

# Medienlinguistische Methodik

## Modellierung

Arne Rubehn

Lehrstuhl für Multilinguale Computerlinguistik  
Universität Passau

21.10.2025



# Überblick

In dieser Sitzung widmen wir uns der **Modellierung von Objekten und Prozessen**.

Was sind Objekte und Prozesse in der Wissenschaft?

Wie können Objekte und Prozesse modelliert werden und zu welchem Zweck?

Was verstehen wir unter Daten und wie können wir sie analysieren?



# Objekte und Prozesse

Die Wissenschaft beschäftigt sich mit **Prozessen** (z.B. *Inflation, Revolution, Klimawandel*) und **Objekten** (z.B. *Buch, Werkzeug, Berg*).

Um Prozesse oder Objekte untersuchen zu können, benötigen wir eine wissenschaftliche **Repräsentation**.

Hierbei sind wissenschaftliche **Konstrukte** notwendig (vereinfacht gesagt: Abstraktionen der tatsächlich zu beobachtenden Prozesse oder Objekte).



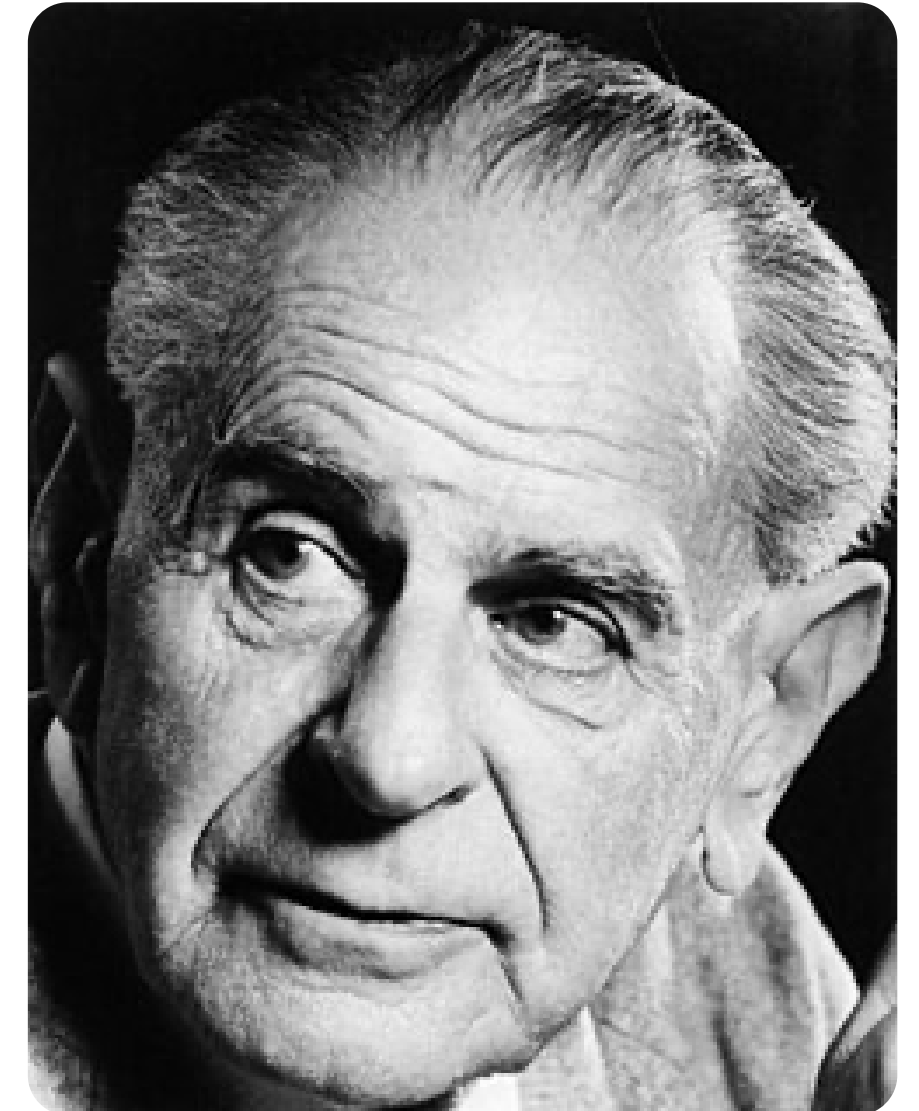
# Poppers Drei Welten

Der Philosoph Karl Popper (1902-1994) macht diese Unterscheidung in seiner Theorie der **Drei Welten** klar:

**Welt 1.** Physikalische Objekte & Prozesse.

**Welt 2.** Individuelle Wahrnehmungen.

**Welt 3.** Kollektive Erkenntnis.



# Poppers Drei Welten

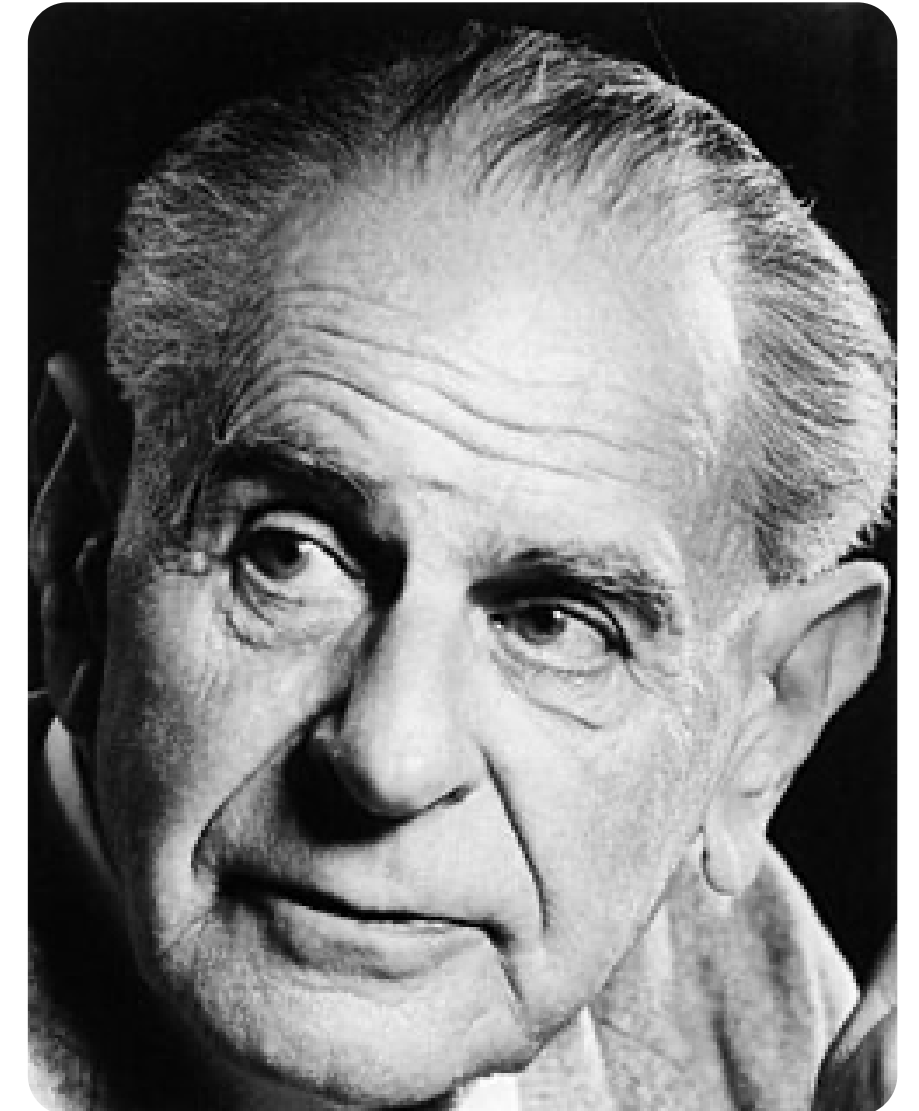
## **Welt 1.** Physikalische Objekte & Prozesse.

*“There is, first, the world that consists of physical bodies: of stones and of stars; of plants and of animals; but also of radiation, and of other forms of physical energy. I will call this physical world ‘world 1.’”*

(Popper, 1978: 143)

## **Welt 2.** Individuelle Wahrnehmungen.

## **Welt 3.** Kollektive Erkenntnis.



# Poppers Drei Welten

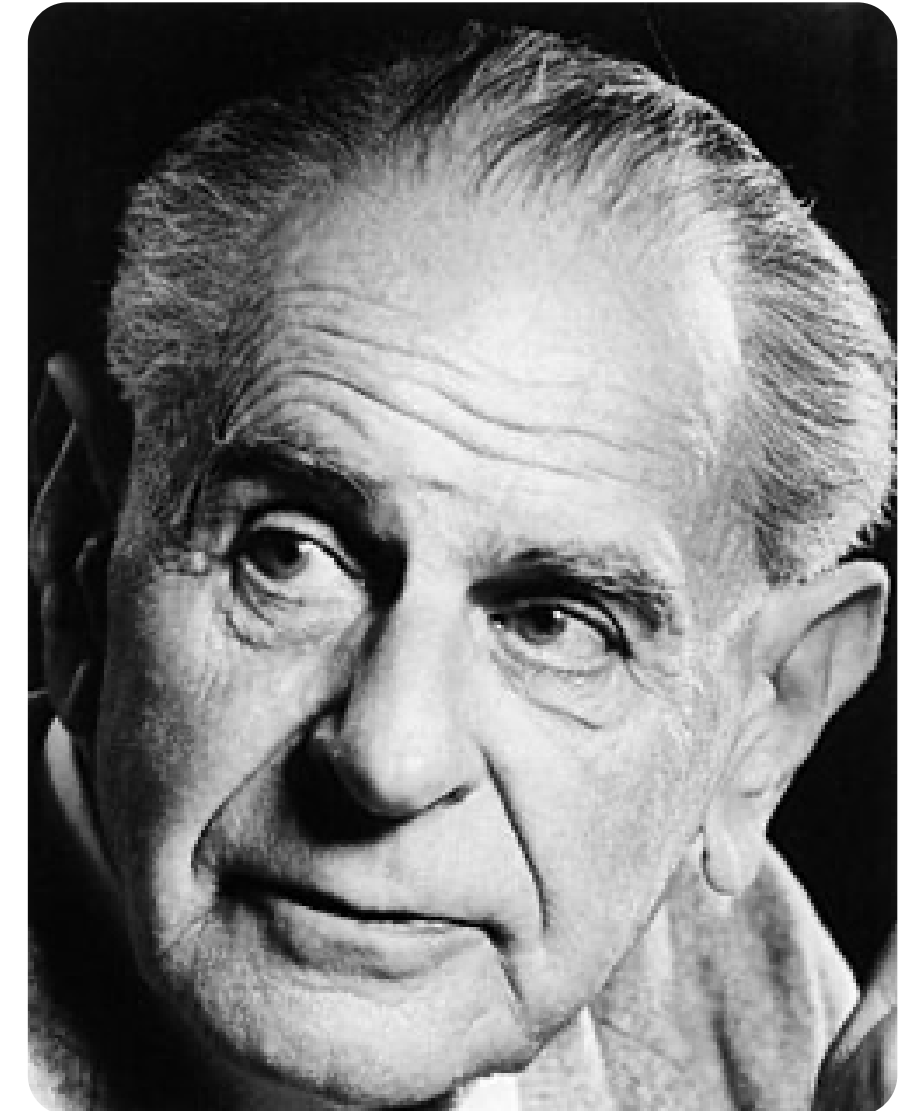
**Welt 1.** Physikalische Objekte & Prozesse.

**Welt 2.** Individuelle Wahrnehmungen.

*“There is, secondly, the mental or psychological world, the world of our feelings of pain and pleasure, of our thoughts, of our decisions, of our perceptions and our observations; in other words, the world of mental or psychological states or processes, or of subjective experiences. I will call it ‘world 2.’”*

(Popper, 1978: 143)

**Welt 3.** Kollektive Erkenntnis.



# Poppers Drei Welten

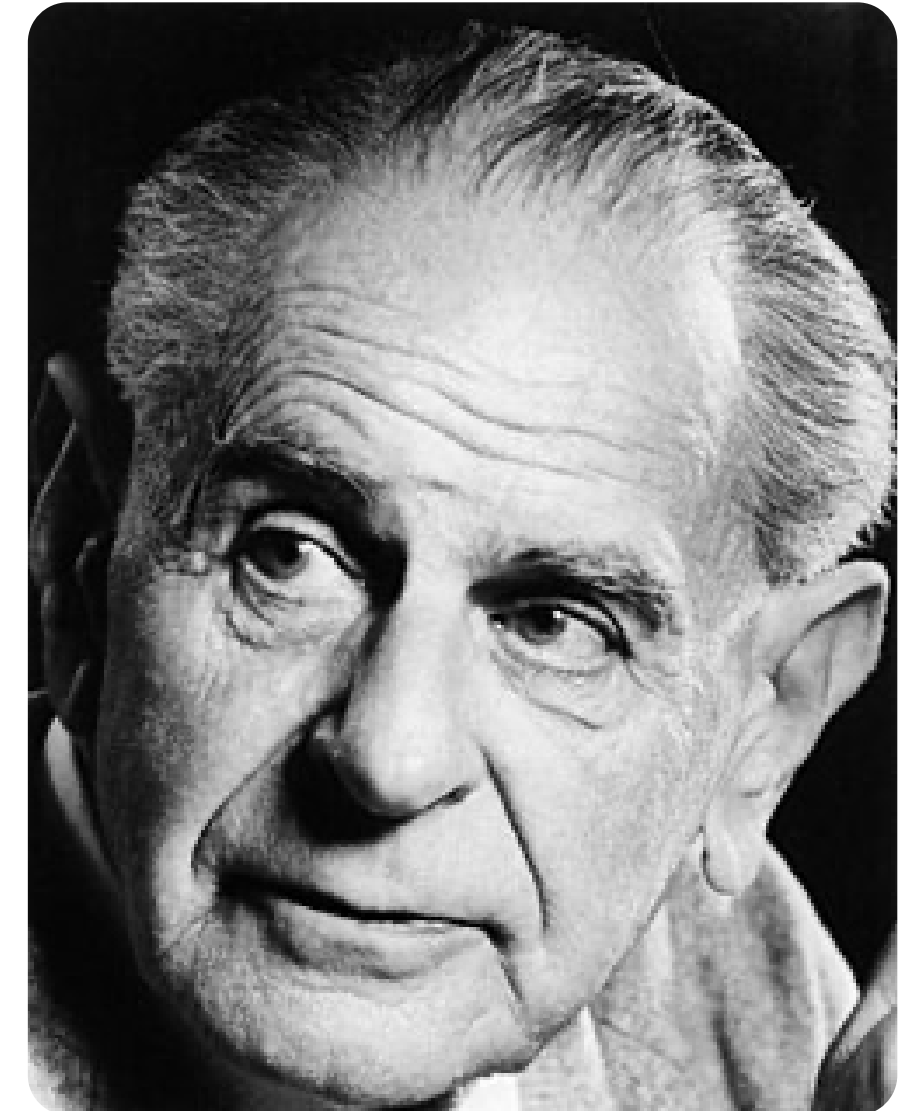
**Welt 1.** Physikalische Objekte & Prozesse.

**Welt 2.** Individuelle Wahrnehmungen.

**Welt 3.** Kollektive Erkenntnis.

*“By world 3 I mean the world of the products of the human mind, such as languages; tales and stories and religious myths; scientific conjectures or theories, and mathematical constructions; songs and symphonies; paintings and sculptures. But also aeroplanes and airports and other feats of engineering.”*

(Popper, 1978: 144)



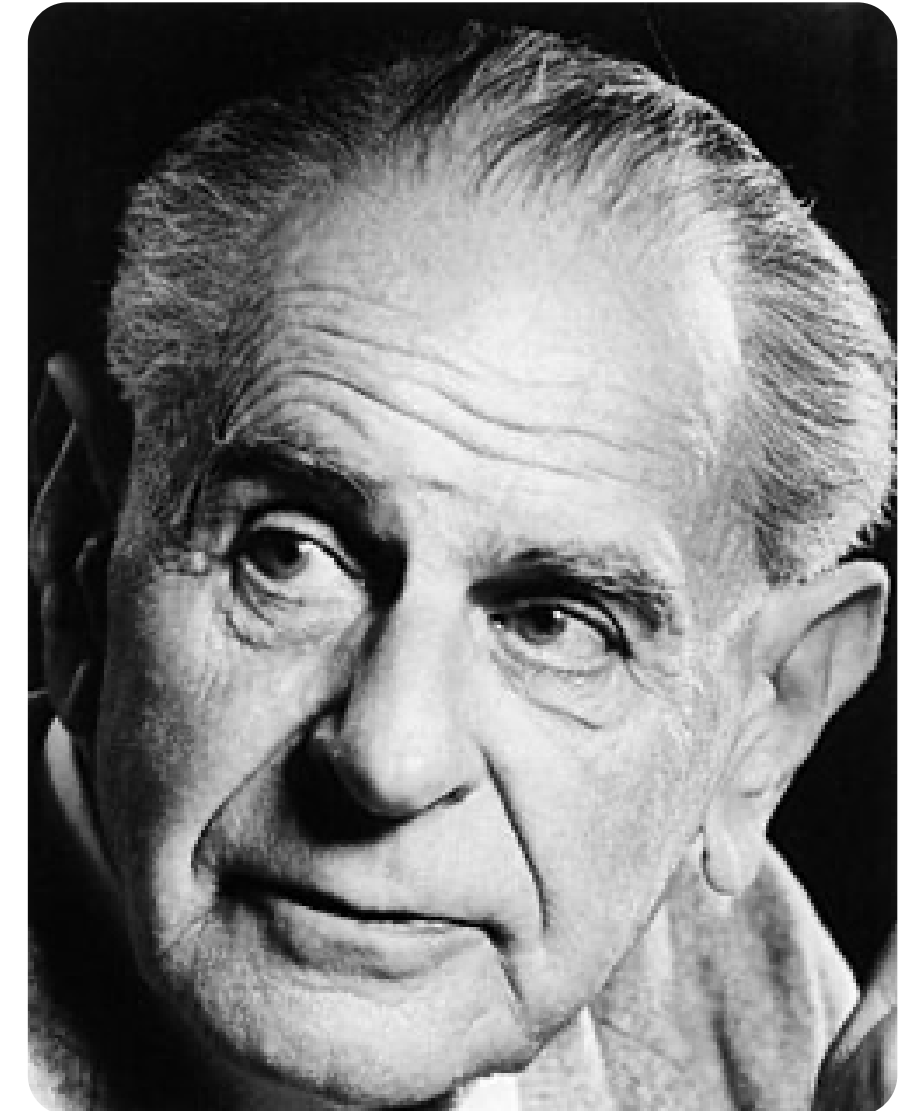
# Poppers Drei Welten

Alle drei Welten sind **real** und haben direkte Auswirkungen aufeinander.

Objekte können auch gleichzeitig zu mehreren Welten gehören:

*“Many of the objects belonging to world 3 belong at the same time also to the physical world 1. Michelangelo’s sculpture The Dying Slave is both a block of marble, belonging to the world 1 of physical objects, and a creation of Michelangelo’s mind, and as such belonging to world 3. The same holds of course for paintings.”*

(Popper, 1978: 144)



# Modellierung

Was sind Modelle?



<https://partici.fi/92489810>



# Modellierung

Modelle sind vereinfachte, rechenbare **Repräsentationen** von Prozessen und Objekten in der Wissenschaft. Sie zeichnen sich durch drei Merkmale aus:


**Abbildung.** Modelle bilden ein Original ab.

**Verkürzung.** Modelle besitzen ausgewählte, aber nicht alle Merkmale des Originals.

**Pragmatik.** Modelle haben einen Verwendungszweck – sie werden verwendet, um Erkenntnisse über Objekte oder Prozesse zu gewinnen.



# Modellierung



**89**  
ST

**Kane**

TEM SCH PAS DRI DEF PHY  
64 92 83 82 48 82

Position  
**ST**

Schwacher Fuß  
★★★★☆

Spezialbewegungen  
★★★★☆

Starker Fuß  
**Rechts**

Größe  
188 cm / 6'2" Fuß

## Harry Kane

<b>Tempo</b>	<b>64</b>	<b>Schüsse</b>	<b>92</b>	<b>Passen</b>	<b>83</b>	<b>Dribbling</b>	<b>82</b>
Beschleunigung	62	Positionsspiel	94	Übersicht	86	Beweglichkeit	67
Sprintgeschwindigkeit	66	Abschluss	93	Flanken	80	Balance	72
		Schusskraft	91	Freistoß-Präzision	69	Reaktionen	94
		Weitschüsse	89	Kurze Pässe	86	Ballkontrolle	87
		Volleys	89	Lange Pässe	79	Dribbling	80
		Elfmeter	95	Effet	82	Ruhe	92
<b>Defensive</b>	<b>48</b>	<b>Physis</b>	<b>82</b>				
Abgefangene Bälle	42	Sprungkraft	87				
Kopfball-Präzision	90	Stamina	76				
Defensives Bewusstsein	46	Stärke	86				
Faire Zweikämpfe	44	Aggressivität	80				



# Modellierung



## Orthographie

Auge  
Nase  
Mund

## IPA

aʊ g ə  
n aː z ə  
m ʊ n t



# Modellierung

Modelle repräsentieren entsprechend Objekte oder Prozesse, die wir in **wissenschaftlichen Konstrukten** denotieren.

Sie beschränken sich bewusst auf bestimmte Merkmale der Konstrukte, ohne zu versuchen, diese vollständig zu erfassen.

Sie werden erstellt, um bestimmten, konkreten Zwecken zu dienen.



# Modelle und Qualität

Es ist nicht einfach, wissenschaftliche Modelle direkt miteinander zu **vergleichen**, vor allem auch deshalb, da uns die klaren Kriterien fehlen, ein Modell einem anderen vorzuziehen.

Wenn Modelle **konkrete, überprüfbare Vorhersagen** treffen können, können diese als direkte Tests dienen.

Sind solche direkten Tests nicht möglich, ist die Evaluierung von Modellen häufig sehr schwierig.



# Modelle und Qualität

Modelle sind streng genommen immer falsch – sie bilden nie die gesamte Realität ab, sondern approximieren sie nur.

Modelle sind also in erster Linie anhand ihrer **Nützlichkeit** zu bewerten.

**Ockhams Rasiermesser:** Die ökonomischste Erklärung für Naturphänomene ist häufig die beste!



**Ockham's Razor**  
*Other Weapon - Dagger*

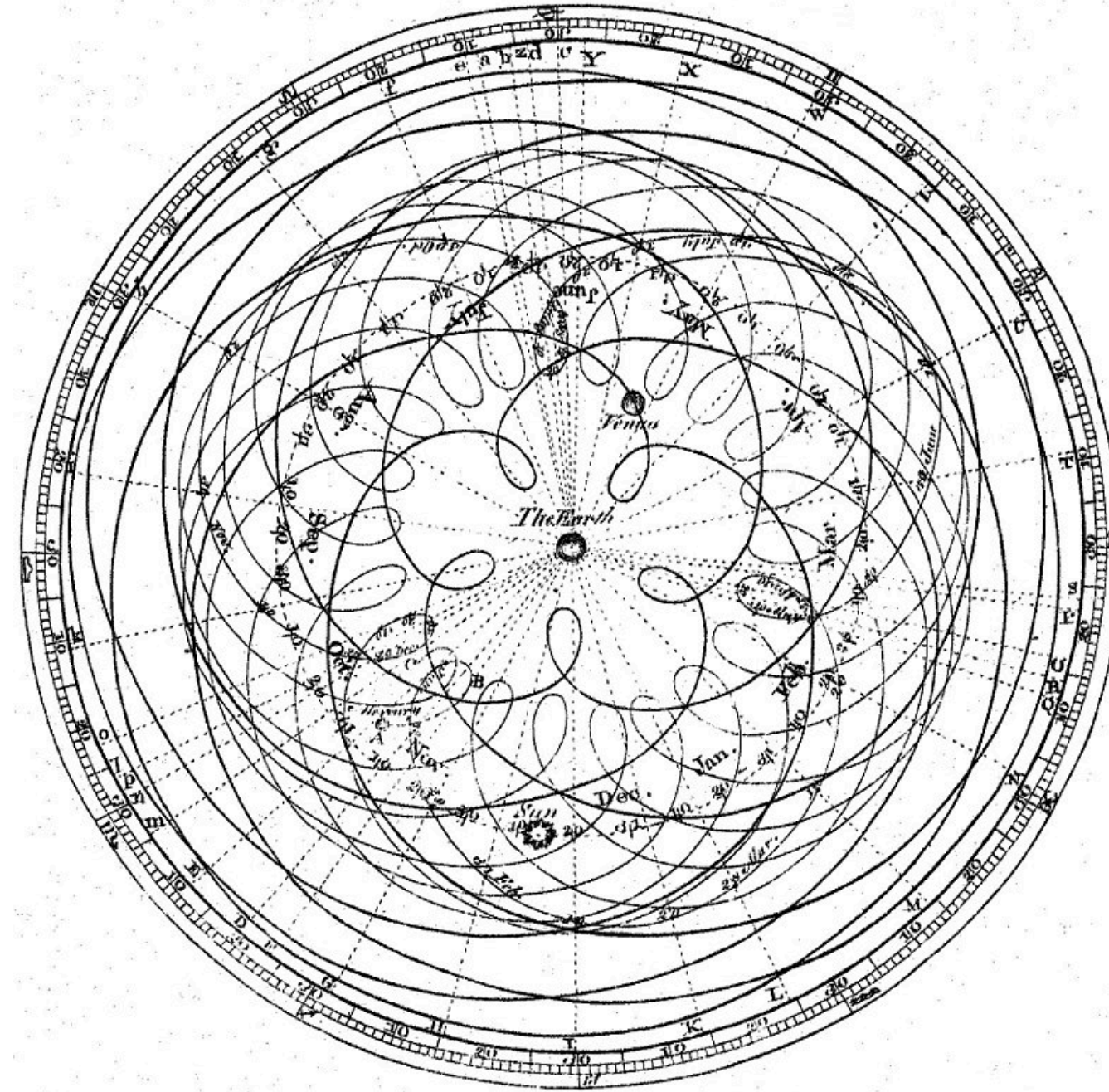
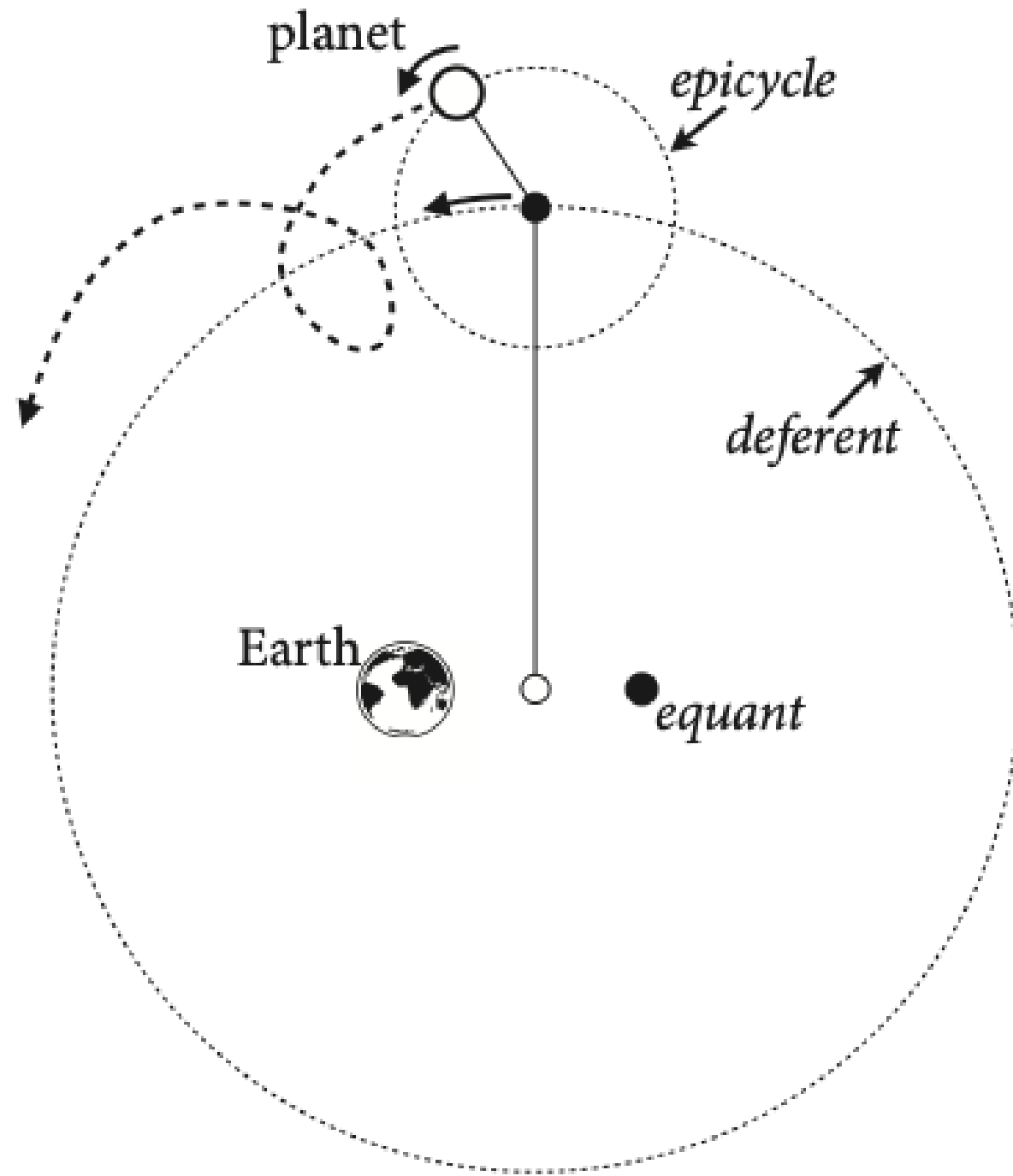
A dagger that can rid the world of superfluous entities and retain only the most essential - God.

Weight	0.8
Price	23

Screenshot aus *Kingdom Come: Deliverance II*

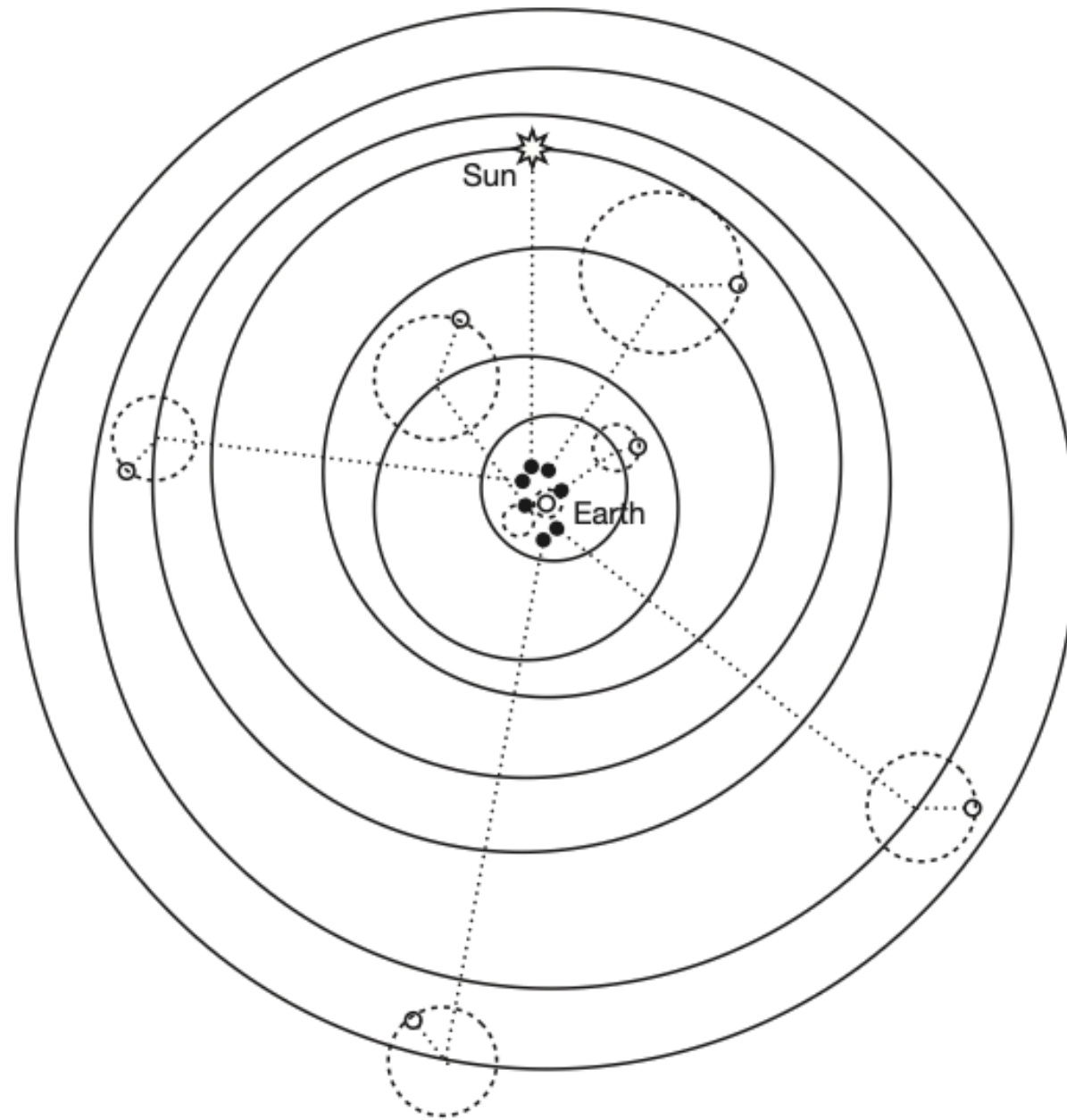


# Beispiel: Epizykeltheorie

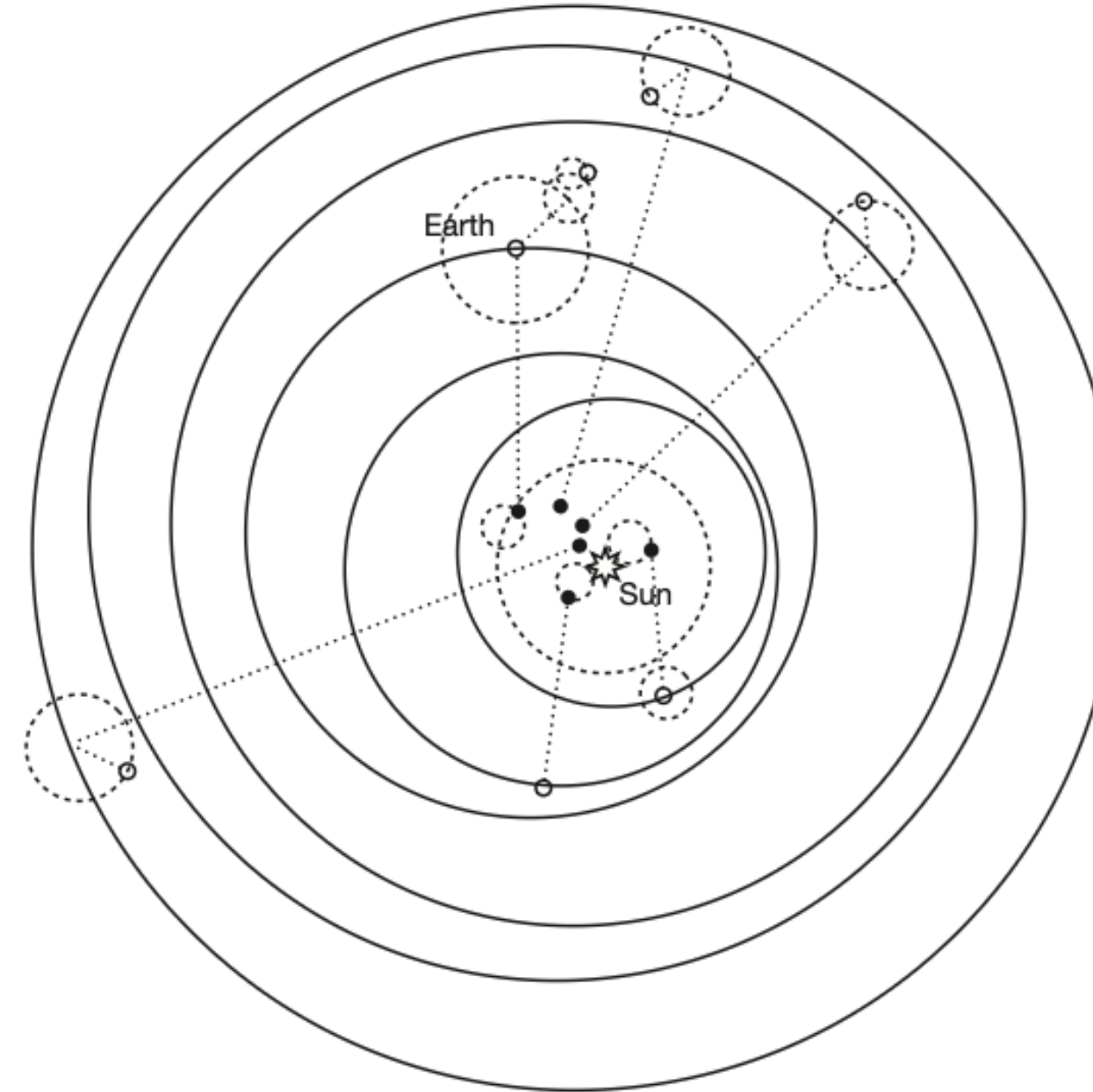


# Beispiel: Epizykeltheorie

Ptolemaic Model



Copernican Model



# Beispiel: Epizykeltheorie

Die **Epizykeltheorie** ist ein Modell, das die Umlaufbahnen von Himmelskörpern anhand von miteinander verbundenen Kreisen modelliert und vorhersagt.

Akkurate Vorhersagen sind sowohl in einem geozentrischen (ptolemäischen) als auch in einem heliozentrischen (kopernikanischen) Weltbild möglich – allerdings kommt letzteres mit deutlich weniger Epizyklen aus!

Beide Modelle sind ohnehin falsch, da die Umlaufbahnen eher elliptisch als rund sind (*Keplerbahnen*) – das kopernikanische Modell ist unter den beiden wohl dennoch das Nützlichere.



# Overfitting und Underfitting

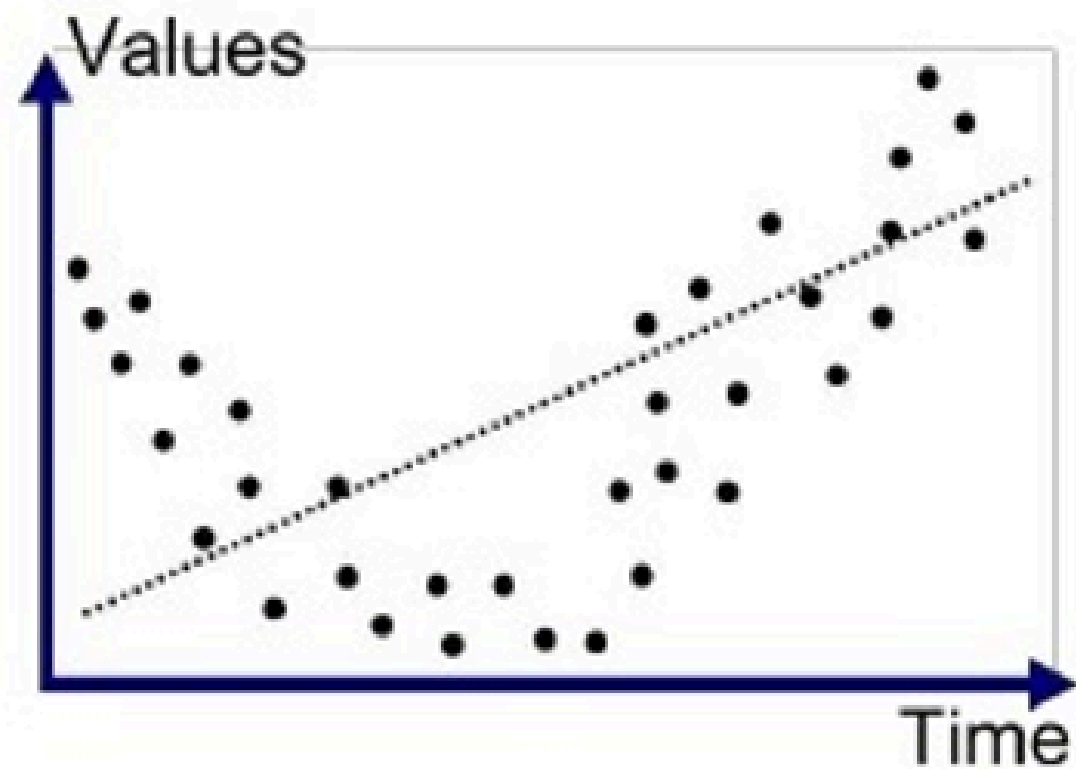
Ein gutes Modell muss einen Kompromiss zwischen den folgenden zwei Fehlerarten finden:

**Underfitting.** Das Modell passt zu schlecht zu den Daten und birgt daher zu wenig Erklärungspotenzial.

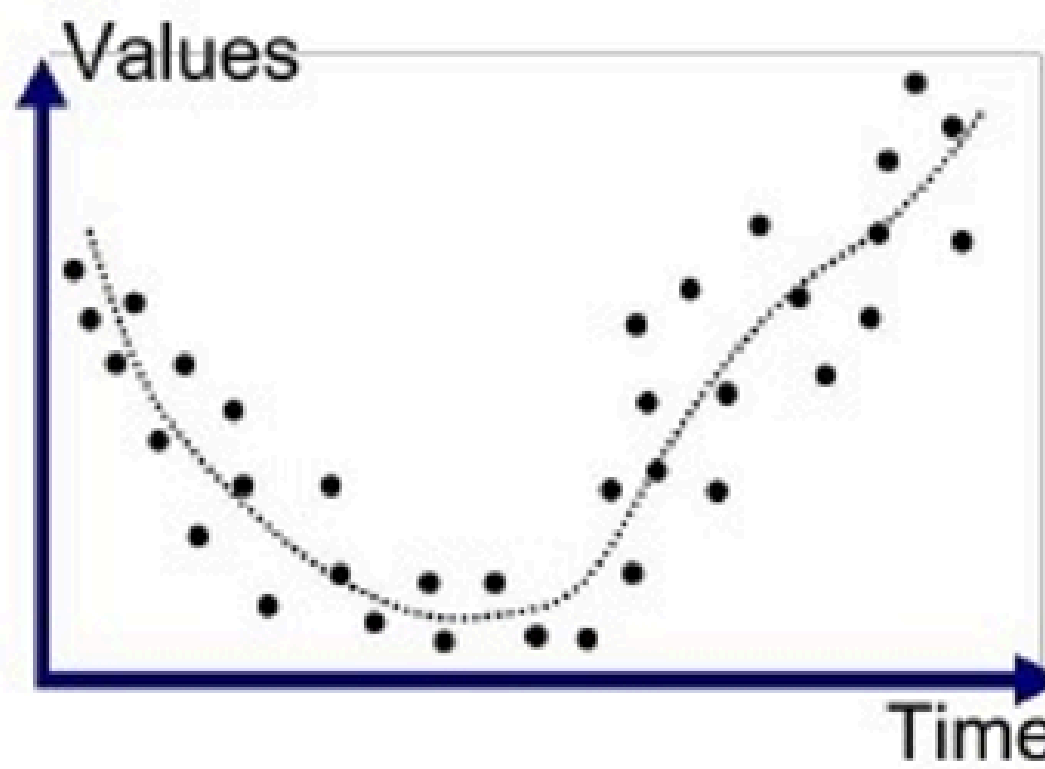
**Overfitting.** Das Modell ist zu gut auf die vorliegenden Daten zugeschnitten und ist daher nicht gut generalisierbar.



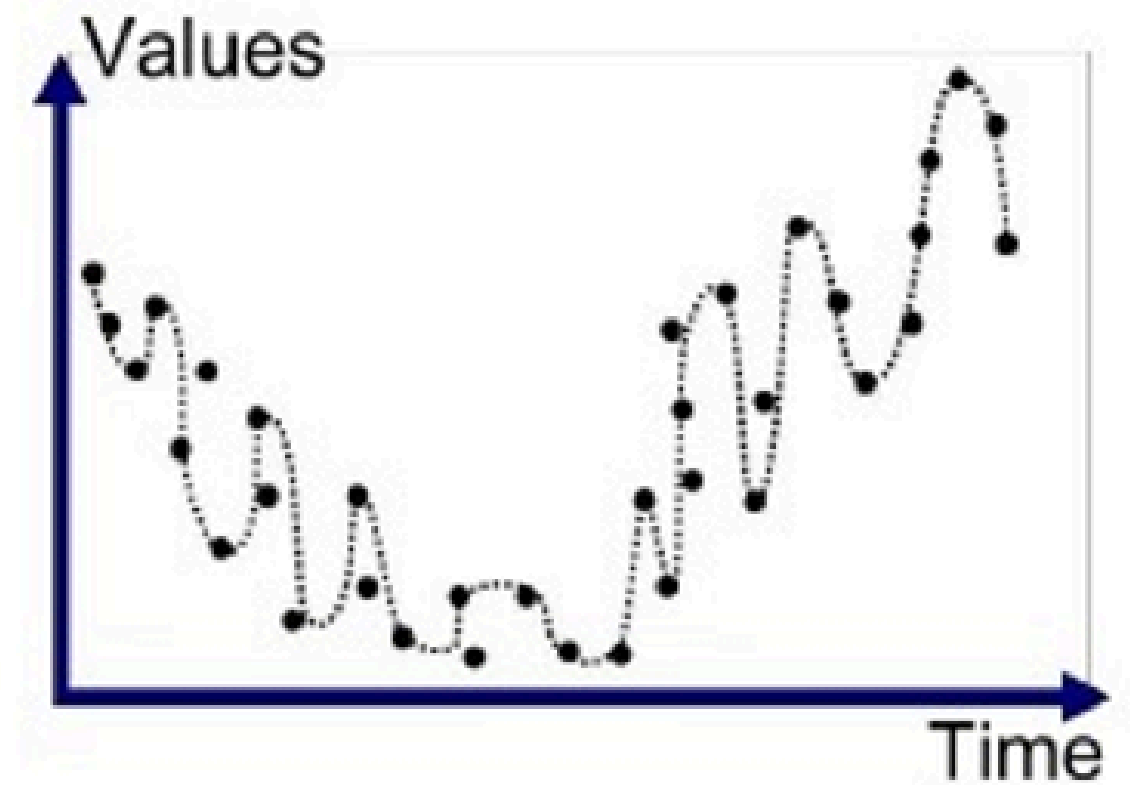
# Overfitting und Underfitting



Underfitted



Good Fit/Robust



Overfitted



# Daten

Was sind Daten?



<https://partici.fi/92489810>



# Daten

Auch beim Konzept der **Daten** gibt es ein Problem mit klaren Definitionen.

Dies hängt damit zusammen, dass der Begriff je nach Wissenschaftszweig sehr unterschiedlich interpretiert wird.

Grundsätzlich lassen sich Daten aber als “eine Art Informationseinheit im Zusammenhang mit Digitaltechnologien” verstehen (Geiger, 2023: 2/8).

In dieser Auffassung besteht gewöhnlich der Anspruch, dass Daten **maschinenlesbar** sind.



# Repräsentation von Daten

Im weiteren Sinne sind Überlegungen sinnvoll, in welcher Form wir Daten **repräsentieren** wollen. Es gibt prinzipiell beliebig viele Möglichkeiten, dieselbe Information zu repräsentieren.

Es gibt zwei Arten von Dateien auf Computern: **Textdateien** (direkt lesbar) und **Binärdateien** (nicht direkt lesbar).

Wir können Textdateien zudem in **tabellarische** und **nicht-tabellarische Daten** unterteilen.



# Repräsentation von Daten

Person	Name	Alter	Studiengang	Lieblingsbuch
1	Marion	33	Medien und Kommunikation	Harry Potter 1
2	Peter	22	Sprach- und Textwissenschaften	Der Vorleser
3	Agnes	23	Musiktheorie	Der Name der Rose
4	Theodor	28	Philosophie	Der Ekel



# Repräsentation von Daten

Person, Name, Studiengang, Lieblingsbuch

1, Marion, 33, Medien und Kommunikation, Harry Potter 1

2, Peter, 22, Sprach- und Textwissenschaften, Der Vorleser

3, Agnes, 23, Musiktheorie, Der Name der Rose

4, Theodor, 28, Philosophie, Der Ekel



# Repräsentation von Daten

Objekt	Variable	Wert
1	Name	Marion
1	Studiengang	Medien und Kommunikation
1	Lieblingsbuch	Harry Potter 1
1	Alter	33
2	Name	Peter
...	...	...



# Repräsentation von Daten

```
{  
  "1": {  
    "Name": "Marion",  
    "Studiengang": "Medien und Kommunikation",  
    "Alter": 33,  
    "Lieblingsbuch": "Harry Potter 1"  
  },  
  "2": {  
    "Name": "Peter",  
    "Studiengang": "Sprach- und Textwissenschaften",  
    "Alter": 22,  
    "Lieblingsbuch": "Der Vorleser"  
  },  
  ...  
}
```



# Analyse von Daten

Mit Daten repräsentieren wir die Objekte und Prozesse, die wir in unseren wissenschaftlichen Modellen untersuchen wollen.

Wir können Daten verwenden, um Hypothesen bzw. Modelle zu testen und zu verfeinern (**hypothesenbasierte Analyse**).

Alternativ können wir Daten anhand von Modellen untersuchen, um Neues über die Daten zu lernen (**explorative Analyse**).



# Beispiel: Berechnung von Inflation

**Inflation** ist ein Konstrukt, das erklärt, wie Geld über Zeit an Wert und an Kaufkraft verliert. Sie ist intuitiv sehr leicht nachzuvollziehen, aber schwer direkt zu messen.

**Modell:** Fiktiver Warenkorb aus 650 verschiedenen Gütern mit unterschiedlicher Gewichtung.

**Daten:** Preise für die o.g. Güter pro Jahr.

**Analyse:** Berechnung der Preisentwicklung des Warenkorbs.  
*(aber auch: Anpassung des Warenkorbs und der Gewichtung!)*



# Beispiel: Berechnung von Inflation

**Inflation** ist also ein sehr komplexes wirtschaftliches Konstrukt, dem sich die Forschung systematisch versucht anzunähern. Wir können aber nicht direkt evaluieren, wie akkurat oder nützlich das Modell ist.

Wie genau die Daten hierbei repräsentiert und verarbeitet werden, ist nicht bekannt. Es sind aber durchaus einige Möglichkeiten denkbar. (*Excel-Tabellen, SQL-Datenbanken, JSON-Objekte, ...*)



# Aufgabe: Modelle zur Textverarbeitung

*“Wenn Robben hinter Robben robben, robben Robben hinter Robben.”*

Wie viele Wörter stecken in diesem Satz?

Wie häufig kommt welches Wort vor?

Können wir solche Fragen für gesamte Bucherreihen beantworten?

**Erstelle ein Notebook auf [kaggle.com](https://www.kaggle.com) und versuche, eine Python-Implementierung zur Beantwortung dieser Fragen zu erstellen.**

