

Laborexperimente in der sozialwissenschaftlichen Forschung

Joël Berger

Forschungsseminar “Just My Luck!”
Institut für Soziologie, Universität Zürich
13.03.2017



Was ist ein Experiment?

Kausalität: Kontrafaktisches Modell



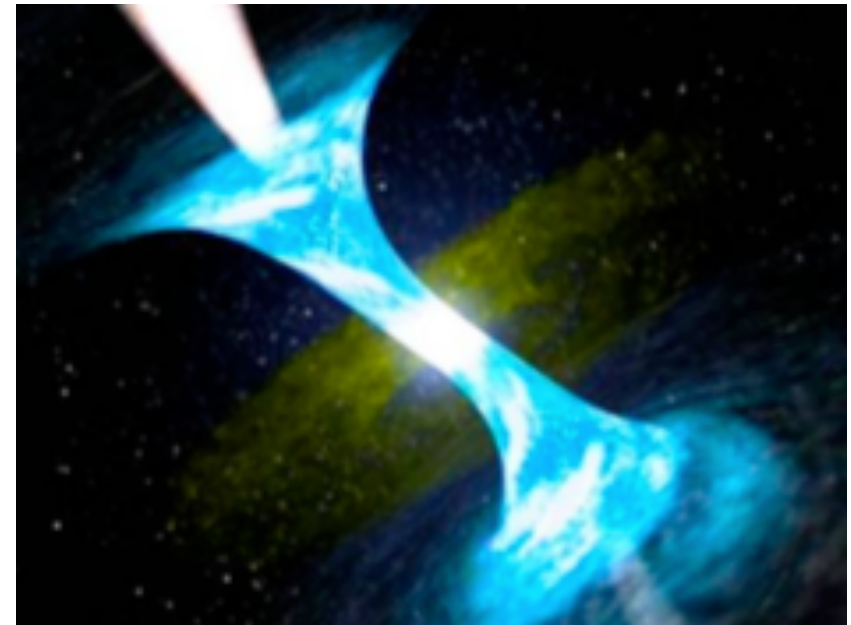
John Stuart Mill (1872):

»If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or the cause, or an indisputable part of the cause, of the phenomenon.«

Kontrafaktisches Modell: Idee (*Rubin causal model*)

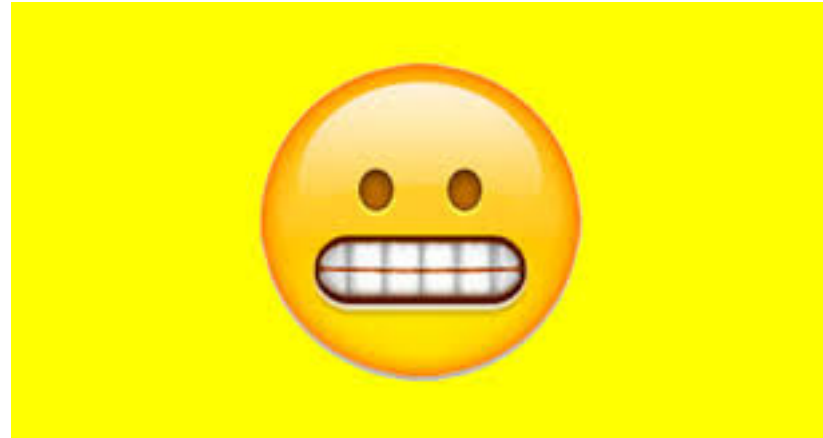
- Zwei Zustände sind notwendig für einen kausalen Schluss
- “If an hour ago, I had taken two aspirins instead of just a glass of water, my headache would now be gone.”
- “Because an hour ago I took two aspirins instead of just a glass of water, my headache is now gone.”

(Rubin 1974, vgl. Morton & Williams 2010)

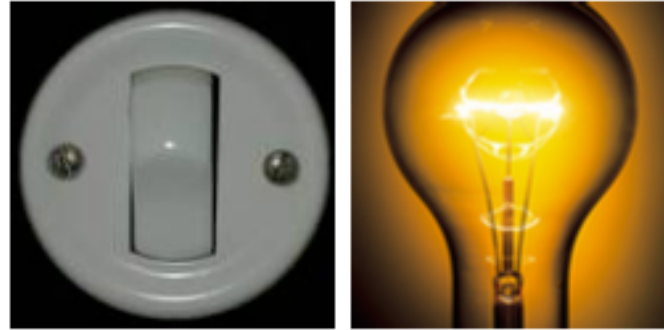


Fundamentales Problem der kausalen Inferenz

- Beide Zustände können nicht zugleich beobachtet werden.



Mögliche Lösung 1: Annahme der Einheitshomogenität (*unit homogeneity*)

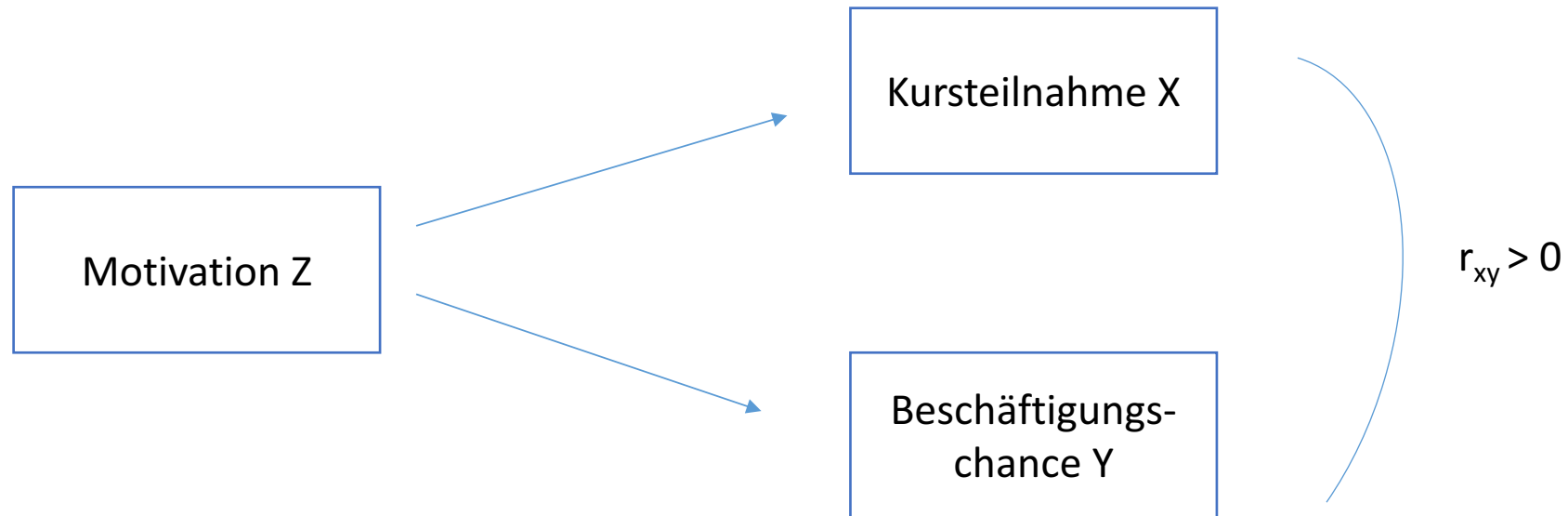


Das neue Arbeitsmarktintegrations-Programm

Die klassische soziologische Herangehensweise

		Programmteilnahme	
		0 (nein)	1 (ja)
Arbeitslosigkeitsdauer	> 6 Monate	500	0
	max. 6 Monate	0	500

Das neue Arbeitsmarktintegrations-Programm

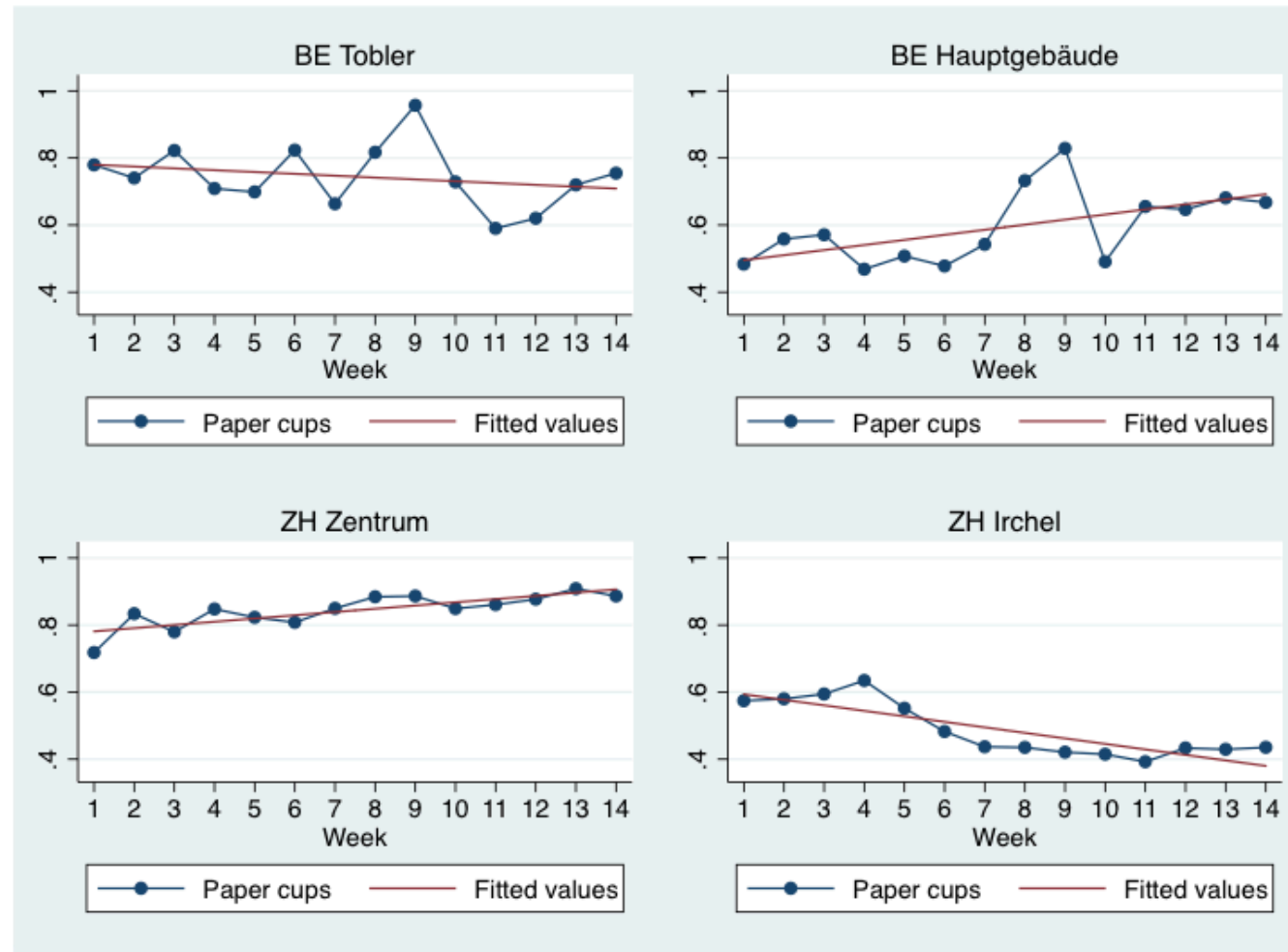


- *Unit homogeneity* verletzt
- Z ist ein *confounder* (Störvariable)

Mögliche Lösung 2:
Annahme der Zeitkonstanz (*temporal stability*)



Beispiel: Tasse statt Becher



Problem

- Beide Annahmen (*unit homogeneity, temporal stability*) sind nicht beweisbar.
- *Unit homogeneity*: Unbeobachtete oder gar unbekannte Störvariablen?
- *Temporal stability*: Reifungsprozesse, parallel ablaufende Prozesse, Reihenfolge-Effekte

(Diekmann 2006)

Lösung: Randomisierung



- Werden die VP auf eine Treatment und eine Kontrollgruppe zufällig zugeteilt, sind im Durchschnitt in beiden Gruppen für Y kausal relevante Faktoren Z_i gleich verteilt.
- Die Gruppen unterscheiden sich in der Erwartung nicht bis auf den Prüffaktor X (*unit homogeneity* auf Gruppenebene gegeben).

Lösung: Randomisierung

- Geprüft wird die kausale Relevanz eines Prüffaktors X für das interessierende Merkmal Y auf Gruppenebene.
- Geprüft wird somit der durchschnittliche kausale Effekt (*average treatment effect*).

Was ist ein Experiment

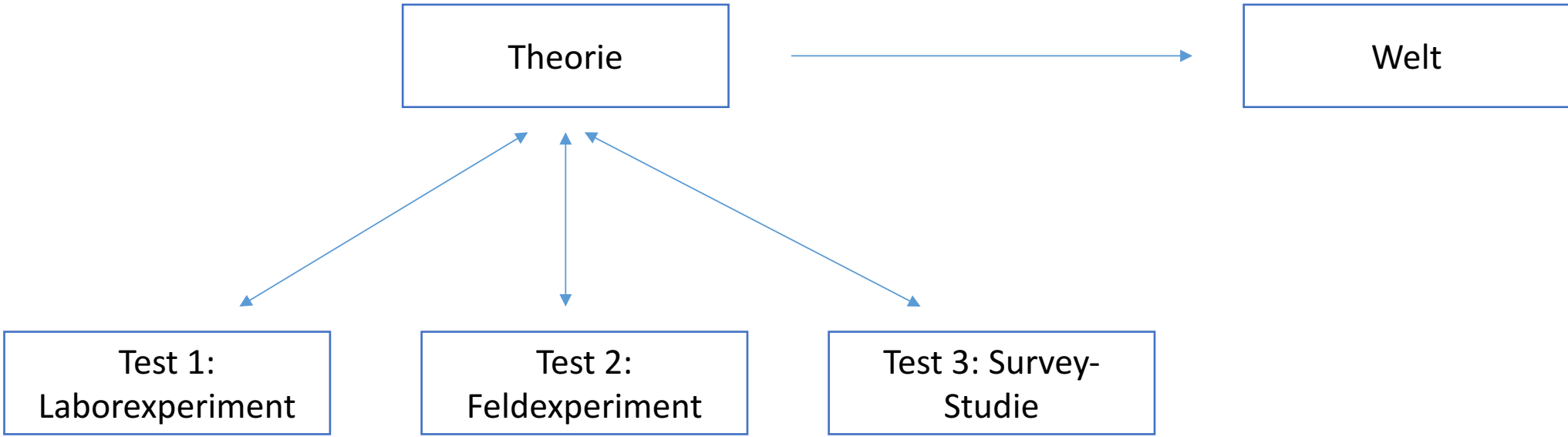
Definition

1. Mindestens zwei Gruppen (Versuchs- und Kontrollgruppe)
2. Randomisierung der VP auf die beiden Gruppen
3. Manipulation der unabhängigen Variablen durch Forscher/in

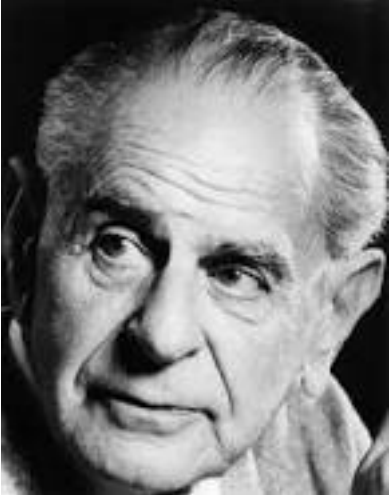
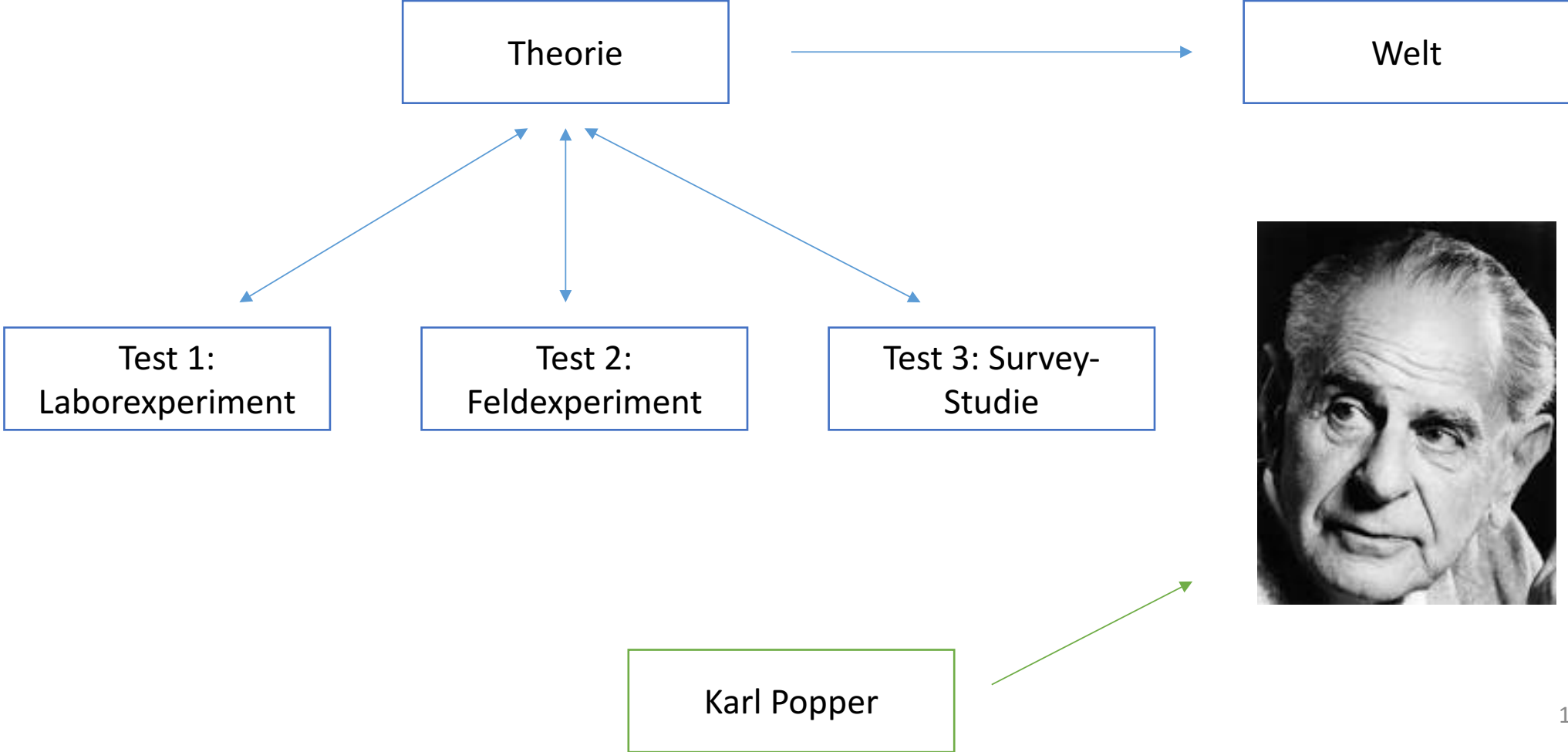
(Diekmann 2006)

R		X	O	Versuchsgruppe
R			O	Kontrollgruppe

Theorie, Experiment, Welt

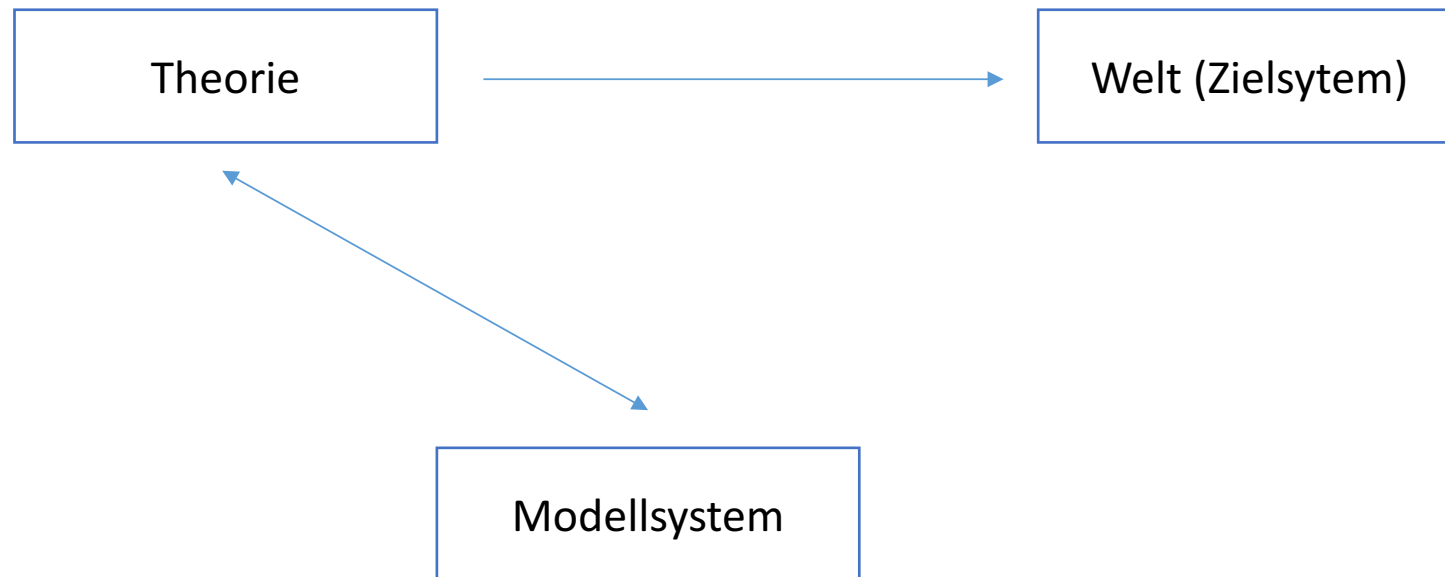


Theorie, Experiment, Welt



Experimente und Modelle

- Oft kann ein Zielsystem nicht experimentell untersucht werden.
- Ein strukturell ähnliches Modell-System wird untersucht.



Experimente und Modelle

Modell-System (*model system*)



(Porsolt et al. 1977)

Ziel-System (*target system*)



Gütekriterien von Modellen

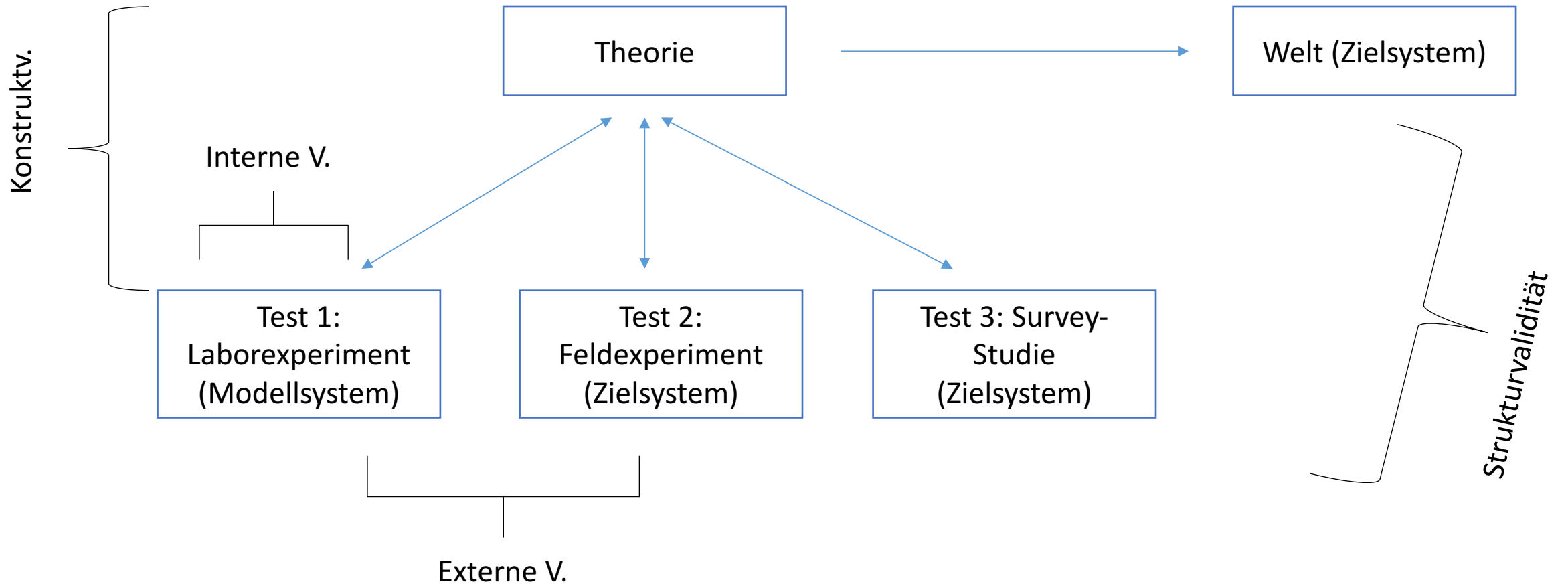
- **Face validity**: Ausmass der Ähnlichkeit zwischen Modellsystem und Zielsystem.
 - Maus und Mensch haben zwei Augen.
- **Structural validity**: Ausmass der Ähnlichkeit der kausalen Strukturen im Modellsystem und im Zielsystem.
 - Substanz X erhöht den Antrieb in Maus und Mensch.
- Relevant ist die strukturelle Validität, nicht die *face validity*.

(vgl. Frigg & Hartmann 2009, Suárez 2010, Willer & Walker 2007)

Gütekriterien von Experimenten

- *We use the term **intern validity** to refer to inferences about whether observed covariation between A and B reflects a causal relationship from A to B.*
(Shadish et al. 2002: 83)
- ***External validity** concerns inferences about the extent to which a causal relationship holds over variations in person, settings, treatments, and outcomes.*
(Shadish et al. 2002: 53)
- **Construct validity:** Ausmass, zu dem die im Experiment gewählte Operationalisierung eines theoretischen Begriffs dem theoretischen Begriff tatsächlich entspricht.

Theorie, Experiment, Welt



Late School Tracking, Less Class Bias in Educational Decision-Making? The Uncertainty Reduction Mechanism and Its Experimental Testing

Joël Berger^{1,2,*} and Benita Combet^{3,*}

¹Department of Sociology, Utrecht University, The Netherlands, ²Institute of Sociology, University of Zurich, Switzerland and ³Institut des Sciences Sociales, LIVES—Swiss National Centre of Competence in Research, University of Lausanne, Switzerland

*Corresponding author. Email: j.berger@uu.nl; benita.combet@unil.ch

Note: Shared first authorship (the ordering of names has been determined randomly).

Submitted March 2016; revised October 2016; accepted October 2016

Abstract

A long-standing hypothesis in the sociology of education is that the timing of ability tracking impacts the inequality of educational opportunity. While earlier studies mainly focused on how early tracking impacts the primary effect of social origin (systematic performance differences due to social background), the impact of early tracking on the secondary effect of social origin (class-specific educational decision-making) has been neglected. Recently, the idea has been put forward that late tracking decreases uncertainty in educational decision-making, thus enabling more rational decision-making by lower-class individuals. Extending this idea, we propose the uncertainty reduction mechanism (URM) as a theoretical foundation; this mechanism can be derived from a decision-theoretic model on educational decision-making based on prospect theory. Moreover, we perform a first empirical test of the URM by means of a computerized laboratory experiment. The evidence is in line with the model predictions and the results support the intuition that a postponement of tracking could reduce the negative bias of lower-class individuals in educational decision-making, thereby reducing educational inequality with respect to social background.

Schichtspezifische Bildungsungleichheit

- Personen aus höheren sozialen Schichten nehmen öfters an Hochschulbildung teil als Personen aus tieferen Schichten.
- Der Zusammenhang ist nur teilweise durch schichtspezifische Performanz (primärer Herkunftseffekt) erklärbar.
- Ebenso relevant: Schichtspezifische Bildungsentscheidungen (sekundärer Herkunftseffekt)
- Problem:
 - Konflikt mit Ideal der Meritokratie
 - Verlust von Talent/Humankapital

Rolle des Bildungssystems

- Zusammenhang zwischen sozialer Herkunft und Teilnahme an Hochschulbildung variiert im internationalen Vergleich.
- Sekundärer Herkunftseffekt variiert stärker als der primäre.
- Je später die leistungsabhängige Zuteilung der Schüler/innen auf verschiedene Schultypen (Tracking), desto geringer ist der Zusammenhang zwischen sozialer Schicht und Bildungserfolg.
- Problem: Bekannte (und unbekannte) Störvariablen (*confounders*)

Mögliche Erklärung: Unsicherheitsreduktion

- Der Unsicherheitsreduktions-Mechanismus (URM):

We would expect transitions taken at older ages to be less susceptible to secondary effects because the amount of information available to students about their own abilities and their chances of success at higher levels of education must only increase as their exposure to formal education increases.

(Jackson & Jonsson 2013: 332)

Theoretisches Modell I: Statuserhalt und Framing

- Soziale Schicht als Referenzpunkt bei der Bildungsentscheidung
- Annahme: Zwei soziale Schichten, zwei Tracks
- Höhere Schicht
 - Höhere Bildung notwendig für Statuserhalt → Verlust-Frame → Risikosuchendes Entscheidungsverhalten
- Tiefere Schicht
 - Höhere Bildung nicht notwendig für Statuserhalt → Gewinn-Frame → Risikoaverses Entscheidungsverhalten
- $B_B = V(B_B)$ und $B_H = V(B_{HS}, \pi[P]; BHF, 1 - \pi[P])$

(Page 2005, Kahneman and Tversky 1979)

Theoretisches Modell II: Tracking und Unsicherheit

- Je früher das Tracking, desto weniger Information über seine Performanz hat ein Akteur und desto länger der Zeitraum, über den eine Vorhersage getroffen werden muss.
- Bei einer Entscheidung für oder gegen weiterführende Bildung basiert die subjektive Erfolgswahrscheinlichkeit auf der bisherigen Performanz → Grössere Unsicherheit bei frühem Tracking
- Personen in einem Gewinn-Frame (tiefe Schicht) haben eine grössere Abneigung gegenüber Unsicherheit als Personen in einem Verlust-Frame. → Personen im Gewinn-Frame profitieren überproportional von spätem Tracking
- $B_H = V(B_{HS}, \pi[P]; B_{HF}, 1 - \pi[P])$ mit $\pi[P] = E(P) - \lambda\sigma$

Experimentelles Design: Task



- Performanz-Mass: Zahl der gelösten Anagramme

Level 1 - Set 1

DELH	<input type="text" value="Held"/>	<input type="button" value="Speichern"/>	gespeichert
LEAI	<input type="text" value="Laie"/>	<input type="button" value="Speichern"/>	nicht gespeichert
OKCD	<input type="text"/>	<input type="button" value="Speichern"/>	nicht gespeichert
OLEFI	<input type="text"/>	<input type="button" value="Speichern"/>	nicht gespeichert
AOSR	<input type="text"/>	<input type="button" value="Speichern"/>	nicht gespeichert

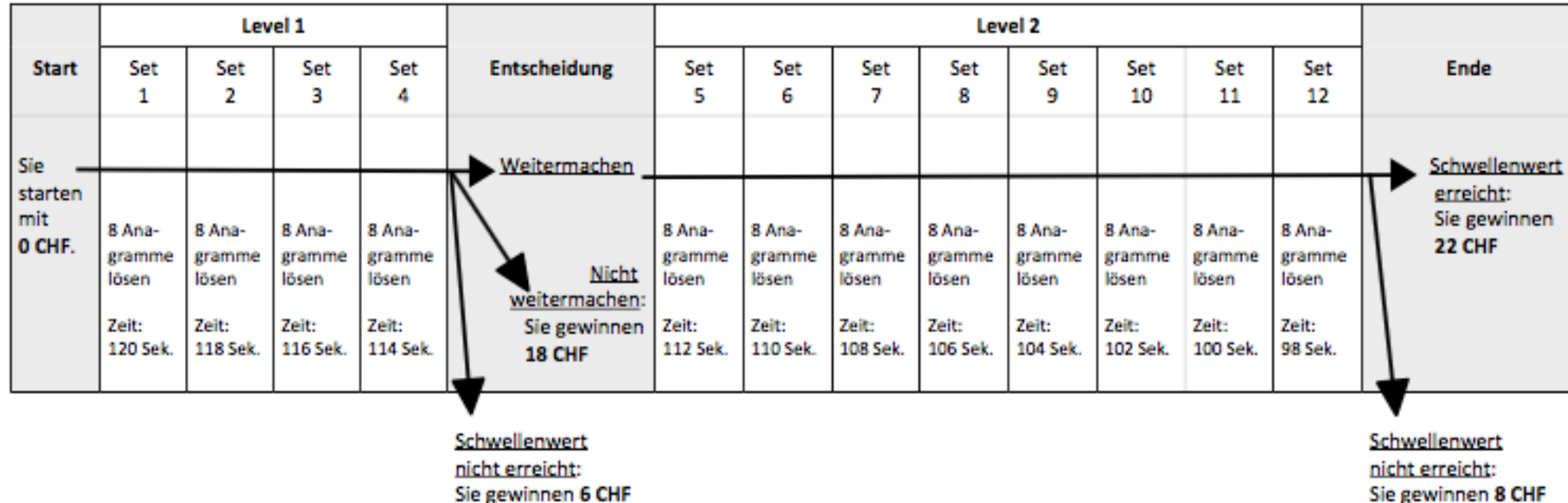
Experimentelles Design: Faktoren

		Frame	
		gain	loss
Decision	early	40	41
	late	42	42

N=165

Experimentelles Design: Instruktionen

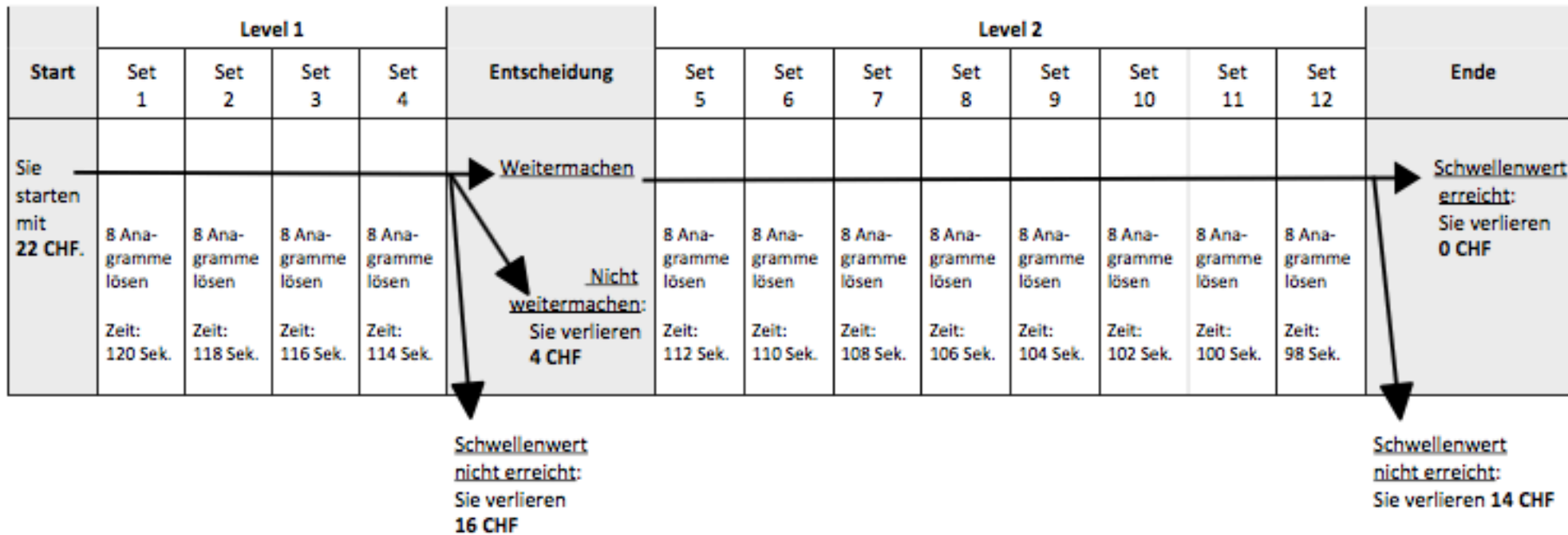
Early decision/gain frame-Versuchsbedingung



→ Zusätzlich erhalten Sie 15 CHF Show-Up-Fee

Experimentelles Design: Instruktionen

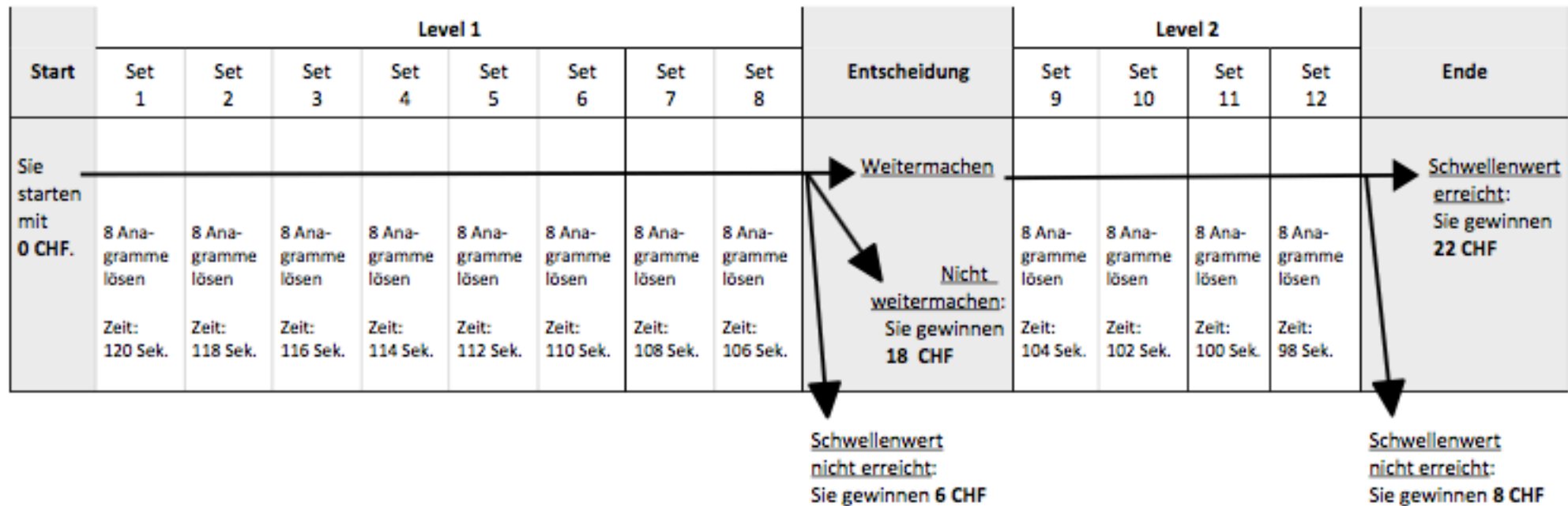
Early decision/loss frame-Versuchsbedingung



→ Zusätzlich erhalten Sie 15 CHF Show-Up-Fee

Experimentelles Design: Instruktionen

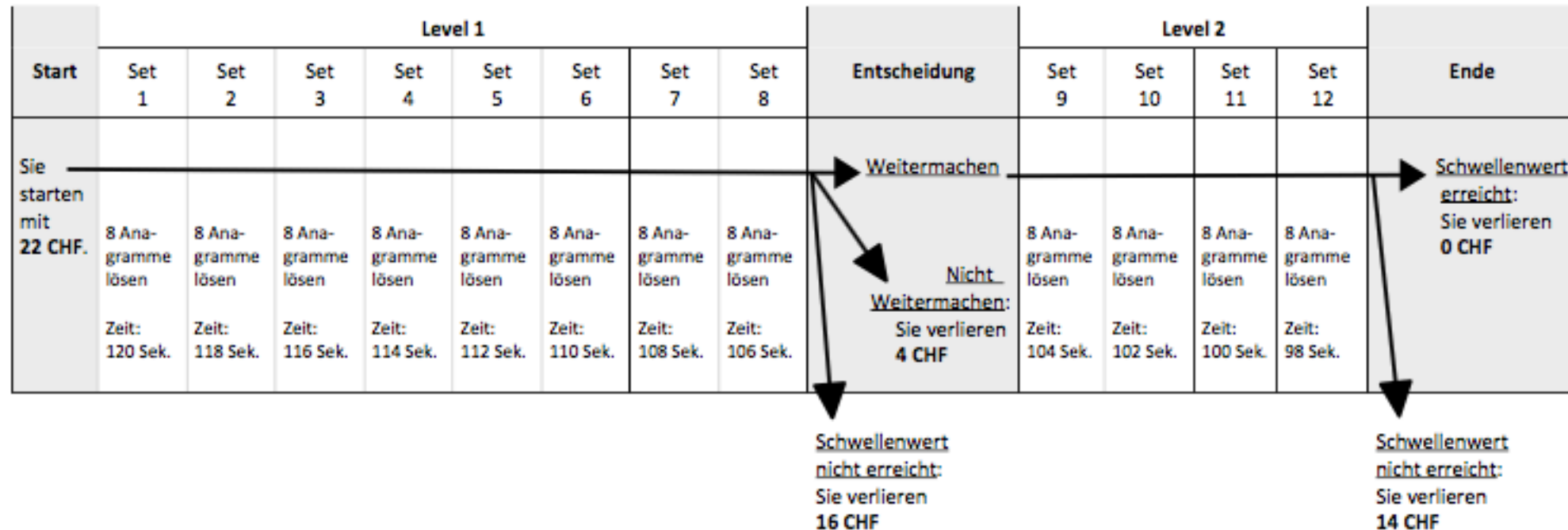
Late decision/gain frame-Versuchsbedingung



→ Zusätzlich erhalten Sie 15 CHF Show-Up-Fee

Experimentelles Design: Instruktionen

Late decision/loss frame-Versuchsbedingung



→ Zusätzlich erhalten Sie 15 CHF Show-Up-Fee

Aspect of the design	Real-world counterpart	Note
Level	School type	
Stage	School year	
Anagram	Subject material	
Decreasing time	Increasing standard	
Payoffs	Utility	Utility derived from monetary and non-monetary benefits (e.g. social status) minus monetary costs (including opportunity costs) and non-monetary costs (including psychological costs, e.g. when failing to complete the more demanding track)
Loss frame	High aspiration level/motive for maintaining economic, social and cultural status quo	Higher education necessary for maintenance of status
Gain frame	Low aspiration level	No higher education necessary for maintenance of status Frame in general: social classes differ across a huge variety of factors and sociologists usually distinguish between economic and cultural factors of social origin. Prospect theory is very general; it predicts that the loss of a given amount of a certain resource (be it economic, such as money, or cultural, such as social capital) leads to a greater negative utility than the degree to which the gain of the very same amount of the same resource leads to positive utility. In the laboratory, a different aspiration level is induced by framing monetary rewards either as gains or losses.

Resultate: Haupteffekte

Probability of continuing (AME)	
Loss frame	.18*
(ref.: gain frame)	(.08)
Late decision	.03
(ref.: early decision)	(.08)
N=137, * $p < .05$, se in parentheses	

- **Frame:** Starker Haupteffekt
 - VP im Verlust-Frame machen mit einer um 18 Prozentpunkte höheren Wahrscheinlichkeit weiter ($p = .028$).
- **Zeitpunkt der Entscheidung:** kein Effekt
 - $p = .752$

Resultate: H1

		Overall	Performance	
			low	high
Early decision	Gain frame	0.249 (0.074)	0.164 (0.084)	0.364 (0.126)
	Loss frame	0.457 (0.084)	0.315 (0.097)	0.742 (0.129)
Late decision	Gain frame	0.302 (0.078)	0.168 (0.076)	0.661 (0.161)
	Loss frame	0.440 (0.079)	0.452 (0.098)	0.419 (0.142)

- **H1:** VP mit hoher Performanz im Gain-Frame machen mit höherer Wahrscheinlichkeit bei später als bei früher Entscheidung weiter.
- VP mit hoher Performanz im Gain-Frame machen bei später Entscheidung fast doppelt so häufig weiter (66.1% vs. 36.4%, $p=.087$).

Resultate: H2

		Overall	Performance	
			low	high
Early decision	Gain frame	0.249 (0.074)	0.164 (0.084)	0.364 (0.126)
	Loss frame	0.457 (0.084)	0.315 (0.097)	0.742 (0.129)
Late decision	Gain frame	0.302 (0.078)	0.168 (0.076)	0.661 (0.161)
	Loss frame	0.440 (0.079)	0.452 (0.098)	0.419 (0.142)

- **H2:** VP mit tiefer Performanz im Gain-Frame machen nicht mit höherer Wahrscheinlichkeit bei später als bei früher Entscheidung weiter.
- Kein Unterschied: (16.8% vs. 16.4%, $p=.49$)

Resultate: H3

		Overall	Performance	
			low	high
Early decision	Gain frame	0.249 (0.074)	0.164 (0.084)	0.364 (0.126)
	Loss frame	0.457 (0.084)	0.315 (0.097)	0.742 (0.129)
Late decision	Gain frame	0.302 (0.078)	0.168 (0.076)	0.661 (0.161)
	Loss frame	0.440 (0.079)	0.452 (0.098)	0.419 (0.142)

- **H3:** Die Wahrscheinlichkeit weiterzumachen unterscheidet sich für VP im Loss-Frame nicht zwischen den Entscheidungszeitpunkten.
- Kein Unterschied im Gesamtsample oder bei VP mit tiefer Performanz.
- Unerwartet: Personen mit hoher Leistung machen mit einer geringeren Wahrscheinlichkeit weiter bei später Entscheidung.

Resultate: H4

		Overall	Performance	
			low	high
Early decision	Gain frame	0.249 (0.074)	0.164 (0.084)	0.364 (0.126)
	Loss frame	0.457 (0.084)	0.315 (0.097)	0.742 (0.129)
Late decision	Gain frame	0.302 (0.078)	0.168 (0.076)	0.661 (0.161)
	Loss frame	0.440 (0.079)	0.452 (0.098)	0.419 (0.142)

- **H4:** Unter den VP mit hoher Leistung ist die Differenz zwischen den VP im Loss-Frame die weitermachen und den VP im Gain-Frame die weitermachen geringer bei später als bei früher Entscheidung.
- Signifikanter Unterschied zwischen VP mit hoher Leistung bei früher ($p=.03$), aber nicht bei später Entscheidung ($p=.14$).

Diskussion

- Hauptbefund: VP mit hoher Leistung im Gain-Frame machen eher weiter wenn Entscheidungen spät getroffen werden.
- Befund stützt die URM-Hypothese, und damit die Hypothese, dass spätes Tracking schichtspezifische Ungleichheiten reduzieren könnte.
- Limitation: Geringe Fallzahl, erster Test unter idealisierten Laborbedingungen
- Weiterführende Forschung: Feld- und Surveystudien

Laborexperiment: Charakterisierung



- Maximale **Kontrolle** von Störfaktoren: Experimentaldesign als Computerprogramm; soziale Interaktion anonym über Computernetzwerk
- **Verhaltensbeobachtung**: Incentives als Verhaltensanreiz
- **Randomisierung**
- Nähe Theorie – Design, kaum Kontext
- VP: **Studierende**

Laborexperiment: Vor- und Nachteile

+	-
Kontrollierte Umgebung, relativ homogene Gruppe der Studierenden als VP, Randomisierung → Hohe Kontrolle von Störfaktoren (hohe interne Validität)	Lebensfern, artifiziell → Gefahr von Artefakten
Theoretisch relevante Situationen können fast beliebig hergestellt werden.	Lebensfern, artifiziell → Gefahr von Artefakten
Hohe Standardisierung → Reproduzierbarkeit	Lebensfern, artifiziell → Gefahr von Artefakten
Anreizkompatibles Verhalten	Geld-Priming
Umfangreiche Datensammlung möglich (bewusste Entscheidungen, Verhaltensspuren, Befragung)	Reaktivität

Bildungsentscheidungen



Aufgabe

1. Modell- und Zielsystem in der Studie?
2. Beurteilen Sie
 - Konstruktvalidität
 - *Face validity*
 - Strukturelle Validität
 - Interne Validität
 - Externe Validität
3. Was könnten weiterführende komplementäre Studien sein?

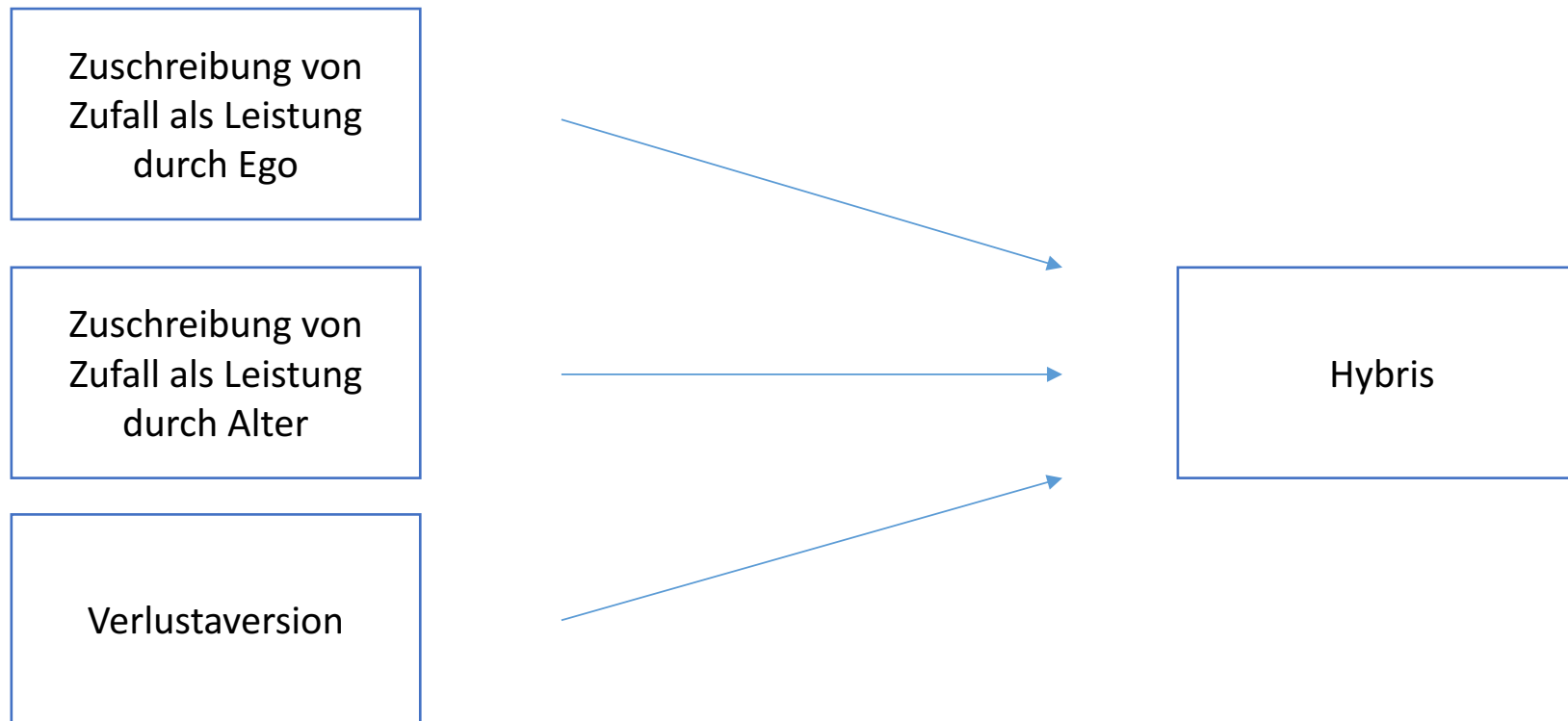
Take Home: Laborexperimente

- Randomisierung → Kausalität
- Verhalten
- Test von Theorien
- Experimente: oft stärker modellbasiert als andere Verfahren
- Gütekriterien: Konstruktvalidität, interne Validität, externe Validität
- I.d.R. Beginn eines Forschungsprogramms
- Schwächen: Artifizuell, Reaktivität

Übung: Experimentaldesign zur Hybris-Problematik

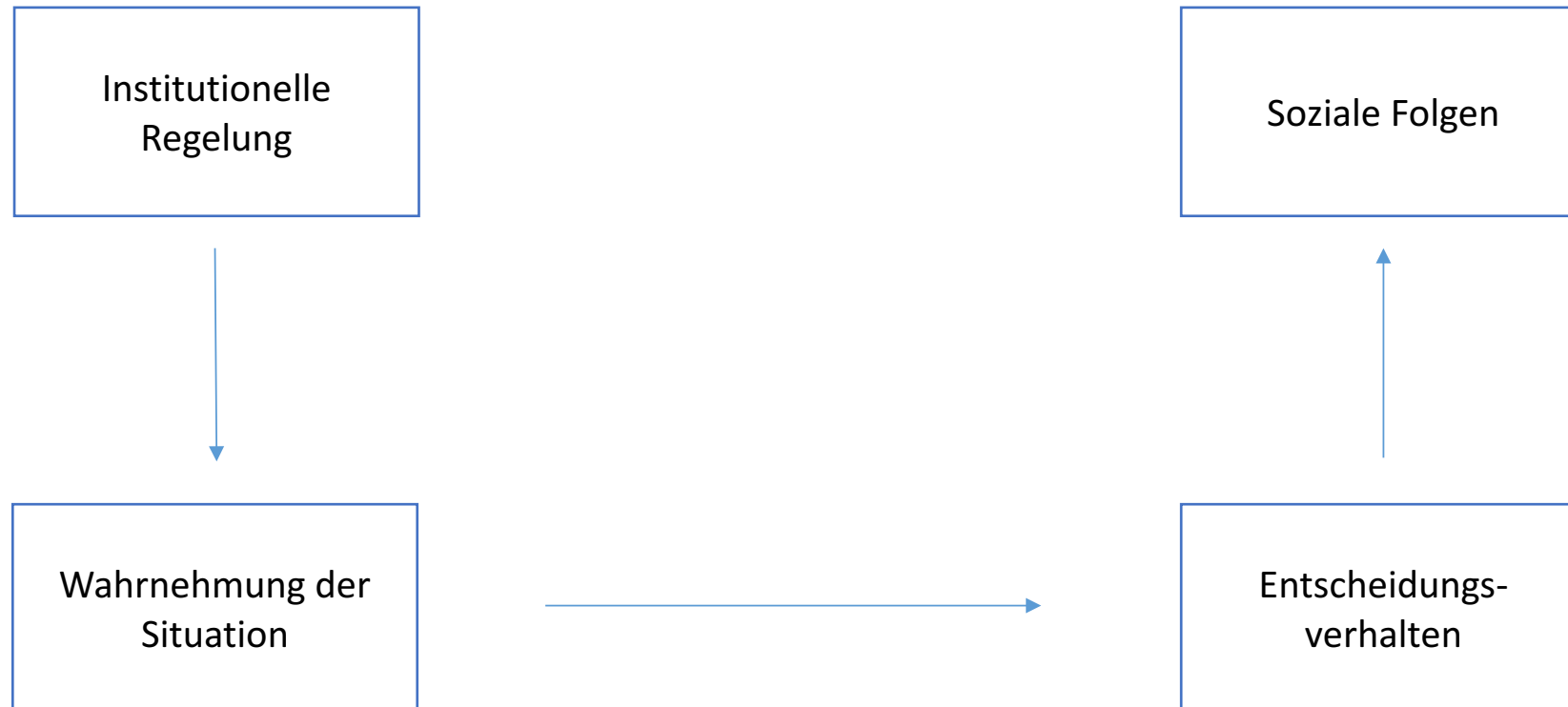
- Stärke von Laborexperimenten: Test von Theorien.
- Bedingung: Theorie ist exakt formuliert.
 - Ideal: Mathematisches Modell
 - Alternative: Kausaldiagramm

Übung: Experimentaldesign zur Hybris-Problematik



Übung: Experimentaldesign zur Hybris-Problematik

- Parallelen zum Experiment von Berger & Combet (2017)



Übung: Experimentaldesign zur Hybris-Problematik

Aufbau des Referats

- Kontext/Framing
- Theorie (möglichst präzise, eventuell Kausaldiagramm)
- Experimentelles Design
 - Versuchsaufbau: Leistungstask, Framing des Experiments, Parallelen Modellsystem/Zielsystem
 - Versuchsbedingungen
 - Ungefähre Zahl der Versuchspersonen
 - Höhe der Auszahlung (insgesamt 30 Fr., 10 Fr. Show up fee).
 - Beachte: *Online Supplementary Data* zu Berger & Combet (2017)

Literatur I

- Berger J, Combet B (2017): Late School Tracking, Less Class Bias in Educational Decision-Making? The Uncertainty Reduction Mechanism and Its Experimental Testing. *European Sociological Review* 33: 124–136.
- Diekmann A (2006): *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*. Reinbek bei Hamburg: Rowohlt.
- Frigg R, Hartmann S (2009): Models in Science. *The Stanford Encyclopedia of Philosophy*. Edited by Zalta, Edward N. <http://plato.stanford.edu/archives/sum2009/entries/models-science/>
- Morton R, Williams K (2010): *Experimental Political Science and the Study of Causality: From Nature to the Lab*. Cambridge, Cambridge University Press.
- Porsolt RD, Bertin A, Jalfre M (1977): Behavioral Despair in Mice: A Primary Screening Test for Antidepressants. *Archives Internationales De Pharmacodynamie Et De Therapie* 229: 327-336.

Literatur II

- Rubin DB (1974): Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology* 66: 688-701.
- Shadish WR, Cook TD, Campbell D (2002): *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.
- Suárez M (2010): Scientific Representation. *Philosophy Compass* 5: 91-101.
- Webster M, Sell J (2007): *Laboratory Experiments in the Social Sciences*. Amsterdam, Elsevier.
- Willer D, Walker HA (2007): *Building Experiments. Testing Social Theory*. Stanford: Stanford University Press.

Appendix

		Number of anagrams solved L1	Frame	
			Gain	Loss
Performance				
Timing of decision	Early	$Nr \geq Q_{0.75}$	14	11
		$Nr < Q_{0.75}$	26	30
	Late	$Nr \geq Q_{0.75}$	9	12
		$Nr < Q_{0.75}$	33	30
Training at home				
Timing of decision	Early	Yes	10	10
		No	30	31
	Late	Yes	10	16
		No	32	26
Experimental training				
Timing of decision	Early	Yes	23	24
		No	17	17
	Late	Yes	22	22
		No	20	20