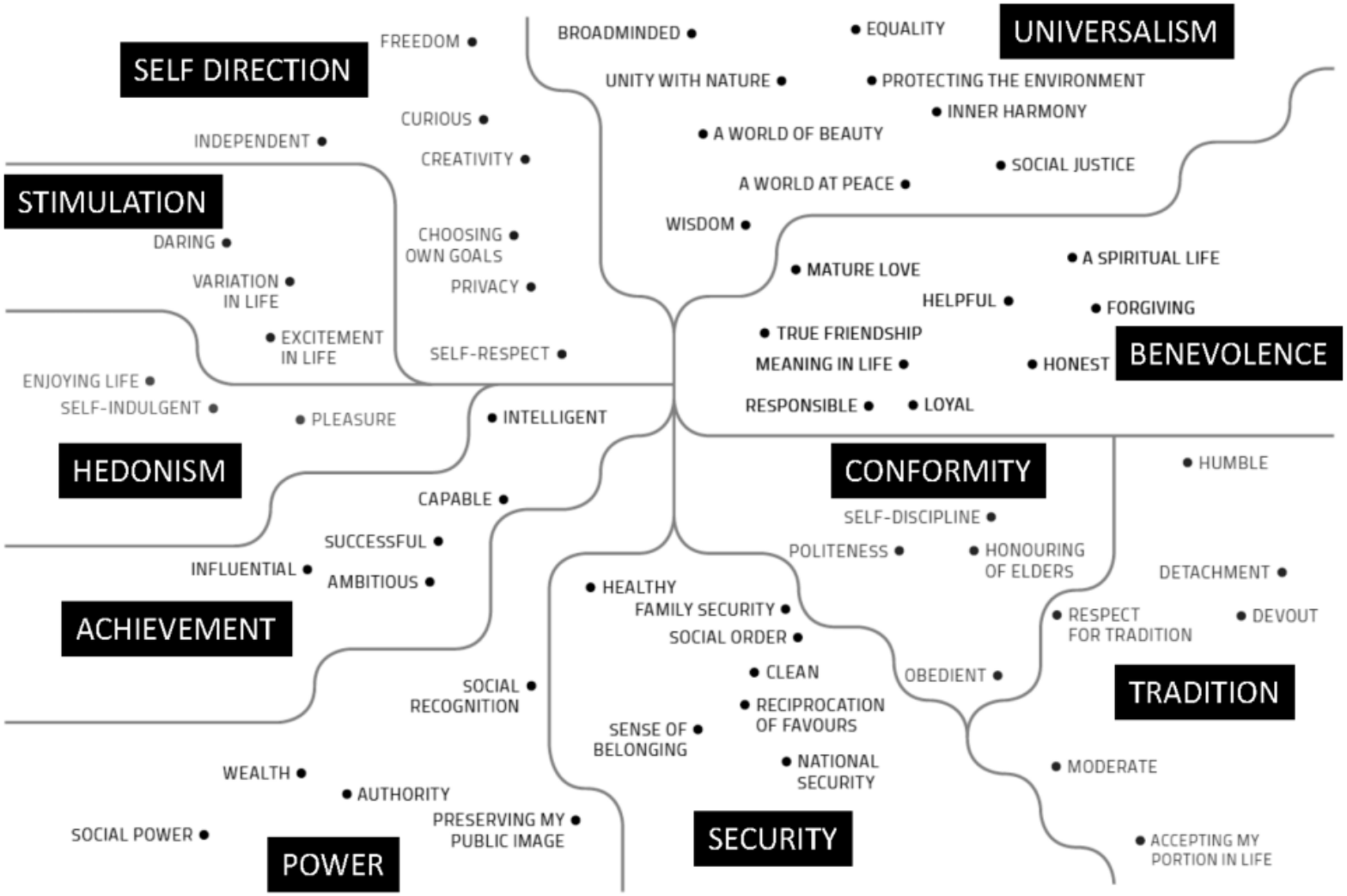

Figure: Schwartz's theory of basic human values [3].

Approach

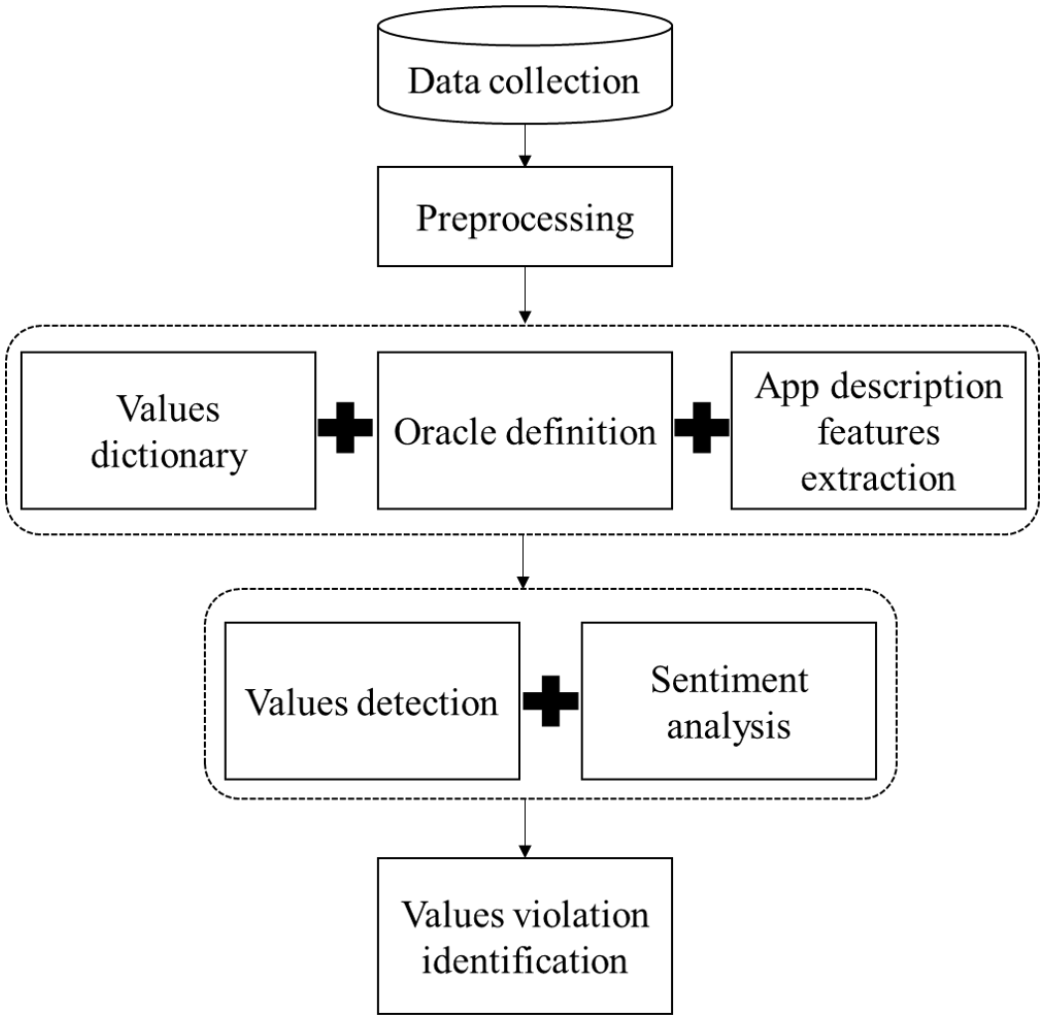

Figure: Overall approach for values violation identification.

Threats to Validity

- Threats to construct validity: definition of truthset
- Threats to internal validity: detection errors
- Threats to external validity: selection of apps

Outline

1 Background

2 Methodology

3 Results

Results

RQ1

What are the most common human values violations perceived by app users as documented in their mobile app reviews?

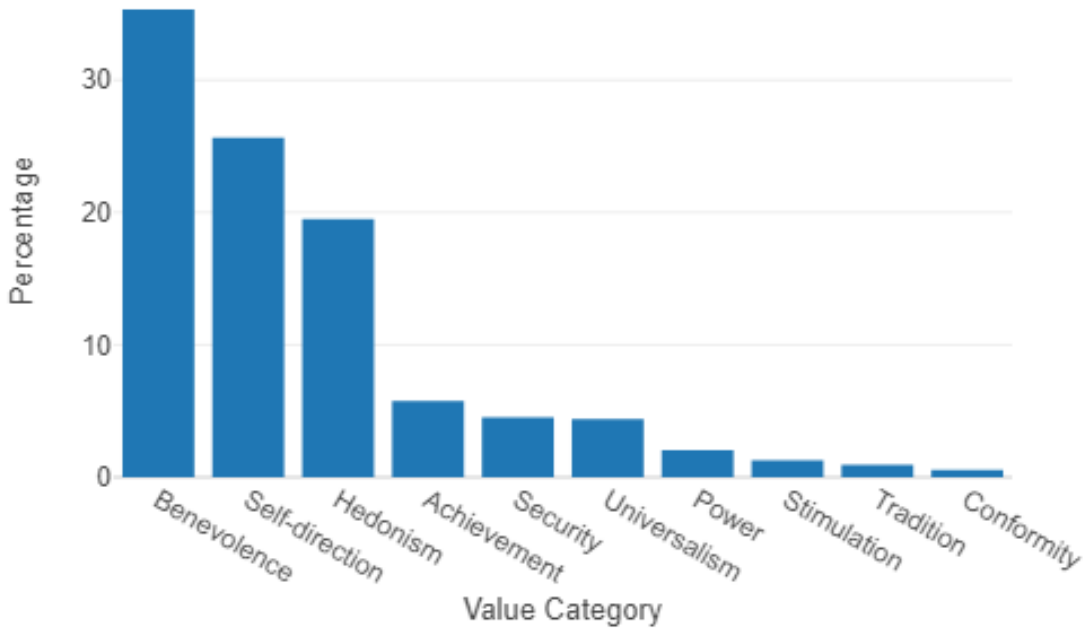

Figure: Percentage of violated values category.

Results

RQ2

Which reviews indicating human values violations by a mobile app are the most supported by other end users?

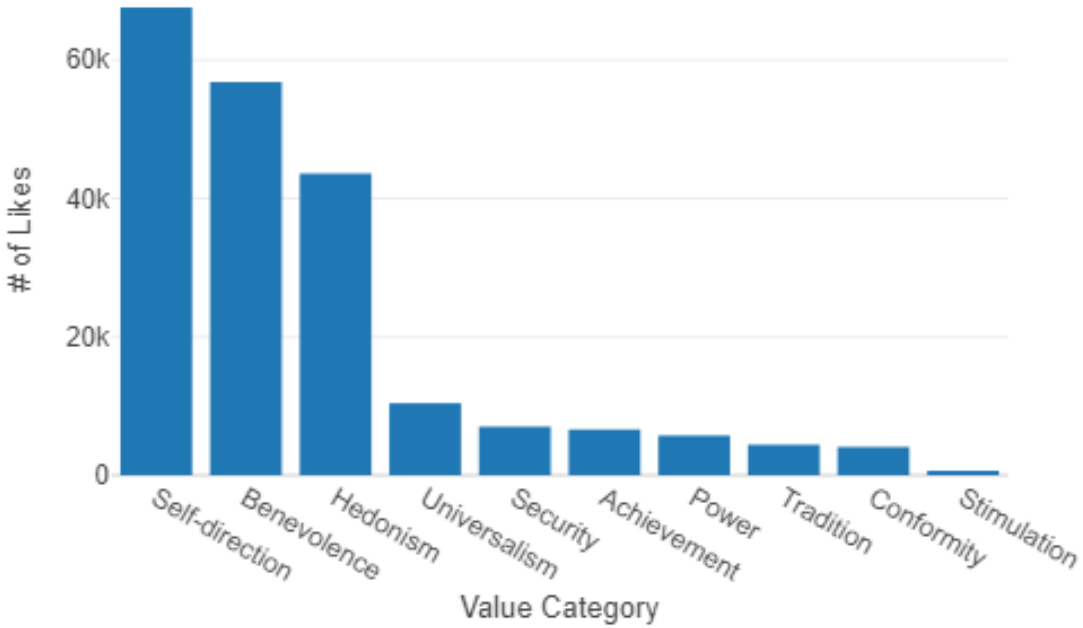

Figure: Number of likes for each violated values category.

Results

RQ3

Are app description features related to specific human values violations?

Table: Sample app features and related value-violations.

Feature	App Name	Value-violation
Add stickers	PicsArt	Pleasure
Set reminders	Any.do	Responsibility
Add review	Tripadvisor	Helpfulness, Honesty
...	...	...

Implications and Recommendations

- Mining values requirements
- Values alignment between stakeholders
- Critical technical practice in (mobile) SE

Summary

- Values are embodied and violated in mobile apps
- Analysis of 22k app reviews from 12 apps
- 26.5% of all reviews contained (perceived) values-violation
- Need for proactive approaches in supporting values in (mobile) SE

References

[1] A. A. Principles, "Asilomar AI principles," [Online.] Available: https://futureoflife.org/ai-principles/, accessed July, 2020.

[2] A.-S. Cheng and K. R. Fleischmann, "Developing a meta-inventory of human values," in *ASIS&T*, vol. 47, 2010.

[3] S. H. Schwartz, "Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries," *Advances in Experimental Social Psychology*, vol. 25, 1992.