Cross-Cultural Privacy Prediction

Yao Li
Univ. of California, Irvine
yao.li@uci.edu

Alfred Kobsa
Univ. of California, Irvine
kobsa@uci.edu

Bart P. Knijnenburg
Clemson University
bartk@clemson.edu

M-H. Carolyn Nguyen
Microsoft Corporation
cnguyen@microsoft.com

Motivation:

- Previous privacy research focuses on narrow population from a single country;
- As organizations run their business globally, they need to customize their data collection strategies to match the privacy preferences of users from different cultural backgrounds.

Research Questions:

1. Will the addition of culture-related features improve the privacy predictions?
2. Which measurement of culture yields better prediction accuracy: country of residence, native language, or Hofstede's cultural dimensions?
3. Do contextual or attitudinal factors influence users' privacy decisions differently in different culture?
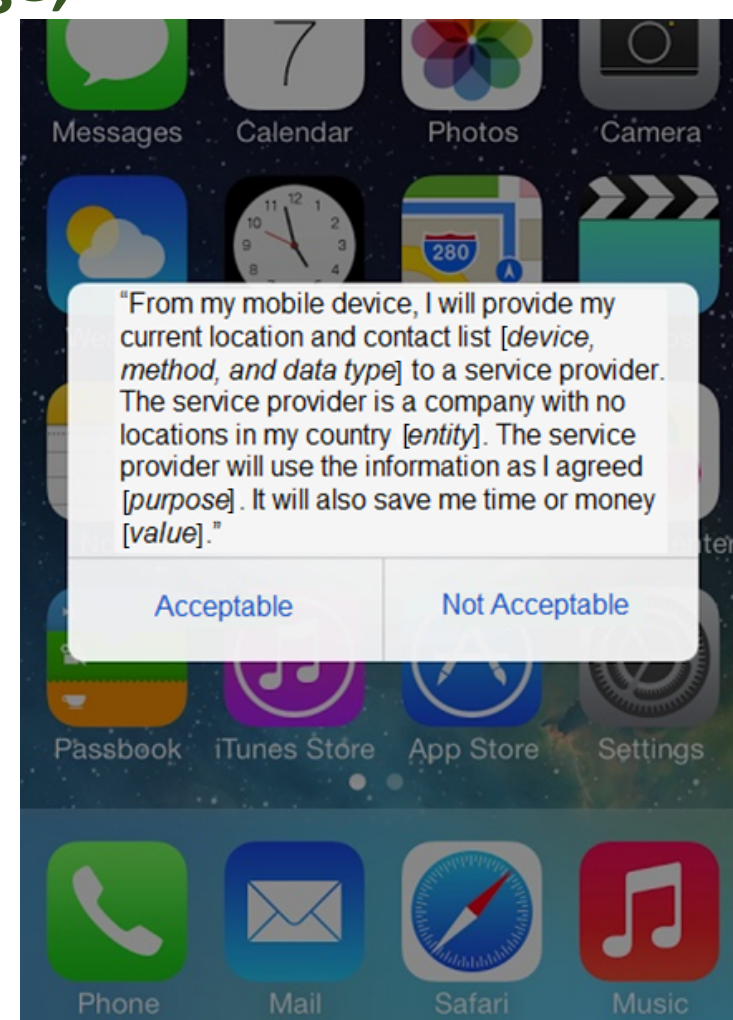
Data Collection:

9625 online survey responses from eight countries, which collect:

1. Participants' **demographics**, **country**, **language,** and **attitudes** wrt. information privacy.
2. Participants' **disclosure intent** in 6 different scenarios, each containing a permutation of 6 contextual variables:
   - **Type of personal data being collected**
   - **Collection method**
   - **Device from which data is collected**
   - **Entity that is collecting the data**
   - **Usage purpose of the data**
   - **Value exchange from data collection**

Cultural dimensions for countries (Geert Hofstede)

- **Power Distance**
- **Individualism**
- **Uncertainty Avoidance**
- **Indulgence**
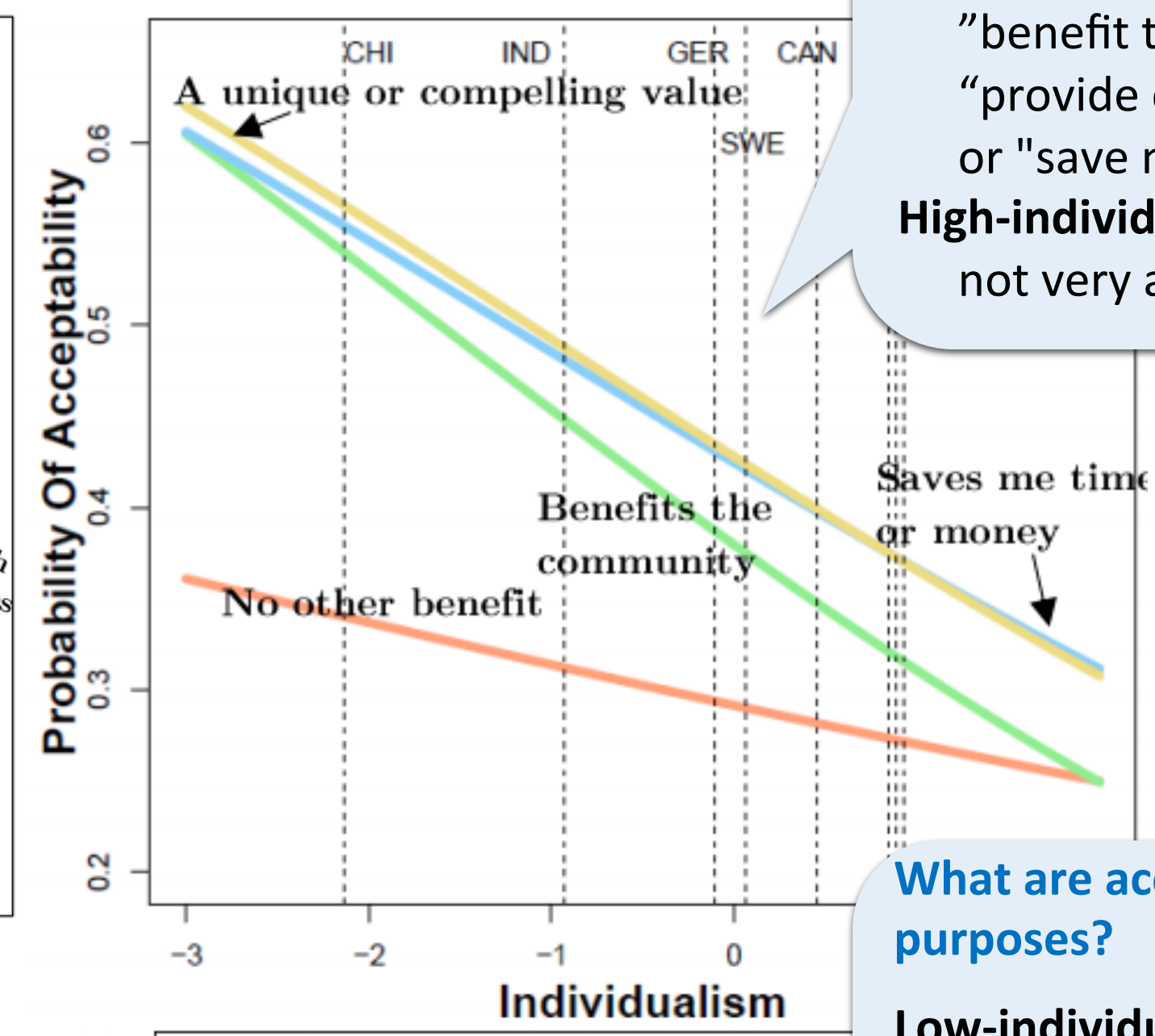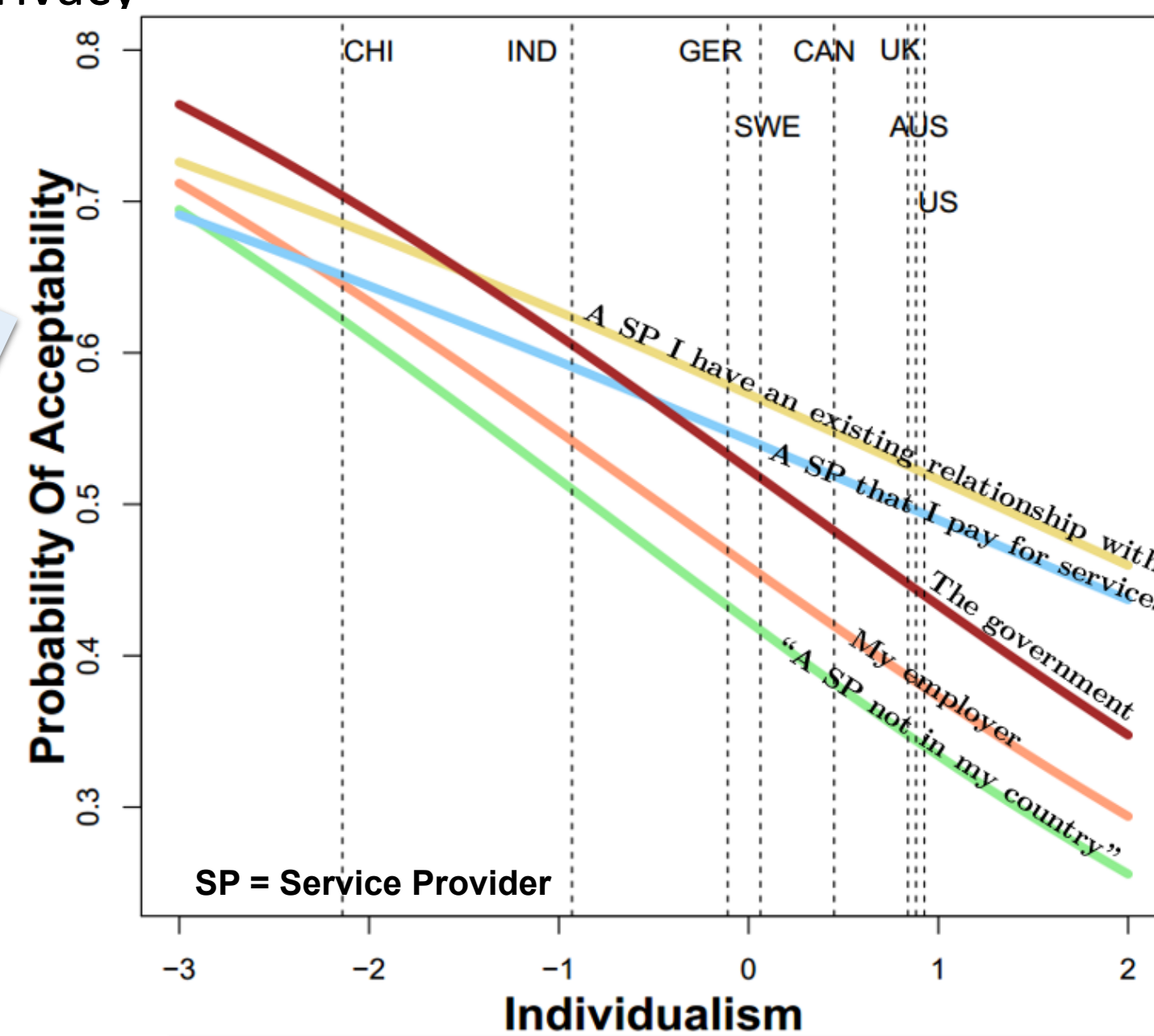- **Masculinity**
- **Pragmatism**

Independent Variables
Dependent Variable

Methods:

- Supervised Machine Learning with Decision Trees
- Generalized Linear Mixed-Effects Regression

Accuracies and F measures from Decision Trees

Baseline Feature	Culture-Related Feature	Acc. (%)	F
demographic + contextual + attitudinal		76.1580	0.722
	+country	76.3545	0.730
	+language	76.5286	0.731
	+Hofstede	**76.7978**	**0.741**
	+country+language	76.5333	0.731
	+country+Hofstede	**76.7978**	**0.741**
	+language +Hofstede	**76.7978**	**0.741**
	+country+language+ Hofstede	**76.7978**	**0.741**
demographic + contextual + attitudinal	+power distance	76.7394	0.740
	+individualism	**76.7978**	**0.741**
	+masculinity	76.1641	0.721
	+uncertainty avoidance	76.5861	0.729
	+pragmatism	76.6104	0.734
	+indulgence	**76.7978**	**0.741**

Top 9 most relevant features from feature selection

IG	Feature Name
0.043	device, method, and data type
0.042	**language**
0.042	**uncertainty avoidance**
0.042	**country**
0.042	**individualism**
0.042	**indulgence**
0.041	**power distance**
0.034	importance of notice and control
0.033	**pragmatism**

- Cultural features can aid in privacy prediction.
- Cultural values are better than language and country in predicting.
- Individualism and indulgence outperform other cultural values.

Who are acceptable data collectors?

Low-individualism countries: government, employer, foreign service provider
High-individualism countries: service providers "that I pay for" or "have an existing relationship with"
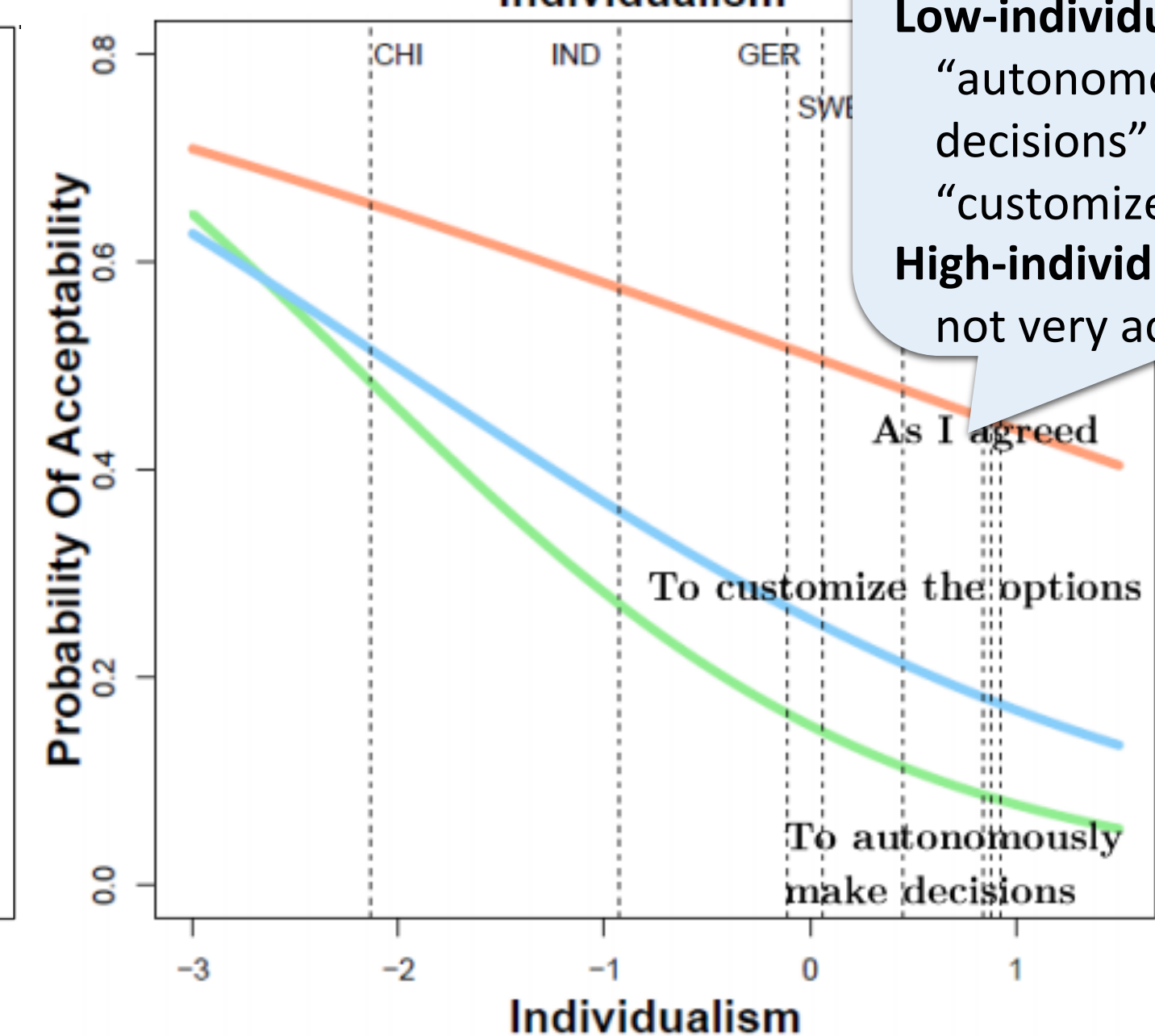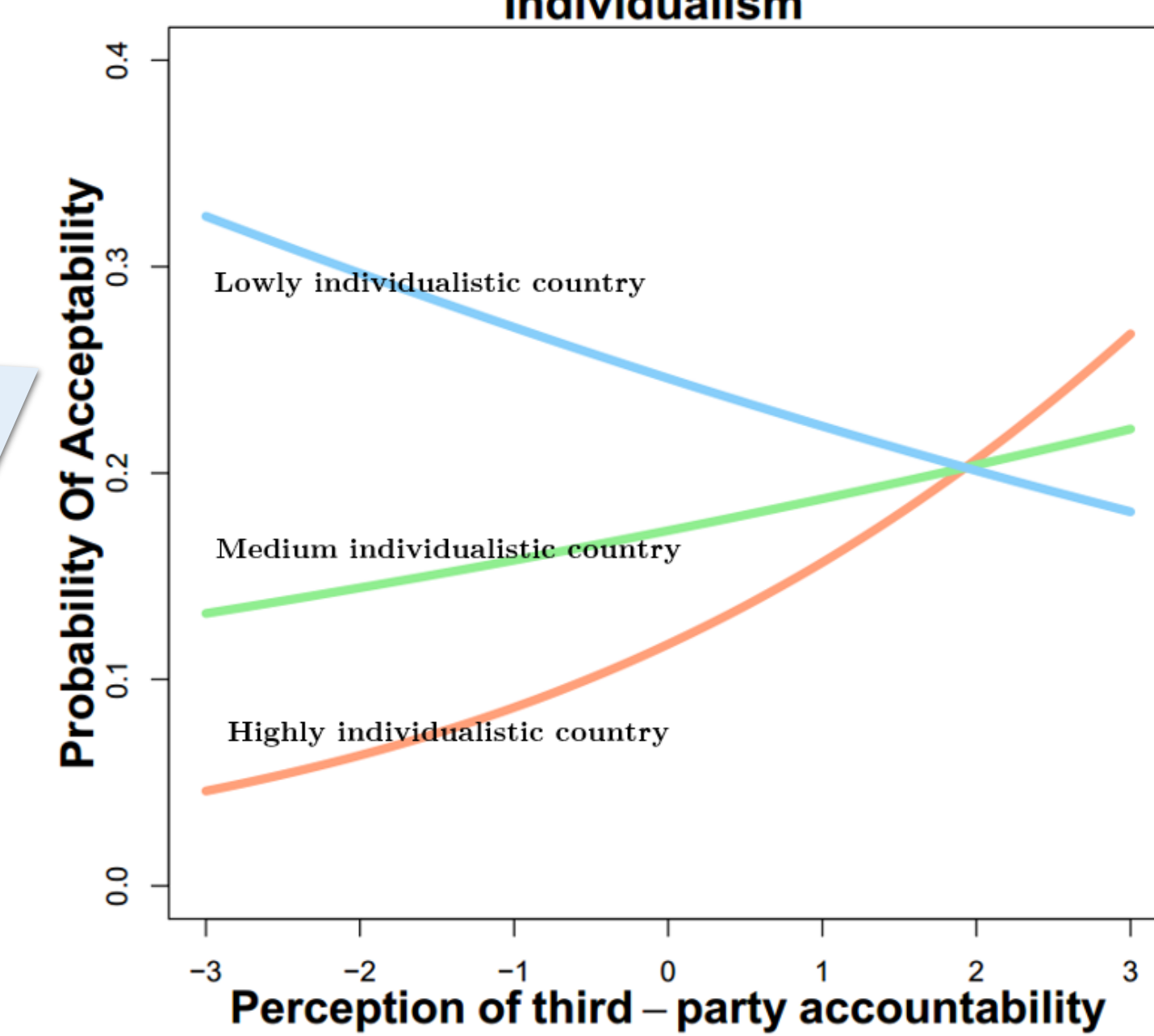
What are acceptable value exchanges?

Low-individualism countries: "benefit the community", "provide compelling value" or "save me time or money"
High-individualism countries: not very acceptable

When users believe a third party is accountable for data collection:

Low-individualism countries tend not to disclose
High-individualism countries tend to disclose

What are acceptable purposes?

Low-individualism countries: "autonomously make decisions" or "customize the options"
High-individualism countries: not very acceptable