Trinity Business School

Transforming Business for Good

Fair and Socially Acceptable AI in HR Management

Swiss HR Analytics Event, Olten
June 27, 2024

DR ULRICH LEICHT-DEOBALD
Associate Professor in Responsible Leadership

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

AACSB ACCREDITED | ASSOCIATION OF MBAS ACCREDITED | EFMD EQUIS ACCREDITED

Short Bio

Akademisch

- 2018 Gastwissenschaftler in INSEAD, Fontainebleau und Groningen
- 2015-2018 Senior Research Fellow am Institut für Wirtschaftsethik der Universität St.Gallen
- 2014-2015 Senior Research Fellow am Institut für Führung und HRM der Universität St.Gallen
- 2011-2014 Doktorat in Management, Universität St.Gallen
- 2005-2010 Master in Psychologie an der Universität Bremen

Short Bio

Theater

- 1996-2005 Nach der Schauspielschule in Hamburg Engagements als Schauspieler am Deutschen Schauspielhaus Hamburg, am Münchner Volkstheater und am Landestheater Detmold

Theater, an denen ich gespielt habe

“Socially Acceptable AI and Fairness Trade-offs in Predictive Analytics” (NRP-77: Digital Transformation)


Corinna Hertweck
ZHAW


Michele Loi
University of Zurich


Ulrich Leicht-Deobald
Trinity College Dublin


Markus Christen
University of Zurich


Serhiy Kandul
University of Zurich


Eleonora Viganò
University of Zurich


Christoph Heitz
ZHAW

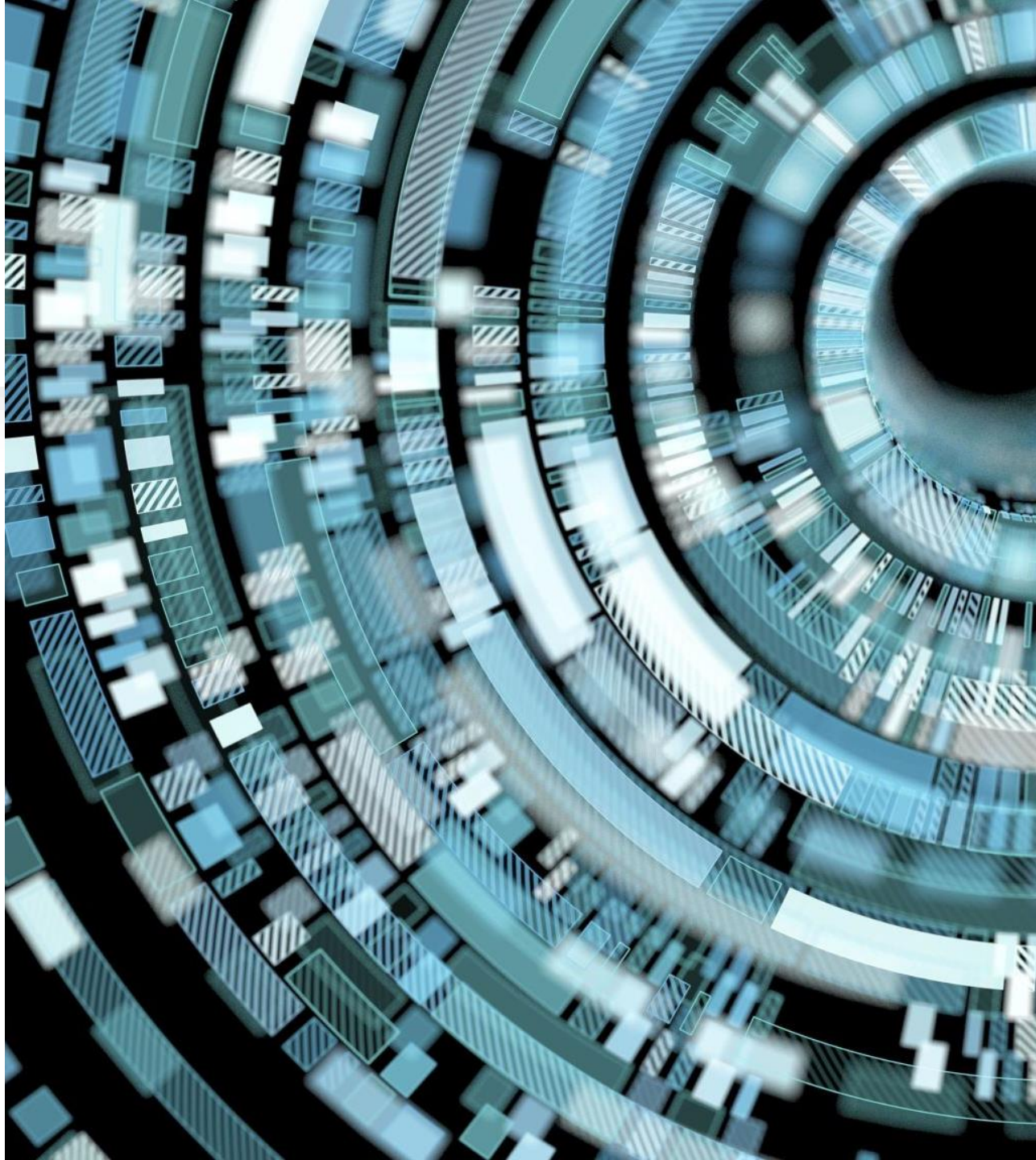

Joachim Baumann
ZHAW

Research Project Algorithmic Fairness

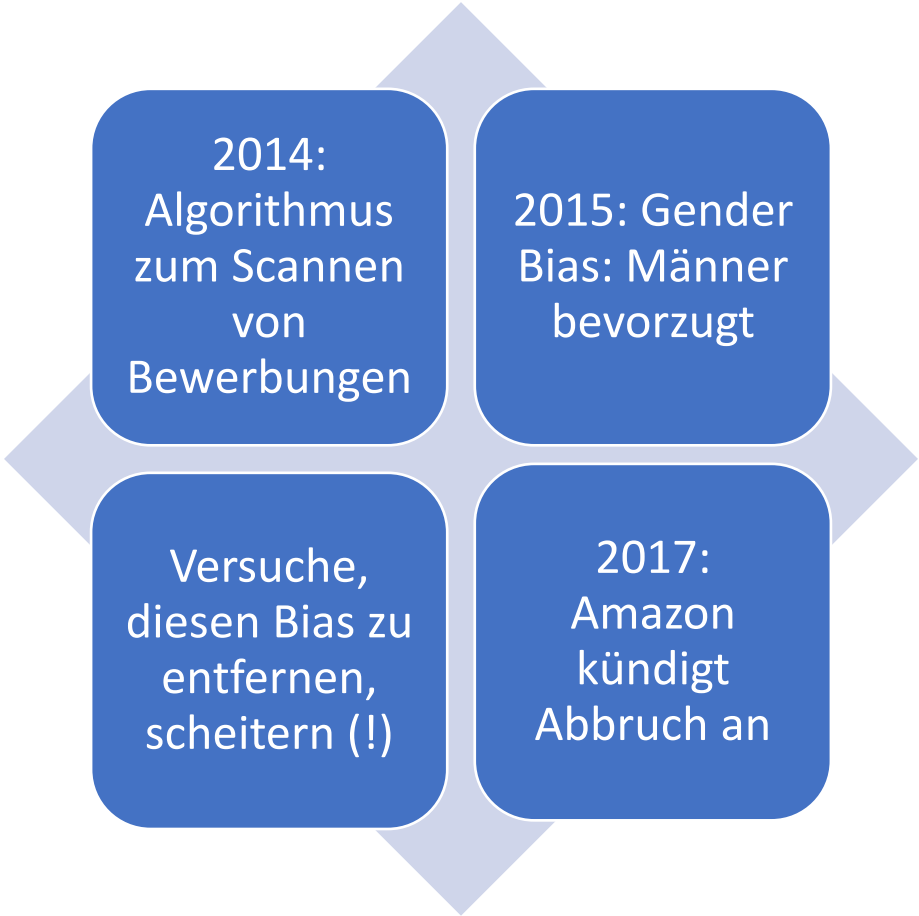
Interdisciplinary team: computer science, philosophy/ethics, social sciences:

- Ethics: "What kind of fairness should an algorithm have?“
- Computer Science: "How do you build fairness into algorithms?
- HR Management: "Algorithm ethics in a corporate context“

- What fairness means depends on the situation is usually controversial.
- Fairness of all decision systems has to be defined by society.



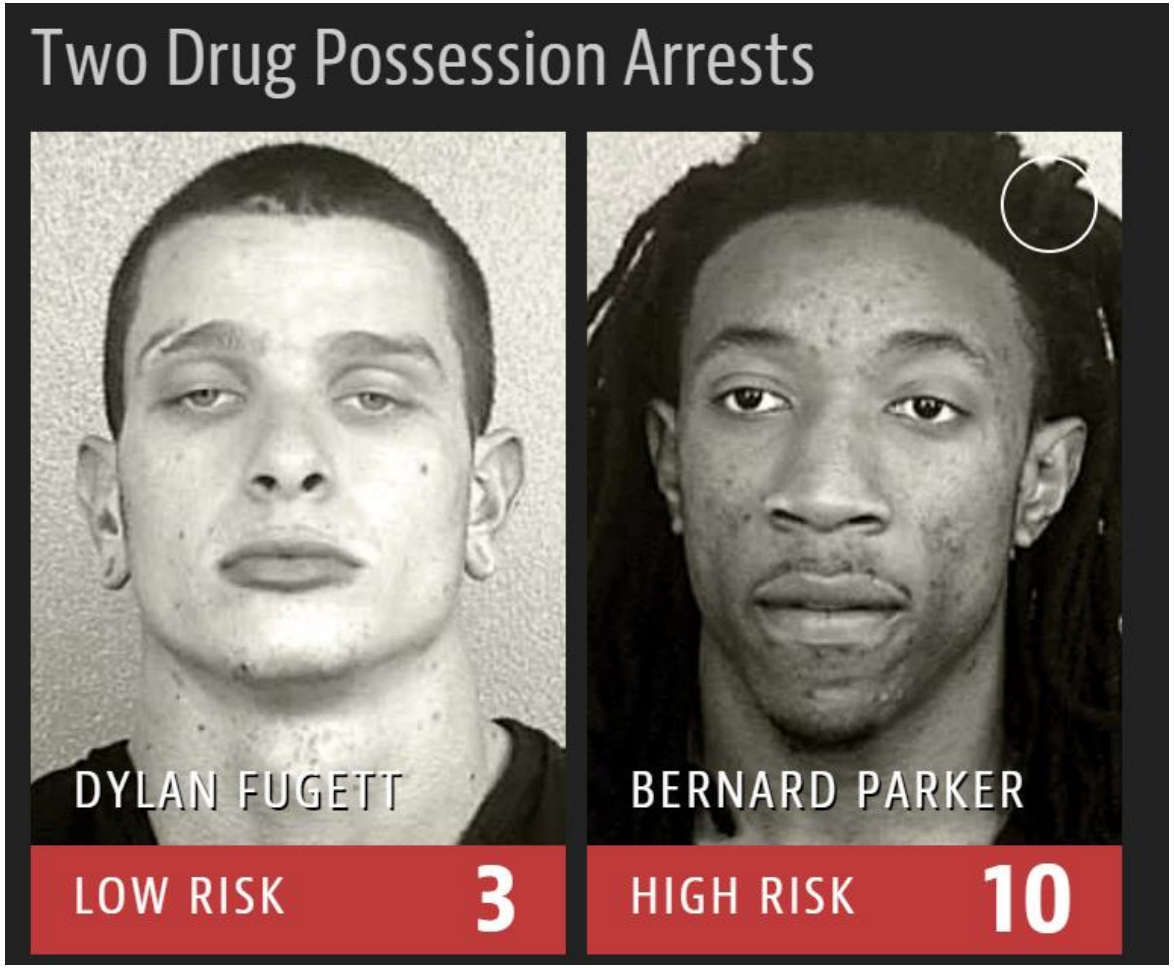
Amazons sexistischer Bewerbungsalgorithmus



amazon

Reuters. 2018. “Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women,” October 10, 2018. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

COMPAS (2016)


Two Drug Possession Arrests
DYLAN FUGETT
LOW RISK 3
BERNARD PARKER
HIGH RISK 10

Fairness von Algorithmen

Die guten Nachrichten:

Fairness kann man messen

z.B.: Anteil von Männern, die einen Job erhalten –

Anteil von Frauen, die einen Job erhalten

Fairness kann man nachprüfen

… geht bei Menschen in der Regel nicht

Fairness von Algorithmen

Die schlechte Nachricht

Datenbasierte Algorithmen sind in der Regel nicht fair

Gründe:

- Probleme mit Lerndaten (nicht repräsentativ, tragen schon Unfairness in sich)
- Algorithmen sind optimiert auf Nutzenmaximierung, nicht auf Fairness

Fairness von Algorithmen

- Fairness-Probleme entstehen in der Regel ohne Absicht
  - ➢Keine bösartigen Entwickler

- Nebenprodukt der ML-Logik: „Maximiere den Nutzen der Entscheidung“

Welche Use Cases im HR-Bereich können sie sich vorstellen, bei denen Fairness eine Rolle spielt?

5 min

Hintergrund

<u>Diskriminierungsverbot im neuen EU AI Act:</u>

«Zu den Zu den sieben Grundsätzen gehören: menschliches Handeln und menschliche Aufsicht, technische Robustheit und Sicherheit, Privatsphäre und Daten-Governance, Transparenz, Vielfalt, **Nichtdiskriminierung und Fairness**, soziales und ökologisches Wohlergehen sowie Rechenschaftspflicht.

<u>Art. 8 Bundesverfassung:</u>

„2 Niemand darf diskriminiert werden, namentlich nicht wegen der Herkunft, der Rasse, des Geschlechts, des Alters, der Sprache, der sozialen Stellung, der Lebensform, der religiösen, weltanschaulichen oder politischen Überzeugung oder wegen einer körperlichen, geistigen oder psychischen Behinderung.

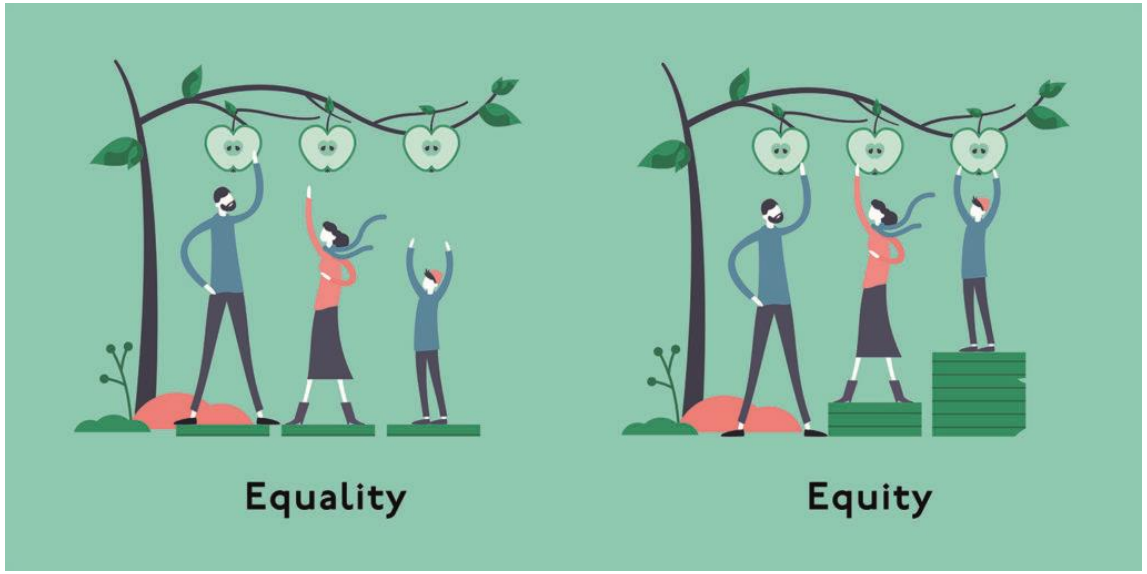
Was genau ist „fair“?

Fairness-Definitionen

Fairness ist…..

- … gleiche Regeln für alle
- … gleiche Chancen für alle
- … gleiche Chancen für alle, die diese Chancen auch verdienen
- … usw.

Mehr als 80 Definitionen für Fairness in der ML Literatur!



Zwischenfazit

Was Fairness bedeutet

- hängt von der Situation ab.
- ist in der Regel umstritten.

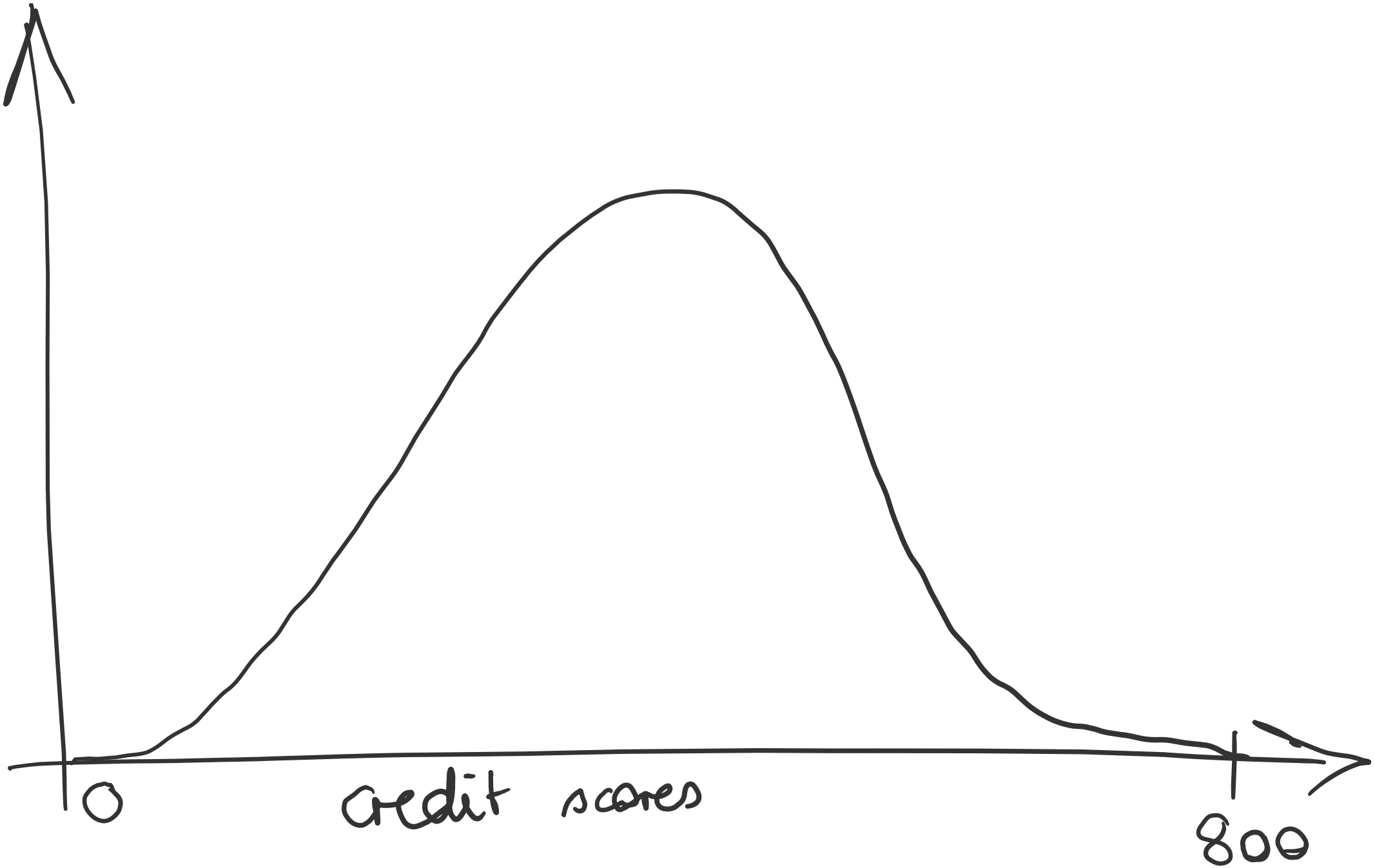
Aber: es ist nicht egal!

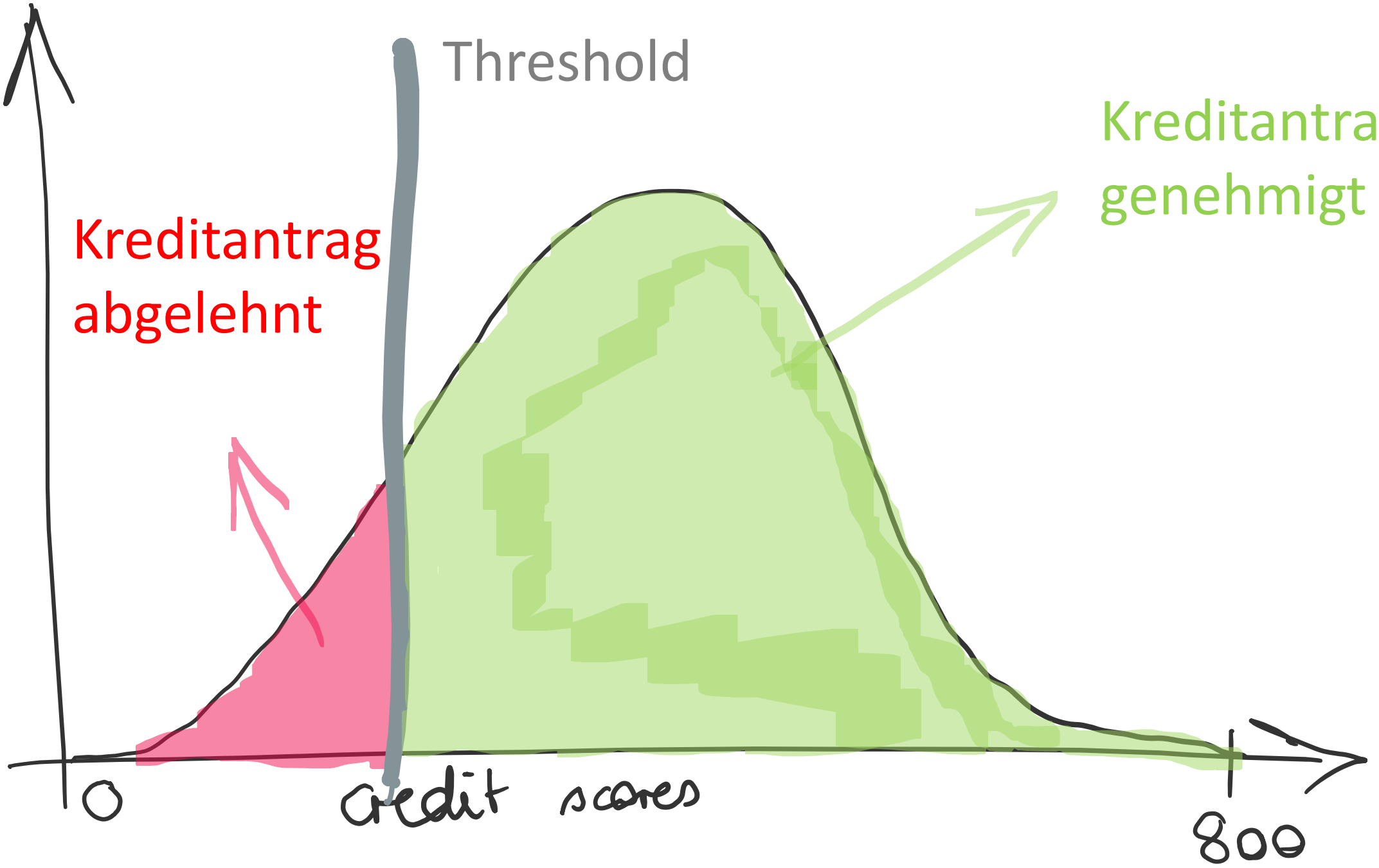
Fairness von allen Entscheidungssystemen muss von der Gesellschaft definiert werden.

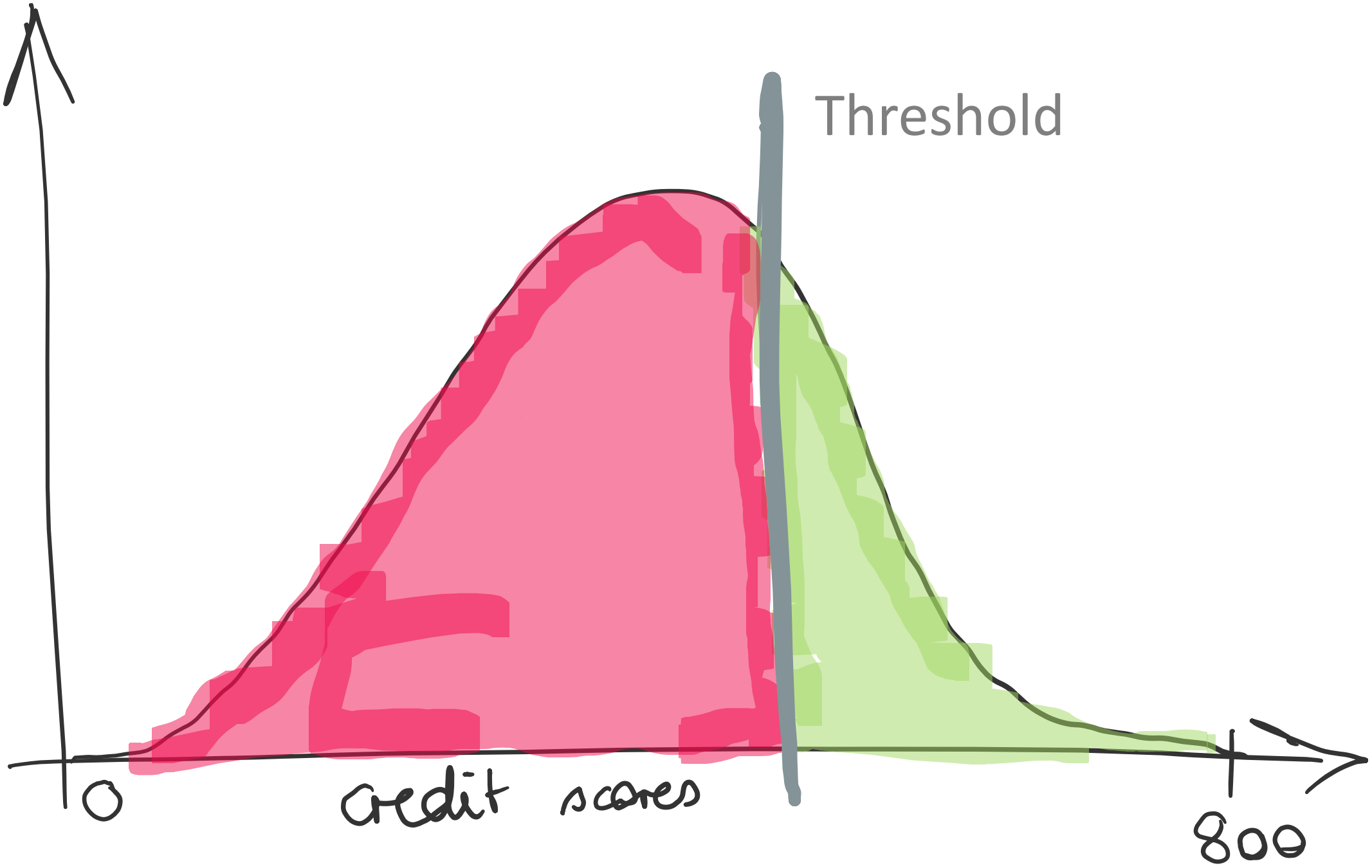
Dies gilt auch für algorithmische Entscheidungssysteme.

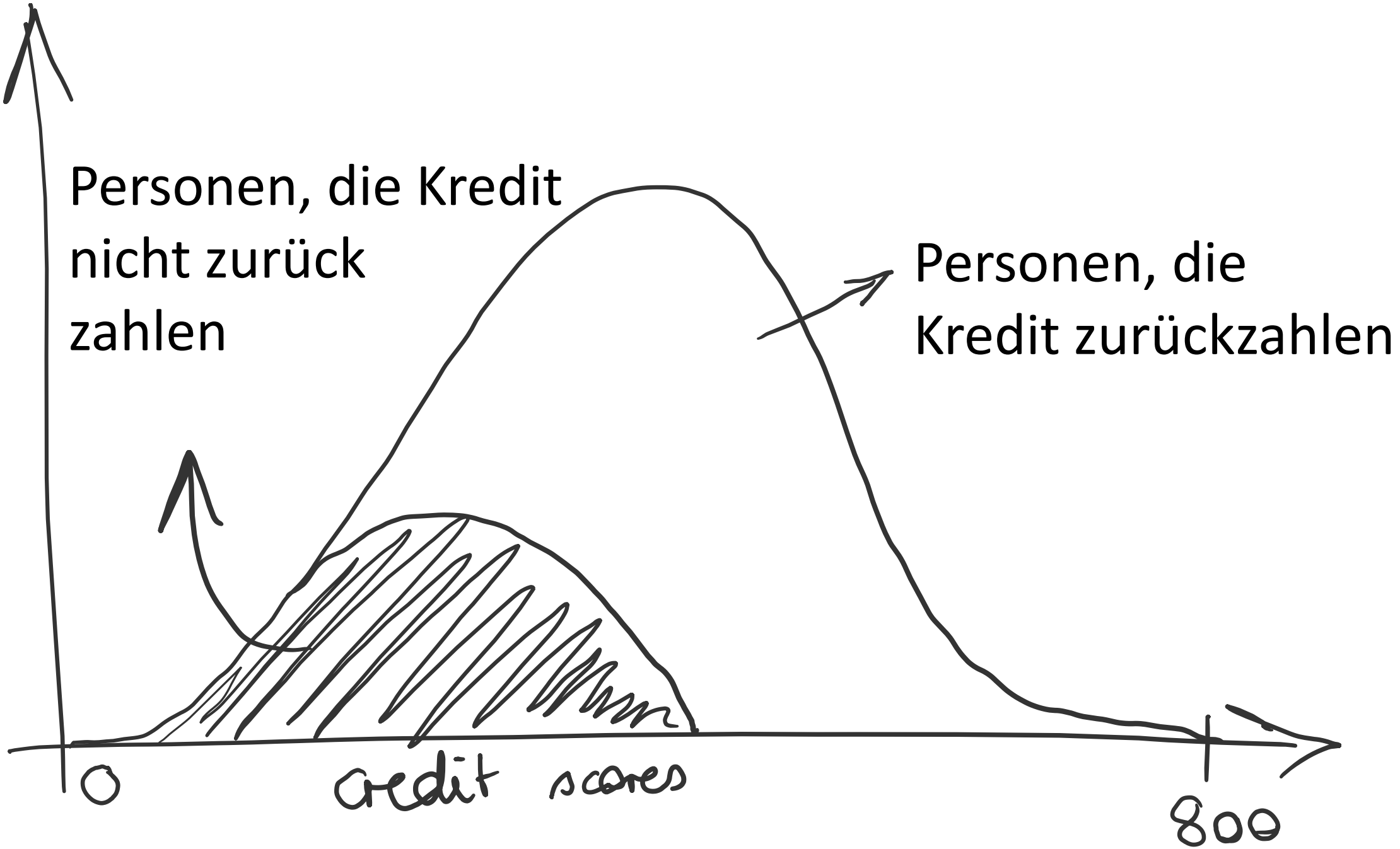
Beispiel Kreditvergabe

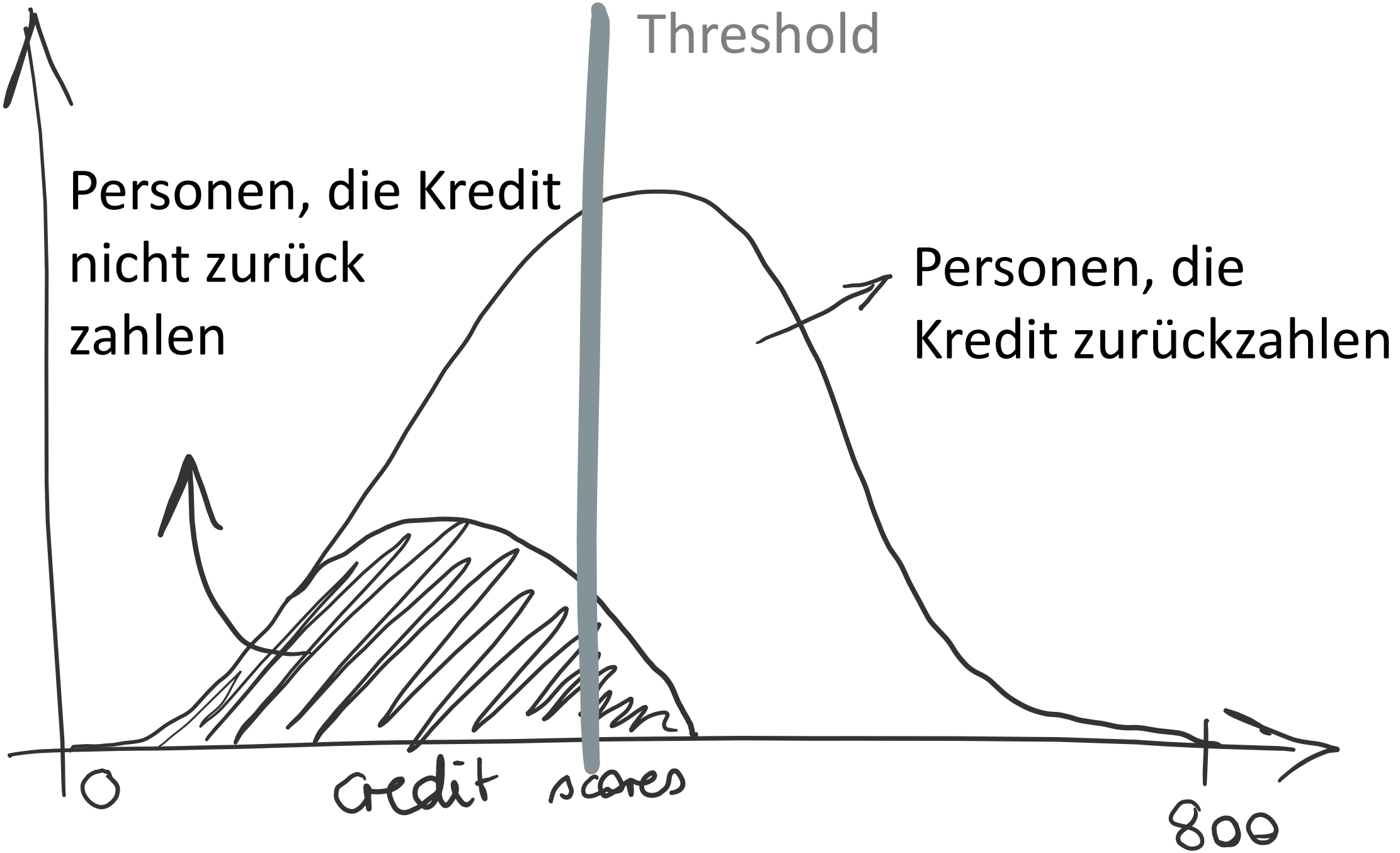
Kreditvergabe basierend auf Bonitätswerten

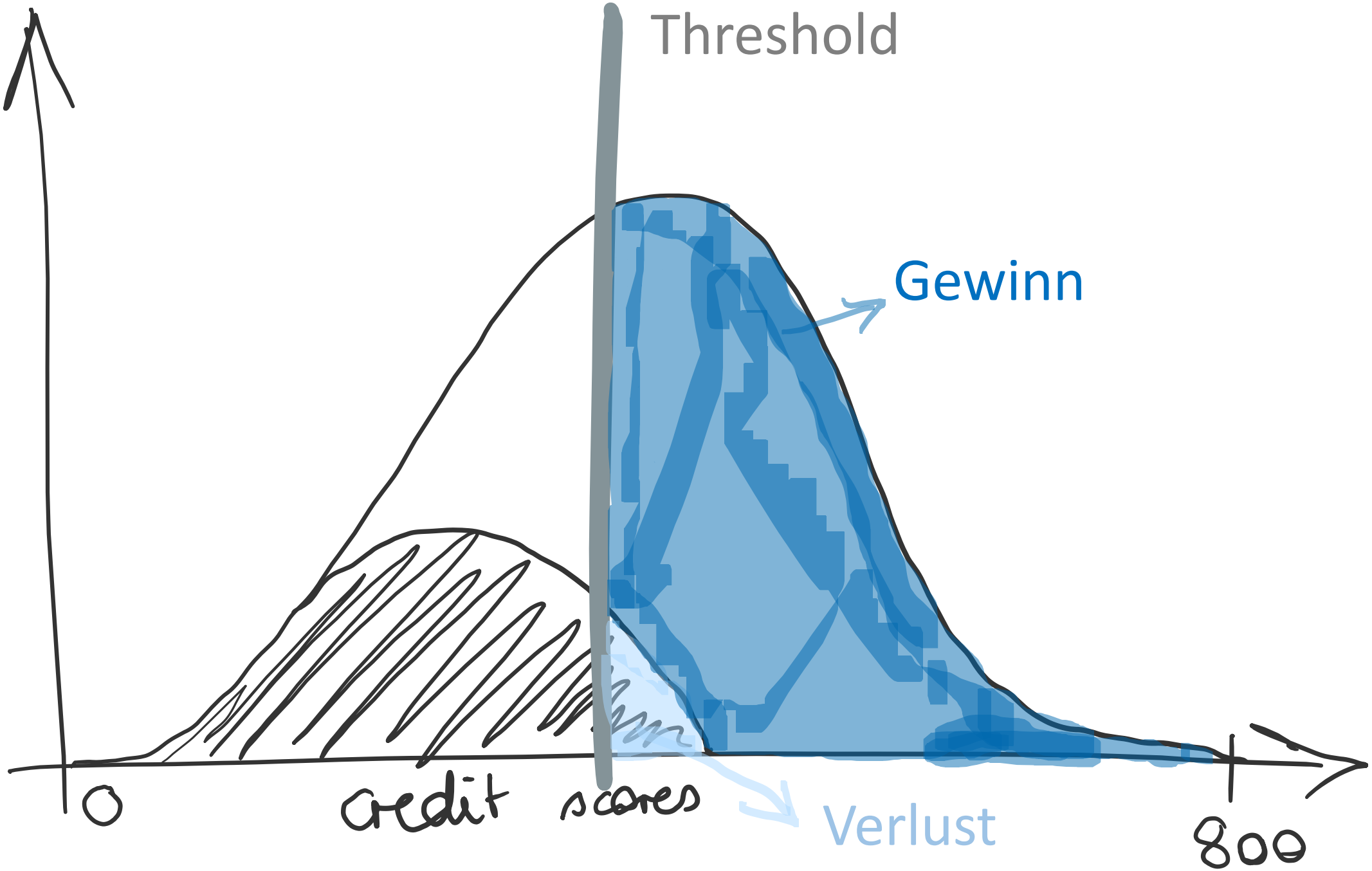

0
credit scores
800

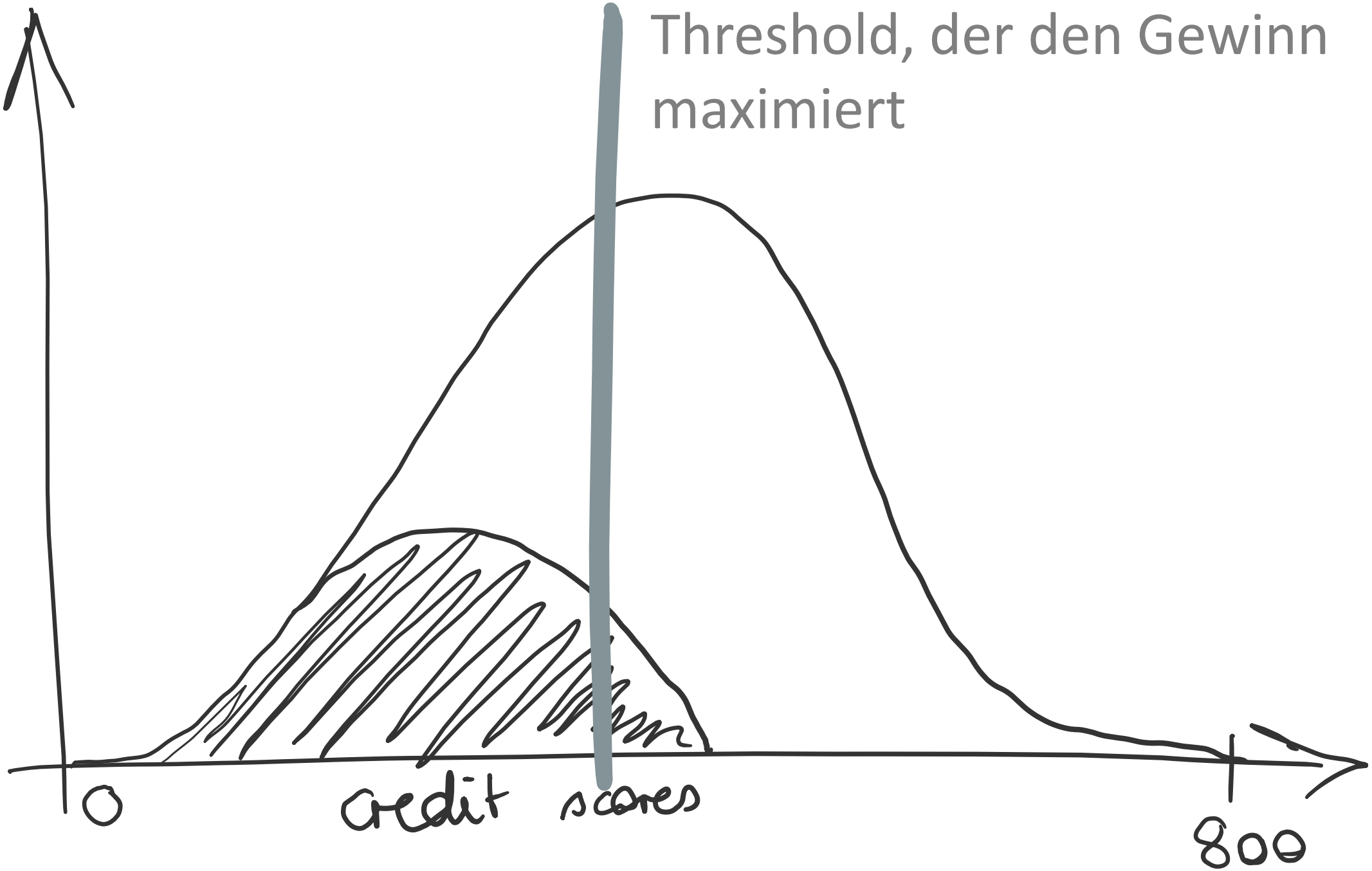

Threshold
Kreditantrag genehmigt
Kreditantrag abgelehnt
0
credit scores
800

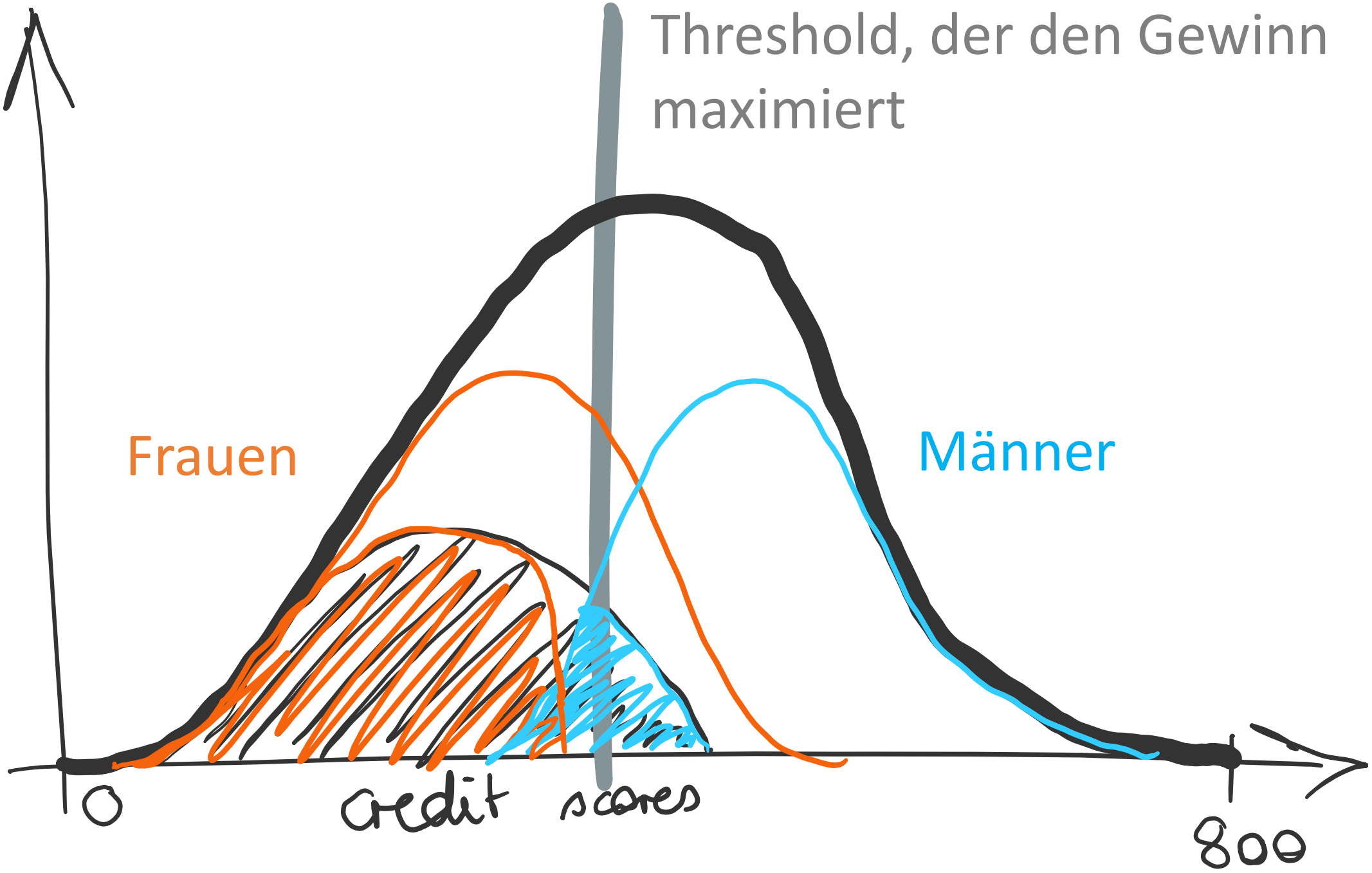

Threshold
0
credit scores
800

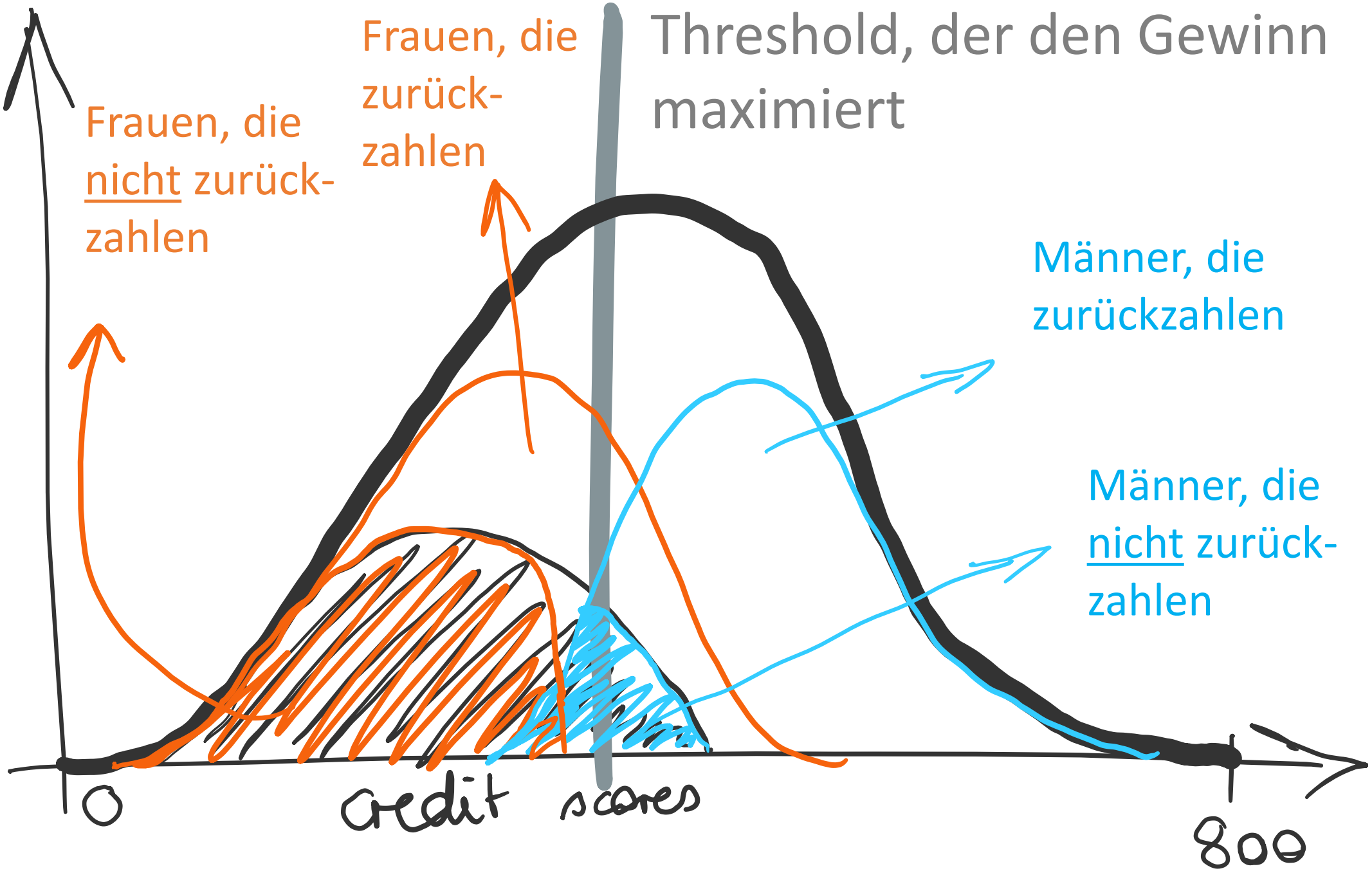

Personen, die Kredit nicht zurück zahlen
Personen, die Kredit zurückzahlen
0
credit scores
800

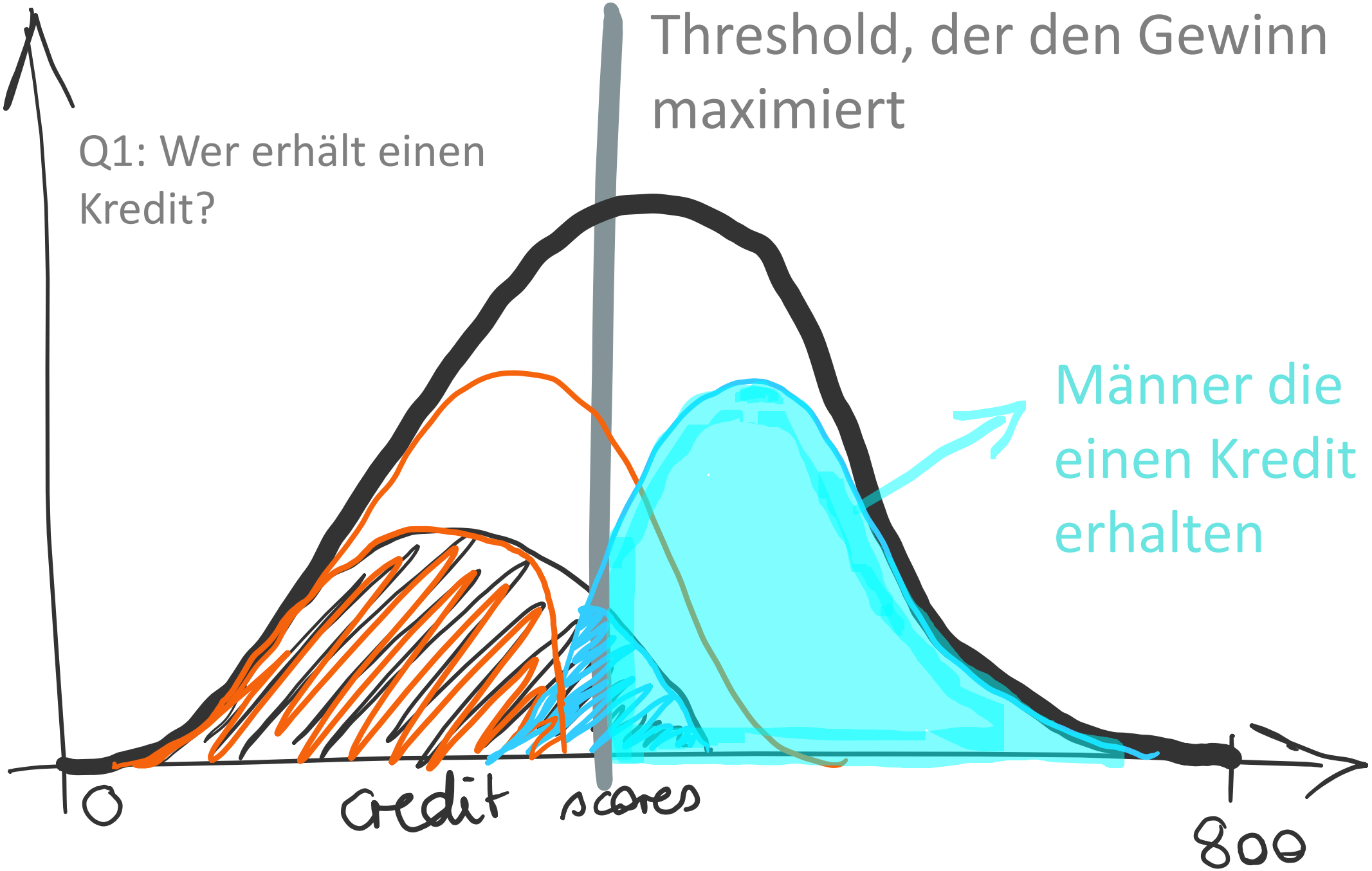

Threshold
Personen, die Kredit nicht zurück zahlen
Personen, die Kredit zurückzahlen
0
credit scores
800

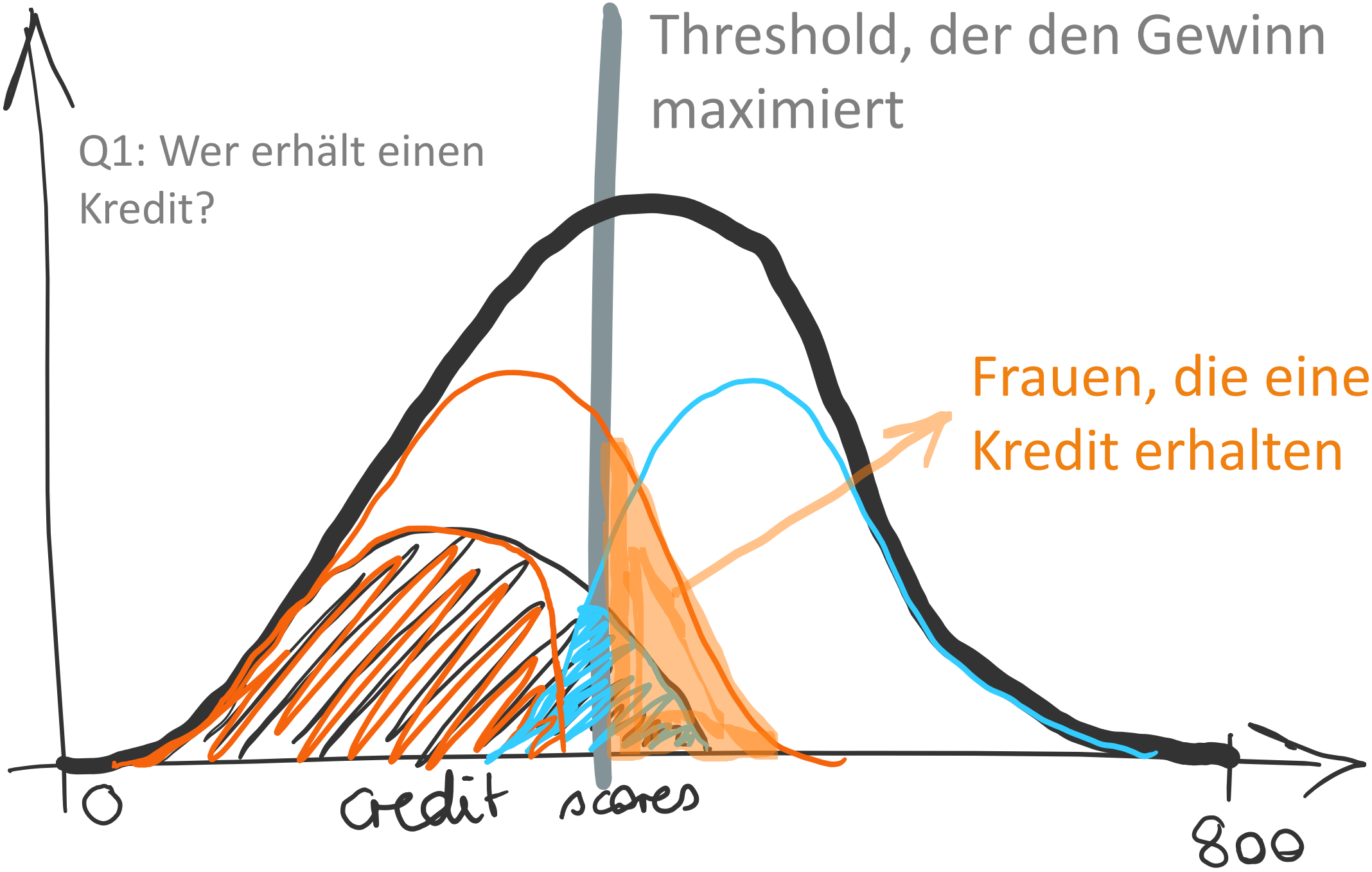

Threshold
Gewinn
0
credit scores
Verlust
800

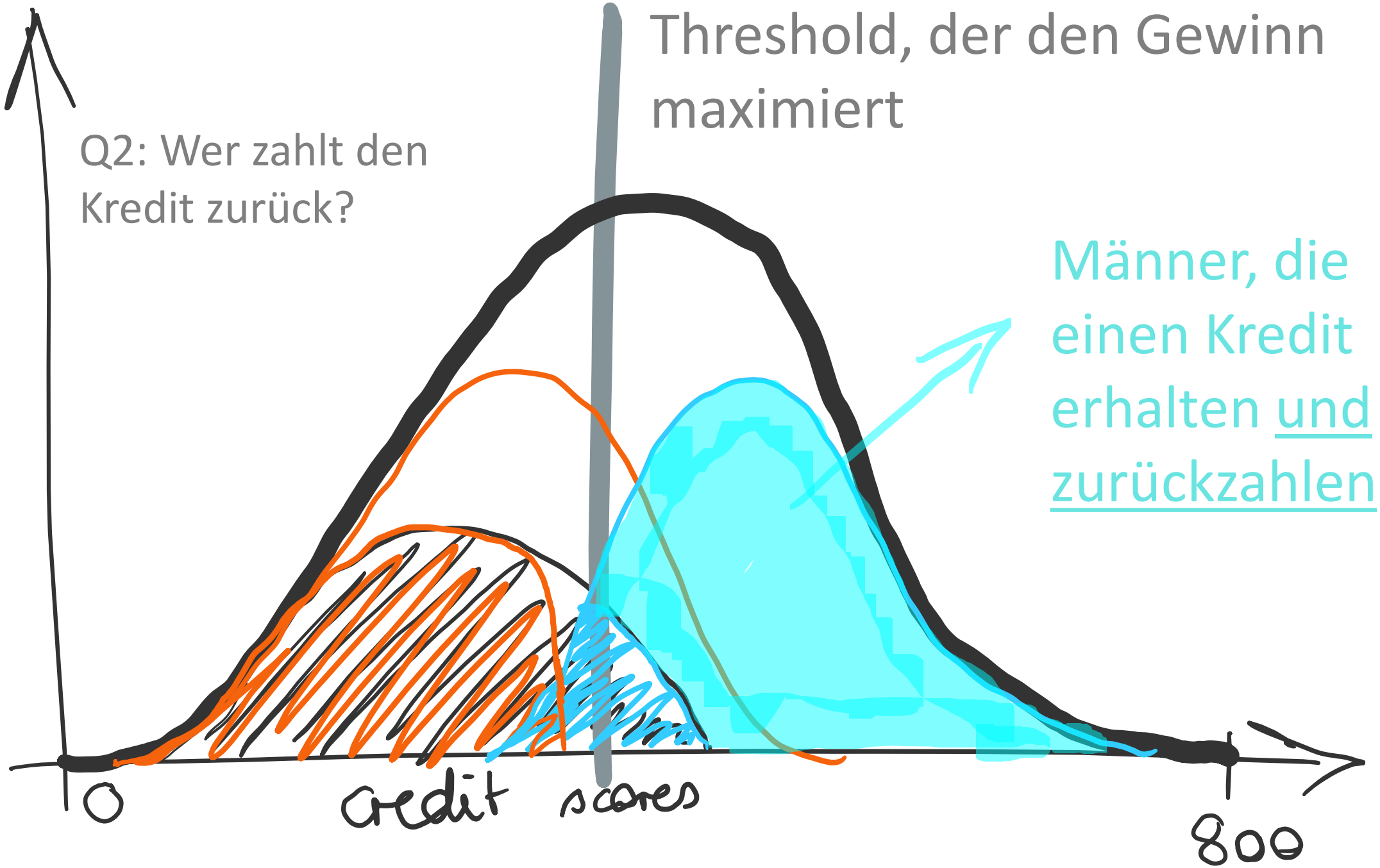

Threshold, der den Gewinn maximiert
0
credit scores
800

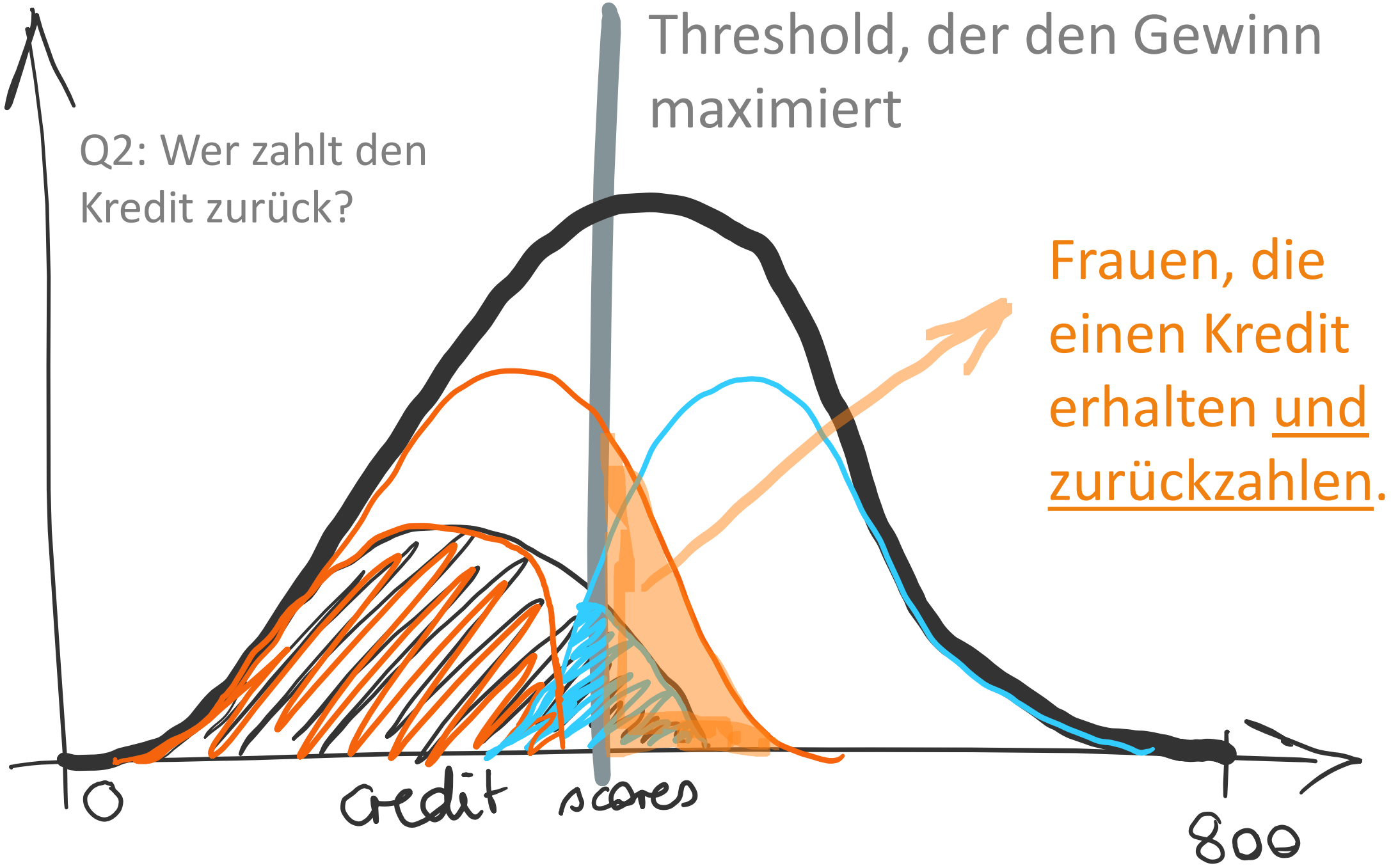

Threshold, der den Gewinn maximiert
Frauen
Männer
0
credit scores
800


Frauen, die nicht zurück-zahlen
Frauen, die zurück-zahlen
Threshold, der den Gewinn maximiert
Männer, die zurückzahlen
Männer, die nicht zurück-zahlen
0
credit scores
800


Threshold, der den Gewinn maximiert
Q1: Wer erhält einen Kredit?
Männer die einen Kredit erhalten
0
credit scores
800


Threshold, der den Gewinn maximiert
Q1: Wer erhält einen Kredit?
Frauen, die einen Kredit erhalten
0
credit scores
800


Threshold, der den Gewinn maximiert
Q2: Wer zahlt den Kredit zurück?
Männer, die einen Kredit erhalten und zurückzahlen.
0
credit scores
800


Threshold, der den Gewinn maximiert
Q2: Wer zahlt den Kredit zurück?
Frauen, die einen Kredit erhalten und zurückzahlen.
0
credit scores
800

Zusammen-fassung

Wer erhält einen Kredit?

90% der Männer

20% der Frauen

Wer von den zurückzahlenden Bewerbern erhält einen Kredit?

100% der zurückzahlenden Männer

40% der zurückzahlenden Frauen

Die **Standard-Methode zur Maximierung des Profits** durch Entscheidungen auf der Grundlage von Prognosen führt zu **sozialer Ungerechtigkeit.**

Was tun bzgl. der Ungerechtigkeit?

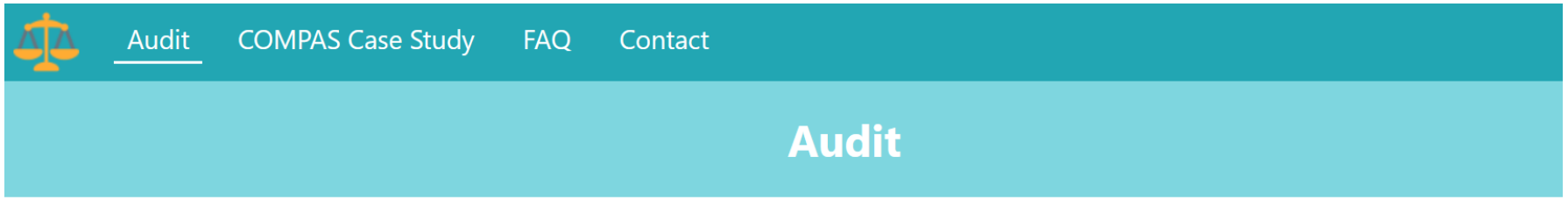
Fairness Lab

- Kann für einen Fairness Audit benutzt werden
- Hilft mit einem strukturiertem Step-by-Step Verfahren, eine begründete Entscheidung zur Fairness eines Algorithmus zu treffen.
- Ist frei verfügbar: **https://joebaumann.github.io/FairnessLab/#/**
- **Kann Unternehmen helfen, die Anforderungen des neuen EU AI Acts zu erfüllen.**
- Bei Interesse, wendet Euch gerne an Christoph Heitz (christoph.heitz@zhaw.ch)

Christoph Heitz

ZHAW

Fairness Lab



Dataset

Choose a dataset that you want to audit.

COMPAS

The COMPAS dataset was collected by ProPublica for their article "Machine Bias." We preprocessed the dataset to make it usable for this demo. The predicted scores are the original (decimal) scores from COMPAS.

- Y=0: Was arrested within two years
- Y=1: Was not arrested within two years
- D=0: Predicted to be rearrested
- D=1: Predicted not to be rearrested
- Group A: Black
- Group B: white

You can find the notebook here to see how we prepared the data.

Credit lending (UCI German Credit)

The German Credit dataset is available in the UCI repository. It is a small dataset of German credit loans from the 1970s. The scores have been predicted with a vanilla logistic regression.

Zusammenfassung

- Datenbasierte Algorithmen sind in der Regel nicht fair
- Fairness-Probleme entstehen in der Regel ohne Absicht beim Maximieren des Nutzens eines Algorithmus
- Ohne explizites Fairness Design bleibt ein Algorithmus diskriminiert gegenüber geschützten Attributen wie Alter, Geschlecht und Ethnie
- Das Fairness Lab kann mit einem geführten Step-by-Step Verfahren helfen, eine passendes Fairness Design zu finden.
- Es gibt nicht eine richtige Antwort in Bezug auf Fairness. Daher benötigt es einen Stakeholder Dialog.

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Vielen Dank fürs Teilnehmen!

Meldet Euch gerne per Email (ulrich.leicht-deobald@tcd.ie)
LinkedIn (https://www.linkedin.com/in/ulrich-leicht-deobald-a62178b2/?originalSubdomain=ie)

für Fragen, Anregungen und mögliche Kooperationen.

Dr. Leicht-Deobald
Associate Professor in Responsible Leadership
Trinity Business School, Trinity College Dublin