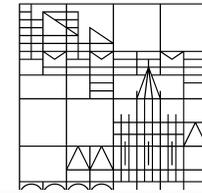


Universität
Konstanz



Workshop II

Lab Experiments



Prof. Dr. Ulrike Stefani

Hamburg, February 8, 2017

Content

Experimental Economics

- History
- Definition
- Control
- Research Questions
- Points of Criticism

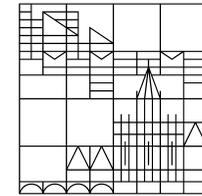
Behavioral Accounting

- Experiments in Accounting
- Design Choices
- Collecting Process Data
- Achieving a Good Design

Examples

- Strategic Interaction and Testing Theories (Lab): Fischbacher / Stefani (2007), TAR
- Strategic Interaction and Comparison of Institutions (Lab):
 - Hattenbach / Stefani (2016), Working Paper
- Individual Decision-Making and Comparison of Rules (Paper & Pencil):
 - Lachmann / Stefani / Wöhrmann (2015), AOS
- Individual Decision-Making and Comparison of Rules (Mobile Lab & Eye-Tracking):
 - Hattenbach / Lachmann / Stefani

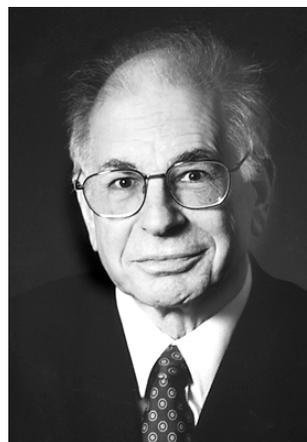
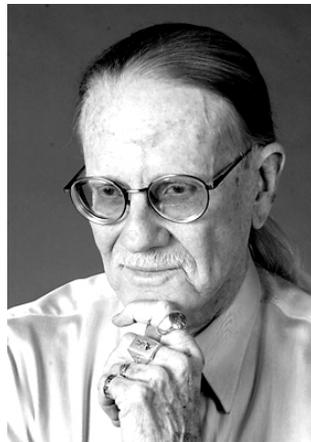
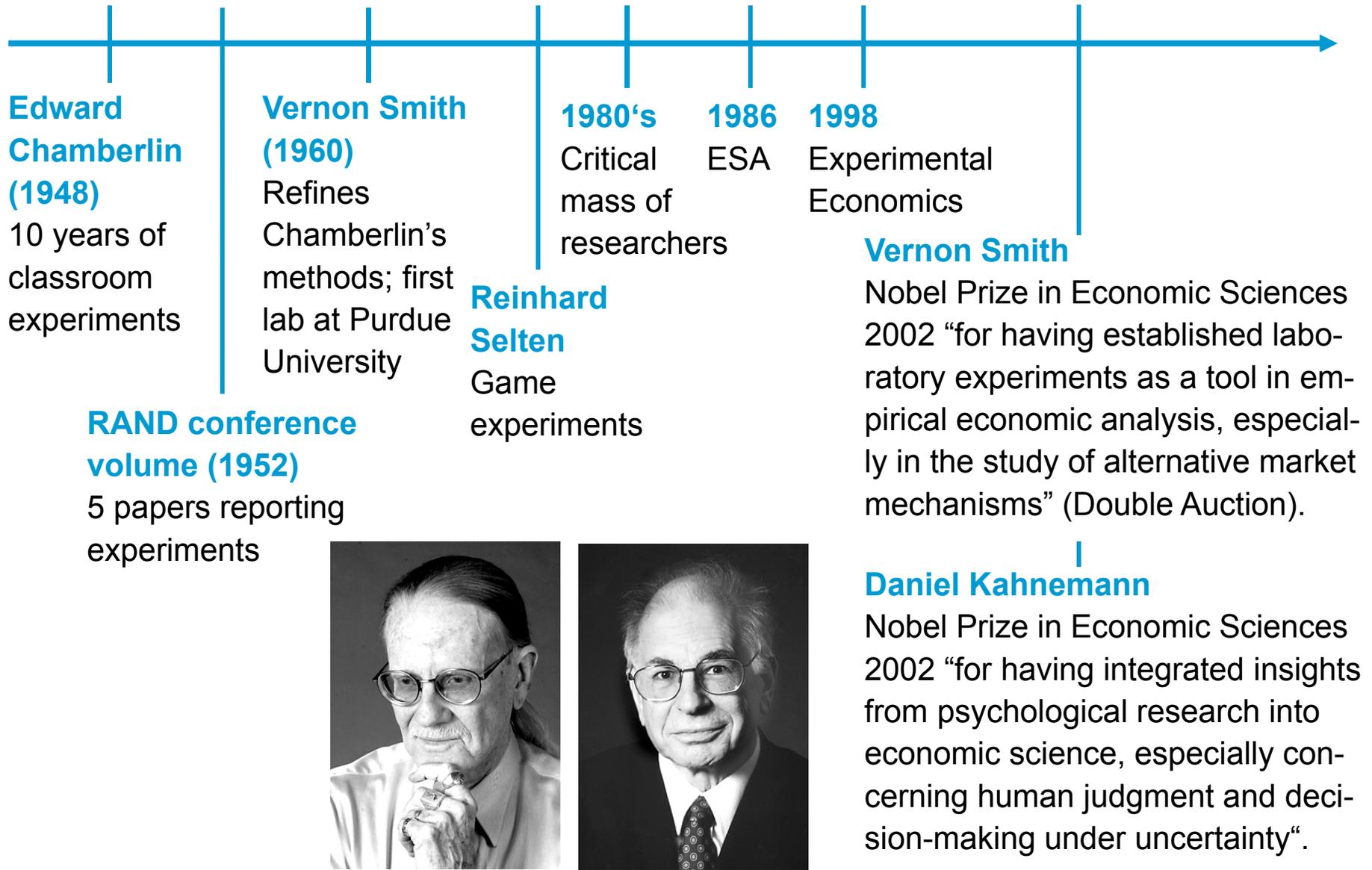
Universität
Konstanz



Experimental Economics



The History of Experimental Economics



What are Economic Experiments?

“Definition”

- An experiment is “a test under controlled conditions that is made to examine the validity of a hypothesis” (Shadish et al. (2002)).
- Hypotheses predict causal effects of independent variables on dependent variables.
- Experiments systematically vary the independent variable(s) and observe the resulting differences in the dependent variable(s).
- The key element of an experiment is control.
- All factors that could affect the dependent variable are either:
 - manipulated (systematically varied, ideally across the treatments of an experiment),
 - measured,
 - held constant (standardization), or
 - randomized.

Control is Essential

Measurement

- Each individual characteristic of the participants that could potentially affect the dependent variable is measured.
 - Demographic data (collected after the participants have completed the main task)
 - Experience (recruiting participants with certain skills; avoiding subjects that have already participated in similar experiments)
 - Understanding of the task
 - **Control questions** in lab experiments **before** the participants can enter the task (usually used in game experiments).
 - **Manipulation check questions** in case-based experiments **after** the task is completed.
 - (How to) use data from those participants who failed to correctly answer the manipulation check questions?
 - Report main results based on all subjects who correctly answered the manipulation checks, but include additional results based on all subjects (including those who failed to answer correctly).
 - Cognitive abilities, individual preferences, etc. (**control experiments**)

Control is Essential

Examples: Control experiments

- **Risk preferences: The Devil's Task** (adapted from Slovic (1966))
- Procedure and instructions:
 - Each participant receives **a plastic bag**.
 - The plastic bag contains **4 already opened paper slips** with a money bag.
 - The plastic bag also contains **12 closed paper slips**:
 - 11 of the 12 closed paper slips also show a money bag.
 - 1 of the 12 closed paper slips shows a devil:
 - Each participant can open as many paper slips as he/she would like.
- **Payoffs**:
 - 100 Points for each opened paper slip showing a money bag.
 - 2 Points for each paper slip remaining unopened (to avoid "gaming").
 - If the participant finds the devil, all the points from all paper slips showing a money bag are lost.
- The task is finished when the participant draws the devil or when he/she decides to stop opening paper slips.



Control is Essential

Examples: Control experiments

- Risk preferences: The Devil's Task (adapted from Slovic (1966))
- Classification:

	Prob. for money bag	Prob. for Devil	Closed paper slips	Open		Stop	
Slip 1	1		12	$400 + 12 \cdot 2$	(424,0)		
Slip 2	1		12	$400 + 12 \cdot 2$	(424,0)		
Slip 3	1		12	$400 + 12 \cdot 2$	(424,0)		
Slip 4	1		12	$400 + 12 \cdot 2$	(424,0)		
Slip 5	11/12	1/12	11	$11/12 \cdot 500 + 11 \cdot 2$	480,3	$400 + 12 \cdot 2$	424,0
Slip 6	10/11	1/11	10	$10/11 \cdot 600 + 10 \cdot 2$	565,5	$500 + 11 \cdot 2$	522,0
Slip 7	9/10	1/10	9	$9/10 \cdot 700 + 9 \cdot 2$	648,0	$600 + 10 \cdot 2$	620,0
Slip 8	8/9	1/9	8	$8/9 \cdot 800 + 8 \cdot 2$	727,1	$700 + 9 \cdot 2$	718,0
Slip 9	7/8	1/8	7	$7/8 \cdot 900 + 7 \cdot 2$	801,5	$800 + 8 \cdot 2$	816,0
Slip 10	6/7	1/7	6	$6/7 \cdot 1000 + 6 \cdot 2$	869,1	$900 + 7 \cdot 2$	914,0
Slip 11	5/6	1/6	5	$5/6 \cdot 1100 + 5 \cdot 2$	926,7	$1000 + 6 \cdot 2$	1012,0
Slip 12	4/5	1/5	4	$4/5 \cdot 1200 + 4 \cdot 2$	968,0	$1100 + 5 \cdot 2$	1110,0
Slip 13	3/4	1/4	3	$3/4 \cdot 1300 + 3 \cdot 2$	981,0	$1200 + 4 \cdot 2$	1208,0
Slip 14	2/3	1/3	2	$2/3 \cdot 1400 + 2 \cdot 2$	937,3	$1300 + 3 \cdot 2$	1306,0
Slip 15	1/2	1/2	1	$1/2 \cdot 1500 + 1 \cdot 2$	752,0	$1400 + 2 \cdot 2$	1404,0
Slip 16	0	1	0		0,0	$1500 + 1 \cdot 2$	1502,0

Risk-averse

Risk-taking

Control is Essential

Examples: Control experiments

- **Risk preferences: The Devil's Task** (adapted from Slovic (1966))
 - The Devil's Task is **easy** to understand (suitable even for children) and feels quite realistic (in particular if done with paper slips, which needs a lot of preparation, instead of computerized).
 - **Finding the devil stops the experiment** (i.e., using the number of opened paper slips dilutes the classification of the participants).
 - Solution: Use three plastic bags; at the end of the experiment, one of the participants rolls a die to determine which plastic bag is payoff-relevant.
 - Classification:
 - Risk-neutral participants should open 8 paper slips (and then stop).
 - Risk-taking: More than a total of 8 paper slips opened in at least one of the 3 trials.
 - Risk-averse: A total of 8 or fewer paper slips opened in all of the 3 trials and stopped the experiment in at least one of the 3 trials.
 - **Endowment-effect** might play a role.
 - **Gambling** can not entirely be ruled out.

Control is Essential

Examples: Control experiments

- **Risk preferences: The Lottery Task** (Dohmen et al. (2011))
- Procedure and instructions:
 - Participants choose between the two options “Lottery” and “Safe payment” for each of the 20 situations.
 - A 20-sided die determines which of the 20 situations is relevant; a 6-sided die determines the result of the lottery.

	Lottery	Safe Payment
1	(300; 0)	0
2	(300; 0)	10
3	(300; 0)	20
4	(300; 0)	30
5	(300; 0)	40
6	(300; 0)	50
7	(300; 0)	60
8	(300; 0)	70
9	(300; 0)	80
10	(300; 0)	90
11	(300; 0)	100
12	(300; 0)	110
13	(300; 0)	120
14	(300; 0)	130
15	(300; 0)	140
16	(300; 0)	150
17	(300; 0)	160
18	(300; 0)	170
19	(300; 0)	180
20	(300; 0)	190

Risk-averse (78%)

Risk-neutral (13%)

Risk-taking (9%)

Control is Essential

Examples: Control experiments

- Risk preferences: Which task should be used?
 - A certain percentage of the participants remains unclassified if the Lottery Task is used.
 - Both the Devil's Task and the Lottery Task lead to different classifications.
 - Both the Devil's Task and the Lottery Task make sense only if the participants receive compensation.

Control is Essential

Examples: Control experiments

- **Risk preferences: Which task should be used?**
 - Does the test really measure risk preferences?
 - **Devil's Task**: Hoffrage et al. (2003) document a link between children's risky and potentially harmful decisions to cross the street and behavior in the Devil's Task.
 - **Lottery Task**: Dohmen et al. (2011): Responses from 450 subjects (adult population in Germany) to a general risk question in a survey are a reliable predictor of actual behavior in a real-stakes lottery experiment.
 - **Lottery Task**: Dohmen et al. (2010): People with lower cognitive ability are significantly less willing to take risks, and are significantly less patient.
 - Schonberg / Fox / Poldrack (2011):
 - Neuro-economic studies using simple behavioral tasks have identified a network of brain regions that respond to economic risk, but failed to predict naturalistic risk-taking.
 - More complex behavioral tasks (e.g., BART and Devil's Task) correlate with naturalistic risk-taking but can not be decomposed into distinct cognitive constructs.

Control is Essential

Examples: Control experiments

- Cognitive Ability: The Cognitive Reflection Test (Frederick (2005))

A bat and a ball cost 1.10 € in total. The bat costs 1.00 € more than the ball. How much does the ball cost?

intuitive	reflected	
10	5	cents

If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?

100	5	minutes
-----	---	---------

In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

24	47	days
----	----	------

- What does the test measure?

- The test measures an individual's tendency to **override a prepotent (but incorrect) response** and to further reflect on the correct answer.
- The framework of an incorrectly primed initial response that must be overridden is connected to the **dual-process models** (Kahneman / Frederick (2002)).

Control is Essential

Examples: Control experiments

- **Cognitive Ability: The Cognitive Reflection Test** (Frederick (2005))
 - Frederick (2005): Large numbers of highly select university students at MIT, Princeton, and Harvard are cognitive misers (the CRT does not only measure IQ).
 - The CRT does **not** measure **mathematical ability**: Most participants that incorrectly answer the “bat and ball” problem are able to correctly solve the “banana and bagel” problem: “A banana and a bagel cost 37 cents. The banana costs 13 cents more than the bagel. How much does the bagel cost?” → 12 cents
 - **Heuristics-and-biases tasks**: The CRT score is correlated to cognitive ability, but is also a unique predictor of performance in heuristics-and-biases tasks (Toplak / West / Stanovich (2011)); it is related to the base rate fallacy, conservatism, and overconfidence, but not to the endowment-effect (Hoppe / Kusterer (2011)).
 - Nagin / Pogarsky (2003): Performance in the “bat and ball” problem is negatively correlated with **cheating**.
 - Frederick (2005): Positive correlation between the CRT score and **time preference** (patience).
 - **Problems**:
 - The task works only if the **participants do not receive compensation**.
 - The task works **only once** (recruiting!)

Control is Essential

Examples: Control experiments

- Cognitive Ability: More CRT Questions (Toplak / West / Stanovich (2014))

If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together?

intuitive	reflected	
9	4	days

Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class?

30	29	students
----	----	----------

A man buys a pig for \$60, sells it for \$70, buys it back for \$80, and sells it finally for \$90. How much has he made?

10	20	\$
----	----	----

Simon decided to invest \$8,000 in the stock market one day early in 2008. Six months after he invested, on July 17, the stocks he had purchased were down 50%. Fortunately for Simon, from July 17 to October 17, the stocks he had purchased went up 75%. At this point, Simon has: a. broken even in the stock market, b. is ahead of where he began, c. has lost money

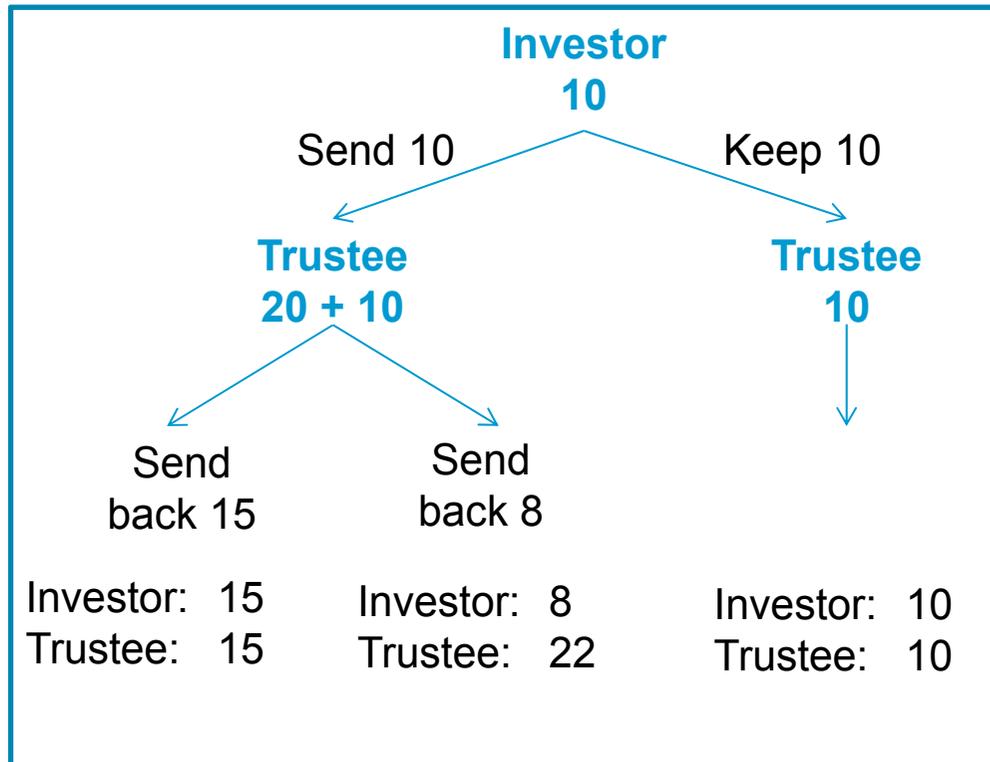
b	c
---	---

Control is Essential

Examples: Control experiments

- Trust and Trustworthiness: Trust Games

- Player 1 places resources at the disposal of Player 2 (trust); the resources are multiplied; Player 2 transfers an amount to Player 1 (trustworthiness)
- Binary Trust Game (Bohnet / Zeckhauser (2004))



Trust: The investor's willingness to make herself vulnerable to the trustee's actions.

Control is Essential

Examples: Control experiments

- Trust and Trustworthiness: Trust Games

- An individual's risk preference plays a role in trusting behavior (Karlan (2005); Schechter (2006); Bohnet et al. (2008); Fehr (2009)).
- Kosfeld et al. (2005): Oxytocin increases trust (but not the trustees' trustworthiness, the investors' beliefs, or the subjects' risk preferences): Trust is not a special case of risk taking or decision making under ambiguity (Fehr (2009)), but a social preference.
- Betrayal aversion (i.e., people dislike non-reciprocated trust) plays a role in trusting behavior (Bohnet et al. (2008); Fehr (2009)).
- Problem:
 - The Trust Game might measure altruism instead of trust (Fehr (2009)).
 - The Trust Game does not enable a clean separation between behavioral trust and expected trustworthiness.
 - Binary decision is a coarse measure of trust.

Control is Essential

Examples: Control experiments

- Trust and Trustworthiness: Trust Games
 - Continuous Trust Game (Berg / Joyce / Dickhaut / McCabe (1995))
 - $\text{Payoff}_1 = 10 - t_1 + r_2$
 - $\text{Payoff}_2 = 10 + 3t_1 - r_2$
 - Problem:
 - Results imply that r_2 is linear in t_1 (but trustworthiness could be non-linear).
 - Corner-solutions for risk-neutral players.
 - Expected trustworthiness affects behavioral trust.

Control is Essential

Examples: Control experiments

- **Trust and Trustworthiness: Trust Games**
 - Fischbacher / Föllmi-Heusi / Stefani:
 - B makes a non-binding promise of Allocation I (fair), Allocation II (unfair), or does not make a promise.
 - A decides how many Points to invest. **A's investment level is a measure of A's belief regarding B's trustworthiness** (if A is risk neutral).

	Allocation I		Allocation II			A's belief that B keeps the promise to choose A I
Points	Payoff A	Payoff B	Payoff A	Payoff B		
200	100	100	99	101		0%
240	120	120	98	142		10%
280	140	140	95	185		20%
320	160	160	89	231		30%
360	180	180	80	280		40%
400	200	200	65	335		50%
440	220	220	44	396		60%

Control is Essential

Examples: Control experiments

- **Trust and Trustworthiness: Survey Questions** (Fehr (2009))
 - American General Social Survey (GSS) (has measured trust annually since 1972) and World Values Survey (WVS) (applied to measure cross-cultural differences in trust):
“Generally speaking, would you say that most people can be trusted or that you can’t be too careful in dealing with people?”
 - Binary answer: **“Most people can be trusted”** or **“Can’t be too careful”**
 - “Confound” (important in particular in cross-country studies):
One could agree with both options (Miller / Mitamura (2003))
 - “Most people can be trusted” (experience in the past)
 - “Can’t be too careful” (prudence, unwillingness to take social risks).
 - Miller / Mitamura (2003): **“Do you think that most people can be trusted?”**
(answer categories on a 7-point Likert Scale from “not at all” (1) to “complete trust” (7))
 - German Socio-Economic Panel (SOEP): **“In general, one can trust people”**
(answers: “agree fully”, “agree somewhat”, “disagree somewhat”, “disagree fully”).

Control is Essential

Standardization

- Everything but the treatment variable is held constant.
 - The procedure of the experiment (document precisely!)
 - The instructions (be prepared to make them available!)
 - Temperature, sounds, distractions, time of the day, subject pool, ..., (what to do if reviewer/editor demands additional treatments?)

Randomization

- Random allocation of participants to treatments
- Random allocation of treatments to sessions
- Random allocation of participants to roles
- ...

Control is Essential

- Control is important because ...
 - ... control reduces noise (statistical reason).
 - ... control allows the replication of an experiment (by others).
 - ... control allows causal inferences.

Usage and Research Questions

1 Verification / falsification of theories

- Implementing the assumptions of a theory
- Testing the predictions of a theory
- Finding out the conditions under which a theory works / does not work.



Economics
(Game Theory)



2 Behavioral patterns as a basis for the development of new theories

- Empirically documented behavioral patterns can serve as a basis for the development of new theories / models.
- For games in which there are multiple equilibria, experiments can help to sort out the behaviorally relevant equilibrium (e.g., experiments can help in cases where theory fails).
- The finding that the „homo oeconomicus“ assumption does not explain behavior in many settings led to the development of new behavioral theories (e.g., bounded rationality, learning, social preferences like fairness, altruism, or lying aversion ...).



Interdisciplinary
research
(Economics,
Psychology,
Sociology)

Usage and Research Questions

3 Elicitation of preferences

- Individual preferences like risk aversion, cognitive abilities, intuitive vs. controlled behavior, inequity aversion, etc.
- Social norms



Interdisciplinary
research
(Economics,
Psychology,
Sociology)



1 Verification / falsification of theories

2 Behavioral patterns as a basis for the development of new theories



4 Comparison of institutions / laws / rules

5 Wind tunnel experiments

Usage and Research Questions

4 Comparison of institutions / laws / rules

- Comparison of rules / laws / contractual arrangements within an institution (e.g., market)
- Even rules / laws / contractual arrangements that **do not (yet) exist in reality** can be analyzed (empirical analyses based on archival data are not possible).



Application to
real-world
problems (also
Accounting)

5 Wind tunnel experiments

- Testing the effects of mechanisms / laws / regulations on behavior and / or social welfare
- Deriving implications of changes in mechanisms / laws / regulations
- **Fine-tuning of mechanisms / laws / regulations**
(control treatment and variation of single variables)
- This is the most frequently used application of experiments
 - Example: „Ebay-Professor“ Axel Ockenfels (effect of ratings given by buyers and sellers)



Application to
real-world
problems (also
Accounting)

Usage and Research Questions

6 Classroom experiments

- Facilitate the understanding of economic processes / phenomena
- Examples: Markets, bubbles, social dilemmas

- Learning by doing
 - Lack of rationality
 - Social behavior
 - Different perspectives in judging economic problems

- Examples (including detailed instructions):
 - Economics
 - Auditing: Schwartz / Spires / Young (2002)
 - Asset markets in which inefficiencies result from asymmetric information: societal benefits of auditing.
 - Acquisition of new clients by auditors, relationship between lowballing and auditor independence.
 - Threat of litigation faced by auditors.



Points of Criticism at Economic Experiments

Internal and external validity

- **Internal validity**
 - Does the observed behavior allow for causal inferences? Can alternative explanations for the observed effect be ruled out?
 - Internal validity can be achieved by a good experimental design (including the implementation of control) and correct analysis of the data.
- **External validity**
 - Can the findings from the lab be transferred to the real-world?
 - Do the inferences about cause-effect relationships hold over variation in persons, settings, manipulations, and measures?

Experiments are unrealistic

- **Experimental realism**: Participating in the experiment feels like acting in the real world.
- **Mundane realism**: Experimental setting is similar to the real world.
- Sometimes it's necessary and reasonable to intentionally neglect certain factors.
- Simplifications often help in understanding the main causalities (this point is standard in analyzing theoretical models).
- Whether „reality“ is a necessary aspect depends on the research question. Often mundane realism is considered relatively unimportant.

Points of Criticism at Economic Experiments

Using students as subjects biases the results

- Students are cheap.
- Students are intelligent; they learn fast.
- There are studies investigating behavioral differences across subject pools:
 - List (2003) (experienced vs. unexperienced market traders);
 - Fehr and List (2003) (students are more egoistic than managers);
 - Cooper (1999) (students converge faster to the equilibrium than managers).
- Whether students are useful as subjects depends on the task.

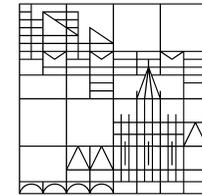
The payoffs are too low to be regarded as relevant

- Higher payoffs reduce the variance in decisions (Camerer and Hogarth (1999)).
- The payoff level is relevant if risk preferences have an impact on decision-making (Holt and Laury (2002)).
- The payoff level does not have a significant impact on fairness preferences (Cameron (1999), Slonim and Roth (1998), Hofmann et al. (1996), Fehr and Tougareva (1995)).

Points of Criticism at Economic Experiments

Experiments generate too few data to derive meaningful results

- Generating more data is always possible (but expensive).
- Of course experiments can also be done at a large scale
 - Example: Becker et al. (2015): Global Preference Survey (globally representative dataset on risk and time preferences, positive and negative reciprocity, altruism, and trust; contains 80,000 individuals, drawn as representative samples from 76 countries around the world).
- Sometimes small samples are sufficient if the data is correctly analyzed. If the results are significant using a small sample, it does not make sense to collect more data.



Behavioral Accounting



(How) Can Experiments be Used to Study Problems in the Fields of Accounting and Auditing?

Application of experiments to Accounting issues

- Experiments related to Accounting issues could have each of these focuses:
 - Verification / falsification of theories
 - Finding behavioral patterns as a basis for the development of new theories
 - Elicitation of preferences
 - **Comparison of institutions / laws / rules**
 - **Wind tunnel experiments**
 - Classroom experiments
- Deriving **policy implications** is interesting and important
 - What are the effects of a new IFRS that is not yet effective?
 - What effects can be expected from new regulations (e.g., mandatory audit firm rotation)?
 - Has a new incentive scheme the desired effects within a company?
- **Advantage over empirical studies based on archival data:**
 - Data availability
 - Empirical studies: Comparison of outcome variables before / after the implementation of a new setting (is everything else constant? Powerful statistical methods are needed to isolate the effects).
 - Experiments allow deriving causal effects.
 - Experiments allow observing the decision process in addition to outcome variables.

The number of experimental studies published in international journals is increasing.

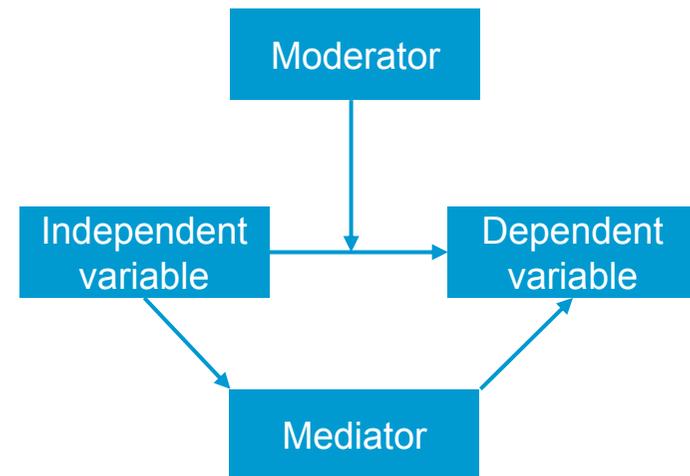
Controlled Test of Hypotheses

Hypotheses are based on theory

- A theory is a coherent set of ideas that can be used to explain and predict empirical observations.
- To derive hypotheses for Accounting issues, theories from (Behavioral) Economics, Psychology, and Sociology are used.
- Experiments are not the most appropriate research method to do exploratory research.

Research question:

- The **dependent variable** measures judgment, decision-making, or behavior.
- Which **independent variables** are relevant in affecting the dependent variable?
- What are the factors that **mediate** the relation between the dependent and the independent variable?
- What are the factors that **moderate** the relation between the dependent and the independent variable?



Design Choices: Terms

Experiment

- Comprises all sessions, treatments, etc.; is often used as a term for the research project.

Subjects

- Participants at the experiment (in most of the cases, students are used as subjects).

Treatment

- Variants that differ only in one variable (the “treatment variable”).
- Often there is a “control treatment”; the other treatments are compared against the control treatment to test whether the treatment variable affects the subjects’ decision-making.

Session

- A single run of one of the treatments

Period / round

- In most cases, there are several rounds (i.e., replications) of the same task ...
 - ... to allow learning of the subjects.
 - ... to allow reputation building.
- There are also “one-shot games” where only a single round is implemented.

Design Choices: Between- or Within-Subjects Design?

- **Between-subjects design:**
 - Each participant experiences only one treatment.
- **Within-subjects design:**
 - Each participant experiences two (or more) treatments.
 - Controls for individual idiosyncrasies (the same effect could be achieved in a between-subjects design where the participants are randomly allocated to the treatments).
 - Drawback:
 - Learning across the treatments (this problem could be solved by randomizing the treatment order).
 - Order-effects
 - „Balanced Design“

Design Choices: Types of Experiments

Laboratory experiments

- Conducted in the laboratory under controlled conditions.
- Computer-based; sometimes additional equipment is used (eye-tracking, Transcranial Magnetic Stimulation (TMS), Electroencephalography (EEG), ...).
- Pro's:
 - It's easier to make sure that the participants are not distracted by anything.
 - The information that the participants have access to (e.g., instructions, internet, calculator) is precisely controlled.
 - It's easy to allow (and control) interaction among the participants and between the experimenter and the participants (use, for example, the Ztree software; Fischbacher (2007)).
 - It's easier to control time.
 - It's easier to recruit adequate subjects (use ORSEE; Greiner (2004)).
 - Payments are easily managed.
 - Each lab has its rules (deception, payments, participants' expectations, etc.)
- Con's:
 - An artificial situation might cause unusual behavior.
 - It's nearly impossible (and extremely expensive) to get experienced professionals into the lab. Alternative: Mobile lab.

Design Choices: Types of Experiments

Scenario-based experiments

- Often “paper & pencil”, sometimes computerized.
- Participants work through a case.
- The dependent variable is a judgment or decision that is made in the setting provided (usually there is no strategic interaction).
- The independent variables are embedded in the case materials (i.e., treatment-conditions).
- Pro's:
 - It's easier to describe a setting in a realistic way (mundane realism).
 - Scenario-based paper & pencil experiments are relatively easy to run.
 - If a decision experiment is done, participants need not (necessarily) to be at a specific place at a specific point in time (although this would be the preferable solution).
 - Depending on the research question, it can be easier to recruit professionals.
- Con's:
 - It's difficult to allow for interaction among the participants. Alternative: Strategy method.
 - It's difficult to control the interaction between the experimenter and the participants (Protocol!).

The community seems to be split: Some researchers do only lab experiments, some do only scenario-based experiments. Both types of experiments get published.

Design Choices: Types of Experiments

Field experiments

- Experiment in real-world setting.
- Used only infrequently in Accounting research.

Newspaper experiments

- A (simple) task is described on an insert to a newspaper or journal; a reply card is included.
- Decision experiments are easier to implement; strategic interaction can be implemented by using the strategy method.
- Limited control over subjects' characteristics (type of newspaper chosen can be used to address a specific subject pool).
- Could be a good way to recruit professionals or subjects with certain skills.
- Too expensive for conducting scenario-based experiments.

Design Choices: Types of Experiments

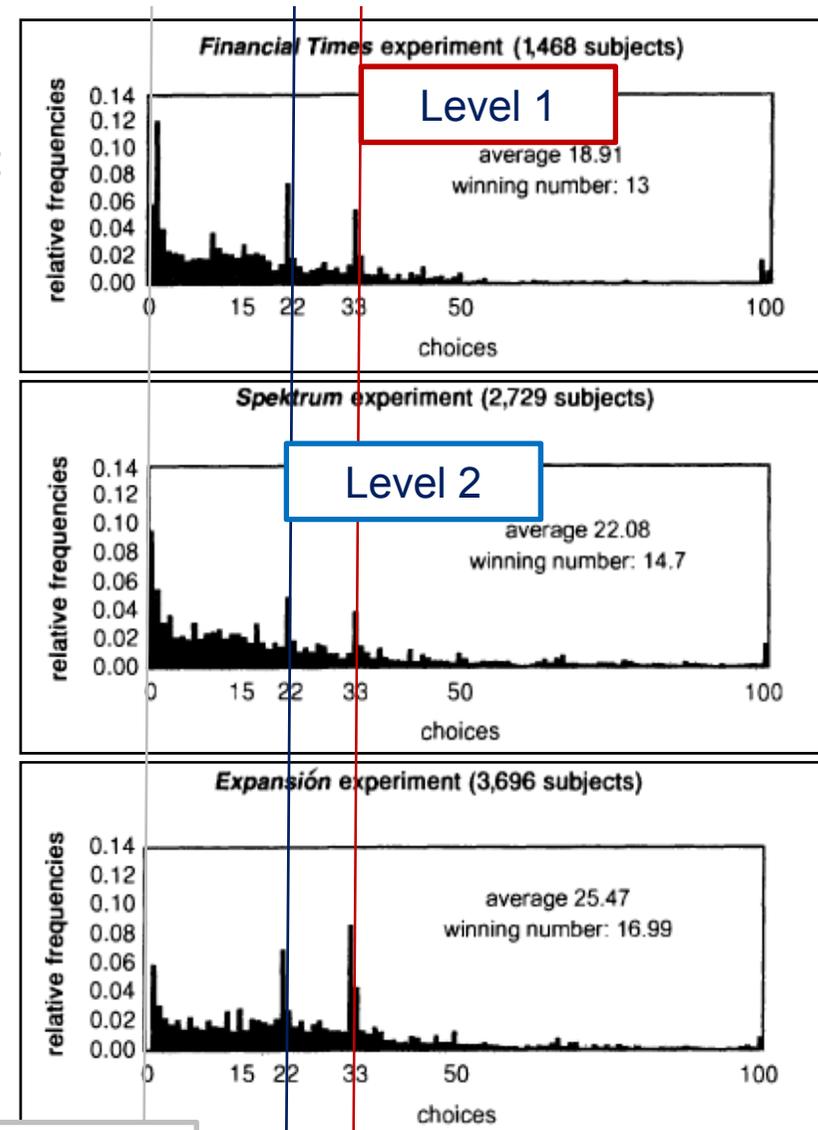
Newspaper experiments

- Example: The Guessing Game (Nagel (1995)):
 - Instructions: Choose a number between 0 and 100. The average (mean) of all numbers chosen in the experiment is calculated. The participant with the number closest to two thirds of this average receives 10.00 €.
 - Theoretical Prediction (iterated best response): Participants start at a uniform prior over other players' choices. Subjects can be classified according to the number of levels of reasoning. At each level, every player has the (incorrect) belief that he/she is one level of reasoning deeper than the rest.
 - If participants expect that all numbers are randomly chosen, the average is 50. → Level 0 of rationality
 - Participants expecting the others to choose their numbers randomly will choose $50 \cdot 2/3 =$ 33.33 → Level 1 of rationality
 - Participants expecting the others to have level 1 of rationality will choose $33.33 \cdot 2/3 =$ 22.22 → Level 2 of rationality
 - Participants expecting the others to have level 2 of rationality will choose $22.22 \cdot 2/3 =$ 14.81 → Level 3 of rationality
 - 0 → Level inf of rationality

Design Choices: Types of Experiments

Newspaper experiments

- Example: The Guessing Game (Nagel (1995)):
 - Results
 - Similar behavior across subject pools
 - Spikes at number choices 33.33, 22.22, and 0.



Design Choices: Types of Experiments

Experiments on Amazon's Mechanical Turk (MTurk)



- Relatively new website that contains
 - an integrated participant (“worker”) **compensation system**;
 - a **large subject pool**: over 500,000 “workers” from over 100 countries completing over 40.000 “human intelligence tasks” (HITs) daily;
 - a **streamlined process** of study design, participant recruitment, and data collection
- “Labor market for micro-tasks”.
- **General procedure**:
 - “**Requesters**” can post any task that can be done at a computer (i.e., surveys, experiments, writing, etc.) using simple templates or technical scripts or linking workers to external online survey tools (e.g., SurveyMonkey, Qualtrics).
 - “**Workers**” can browse available tasks; they are paid upon successful completion of each HIT (Requester can refuse to pay).
 - **Amazon** charges the Requester a 10% commission of the total pay granted.
 - **Compensation** in MTurk is monetary (to be spent on the Amazon.com website or to transfer to a U.S. or Indian bank account), but the amount awarded is typically small. The workers (with partially high yearly income) are mainly internally motivated (they complete HITs for enjoyment).

Design Choices: Types of Experiments

Experiments on Amazon's Mechanical Turk (MTurk)

- Experiments run on MTurk

- MTurk has been designed as a platform for humans to do tasks that are difficult or impossible for computers.
- Tasks that can be run on MTurk
 - Surveys and decision experiments
 - Experiments with random assignments
 - Synchronous experiments with interaction: Building a subject panel, notifying the panel of upcoming experiments, providing a “lobby” for queuing participants, handling exogenous and endogenous dropout during the experiment (Mason / Suri (2010))
- Differences between lab experiments and experiments conducted on MTurk (Arechar / Gächter / Molleman (2016)):
 - Deception is not used in the lab / MTurk workers could have experienced deception
 - Show-up fee and performance-contingent payment in the lab / fixum plus bonus on MTurk (Requester can refuse payment)
 - Subjects are not informed about the content of a lab experiment before the experiment starts / self-selection of Workers into HITs
 - Recruitment (criteria, experienced „lab rats“, duplicate subjects)

Design Choices: Types of Experiments

Experiments on Amazon's Mechanical Turk (MTurk)

- Advantages of MTurk (Mason / Suri (2010)):
 - Subject pool:
 - Large and stable
 - Diverse: Facilitates cross-cultural and international research, helps broadening the validity of experimental findings beyond student subjects.
 - Amazon forbids workers from using “bots”; “Spammers” are sanctioned by MTurk's built-in reputation system; Duplicates can be ruled out.
 - Low cost (built-in mechanism to pay “workers”, both fixed and bonuses)
 - Fast!

- Problems connected to MTurk (Mason / Suri (2010)):
 - Conducting simultaneous experiments
 - Security
 - Legal and ethical issues of crowdsourcing sites (Felstiner (2010))
 - Informed consent
 - Debriefing
 - Restricted populations
 - Payment (amount, taxes, ...)
 - Confidentiality

Design Choices: Types of Experiments

Experiments on Amazon's Mechanical Turk (MTurk)

- Evaluation:

- MTurk **participants** are slightly more demographically **diverse** than standard Internet samples and significantly more diverse than typical American college samples (Buhrmester / Kwang / Gosling (2011)).
- Participation is affected by compensation rate and task length, but participants can still be **recruited rapidly and inexpensively** (Buhrmester / Kwang / Gosling (2011)).
- **Compensation rates**
 - The compensation level does not affect **data quality** (Mason / Watts (2009); Buhrmester / Kwang / Gosling (2011)).
 - Low compensation levels lower **data collection speed** (Huang et al. (2010); Mason / Suri (2010); Buhrmester / Kwang / Gosling (2011)) and the **amount of work done** (Mason / Watts (2009)).
- Data obtained is at least as **reliable** as data obtained via traditional internet survey methods (Buhrmester / Kwang / Gosling (2011)).
- Behavior of participants on MTurk is **comparable** to behavior in lab experiments (Paolacci / Chandler / Ipeirotis (2010); Horton / Rand / Zeckhauser (2011); Suri / Watts (2011); Goodman et al. (2013); Krische (2014); Arechar / Gächter / Molleman (2016)).

MTurk provides access to a persistently available, large set of people who are willing to participate in experiments. MTurk can be used to obtain high-quality data inexpensively and rapidly.

Design Choices: Types of Experiments

Experiments on Amazon's Mechanical Turk (MTurk)

- MTurk experiments in an Accounting context:
 - Koonce / Miller / Winchel (2013) (CAR): How do investors respond to firms' derivative use decisions? How do norms influence investors' judgments in financial reporting settings (participants had to have taken at least two accounting and/or finance classes and to have experience in reading financial statements).
 - Rennekamp (2012) (JAR): Readability of disclosures and investors' reactions (participants had to live in the United States and consider English to be their native language; additional screenings for investment and accounting experience).
 - Krische (2014): What characteristics does the average individual investor have? Are results obtained in 4 Accounting experiments driven by these characteristics?

Design Choices: Incentives

Incentives

- Compensation for participating
 - Show up fee
 - A small gift / payment for every participant
 - A lottery prize for one randomly drawn participant
 - A donation to charity (could be problematic)
 - Course credits (I would not recommend to do that)
 - Personalized feedback (e.g., personality characteristics, preferences, abilities)
- Implementing incentives
 - Payments that are too low do not provide an incentive, payments that are too large can alter the preferences.
 - Rules of the lab (minimum and average payment; normally based on the hourly wage of student assistants).

Types of Experiments Used in Accounting Research

Task	Strategic Interaction	Decision-Making
Level	2, 3, ..., persons, markets	Individual
Theory	(Applied) Game-Theory	Decision-making paradigm, learning, information-processing, heuristics, theories from Psychology, ...
Type of experiment	Laboratory experiment	Scenario-based or decision task (laboratory experiment or Paper & Pencil)
Subjects	Students	Students, professionals
Payment	Yes	Normally not

Collecting Process Data

Example for decision experiments: Financial Reporting

- Changes in the way information has to be reported
- Increasing complexity and amount of what information has to be reported
- Does the presentation format and/or the information content affect addressees' decisions? And how.....?

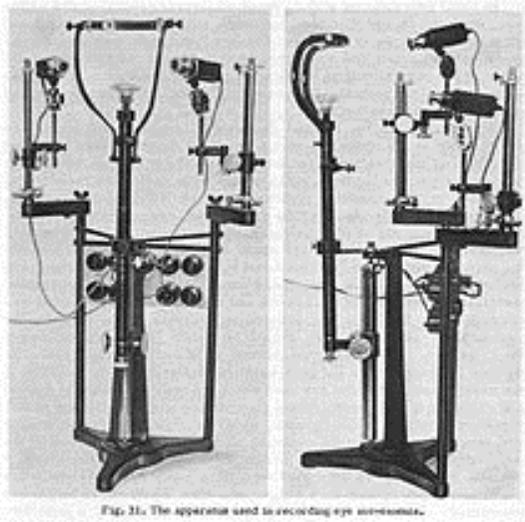
Observing decision processes

- Typically, it is only possible to observe the outcome of a decision.
- With process data,
 - the relative importance of pieces of information can be analyzed.
 - Individual search strategies can be analyzed.

Collecting Process Data: Eye-Tracking

Eye-tracking allows the precise observation of decision-processes

- Eye-tracking collects data on individual eye-movement.
- Hypotheses: There is a link between eye movements and cognitive processes.
- Eye-tracking provides valuable process data in an unobtrusive way.
- Only very few papers in an Accounting context
 - Sirois et al. (2016): Does the presence of key / critical audit matters (KAM) in the auditor's report affect users' attention to financial statement information?
 - Grigg / Griffin (2014): How do (4) participants solve key issues in financial analysis?

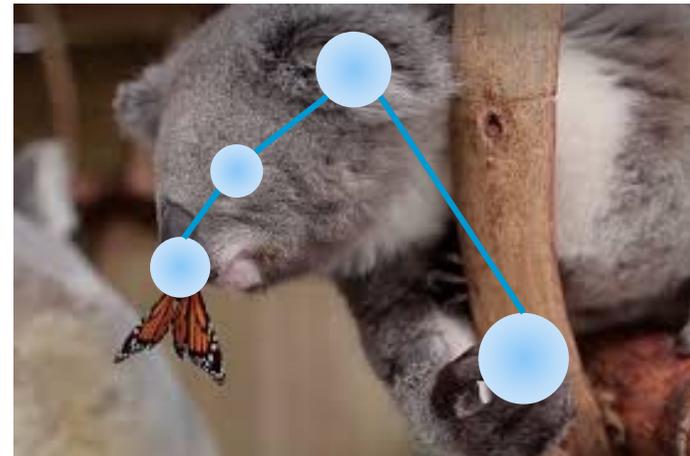


- Eye-tracking devices use a non-contact, optical method for measuring eye motion; they get „cheap“, are small, and can be taken nearly everywhere.

Collecting Process Data: Eye-Tracking

Terms

- **Calibration**
 - The eye-trackers first have to be adjusted to the participant's individual attributes.
- **Area of Interest (AOI)**
 - Pre-defined area on the screen
- **Transition**
 - Movement from one AOI to the next
- **Fixation** (fixation length is 250-300 ms)
 - Period for which gaze is hold still; stabilized position of the retina
 - All the information is mainly acquired during a fixation.
 - 40ms for reading (the perceptual span is 3-4 letter spaces); 200ms for pictures
 - The length of a fixation is usually an indication of information-processing or cognitive activities.
- **Saccades** (extremely fast; av. length is 20ms to 40ms)
 - Rapid movements from one fixation to the next, can be up to $600^{\circ}/s$.
 - When reading English, the mean saccade size is 6-7 letters.
 - Vision is largely suppressed during a saccade.

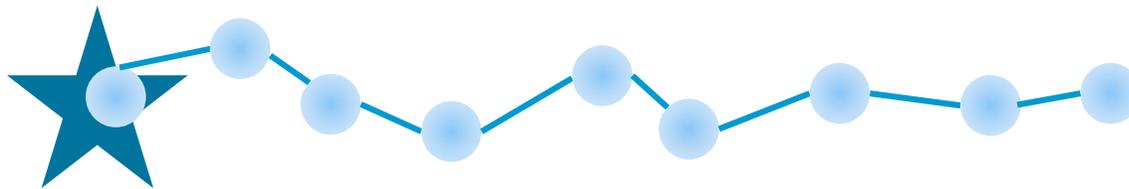


Collecting Process Data: Eye-Tracking

Terms

- Smooth Pursuit

- Eye movements necessary for stabilizing the image of a slowly moving target (slower than saccades)



- Speed is normally $<30^\circ/s$, initiated within 90ms to 150ms.
- Smooth pursuits that are faster than $30^\circ/s$ require additional saccades to catch up.
- Fast objects can be followed with a mixture of smooth pursuits and saccades.

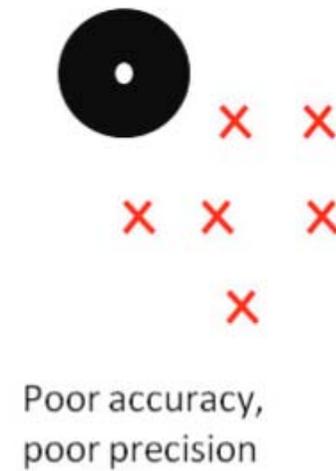
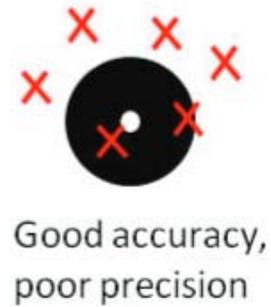
Collecting Process Data: Eye-Tracking

Some variables that can be used for testing the hypotheses

- Duration Before
 - Time to first fixation within AOI
- (Net) Dwell (time)
 - Visit duration, time within an AOI
- Fixation Count
- Hit Ratio
 - Number of participants who have looked in the AOI at least once

Collecting Process Data: Eye-Tracking

„Accuracy“ and „precision“



Collecting Process Data: MouselabWEB

MouselabWEB



- Drawback of eye-tracking: **Participants have to come to a (mobile) lab.**
- MouselabWEB allows process tracing in ordinary **web browsing.**
- Information is hidden behind boxes; to access the information, participants move the mouse pointer over the box (the information can be seen as long as the pointer is over the box).

MouselabWEB records the time of each box opening.

- **Features** of MouselabWEB:

- **Information boxes**: active (i.e., respond to mouse pointer) or inactive (i.e., always open). Acquisition of labeling information cannot be observed in MouselabWeb.
- **Choice buttons**
- Implementation using HTML and javascript (i.e., use over the internet in standard browsers; no plugins required)
- Data storing over the internet.
- Participants need a computer with internet connection and a recent browser.
- Experimenters: “if you are not very sophisticated with these things, buying a few drinks for a friend who knows this stuff is a good idea.”

Camera A	Camera B
Accessories option A	Accessories option B
Features option A	Features option B
Price option A	\$235 excl. Shipping
Camera A	Camera B

Collecting Process Data: MouselabWEB

MouselabWEB

- Data:

- **Basic measures** (i.e., proportion of choices for each task or condition in an experiment; time spent on a cell or total task length; number of clicks for each cell or for a task)
- **Transitions between cells** as a measure of information search.
 - Type I transition: no transition, the same information item is inspected again.
 - Type II transition: **alternative-wise search pattern** (the transition moves within the same alternative but changes the attribute).
 - Type III transition: **attribute-based search pattern** (the transition stays within the same attribute but changes the alternative).
 - Type IV transition: diagonal transition which switches alternative as well as attribute.
 - **Search Index** (SI) (Payne (1976)):

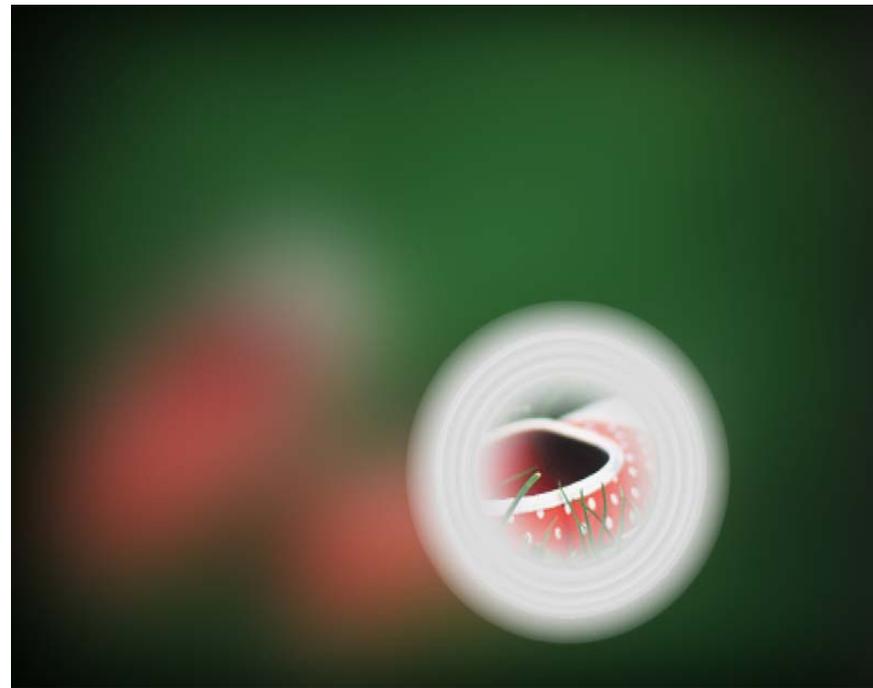
$$SI = \frac{N_{Type II} - N_{Type III}}{N_{Type II} + N_{Type III}}$$

- SI ranges from -1.00 (completely attribute-based information search) to 1.00 (completely alternative-based information search).

Collecting Process Data: Flashlight

Flashlight

- Allows process tracing in ordinary web browsing.
- Flashlight uses a bottom layer with the actual stimulus and a top layer with a blurred version of the stimulus. When participants move the mouse pointer, a pre-defined area around the mouse pointer reveals the bottom layer.
- The position of the mouse pointer (x,y coordinates) is measured with 10Hz.
- Any static (not dynamic!) visual stimulus can be presented.



Collecting Process Data

Do the different methods lead to different results?

	<u>Eye-tracking</u>	<u>Flashlight / Mouselab</u>
Measurement	Measures eye movement	Hand and eye movement is necessary for information to be recorded. <u>“Does the hand know what the eyes see?”</u>
Controlled vs. unintentional processes	Participants can not intentionally filter the information processed; eye movements caused by unconscious mental processes are also recorded.	Participants might pay particular attention to the information search.
Stimulus	Any stimulus	Flashlight: any static stimulus Mouselab: Matrices
Sampling rates	50 HZ to > 1.000 HZ	10 HZ (saccades can not be detected)
Location	Lab	Lab / web
Concurrent participants	One	Many
Availability	Commercial	Open source / for free

Collecting Process Data

Do the different methods lead to different results?

- Mouselab vs. eye-tracking:

- Meißner / Decker / Pfeiffer (2017) (eliciting preferences using choice-based conjoint analysis):
 - With eye tracking, participants are faster and direct their search more towards attributes (instead of alternatives) than with Mouselab.
 - There are no differences regarding the number of pieces of information processed or the decisions made.
 - Mouselab seems not suitable when information processes (rather than decisions) are studied.
- Norman (2010): While Mouselab is more suitable for assessing deliberate processes, eye-tracking allows insights in the intuitive, automatic domain.

Collecting Process Data

Do the different methods lead to different results?

- Flashlight vs. eye-tracking:

- Schulte-Mecklenbeck / Murphy / Hutzler (2011): Arithmetic task (adding five numbers), risky choice task (deciding between two gamble options), and reading task (reading a novel paragraph of text).
 - No differences measuring simple outcome data between Flashlight and eye-tracking.
 - Both methods result in comparable numbers of fixations and similar patterns of acquisition.
 - Task completion takes longer with Flashlight than with an eye-tracking system.
 - **“Flashlight is suited for studying the order and duration of information acquisition, but without the funding or apparatus to purchase eye-tracking equipment.”**

Questions to Achieve a Good Design (Shyam Sunder)

Questions that are important for finding the design of an experiment (and for evaluating experimental studies)

- 1 **What is the question whose answer you would like to find as a result of your experiment? (Hint: A question is a single sentence with a question mark at the end.)**
 - Not: I'd like to explore whether the mandatory audit firm rotation has any effects.
 - But: Is the auditor's independence higher under the mandatory audit firm rotation than in a setting without regulations on the auditor-client contractual relationship?
 - Hypotheses

- 2 **What do you know already about the possible / feasible answers to the question you have stated above?**

- 3 **If there are more than one possible / feasible answer, do you have one that you are inclined to favor? If you asked ten of your friends, are they likely to agree with you?**

- 4 **If there is a significant chance of disagreement among your friends, what are the various possible ways of finding an answer to the question you have stated above? Include both experimental as well as any other methods you can think of.**

Questions to Achieve a Good Design (Shyam Sunder)

5 What are the advantages and disadvantages of using an experiment to find an answer?

- Never do an experiment just for its own sake.
- There are settings in which an empirical study is more suited than an experiment.

6 How important is this question to YOU? What are the chances that the answer you get from your experiment or other research will surprise you or others? What are the chances that it will change someone's mind?

- Is the research question innovative?
- Would the results be useful?
- Is there any contribution?
- Would the results be relevant (for other research questions)?

7 How would you conduct the experiment / research project?

- How should the main figure / table look like?
- Which results would answer the question you stated above?
- Do you need additional treatments?

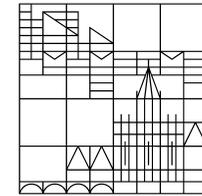
Questions to Achieve a Good Design (Shyam Sunder)

8 Is your experimental / research design the simplest possible design to help answer the question you have stated?

- Simplicity is extremely important!
- A simple experiment has not too many parameters.
- A simple experiment is more comprehensible for the subjects and for the readers of your paper.

9 What are the possible outcomes of the experiment? Do the possible outcomes include at least one outcome that will answer the question you stated above? What is the chance that you will observe this outcome?

At any stage of your thinking, feel free to go back and revise your earlier answers if you wish to.



Examples



Examples

- **Strategic Interaction and Testing Theories (Lab):**

Fischbacher, Urs; Stefani, Ulrike (2007): Strategic Errors and Audit Quality – An Experimental Investigation, in: The Accounting Review 82 (3): 679-704.



- **Strategic Interaction and Comparison of Institutions (Lab):**

Hattenbach, Thomas; Stefani, Ulrike (2016): Audit Firm Rotation vs. Centralized Matching: Experimental Evidence on the Quality of Audited Financial Statements, Working Paper.



- **Individual Decision-Making and Comparison of Rules (Paper & Pencil):**

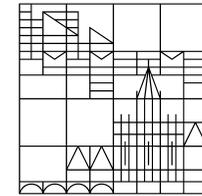
Lachmann, Maik; Stefani, Ulrike; Wöhrmann, Arndt (2015): Fair Value Accounting for Liabilities: Presentation Format of Credit Risk Changes and Individual Information Processing, in: Accounting, Organizations and Society 41 (1): 21-38.



- **Individual Decision-Making and Comparison of Rules (Mobile Lab & Eye-Tracking):**

Hattenbach, Thomas; Lachmann, Maik; Stefani, Ulrike: Auditors' Information Processing and Going-Concern Opinions: Experimental Evidence on IAS 17 vs. IFRS 16.





Strategic Errors and Audit Quality: An Experimental Investigation



Urs Fischbacher / Ulrike Stefani (TAR 2007)

Matching Pennies Game

Literature

- **Matching Pennies Games** in an Accounting and Auditing context:
 - Magee (1980), Fellingham / Newman (1985), Fellingham / Newman / Patterson (1989), Newman / Noel (1989), Patterson (1993), Ewert (1993, 2004), Shibano (1990), ...
- Matching Pennies Games
 - 2 players, each of them has two strategies available:
 - The manager can either prepare correct financial statements or overstate the economic situation of his company.
 - The auditor can either conduct a cursory audit that leaves manipulations undetected, or a comprehensive audit that perfectly detects all errors and irregularities.
 - The actions are chosen simultaneously.
 - The auditor renders an adverse opinion if he has observed that the manager has manipulated the financial statements and a clean opinion otherwise.

Matching Pennies Game

Manager Auditor	Correct financial statements ($1 - \alpha$)	Manipulation (α)
High audit effort ($1 - \beta$)	Z $G - K$	$Z - A$ $G - K - Q + V$
Low audit effort (β)	Z G	$Z + B$ $G - m \cdot p \cdot N \cdot Q$

Manager:

- Basic salary **Z** if he reports truthfully.
- If misreporting is detected, the manager incurs additional costs **A** (loss of reputation, decrease in performance-contingent compensation, adverse reactions of addressees).
- If the auditor fails to detect an overstatement, the manager earns a bonus **B**.

Auditor:

- Basic profit **G**: Fee adequate for choosing a high audit effort, minus the costs for low effort.
- Additional audit costs **K** for high audit effort.
- Economic advantages **Q** lost (quasi-rents) and reputational gains **V** if the auditor reports detected manipulations.
- Loss of $m \cdot N$ clients (**Q**) if the auditor issues a clean opinion when an adverse opinion would have been appropriate, and this is revealed (**p**).

Matching Pennies Game

	Manager	Correct financial statements (1 - α)	Manipulation (α)
Auditor			
High audit effort (1 - β)		Z	Z - A
Low audit effort (β)		Z	Z + B
		G - K	G - K - Q + V
		G	G - m · p · N · Q

Assumptions

- Players are purely opportunistic
- Standard assumptions: Risk neutrality, perfect rationality, and common knowledge about these attributes.

Solution: Unique Nash Equilibrium in mixed strategies

- Probability for **manipulations**: depends on the auditor's payoffs

$$\alpha^* = \frac{K}{Q \cdot (m \cdot p \cdot N - 1) + V} > 0$$

- Probability for **low audit effort**: depends on the manager's payoff

$$\beta^* = \frac{A}{A + B} > 0$$

- **Quality of audited financial statements**: result of the strategic auditor-manager interaction

$$\phi^* = 1 - \alpha^* \cdot \beta^*$$

Experimental Evidence on Matching Pennies Games

Evidence from experiments investigating more general settings

- Observed behavior correlates well with the Nash prediction if certain conditions are fulfilled:
 - Point prediction of the Nash Equilibrium is confirmed for symmetric Matching Pennies Games.
 - The Nash Equilibrium predicts human behavior more precisely if strategic dependence (slope of the best reply functions) is low (Bloomfield (1997)).
- There are situations in which experimental data deviates strongly from the standard prediction:
 - Experimental data for Matching Pennies Games with asymmetric choice probabilities rarely support the Nash prediction (Goeree / Holt (2001)).
 - Experimental data clearly contradicts the comparative static effect predicted by standard theory, because both subjects adjust their behavior in a very intuitive way after a payoff variation for one subject has occurred (Goeree / Holt (2001); Goeree / Holt / Palfrey (2003)) – Own-Payoff Effect (persists even in experiments with repeated games).

Motivation

Aims of this experiment

- Highlight the connections between ...
 - ... behavior as predicted by standard game-theoretic models that analyze problems inherent to Accounting and Auditing (which are frequently used to derive policy recommendations), and
 - experimental results on individual human behavior.
- Adapt standard game-theoretic models to explain and predict these experimental results.
- Practical relevance: To minimize audit risk, the auditor needs a precise estimation of inherent risk, which is endogenously determined by the strategic behavior of the manager (who chooses his reporting strategy contingent on his estimation of the strategic effort his auditor chooses to optimize detection risk).
- Derive policy implications: Predicting behavioral changes as a consequence of the variation of certain conditions of the audit environment:
 - Payoff variations
 - Adding or removing type uncertainty (honest auditors)

Experimental Design

Description of the experimental design

- Subjects are randomly assigned their roles (auditor and manager) which they maintain for the whole session.
- Half of the subjects is in the role of the manager, the other half is in the role of the auditor.
- 15 sessions (within-subjects design); 2 treatments (A and B) in each session, 20 periods per treatment:
 - **Treatment A**: All decisions made are used as a basis for calculating the participants' compensation earned during the experiment (**completely “opportunistic” managers and auditors**).
 - **Treatment B**: One-third of the decisions on audit effort of all participants in the role of auditors are randomly selected and automatically substituted ex post with a high audit effort.
 - All subjects know this process and thus that there is a **one-third probability for a computerized auditor in treatment B**.
 - All of the managers continue to behave opportunistically in treatment B.
- Conversion of Points into money (1 point = US\$ 0.05), US\$ 6.09 show-up fee (US\$ 18.02 per session).

Experimental Design

Sessions

	Session Number	Treatment Order ^a	Framing ^b	Payoff Matrix ^c	Number of Participants	Number of Matching Groups ^d	
Basic experiment	1	A – B	auditor/manager	1	24	1	
	2	A – B	auditor/manager	1	24	3	
	3	A – B	auditor/manager	1	22	3	
	4	A – B	auditor/manager	1	24	3	
	5	A – B	auditor/manager	1	22	3	
						Subtotal: 116	
	6	B – A	auditor/manager	1	24	3	
	7	B – A	auditor/manager	1	24	3	
	8	B – A	auditor/manager	1	24	3	
	9	B – A	auditor/manager	1	24	3	
10	B – A	auditor/manager	1	24	3		
					Subtotal: 120		
Robustness check neutral framing	11	A – B	neutral	1	32	4	
	12	A – B	neutral	1	32	4	
					Subtotal: 64		
Robustness check “Own-Payoff Effect”	13	A – B	auditor/manager	2	32	4	
	14	A – B	auditor/manager	2	32	4	
	15	A – B	auditor/manager	2	30	4	
					Subtotal: 94		
					Total: 394		

Experimental Design

Payoff matrices (payoffs in Points)

Basic Experiment (Payoff Matrix 1, Auditor-Manager Framing)^a

Auditor \ Manager	Correct financial statements	Overstatement of economic condition
Comprehensive audit	2, 6	8, 3
Normal audit effort	7, 6	6, 9

Additional Experiment (Payoff Matrix 2, Auditor-Manager Framing)

Auditor \ Manager	Correct financial statements	Overstatement of economic condition
Comprehensive audit	5, 6	8, 3
Normal audit effort	7, 6	3, 9

^a In sessions 11 and 12 with neutral framing and payoff matrix 1, we denoted the manager and the auditor as “participant A” and “participant B”, respectively, and labeled the actions as “X”, “Y”, “T”, and “II”.

Standard Predictions

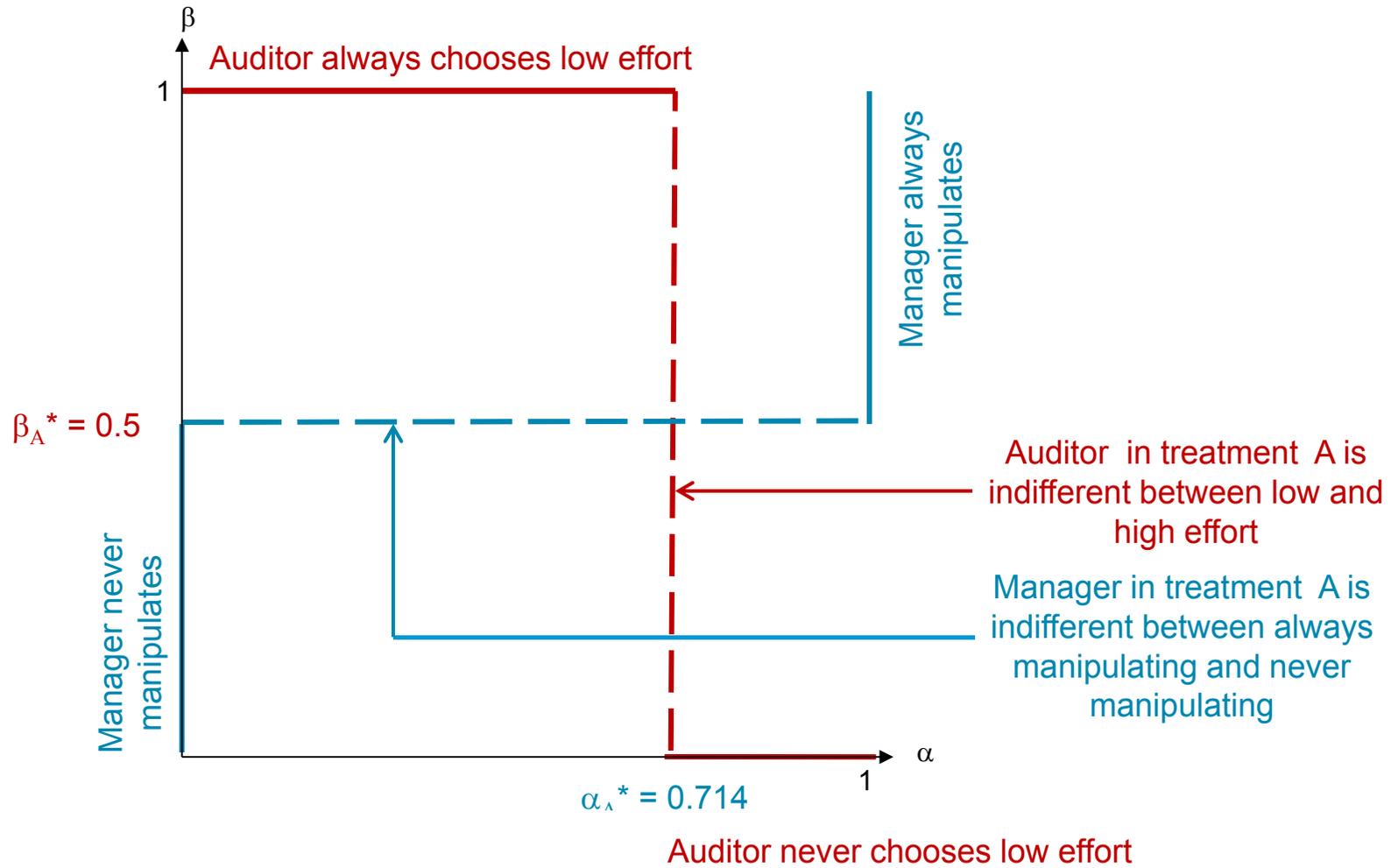
Investigation of comparative static effects: What is the effect of some perfect (honest) auditors on audit quality?

- **Treatment A**: opportunistic auditors only
 - Auditor: $\beta_A^* = 0.5$
 - Manager: $\alpha_A^* = 0.714$
 - Quality: $1 - \alpha_A^* \cdot \beta_A^* = 0.643$

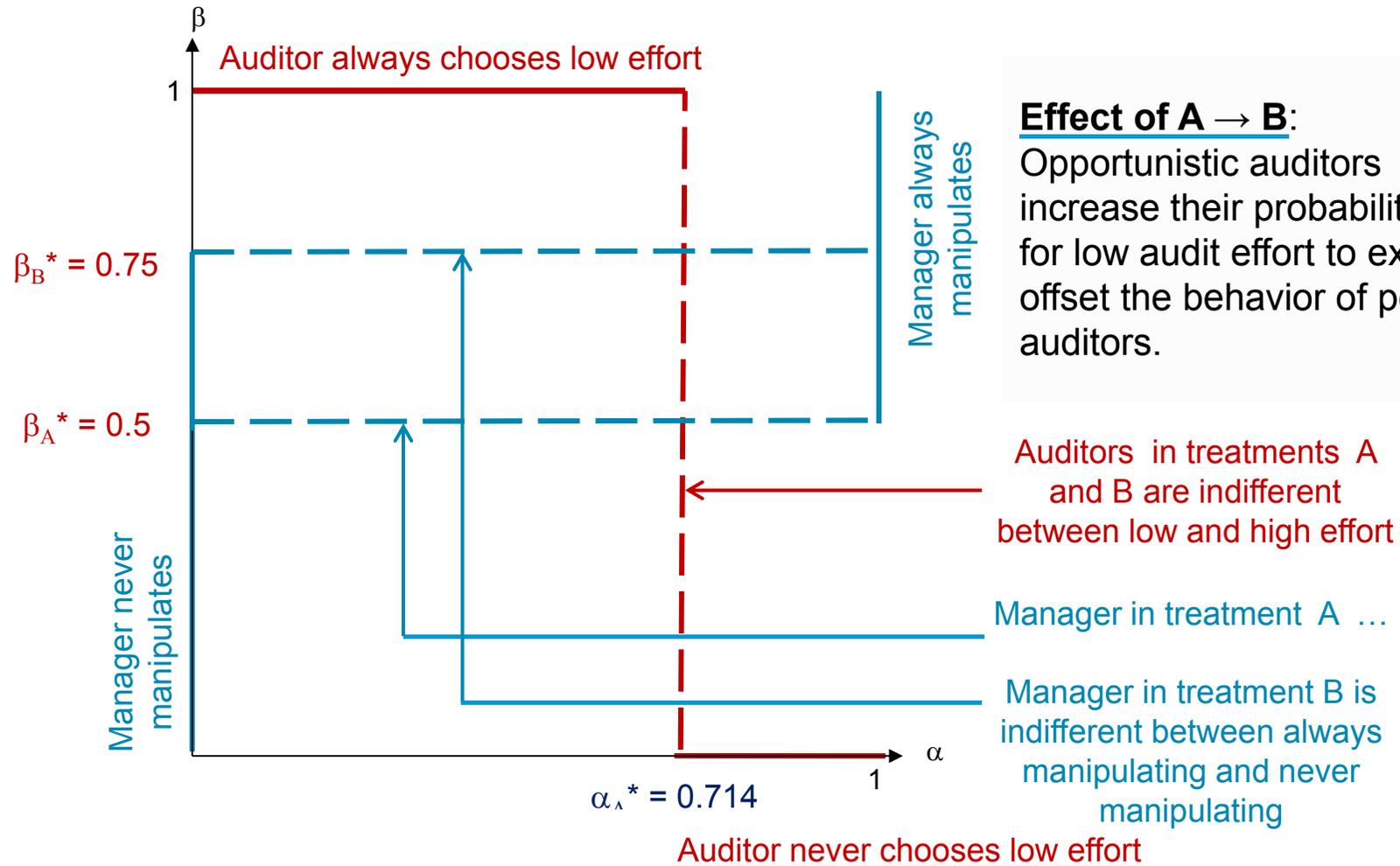
- **Treatment B**: $r = 1/3$ computerized auditors
 - Auditor: $(1 - r) \cdot \beta_B^* = \beta_A^*$: $\beta_B^* = 0.75$
 - Manager: $\alpha_A^* = \alpha_B^*$: $\alpha_B^* = 0.714$
 - Quality: $1 - \alpha_B^* \cdot 2/3 \cdot \beta_B^* = 0.643$

- Standard-prediction:
 - The remaining opportunistic auditors in treatment B completely offset the decisions of the computerized auditors, and the managers anticipate this behavior.
 - **The existence of some „perfect“ auditors does not improve the quality of audited financial statements!**

Standard Predictions



Standard Predictions



Standard Predictions and Quantal Response Equilibrium

Assumptions of this Nash Equilibrium

- Players do not have social preferences.
- Players are risk neutral.
- **Perfect rationality**: Infinitely sensitive reactions, i.e., discontinuous best reply correspondences.

Quantal Response Equilibrium (McKelvey / Palfrey (1995)):

- Captures bounded rational behavior.
- **Error parameter** μ : How sensitive does PR_i^A react to differences between the strategies with regard to expected payoffs?
- Probability that a strategy is chosen:

$$PR_i^A = \frac{\exp(\pi_i^A(s^B) / \mu)}{\sum_{j \in S^A} \exp(\pi_j^A(s^B) / \mu)}$$

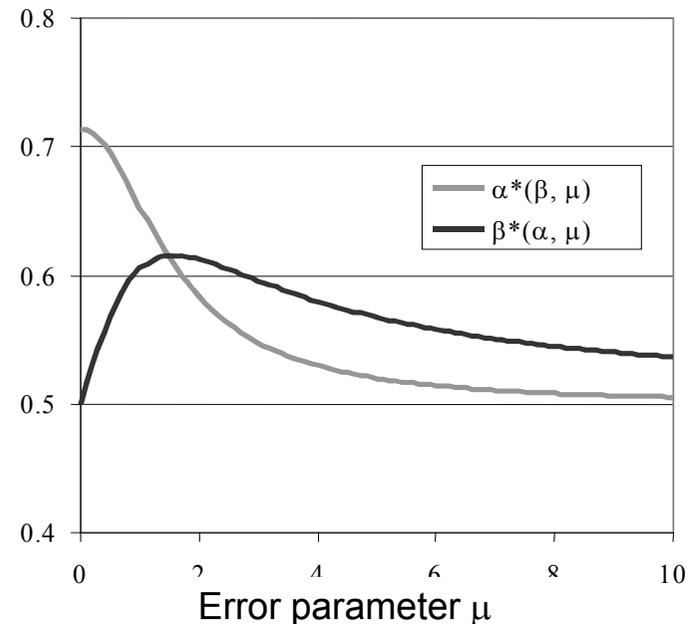
- **Intuition**: Players' decisions are not only subject to errors in a very general sense, but players also take the others' errors into account.
- Players do not always choose their best strategy, but **the probability of playing a strategy increases in the payoff of that strategy choice, given the other player's behavior.**

Quantal Response Equilibrium

Continuous best reply functions

- Strategy choice depends on μ :
 - For $\mu = 0$, the Nash Equilibrium results as a special case.
 - If errors are infinitely large, a player chooses all the strategies with the same probability.

		M		
		Correct FS <i>2/7</i>	Manipulation <i>5/7</i>	
A	High effort <i>1/2</i>	2	8	44/7
	Low effort <i>1/2</i>	7	6	44/7
		6	6	

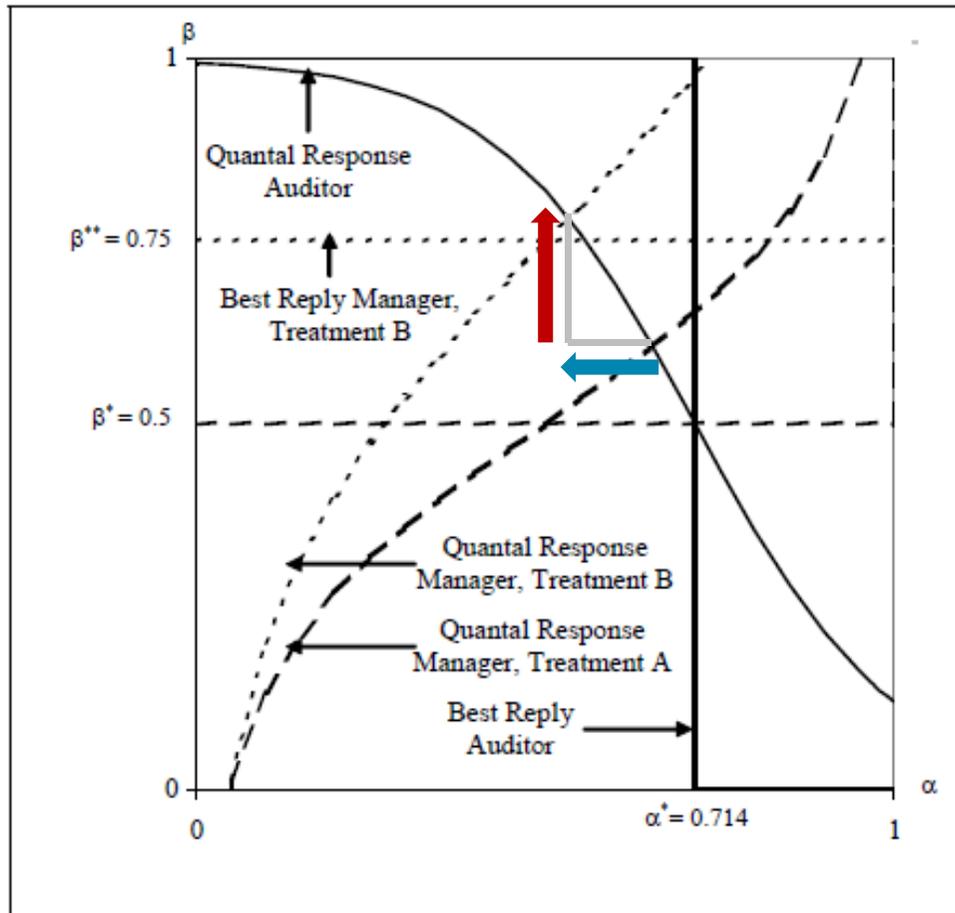


- The QRE is not based on an explicit specification of the errors' psychological nature (strategic uncertainty and heterogeneity in decision makers' preferences could be the main sources of bounded rational behavior).

Quantal Response Equilibrium

Best Replies and Quantal Responses for the treatments A and B

- Payoff matrix 1 and error parameter $\mu = 1$



Effect of A → B:

Opportunistic auditors increase their probability for low audit effort to compensate for the existence of perfect auditors, and managers manipulate less often in the presence of perfect auditors.

Quantal Response Equilibrium

Hypotheses based on the Quantal Response Equilibrium

H1: Managers manipulate less often in treatment B than in treatment A.

- **Contrary to the Nash Equilibrium.**

H2: Opportunistic auditors exert low audit effort more often in treatment B than in treatment A.

- In line with the Nash Equilibrium.
- Opportunistic auditors offset the effect that computerized auditors have on audit quality, i.e., computerized auditors crowd out auditor participants who strategically choose a high audit effort.

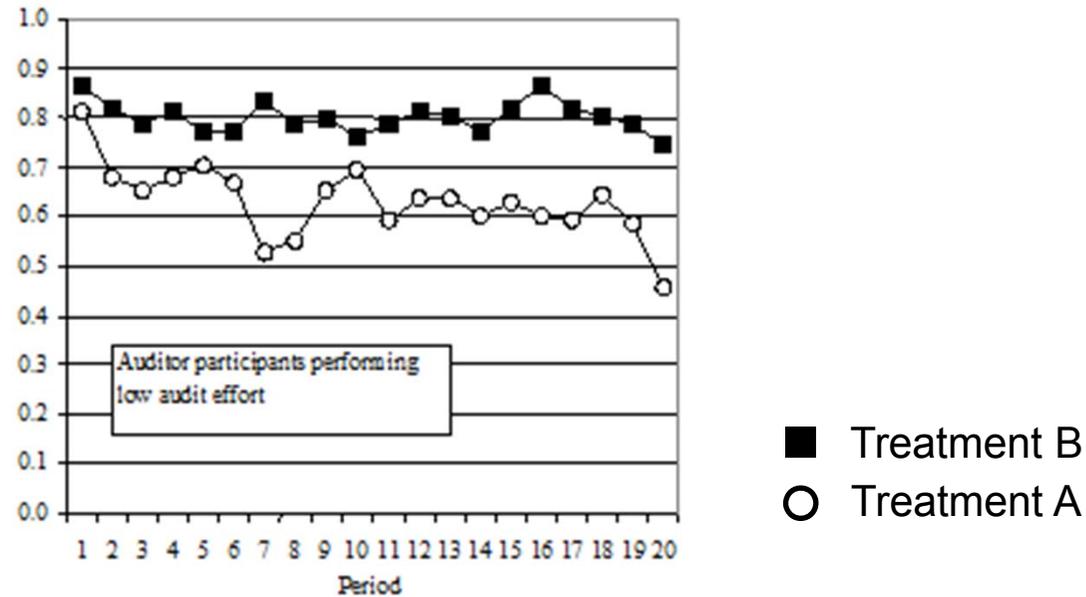
H3: Overall, auditors perform fewer low effort audits in treatment B than in treatment A.

- **Contrary to the Nash Equilibrium.**
- **Crowding out is incomplete** in the sense that overall audit quality increases if computerized auditors are present → **H1**

Results

Result 1: Auditor's behavior (H2)

- Frequencies of low audit effort over time:



- Auditor participants exert low audit effort more often in the presence of computerized auditors (treatment B) than in treatment A.
- **H2 is confirmed.**
- This result is in line with the standard prediction and with the QRE.

Results

Statistical tests for significant differences across the treatments A and B

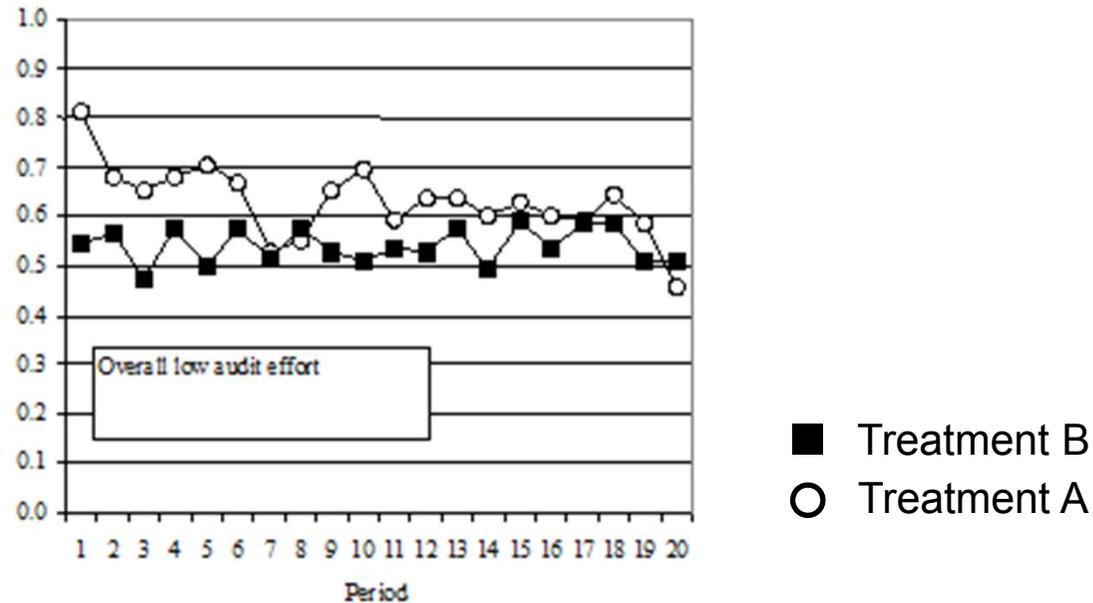
	Tests	Data ^a	All Periods	Periods 11 - 20
<i>Δ Low Audit</i>	Wilcoxon signed-rank test	all data	4.555***	4.362***
		Sessions 1 - 5	3.110***	2.830***
		Sessions 6 - 10	3.352***	3.324***
	Mann-Whitney test	First treatments	4.015***	3.490***
<i>Δ Overall Low Audit</i>	Wilcoxon signed-rank test	all data	-4.158***	-1.062**
		Sessions 1 - 5	-3.111***	-2.521**
		Sessions 6 - 10	-2.814***	-0.369
	Mann-Whitney test	First treatments	-2.887***	-1.342
<i>Δ Manipulation</i>	Wilcoxon signed-rank test	all data	-3.952***	-3.747***
		Sessions 1 - 5	-2.760***	-2.411**
		Sessions 6 - 10	-2.813***	-2.841***
	Mann-Whitney test	First treatments	-3.759***	-3.527***
<i>Δ Quality</i>	Wilcoxon signed-rank test	all data	4.555***	3.976***
		Sessions 1 - 5	3.181***	2.797***
		Sessions 6 - 10	3.239***	2.704***
	Mann-Whitney test	First treatments	4.079***	3.732***

The number shown in the table represents the z-value. It is positive if the corresponding variable is higher in treatment B (with computerized auditors).

Results

Result 2: Auditor's behavior (H3)

- Frequencies of overall low audit effort over time:



- Auditor participants only incompletely compensate for the decisions of the computerized auditors (incomplete crowding out).
- H3 is confirmed.
- This result is in line with the QRE, but contrary to the standard prediction.

Results

Statistical tests for significant differences across the treatments A and B

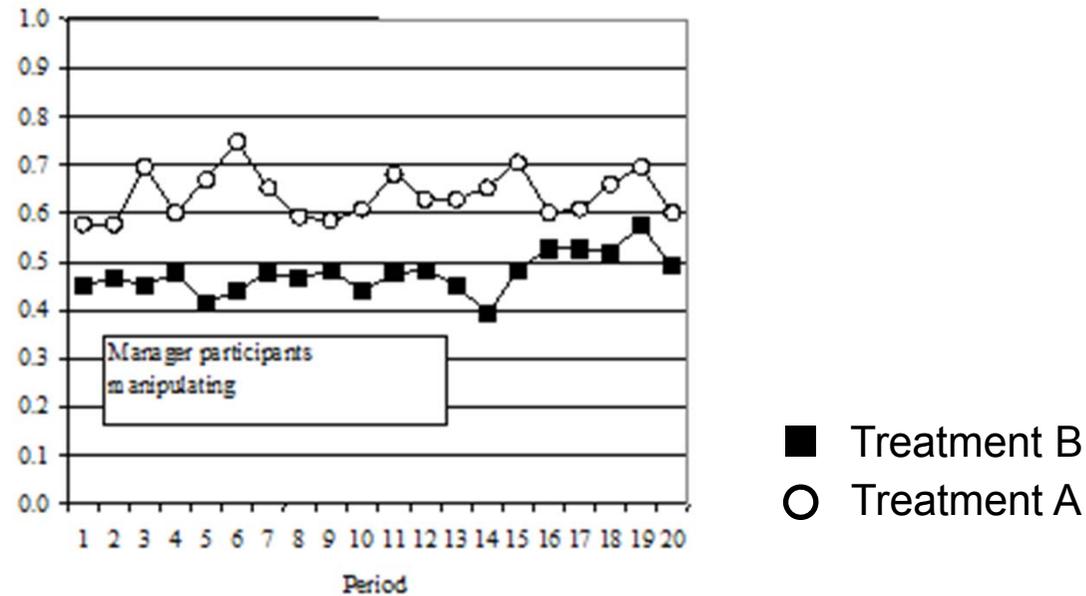
	Tests	Data ^a	All Periods	Periods 11 - 20
<i>Δ Low Audit</i>	Wilcoxon signed-rank test	all data	4.555***	4.362***
		Sessions 1 - 5	3.110***	2.830***
		Sessions 6 - 10	3.352***	3.324***
	Mann-Whitney test	First treatments	4.015***	3.490***
<i>Δ Overall Low Audit</i>	Wilcoxon signed-rank test	all data	-4.158***	-1.062**
		Sessions 1 - 5	-3.111***	-2.521**
		Sessions 6 - 10	-2.814***	-0.369
	Mann-Whitney test	First treatments	-2.887***	-1.342
<i>Δ Manipulation</i>	Wilcoxon signed-rank test	all data	-3.952***	-3.747***
		Sessions 1 - 5	-2.760***	-2.411**
		Sessions 6 - 10	-2.813***	-2.841***
	Mann-Whitney test	First treatments	-3.759***	-3.527***
<i>Δ Quality</i>	Wilcoxon signed-rank test	all data	4.555***	3.976***
		Sessions 1 - 5	3.181***	2.797***
		Sessions 6 - 10	3.239***	2.704***
	Mann-Whitney test	First treatments	4.079***	3.732***

The number shown in the table represents the z-value. It is positive if the corresponding variable is higher in treatment B (with computerized auditors).

Results

Result 3: Manager's behavior (H1)

- Frequencies of manipulations over time:



- Managers manipulate financial statements less often in the presence of computerized auditors.
- H1 is confirmed.
- This result is in line with the QRE, but contrary to the standard prediction.

Results

Statistical tests for significant differences across the treatments A and B

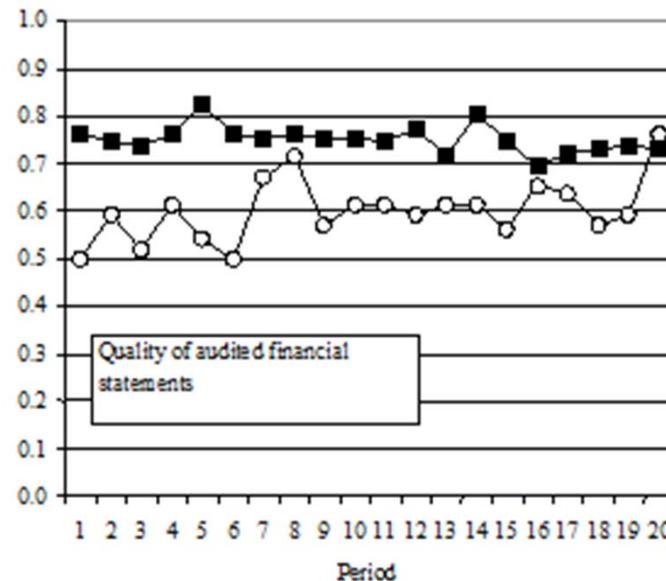
	Tests	Data ^a	All Periods	Periods 11 - 20
<i>Δ Low Audit</i>	Wilcoxon signed-rank test	all data	4.555***	4.362***
		Sessions 1 - 5	3.110***	2.830***
		Sessions 6 - 10	3.352***	3.324***
	Mann-Whitney test	First treatments	4.015***	3.490***
<i>Δ Overall Low Audit</i>	Wilcoxon signed-rank test	all data	-4.158***	-1.062**
		Sessions 1 - 5	-3.111***	-2.521**
		Sessions 6 - 10	-2.814***	-0.369
	Mann-Whitney test	First treatments	-2.887***	-1.342
<i>Δ Manipulation</i>	Wilcoxon signed-rank test	all data	-3.952***	-3.747***
		Sessions 1 - 5	-2.760***	-2.411**
		Sessions 6 - 10	-2.813***	-2.841***
	Mann-Whitney test	First treatments	-3.759***	-3.527***
<i>Δ Quality</i>	Wilcoxon signed-rank test	all data	4.555***	3.976***
		Sessions 1 - 5	3.181***	2.797***
		Sessions 6 - 10	3.239***	2.704***
	Mann-Whitney test	First treatments	4.079***	3.732***

The number shown in the table represents the z-value. It is positive if the corresponding variable is higher in treatment B (with computerized auditors).

Results

Result 4: Financial reporting quality

- Quality of audited financial statements over time



- Treatment B
- Treatment A

- The quality of audited financial statements is higher in the presence of computerized auditors.
- The introduction of computerized auditors improves the overall audit effort that is actually performed. Managers seem to anticipate the auditors' behavior and manipulate less often in the presence of these auditors → the quality of audited financial statements increases.

Results

Statistical tests for significant differences across the treatments A and B

	Tests	Data ^a	All Periods	Periods 11 - 20
<i>Δ Low Audit</i>	Wilcoxon signed-rank test	all data	4.555***	4.362***
		Sessions 1 - 5	3.110***	2.830***
		Sessions 6 - 10	3.352***	3.324***
	Mann-Whitney test	First treatments	4.015***	3.490***
<i>Δ Overall Low Audit</i>	Wilcoxon signed-rank test	all data	-4.158***	-1.062**
		Sessions 1 - 5	-3.111***	-2.521**
		Sessions 6 - 10	-2.814***	-0.369
	Mann-Whitney test	First treatments	-2.887***	-1.342
<i>Δ Manipulation</i>	Wilcoxon signed-rank test	all data	-3.952***	-3.747***
		Sessions 1 - 5	-2.760***	-2.411**
		Sessions 6 - 10	-2.813***	-2.841***
	Mann-Whitney test	First treatments	-3.759***	-3.527***
<i>Δ Quality</i>	Wilcoxon signed-rank test	all data	4.555***	3.976***
		Sessions 1 - 5	3.181***	2.797***
		Sessions 6 - 10	3.239***	2.704***
	Mann-Whitney test	First treatments	4.079***	3.732***

The number shown in the table represents the z-value. It is positive if the corresponding variable is higher in treatment B (with computerized auditors).

Results

Result 5: Decision errors

- **Decision errors are more important for participants' behavior than risk aversion.**
 - We assume that μ is constant across the treatments of our experiment.
 - The result of a maximum likelihood estimation of the pure QRE model is $\mu = 1.11$.
 - Estimation of a model in which we combine the QRE with constant relative risk aversion yields no positive parameter of risk aversion that can improve the estimate of the pure QRE model which assumes that the players are risk neutral.
- The endogenous error parameter μ , rather than risk preferences, is the principal driver of the participants' behavior.

Results

Result 6: QRE vs. Nash Equilibrium

- The QRE outperforms the Nash Equilibrium in predictive power even in static settings.
- For both treatments, the QRE predicts a smaller manipulation probability than does the Nash Equilibrium, and a higher probability of a low audit effort for treatment A without computerized auditors.
- Binomial tests for the variable “sign of the difference between the behavior in a matching group and Nash prediction” shows a significant difference ($p < 0.05$, one-sided test, both sequences used) for both hypotheses.
- Data is closer to the symmetric outcome.

Robustness Checks

Neutral framing

- 2 further sessions (32 subjects each) with a strictly neutral framing
 - “Participant A” and “participant B” instead of “manager” and “auditor”
 - Actions X, Y, I, and II
- The managers’ average manipulation probability does not differ significantly between the two framing conditions.
- In treatment A, however, auditors choose a low audit effort less frequently with the neutral framing than with the non-neutral framing. This result is the opposite of what we would expect from a social desirability argument.
- In treatment B, auditors’ behavior is not significantly different in the two framing conditions.

Robustness Checks

Statistical tests for the robustness checks

Robustness Check 1: Neutral Framing	<i>Low Audit</i>	<i>Overall Low Audit</i>	<i>Manipulation</i>	<i>Quality</i>
Wilcoxon signed rank test (2-sided) that the variable is higher in the presence of computerized auditors in the sessions with neutral framing	2.240**	0.070	-2.457**	2.524**
Mann Whitney test (2-sided) that the variable is higher in the treatment with neutral framing (order A-B; treatment A)	-2.191**	-2.191**	0.766	1.495
Mann Whitney test (2-sided) that the variable is higher in the treatment with neutral framing (order A-B; treatment B)	0.218	0.182	-0.653	0.402
Robustness Check 2: Payoff Matrix 2	<i>Low Audit</i>	<i>Overall Low Audit</i>	<i>Manipulation</i>	<i>Quality</i>
Wilcoxon signed rank test (2-sided) that the variable is higher in the presence of computerized auditors in the sessions with payoff matrix 2	3.061***	2.638***	-2.475**	0.825
Mann Whitney test (2-sided) that the variable is higher with payoff matrix 2 (order A-B; treatment A)	-4.227***	-4.227***	-4.228***	4.247***
Mann Whitney test (2-sided) that the variable is higher with payoff matrix 2 (order A-B; treatment B)	-1.337	-1.227	-3.322***	3.179***

This table presents statistical tests for the control treatments. Row 1 shows that the effect of computerized auditors in the treatment with neutral framing corresponds to the effects observed in the treatment with auditor-manager framing. We report the difference between the neutral framing and the auditor-manager framing in rows 2 and 3. It shows that the framing causes only one significant difference: auditors in treatment A are more likely to perform a low audit effort in the neutral framing. Row 4 shows that subjects respond equally to the introduction of computerized auditors as for payoff matrix 1 if we implement payoff matrix 2. Rows 5 and 6 show that our data replicate the Own Payoff Effect.

Variable Definitions:

- Low Audit* = frequency of low audit effort that auditor participants chose
- Overall Low Audit* = overall low audit effort, which also comprises the computerized audits
- Manipulation* = frequency of manipulated financial statements manager participants chose
- Quality* = resulting quality of audited financial statements

*, **, and *** indicate statistical significance at 10 percent, 5 percent, and 1 percent, respectively.

Robustness Checks

Payoff matrix

- In payoff matrix 1, the auditor's lowest payoff is realized if he chooses a high audit effort and the financial statements are correct.
- Payoff matrix 2 is closer to real-world incentives, because the auditor receives his worst payoff if he fails to detect manipulations.
- 3 sessions with payoff matrix 2 and auditor-manager framing (94 students).

Robustness Checks

Payoff matrices (payoffs in Points)

Basic Experiment (Payoff Matrix 1, Auditor-Manager Framing)^a

Auditor \ Manager	Correct financial statements	Overstatement of economic condition
Comprehensive audit	2, 6	8, 3
Normal audit effort	7, 6	6, 9

Additional Experiment (Payoff Matrix 2, Auditor-Manager Framing)

Auditor \ Manager	Correct financial statements	Overstatement of economic condition
Comprehensive audit	5, 6	8, 3
Normal audit effort	7, 6	3, 9

^a In sessions 11 and 12 with neutral framing and payoff matrix 1, we denoted the manager and the auditor as “participant A” and “participant B”, respectively, and labeled the actions as “X”, “Y”, “T”, and “II”.

Robustness Checks

Payoff matrices

- Payoff matrix 1:

- Auditor: $\beta_A^* = 0.5$ and $\beta_B^* = 0.75$
- Manager: $\alpha_A^* = 0.714$ and $\alpha_B^* = 0.714$
- Quality: $1 - \alpha_A^* \cdot \beta_A^* = 0.643$ and $1 - \alpha_B^* \cdot 2/3 \cdot \beta_B^* = 0.643$

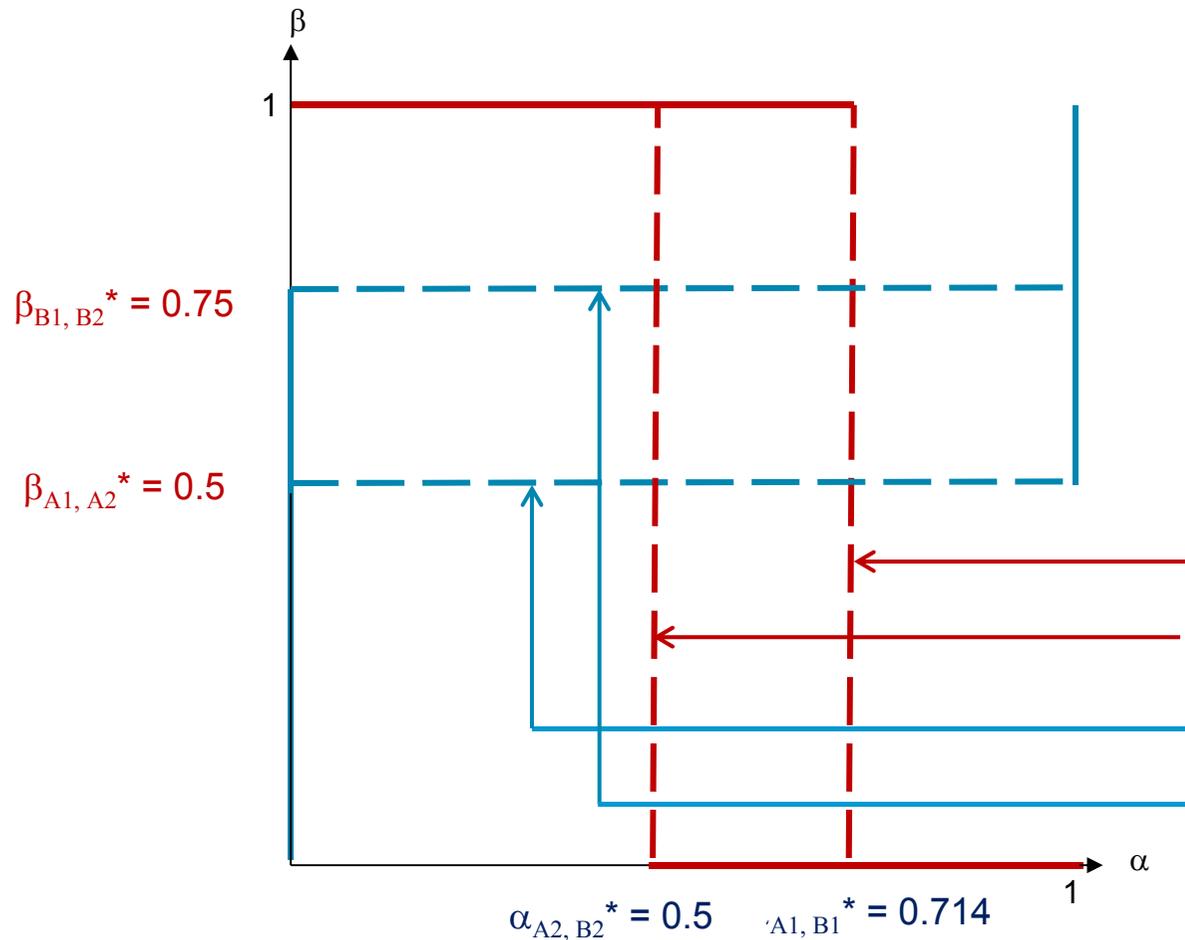
- Payoff matrix 2:

- Auditor: $\beta_A^* = 0.5$ and $\beta_B^* = 0.75$
- Manager: $\alpha_A^* = 0.5$ and $\alpha_B^* = 0.5$
- Quality: $1 - \alpha_A^* \cdot \beta_A^* = 0.75$ and $1 - \alpha_B^* \cdot 2/3 \cdot \beta_B^* = 0.75$

- Standard-Prediction:

- The auditor's best reply does not change if his payoffs change (counter-intuitive)!
- The auditor's Quantal best reply changes if the auditor's payoffs change.
- Whereas the Nash Equilibrium predicts no difference in auditors' behavior between matrix 1 and matrix 2, the QRE predicts fewer low effort audits with payoff matrix 2 (for both treatments A and B).

Robustness Checks



Matrix 1 → matrix 2:
 Managers manipulate less frequently, but auditor's behavior does not change.

Auditor A1 and B1

Auditor A2 and B2

Manager A1 and A2

Manager B1 and B2

Robustness Checks

Statistical tests for the robustness checks

Robustness Check 1: Neutral Framing	<i>Low Audit</i>	<i>Overall Low Audit</i>	<i>Manipulation</i>	<i>Quality</i>
Wilcoxon signed rank test (2-sided) that the variable is higher in the presence of computerized auditors in the sessions with neutral framing	2.240**	0.070	-2.457**	2.524**
Mann Whitney test (2-sided) that the variable is higher in the treatment with neutral framing (order A-B; treatment A)	-2.191**	-2.191**	0.766	1.495
Mann Whitney test (2-sided) that the variable is higher in the treatment with neutral framing (order A-B; treatment B)	0.218	0.182	-0.653	0.402

Robustness Check 2: Payoff Matrix 2	<i>Low Audit</i>	<i>Overall Low Audit</i>	<i>Manipulation</i>	<i>Quality</i>
Wilcoxon signed rank test (2-sided) that the variable is higher in the presence of computerized auditors in the sessions with payoff matrix 2	3.061***	2.638***	-2.475**	0.825
Mann Whitney test (2-sided) that the variable is higher with payoff matrix 2 (order A-B; treatment A)	-4.227***	-4.227***	-4.228***	4.247***
Mann Whitney test (2-sided) that the variable is higher with payoff matrix 2 (order A-B; treatment B)	-1.337	-1.227	-3.322***	3.179***

This table presents statistical tests for the control treatments. Row 1 shows that the effect of computerized auditors in the treatment with neutral framing corresponds to the effects observed in the treatment with auditor-manager framing. We report the difference between the neutral framing and the auditor-manager framing in rows 2 and 3. It shows that the framing causes only one significant difference: auditors in treatment A are more likely to perform a low audit effort in the neutral framing. Row 4 shows that subjects respond equally to the introduction of computerized auditors as for payoff matrix 1 if we implement payoff matrix 2. Rows 5 and 6 show that our data replicate the Own Payoff Effect.

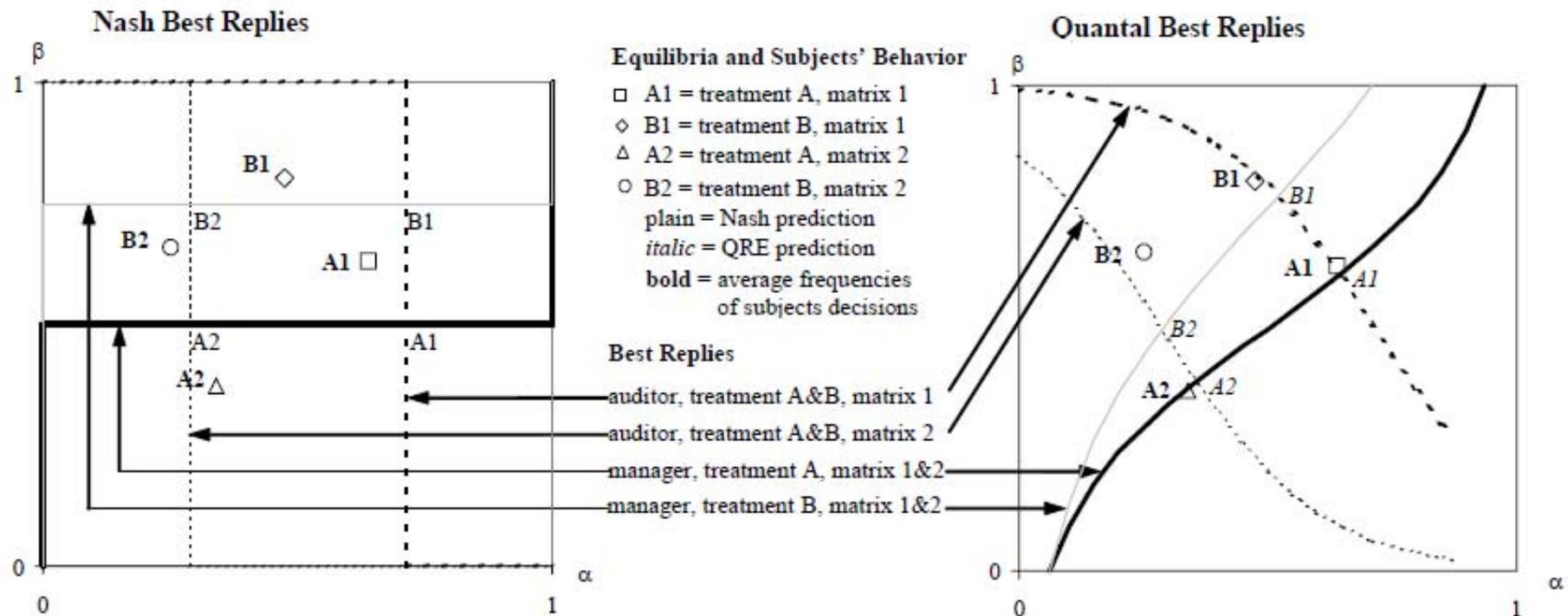
Variable Definitions:

- Low Audit* = frequency of low audit effort that auditor participants chose
- Overall Low Audit* = overall low audit effort, which also comprises the computerized audits
- Manipulation* = frequency of manipulated financial statements manager participants chose
- Quality* = resulting quality of audited financial statements

*, **, and *** indicate statistical significance at 10 percent, 5 percent, and 1 percent, respectively.

Robustness Checks

Nash best replies, Quantal best replies (error parameter $\mu = 1.11$), corresponding equilibria, and experimental data



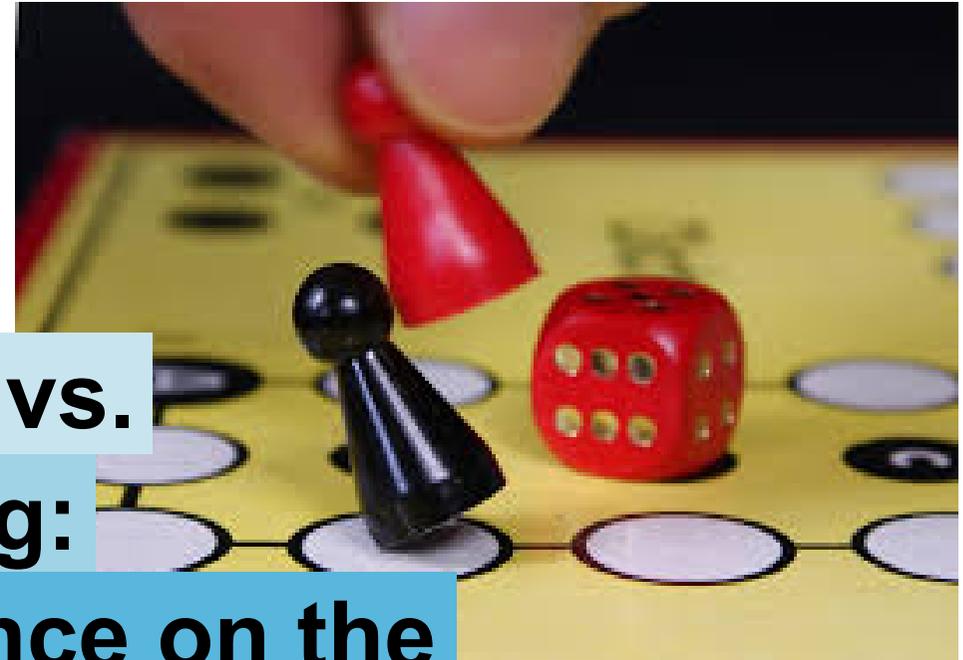
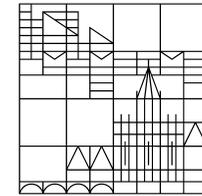
Conclusion

Mixed empirical results

- The QRE outperforms the Nash Equilibrium with regard to predictive power (**point prediction**). This is no big surprise because the QRE has one more parameter.
- The QRE is better than the Nash Equilibrium in **predicting comparative static effects**: The predictions are unambiguous and intuitive.
 - **Framing** does not play a mayor role.
 - We can replicate the **Own-Payoff Effect**.
 - The results are due to **bounded rational behavior**, **not due to risk aversion** ($\mu = 1.11$).
- Recommendations for business practice:
 - A certain percentage of „perfect“ auditors would improve managers‘ behavior and the quality of audited financial statements → professional ethics.
 - Own-Payoff Effect: If the auditor’s payoffs are altered, there is an effect on auditors‘ behavior: Recommendations regarding institutional changes.
- Be aware of counterintuitive predictions based on standard game theory ... ;-)

Examples

- **Strategic Interaction and Testing Theories (Lab):**
Fischbacher, Urs; Stefani, Ulrike (2007): Strategic Errors and Audit Quality – An Experimental Investigation, in: The Accounting Review 82 (3): 679-704. 
- **Strategic Interaction and Comparison of Institutions (Lab):**
Hattenbach, Thomas; Stefani, Ulrike (2016): Audit Firm Rotation vs. Centralized Matching: Experimental Evidence on the Quality of Audited Financial Statements, Working Paper. 
- **Individual Decision-Making and Comparison of Rules (Paper & Pencil):**
Lachmann, Maik; Stefani, Ulrike; Wöhrmann, Arndt (2015): Fair Value Accounting for Liabilities: Presentation Format of Credit Risk Changes and Individual Information Processing, in: Accounting, Organizations and Society 41 (1): 21-38. 
- **Individual Decision-Making and Comparison of Rules (Mobile Lab & Eye-Tracking):**
Hattenbach, Thomas; Lachmann, Maik; Stefani, Ulrike: Auditors' Information Processing and Going-Concern Opinions: Experimental Evidence on IAS 17 vs. IFRS 16. 



**Audit Firm Rotation vs.
Centralized Matching:
Experimental Evidence on the
Quality of Audited Financial Statements**

Thomas Hattenbach / Ulrike Stefani
(Working Paper)

Motivation: The Regulatory Debate

The audit firm rotation (MAR)

- Already is mandatory in some countries / markets.
- In the EU, the MAR after 20 years (4 years cooling-off) is required since 17th June 2016.
- The discussion was highly controversial.

Is a centralized matching mechanism an alternative to the MAR?

- A third party decides about the appointment, the audit fees, and the duration of the mandate.
- Suggested in 2010 (Commission of the European Communities (2010)).
- Similar to the regulations in the UK National Health Service effective until 2012.

Research questions:

- Does the MAR have the effects that the regulator desires (higher degree of independence; less concentrated markets)?
- Would a centralized matching mechanism even outperform the MAR?

Literature

Mixed empirical results

- The quality of audited financial statements is **higher** in the years after an auditor change:
 - Carey / Simnett (2006), Kealey et al. (2007), Dao et al. (2008), Davis et al. (2009), ...
- The quality of audited financial statements is **lower** in the years after an auditor change:
 - St. Pierre / Anderson (1984), Louwers (1998), Johnson et al. (2002), Carcello / Nagy (2004), Mansi et al. (2004), Gul et al. (2007), Chen et al. (2008), Kwon et al. (2014), Cameran et al. (2015), ...
- The MAR **decreases the level of concentration** (Vinten et al. (2005)).
 - The level of market concentration has a **negative** effect on the quality of audited financial statements (Bandyopadhyay / Kao (2001), Boone et al. (2012)).
 - The level of market concentration has a **positive** effect on financial reporting quality (Kallapur et al. (2010), Dunn et al. (2013), Newton et al. (2013), Francis et al. (2013)).

Limited experimental results

- Decision experiments: Jennings et al. (2006), Kaplan / Mauldin (2008), Daniels / Booker (2011), Arel et al. (2006)
- Dopuch et al. (2001): The MAR significantly decreases the auditors' "willingness to bias their reports in favor of management" (Wang / Tuttle (2009); Bowlin et al. (2015)).

Experimental Design

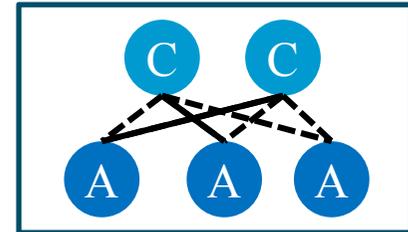
Treatment comparisons

- **NOREST**: Baseline treatment without any regulations regarding the auditor-client matching or the duration of the mandate
- **ROTATION**: Mandatory rotation:
The clients can hire the same auditor only for 4 consecutive periods;
cooling-off of 2 periods.
- **MATCHING**: Computerized matching mechanism:
The experimental software matches the clients who report fraudulently
with the auditors who (check correctly and) report truthfully.

Experimental Design

Procedure within a group / market in each of the 15 periods of a session of the treatment NOREST

- The clients receive information about the behavior of the auditors in their market.
- Each client hires an auditor.
- The auditors are informed whether they are hired; they receive information about the behavior of their client.
- The clients receive a 12x12 matrix containing “0” and “1”; they get to know the true number T of “1” contained in their matrix ($0 \leq T \leq 144$).
- The clients report to their auditor a number X_C of “1” contained in their matrix.
- The auditors receive their client’s matrix and the number X_C reported by their client.
- The auditors can check the number of “1” reported.
- The auditors confirm or correct their client’s report by stating X_A .
- The resulting Points are calculated and displayed.



Experimental Design

Matrix and reports

- Client:

1	0	1	0	1	1	0	1	0	1	1	0
0	1	0	1	1	0	0	1	1	0	1	0
0	1	1	0	0	1	1	0	1	1	0	1
1	0	1	1	1	0	1	1	0	0	1	0
1	1	0	1	1	1	0	1	1	1	0	1
1	0	1	1	0	0	1	0	0	1	1	0
0	0	0	0	1	0	0	0	1	1	0	1
1	0	0	1	0	1	0	1	0	0	0	0
0	0	1	0	1	0	0	1	0	1	0	1
1	0	1	1	1	0	1	0	1	0	1	0
0	1	0	1	0	1	0	1	0	0	0	1

$$T = 67$$

$$X_C = 77$$

- Auditor:

1	0	1	0	1	1	0	1	0	1	1	0
0	1	0	1	1	0	0	1	1	0	1	0
0	1	1	0	0	1	1	0	1	1	0	1
1	0	1	1	1	0	1	1	0	0	1	0
1	1	0	1	1	1	0	1	1	1	0	1
1	0	1	1	0	0	1	0	0	1	1	0
0	0	0	0	1	0	0	0	1	1	0	1
1	0	0	1	0	1	0	1	0	0	0	0
0	0	1	0	1	0	0	1	0	1	0	1
1	0	1	1	1	0	1	0	1	0	1	0
0	1	0	1	0	1	0	1	0	0	0	1

$$X_C = 77$$

Counter 21

$$X_A = 70$$

Experimental Design

Information

- **Client (hiring decision):**
 - number of the current period (1 to 15)
 - auditors' subject-IDs
 - number of periods in which each auditor had been hired
 - number of periods in which the auditors' clients had reported truthfully
 - number of periods in which the auditors had not detected over-reports
 - number of "1" corrected in fraudulent reports, in %
- **Auditor:**
 - number of the current period (1 to 15)
 - subject-ID of the client who had hired the auditor (NOREST and ROTATION) or the auditor was matched with (MATCHING)
 - number of periods in which any hired auditor had corrected this client's report
 - number of periods in which this client's fraudulent reporting had not been corrected
 - number of periods in which this client had been truth-telling

Experimental Design

Payoff structure of the auditor:

$$\text{Points}_A = \begin{cases} 50 - \frac{1}{2} * |T - X_A|, & \text{if hired} \\ 0 & , \text{otherwise} \end{cases}$$

- Fixed fee per client
- Deduction of Points
 - if the auditor does not correctly check the client's report
 - or approves a fraudulent report against better knowledge.

T: number of "1" contained in the matrix
X_A: number of "1" confirmed by the auditor

100 Points ≅ 1€

Experimental Design

Payoff structure of the client:

1. The client reports truthfully ($X_C = T$):

$$\text{Points}_C = T = X_C$$

2. The client reports fraudulently ($X_C > T$):

- a) The auditor collaborates ($X_A = X_C$) or even corrects upwards ($X_A > X_C$):

$$\text{Points}_C = X_C$$

- b) The auditor corrects the client's report downwards ($X_A < X_C$; $X_A \geq T$):

$$\text{Points}_C = X_A - \frac{1}{2} * (X_C - X_A)$$

- c) The auditor corrects the client's report downwards and even goes below the number T of "1" contained in the matrix ($X_A < X_C$; $X_A < T$):

$$\text{Points}_C = T - \frac{1}{2} * (X_C - T)$$

T: number of "1" contained in the matrix
 X_A : number of "1" confirmed by the auditor
 X_C : number of "1" reported by the client

100 Points $\hat{=}$ 1€

Behavioral Predictions: Auditors

- **NOREST**: The auditors try to increase their probability of getting re-hired by confirming $X_A > T$.
- **ROTATION**: Biasing the report (i.e., to choose $X_A > T$) is less beneficial in ROTATION than in NOREST because long-term relationships are ruled out.
- **MATCHING**: By design, the probability of getting matched with a client in the following periods decreases under the central matching mechanism if $X_A > T$: Truth-telling is rewarded.

Hypotheses:

H1: The frequency of auditors' reports stating the true value of "1" contained in the matrix is higher in the treatment ROTATION than in the treatment NOREST.

H2: The frequency of auditors' reports stating the true value of "1" contained in the matrix is higher in the treatment MATCHING than in the treatment ROTATION.

H1x2: The frequency of auditors' reports stating the true value of "1" contained in the matrix is higher in the treatment MATCHING than in the treatment NOREST.

Behavioral Predictions: Clients

- **NOREST**: The clients anticipate the auditors' willingness to bias their reports in order to get re-hired: The clients report $X_C > T$.
- **ROTATION**: The clients anticipate that the auditors' incentives to build a reputation for confirming over-reports are weaker than in NOREST; the clients' relative power over the auditors is lower than in NOREST.
MAR increases the client's uncertainty about the auditor's behavior: Imprecise beliefs are costly.
- **MATCHING**: The central matching makes biased reports "expensive".

Hypotheses:

H3: The magnitude of clients' fraudulent reports is lower in the treatment ROTATION than in the treatment NOREST.

H4: The magnitude of clients' fraudulent reports is lower in the treatment MATCHING than in the treatment ROTATION.

H3x4: The magnitude of clients' fraudulent reports is lower in the treatment MATCHING than in the treatment NOREST.

Behavioral Predictions: Market Concentration

- **NOREST**: Long-term contractual relationships based on “mutual agreements” to misreport.
- **ROTATION**: The rotation results in a more dynamic market.

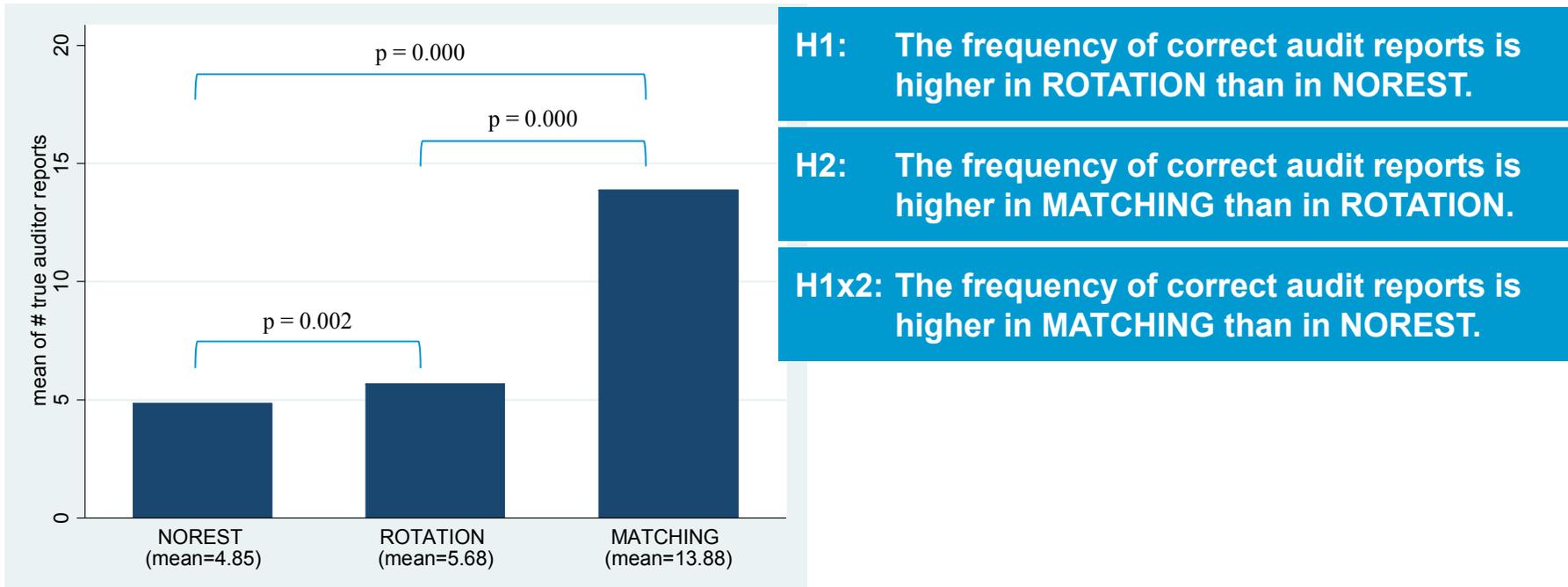
Hypothesis:

H5: The degree of inequality in the distribution of market shares is higher in the treatment NOREST than in the treatment ROTATION.

Note that standard theory does not predict any treatment differences (clients always report truthfully, and auditors always confirm the clients' reports; each auditor has the same market share).

Results: Auditors

The quality of the auditors' reports increases across the treatments



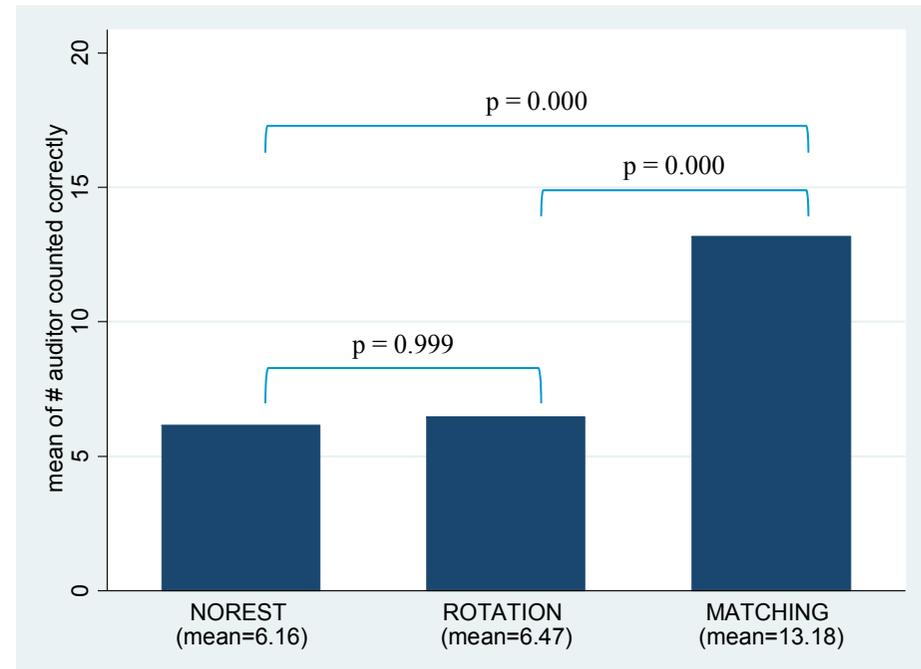
Bar-charts for the mean **number of correct reports issued by the auditor** subjects for the treatments NOREST, ROTATION, and MATCHING. P-Values for the two-sample Kolmogorov-Smirnov test for equality of distributions are displayed.

Evidence for hypotheses H1, H2, and H1x2.

Results: Auditors

The quality of the auditors' reports increases across the treatments

- Higher report quality in ROTATION than in NOREST due to a **higher degree of independence**.
- Highest report quality in MATCHING due to a **higher degree of independence and more intense testing**.

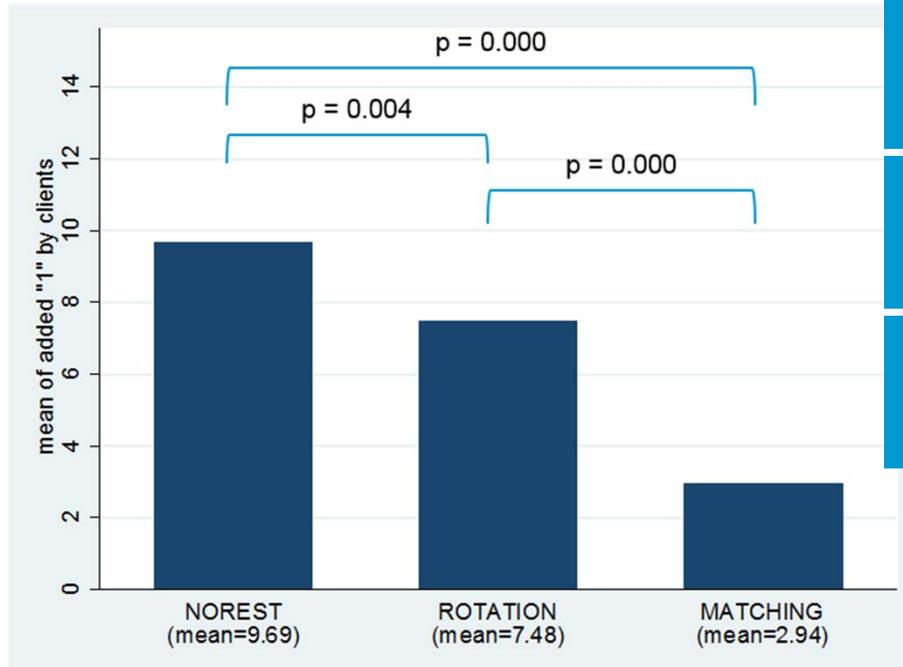


Bar-charts for the mean **number of matrices correctly counted by the auditor** subjects for the treatments NOREST, ROTATION, and MATCHING. P-Values for the two-sample Kolmogorov-Smirnov test for equality of distributions are displayed.

Evidence for hypotheses H1, H2, and H1x2.

Results: Clients

Fraud behavior of the clients decreases across the treatments



H3: The magnitude of clients' fraudulent reports is lower in ROTATION than in NOREST.

H4: The magnitude of clients' fraudulent reports is lower in MATCHING than in ROTATION.

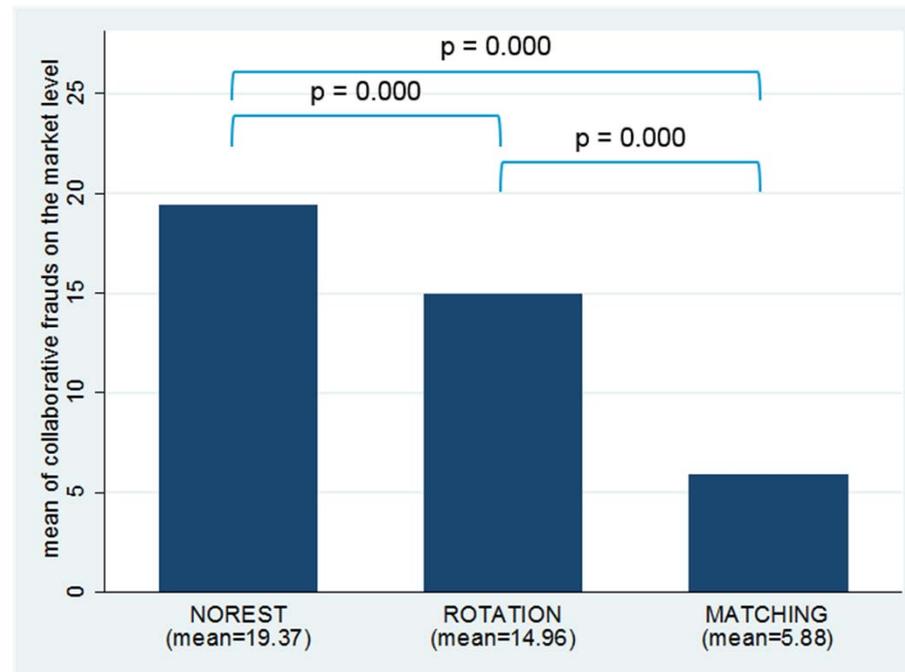
H3x4: The magnitude of clients' fraudulent reports is lower in MATCHING than in NOREST.

Bar-charts for the mean **number of "1" added by the client** subjects for the treatments NOREST, ROTATION, and MATCHING. P-Values for the two-sample Kolmogorov-Smirnov test for equality of distributions are displayed.

Evidence for hypotheses H3, H4, and H3x4.

Results: Collaborative Fraud

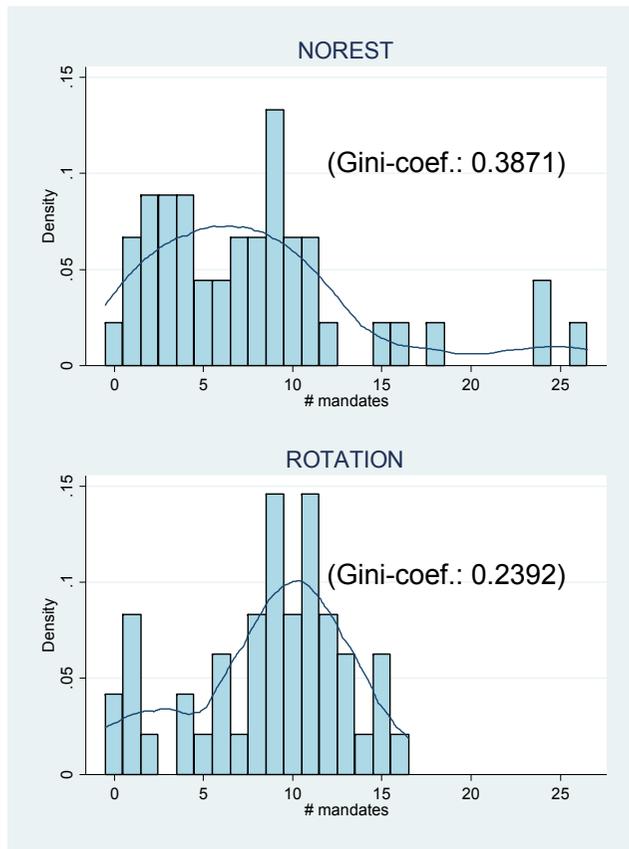
The magnitude of collaborative fraud (aggregated on the market level) is significantly lower in the treatment ROTATION than in the treatment NOREST and lowest in the treatment MATCHING.



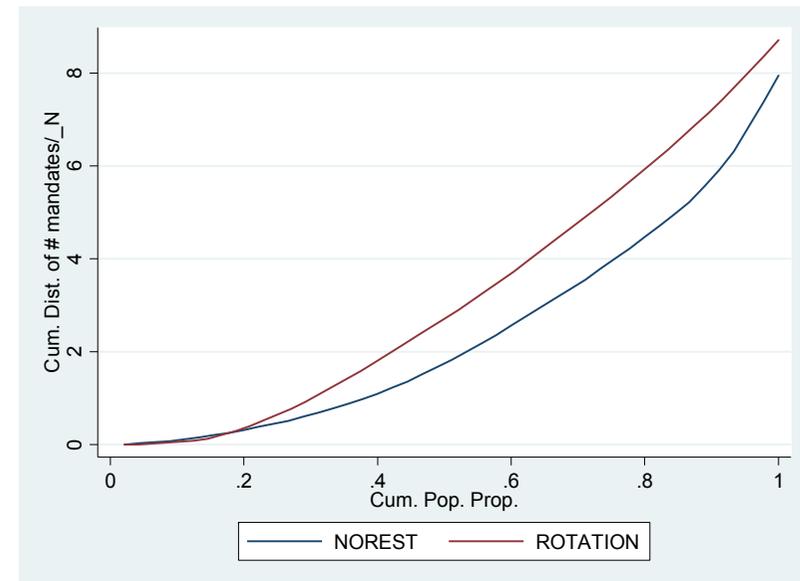
Bar-charts for the number of “1” added by the client subjects which have been approved by the auditors, aggregated on the market level. Bar-charts are displayed for the treatments NOREST, ROTATION, and MATCHING. P-Values for the two-sample Kolmogorov-Smirnov test for equality of distributions are displayed.

Results: Market Concentration

H5: The degree of inequality in the distribution of market shares is higher in the treatment NOREST than in the treatment ROTATION.



Histograms and Kernel-densities for the number of mandates in the treatments NOREST and ROTATION at the end of the experiment.



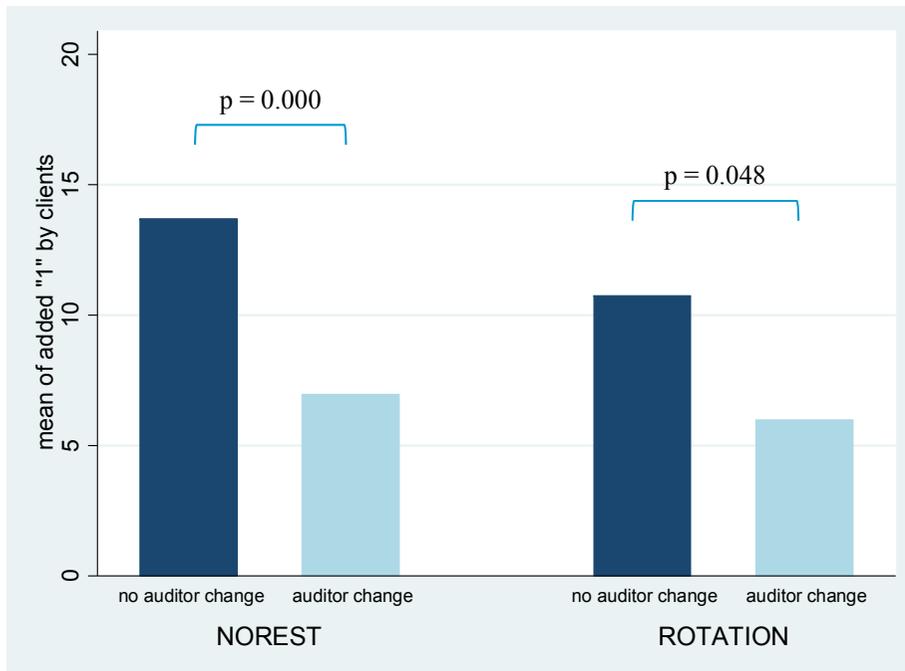
Lorenz-curves for the number of mandates in the treatments NOREST and ROTATION at the end of the experiment.

- The Lorenz-curve representing the market shares in the treatment NOREST is strictly Lorenz-dominated by the corresponding Lorenz-curve representing the treatment ROTATION (p-value: 0.000; bootstrap repetitions: 1000)

Additional Results: Auditor / Client Changes

Clients' reports

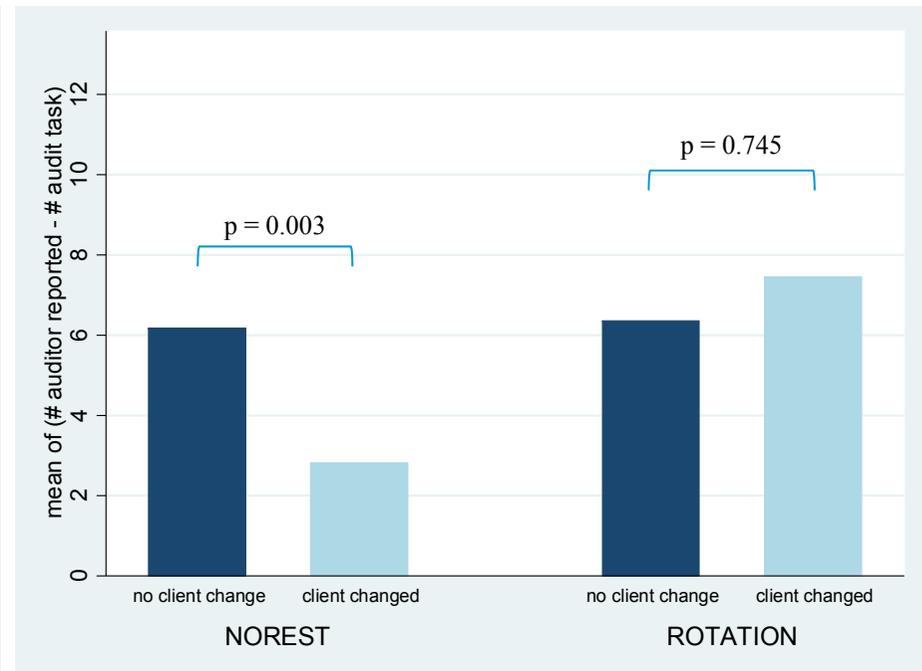
- Clients are more dishonest in ongoing relationships (NOREST, ROTATION): Auditor change reduces over-reporting.



Bar-charts for the mean number of "1" added by the client subjects, separated for either being audited by a newly hired auditor or not. Bar-charts are shown for the treatments NOREST and ROTATION. P-Values for the two-sample Kolmogorov-Smirnov test for equality of distributions are displayed.

Auditors' benevolence

- Auditors are more benevolent in ongoing relationships (NOREST): Auditor change reduces over-reporting.



Bar-charts for the mean number of "1" added by the auditor subjects to the result of their own audit task, separated for either being hired by a new client or not. Bar-charts are shown for the treatments NOREST and ROTATION. P-Values for the two-sample Kolmogorov-Smirnov test for equality of distributions are displayed.

Summary

The frequency of correct auditor reports is higher under the rotation than in an unregulated market and highest under a centralized matching mechanism

- Higher frequency of independent audits (compared to the benchmark NOREST) in both treatments ROTATION and MATCHING.
- More testing in the treatment MATCHING than in NOREST and ROTATION.

The magnitude of clients' fraudulent reporting is lower under the rotation than in an unregulated market and lowest under the centralized matching mechanism

- Overall, fewer frauds in ROTATION than in NOREST and **least and smallest** under MATCHING.

The ROTATION leads to a more equal distribution of market shares.

Research questions:

- Does the MAR have the effects that the regulator desires (higher degree of independence; less concentrated markets)? ► **YES!**
- Would a centralized matching mechanism even outperform the MAR? ► **YES!**

Possible Extensions

Joint audit

- Clients need to hire two auditors each period.
- Each auditor checks half of the matrix; both auditors have to agree on a joint report; they are jointly liable.

Retention

- Retention as an alternative / supplement to the MAR.

Communication

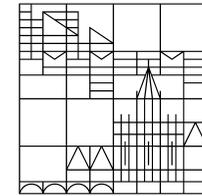
- Allow subjects to chat during the relationship (collaborative fraud?).
- Allow subjects to chat before the hiring decision is made (implicit promises?).
- Establish a group chat for auditors (cartel-building?).

Negotiation

- Include negotiation about the result of the audit.
- Include negotiation about audit fees.

Examples

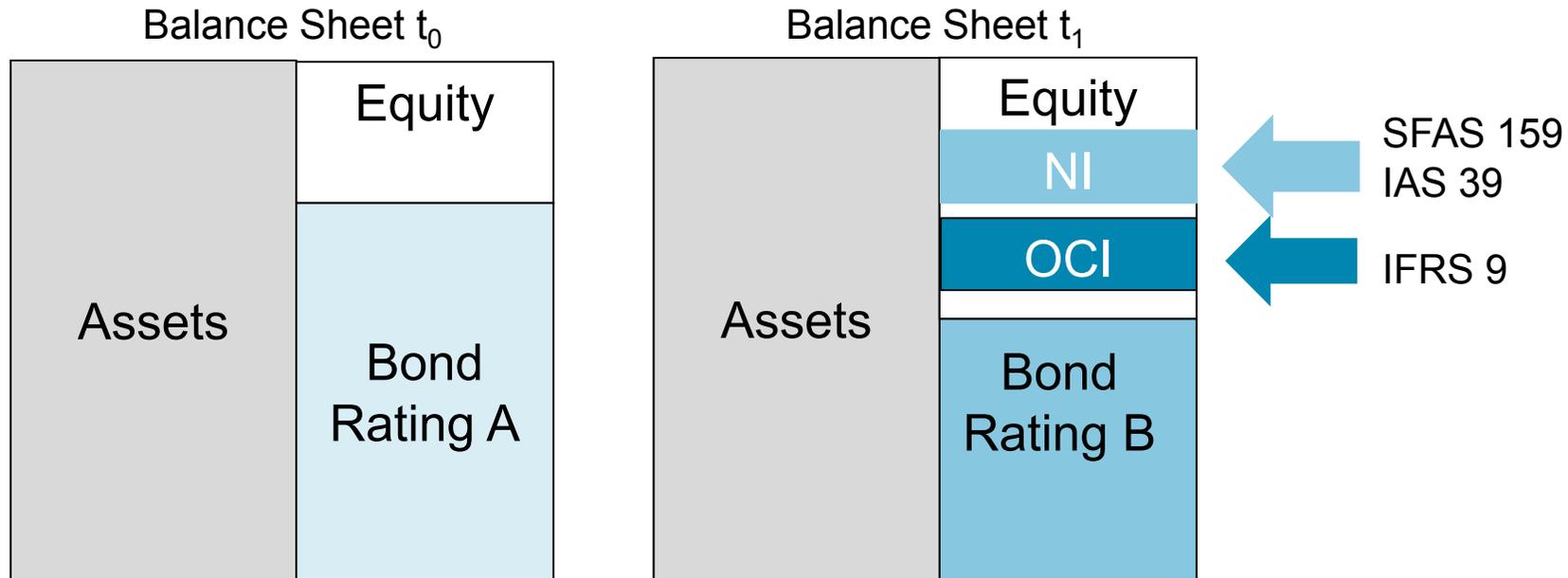
- **Strategic Interaction and Testing Theories (Lab):**
Fischbacher, Urs; Stefani, Ulrike (2007): Strategic Errors and Audit Quality – An Experimental Investigation, in: The Accounting Review 82 (3): 679-704. 
- **Strategic Interaction and Comparison of Institutions (Lab):**
Hattenbach, Thomas; Stefani, Ulrike (2016): Audit Firm Rotation vs. Centralized Matching: Experimental Evidence on the Quality of Audited Financial Statements, Working Paper. 
- **Individual Decision-Making and Comparison of Rules (Paper & Pencil):**
Lachmann, Maik; Stefani, Ulrike; Wöhrmann, Arndt (2015): Fair Value Accounting for Liabilities: Presentation Format of Credit Risk Changes and Individual Information Processing, in: Accounting, Organizations and Society 41 (1): 21-38. 
- **Individual Decision-Making and Comparison of Rules (Mobile Lab & Eye-Tracking):**
Hattenbach, Thomas; Lachmann, Maik; Stefani, Ulrike: Auditors' Information Processing and Going-Concern Opinions: Experimental Evidence on IAS 17 vs. IFRS 16. 



**Fair Value Accounting
for Liabilities: Presentation
Format of Credit Risk Changes
and Individual Information Processing**

Maik Lachmann / Ulrike Stefani / Arndt Wöhrmann (AOS (2015))

Motivation



- Both US-GAAP (SFAS 159) and IFRS (IAS 39) allow measuring liabilities at fair value.
- **Assume a decrease in an entity's creditworthiness**: Decrease in the fair value of its liabilities and increase in its equity (fair value gain).
- **Counter-intuitive effect on NI and on equity**
 - Gaynor et al. (2011): 75% of CPAs misinterpret a fair value gain as a positive signal
- Increase in NI volatility, biased evaluation of firm performance

Biased investment decisions

Research Question

Research question:

- Which presentation format is more appropriate for fulfilling investors' information needs?
- How does the presentation of changes in own credit risk in OCI instead of in NI affect the perceptions of knowledgeable nonprofessional investors (**effect of presentation format**)?
- Information processing as modeled by Maines / McDaniel (2000) (nonprofessional investors apply a sequential search strategy):
 - **Acquisition** (reading and storing) of information on a credit risk increase
 - **Evaluation** (interpretation) of information on a credit risk increase
 - **Weighting** of the credit risk increase to evaluate overall firm performance
 - **Evaluation** of overall firm performance (given the credit risk increase)

Research question:

- How does firm profitability affect investor perceptions of credit risk changes (profit versus loss)?

Rating downgrades...	Profit	Loss
EURO STOXX 50 (2009 -11)	79%	21%
Jiang (2008)	67%	33%

Experimental Design

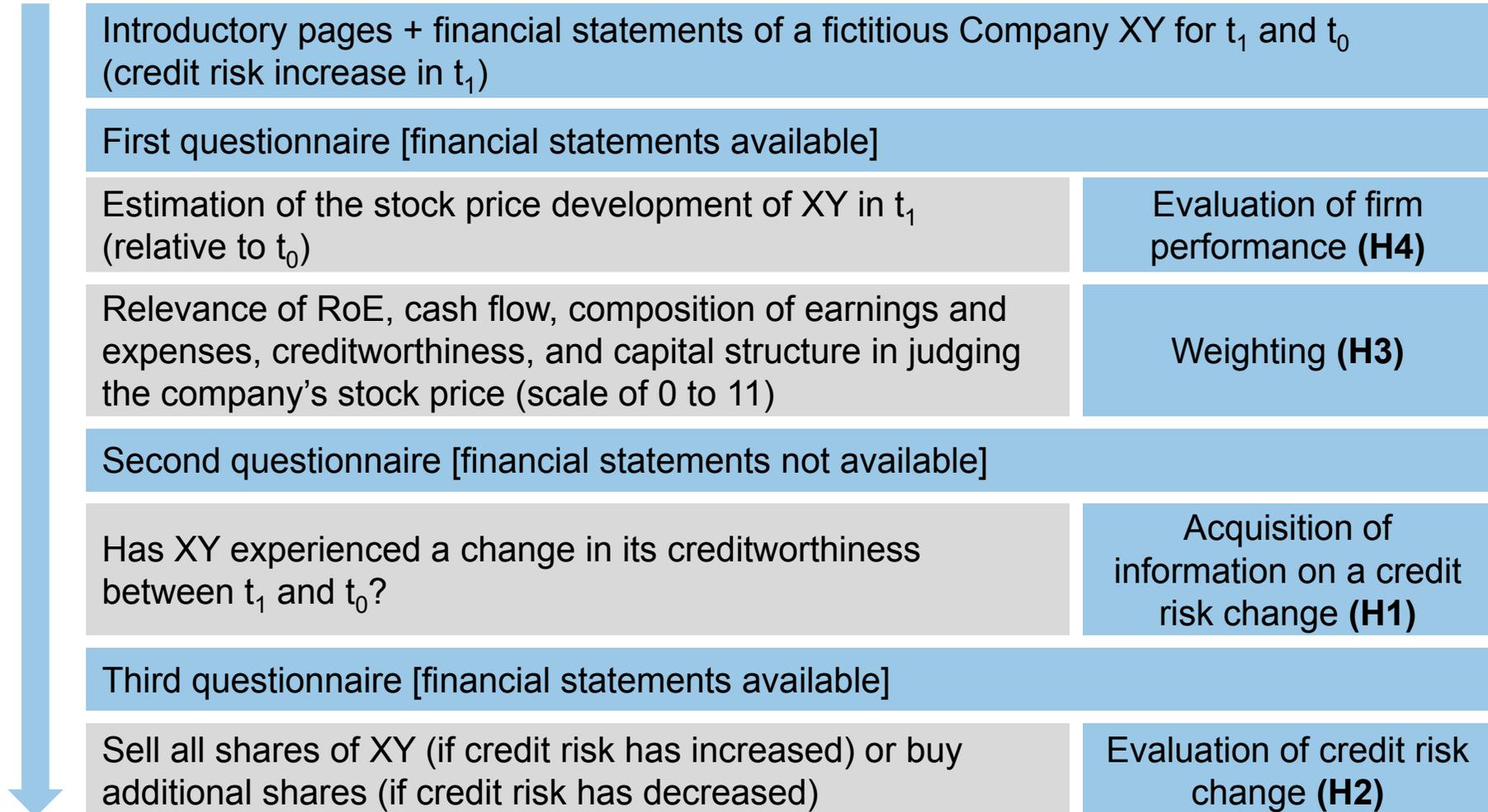
Design

- 2×2 between-subjects design
- Treatment variables:
 - Presentation format for credit risk gains (NI vs. OCI)
 - Profitability (NO_LOSS vs. LOSS due to an R&D asset impairment)
- Paper & pencil experiment conducted during a professional training seminar.
- Participants: **93 auditors** as proxy for **knowledgeable nonprofessional investors**.



Experimental Design

Task



Experimental Design

Task

Cognitive Reflection Test (Frederick (2005))

“A bat and a ball cost € 1.10. The bat costs € 1.00 more than the ball. How much does the ball cost?”

~~10 Cents~~

5 Cents

Disposition to think twice
(40% correct)

Risk Lottery (Dohmen et al. (2011)) with 20 situations

	50/50-Lottery	Safe Payment
15	0/300	140
16	0/300	150
17	0/300	160
18	0/300	170

Risk preferences
(risk-averse 43%,
risk-neutral 43%,
risk-loving 14%)

Manipulation checks (> 91% correct answers)

Demographic data

Experimental Design

Compensation

- € 7 for returning a complete set of questionnaires
- + performance-contingent payment (€ 0 to € 10)
- + payoffs from the Lottery Task (up to € 3)

Treatment	Number of participants	Average total compensation
NI×NO_LOSS	22	€ 16.96
OCI×NO_LOSS	24	€ 17.40
NI×LOSS	24	€ 18.02
OCI×LOSS	23	€ 17.94
Total	93	€ 17.59

Hypotheses

Characteristics of the presentation formats that are likely to affect information processing

	NI	OCI
Prominence	low	high
Nature of income	high relevance	low relevance
Labeling	“gains from changes in the fair value of long-term debt”	

Dow Jones EURO STOXX 50 companies (2010):

- Net income: 13.7 items plus 4.7 (sub-)totals
- OCI: 5.7 items plus 1.5 (sub-)totals

Hypotheses: Acquisition of Credit Risk Information

Prominence:

- The **higher the prominence** of a single piece of information, the **lower the cognitive costs of acquiring** this information.
- Information is more prominent under the OCI presentation format.

H1a: Individuals are more likely to acquire information on credit risk changes when gains from credit risk increases are presented in OCI rather than in net income.

Losses are perceived as a “red flag”

- Individuals are more likely to read financial information (NI) more diligently to identify the reasons for the loss.

H1b: Differences in the likelihood of acquiring information on credit risk changes presented in OCI versus in net income are mitigated when a loss is reported.

Results: Acquisition of Credit Risk Information

Percentage of subjects who acquired the information on a change in credit risk:

	NI	OCI	Σ
NO_LOSS	59%	88%	74%
LOSS	87%	91%	89%
Σ	74%	89%	82%

- Categorical model: Significant main effect for presentation format supports H1a ($\chi^2 = 4.38$, $p = 0.04$).
- Planned contrast test of interaction supports H1b ($\chi^2 = 8.11$, $p < 0.01$).

H1a: Individuals are more likely to acquire information on credit risk changes when gains from credit risk increases are presented in OCI rather than in net income.

H1b: Differences in the likelihood of acquiring information on credit risk changes presented in OCI versus in net income are mitigated when a loss is reported.

Hypothesis: Evaluation of Credit Risk Information

Labeling:

- Credit risk effects on income are labeled identically in the NI and the OCI presentation formats.

H2: Individuals are equally likely to misinterpret information on credit risk changes when gains from credit risk increases are presented in OCI or in net income.

- If a loss has occurred, individuals might expend more cognitive effort in interpreting the fair value gain.
- Information that a loss has occurred primes subjects in the correct direction (i.e., that firm performance has deteriorated).
- The increase in attention and priming affect the two presentation formats equally; H2 does not depend on the sign of net income.

Results: Evaluation of Credit Risk Information

Percentage of subjects who correctly evaluated the change in credit risk [if acquired]:

	NI	OCI	Σ
NO_LOSS	32% [46%]	54% [52%]	43% [50%]
LOSS	50% [52%]	48% [52%]	49% [52%]
Σ	41% [50%]	51% [52%]	46% [51%]

- Categorical model: No significant main effect for presentation format, supports H2.
- Interpretation of changes in credit risk does not depend on firm profitability.
- Wilcoxon-Mann-Whitney Test: Subjects who evaluate changes in credit risk correctly are more likely to correctly answer the first question of the CRT ($Z = 1.64$, $p = 0.05$, one-sided).

H2: Individuals are equally likely to misinterpret information on credit risk changes when gains from credit risk increases are presented in OCI or in net income.

Hypothesis: Weighting of Credit Risk Information

Nature of Income:

- NI is regarded as more relevant for the evaluation of overall firm performance than OCI.

H3a: Presenting credit risk gains in OCI rather than in net income decreases the weight individuals place on this piece of information when assessing overall firm performance.

- When NI becomes negative, investor attention might be distracted from credit risk information. The negative signal of a reported loss outweighs the information on fair value changes of liabilities.

H3b: Differences in the weighting of credit risk changes presented in OCI versus net income are mitigated when a loss is reported.

Results: Weighting of Credit Risk Information

Standardized [original] weight of information on a credit risk change:

	NI	OCI	Σ
NO_LOSS	0.90 [5.62]	-2.68 [3.48]	-1.31 [4.29]
LOSS	-2.94 [3.48]	-2.61 [3.71]	-2.77 [3.60]
Σ	-1.47 [4.29]	-2.64 [3.60]	-2.12 [3.91]

- ANOVA: Only weak effect of presentation format on weight (no significant effect when subjects who did not pass the manipulation checks are dropped), no unambiguous support for H3a.
- Planned contrast test of interaction supports H3b ($\chi^2 = 9.68$, $p = 0.01$).
- Significant main effect of LOSS (the weight attributed to the credit risk information is lower for a loss-making firm versus a no-loss firm).

H3a: Presenting credit risk gains in OCI rather than in net income decreases the weight individuals place on this piece of information when assessing overall firm performance.

H3b: Differences in the weighting of credit risk changes presented in OCI versus net income are mitigated when a loss is reported.

Hypothesis: Evaluation of Overall Firm Performance

Biased estimates of firm performance are less likely when information on an increase in credit risk ...

- ... is **acquired** (which is more likely for OCI (higher prominence of information))
- ... is **misinterpreted**, but **not heavily weighted** (i.e., when presented in OCI (**lower relevance**)).

H4a: Individuals are less likely to make biased estimates regarding overall firm performance when credit risk gains are presented in OCI rather than in net income.

- When NI becomes negative, the change in the economic condition of the firm is obvious.

H4b: Differences in the likelihood of making biased estimates regarding overall firm performance between the presentation formats are mitigated when a loss is reported.

Results: Evaluation of Overall Firm Performance

Percentage of subjects who incorrectly evaluated overall firm performance as positive:

	NI	OCI	Σ
NO_LOSS	64%	17%	39%
LOSS	13%	9%	11%
Σ	37%	13%	25%

- Categorical model: significant main effect of presentation format supports H4a ($\chi^2 = 9.48$, $p < 0.01$).
- Planned contrast test of interaction supports H4b ($\chi^2 = 17.18$, $p < 0.01$).

H4a: Individuals are less likely to make biased estimates regarding overall firm performance when credit risk gains are presented in OCI rather than in net income.

H4b: Differences in the likelihood of making biased estimates regarding overall firm performance between the presentation formats are mitigated when a loss is reported.

Follow-up Experiment

	Critical points		Follow-up experiment
Data	Behavior in the NI×NO_LOSS cell differs considerably from that in the other cells	Only a few observations	Replication of the NI×NO_LOSS cell (increase number of observations, control for timing / subject pool differences)
H3a	When credit risk gains are presented in NI, NI is always higher in the NI treatments than in the OCI treatments	Absolute value of NI as explanation for higher weight in the NI×NO_LOSS cell in comparison to the OCI×NO_LOSS cell (↔ linkage to NI)	Supplemental condition with NI equal to that of the OCI×NO_LOSS cell (lower than in the NI×NO_LOSS cell), higher operating expenses
H3b	A loss was implemented by assuming an R&D asset impairment	Extraordinary items make NI less informative about (future) firm performance → lower weights in the NI×LOSS than in the NI×NO_LOSS cell (↔ shift in attention)	Replication of the NI×LOSS and NI×NO_LOSS cells, asking subjects to rate the importance of NI and of OCI for performance evaluation
Do subjects apply a sequential search strategy?			Question on search 

Follow-up Experiment

Percentage 1×3 between-subjects design (NI only)

	NI×NO_LOSS	NI×LOSS	Supplemental condition	Total
Participants (auditors)	26	23	23	72
Average total compensation	€ 17.15	€ 17.35	€ 16.86	€ 16.96

Results:

- If the absolute value of NI instead of linkage to NI is the reason that we observed a higher weight attributed to credit risk changes in the NI×NO_LOSS cell than in the OCI×NO_LOSS cell, we should find a lower weight in the supplemental condition than in the replicated NI×NO_LOSS condition (Wilcoxon-Mann-Whitney test, $Z = -0.08$, $p = 0.47$, one-sided).
- If the presence of transitory items instead of the loss itself is the reason that we observed a decrease in the differences in weights between the presentation formats, we should find a lower importance of NI for overall performance evaluation in the NI×LOSS than in the NI×NO_LOSS condition (replicated) (Wilcoxon-Mann-Whitney test, $Z = 0.53$, $p = 0.30$, one-sided).
- Answers on the question regarding the search strategy applied indicate that our auditor subjects indeed use a sequential search strategy.

Conclusion

Key findings

- Presenting changes in own credit risk in OCI has a **positive effect on acquisition** of information and firm performance evaluation (particularly when no loss is reported).
- **Risk of misinterpretation is unaffected** by presentation format; subjects scoring higher on the CRT are less likely to misinterpret credit risk changes.
- **Firm profitability is an important moderating factor**.

Highlights

- Running an **incentivized experiment** with **professionals**.
- Investigating **both presentation formats (ex ante)**: Relevant information for standard-setting.
- Investigating the **different phases of information processing**.
- Controlling for the effects of the subjects' **individual characteristics**.
- Testing the influence of a firm's **economic condition** on information processing.
- Disentangling the **presentation format effect** from the **different economic meanings** of NI and OCI items.

Financial Statements of Company XY

NI×NO_LOSS NI×LOSS

Statement of Income	Previous Year	Current Year	Current Year
Revenue	81,790	81,720	81,720
Cost of goods sold	-73,197	-73,587	-73,587
Gross profit	8,593	8,133	8,133
Research and development expenses	-1,200	-1,200	-5,100
Gains from changes in the fair value of long-term debt	-	+1,920	+1,920
Other operating income	+160	+150	+150
Other operating expense	-4,080	-4,980	-4,980
Interest expense	-1,750	-1,800	-1,800
Net income before taxes	1,723	2,223	-1,677
Income taxes	-534	-689	+520
thereof deferred tax expense (+)/income (-)	(-372)	(+223)	(-986)
Net income after taxes	1,189	1,534	-1,157
Statement of Comprehensive Income	Previous Year	Current Year	Current Year
Net income after taxes	1,189	1,534	-1,157
Other comprehensive income			
Items never reclassified subsequently to profit or loss:			
Gains on property, plant, and equipment revaluation	-	+442	+442
Deferred tax expenses for items never reclassified subsequently to profit or loss:	-	-137	-137
Other comprehensive income	0	305	305
Total comprehensive income	1,189	1,839	-852

Financial Statements of Company XY

	OCI×NO_LOSS		OCI×LOSS
	Previous Year	Current Year	Current Year
Statement of Income			
Revenue	81,790	81,720	81,720
Cost of goods sold	-73,197	-73,587	-73,587
Gross profit	8,593	8,133	8,133
Research and development expenses	-1,200	-1,200	-5,100
Other operating income	+160	+150	+150
Other operating expense	-4,080	-4,980	-4,980
Interest expense	-1,750	-1,800	-1,800
Net income before taxes	1,723	303	-3,597
Income taxes	-534	-94	+1,115
thereof deferred tax expense (+)/income (-)	(-372)	(-372)	(-1,581)
Net income after taxes	1,189	209	-2,482
Statement of Comprehensive Income			
Net income after taxes	1,189	209	-2,482
Other comprehensive income			
Items never reclassified subsequently to profit or loss:			
Gains on property, plant, and equipment revaluation	-	+442	+442
Gains from changes in the fair value of long-term debt	-	+1,920	+1,920
Deferred tax expenses for items never reclassified subsequently to profit or loss:	-	-732	-732
Other comprehensive income	0	1,630	1,630
Total comprehensive income	1,189	1,839	-852



Financial Statements of Company XY

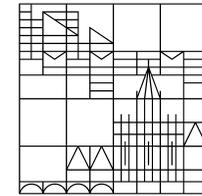
	NI×NO_LOSS		NI×LOSS	Supplemental Condition
Statement of Income	Previous Year	Current Year	Current Year	Current Year
Revenue	81,790	81,720	81,720	81,720
Cost of goods sold	-73,197	-73,587	-73,587	-73,587
Gross profit	8,593	8,133	8,133	8,133
Research and development expenses	-1,200	-1,200	-5,100	-1,200
Gains from changes in the fair value of long-term debt	-	+1,920	+1,920	+1,920
Other operating income	+160	+150	+150	+150
Other operating expense	-4,080	-4,980	-4,980	-6,900
Interest expense	-1,750	-1,800	-1,800	-1,800
Net income before taxes	1,723	2,223	-1,677	303
Income taxes	-534	-689	+520	-94
thereof deferred tax expense (+)/income (-)	(-372)	(+223)	(-986)	(+223)
Net income after taxes	1,189	1,534	-1,157	209
Statement of Comprehensive Income	Previous Year	Current Year	Current Year	Current Year
Net income after taxes	1,189	1,534	-1,157	209
Other comprehensive income				
Items never reclassified subsequently to profit or loss:				
Gains on property, plant, and equipment revaluation	-	+442	+442	+442
Deferred tax expenses for items never reclassified subsequently to profit or loss:	-	-137	-137	-137
Other comprehensive income	0	305	305	305
Total comprehensive income	1,189	1,839	-852	514

Identical to
OCI×
NO_LOSS



Examples

- **Strategic Interaction and Testing Theories (Lab):**
Fischbacher, Urs; Stefani, Ulrike (2007): Strategic Errors and Audit Quality – An Experimental Investigation, in: The Accounting Review 82 (3): 679-704. 
- **Strategic Interaction and Comparison of Institutions (Lab):**
Hattenbach, Thomas; Stefani, Ulrike (2016): Audit Firm Rotation vs. Centralized Matching: Experimental Evidence on the Quality of Audited Financial Statements, Working Paper. 
- **Individual Decision-Making and Comparison of Rules (Paper & Pencil):**
Lachmann, Maik; Stefani, Ulrike; Wöhrmann, Arndt (2015): Fair Value Accounting for Liabilities: Presentation Format of Credit Risk Changes and Individual Information Processing, in: Accounting, Organizations and Society 41 (1): 21-38. 
- **Individual Decision-Making and Comparison of Rules (Mobile Lab & Eye-Tracking):**
Hattenbach, Thomas; Lachmann, Maik; Stefani, Ulrike: Auditors' Information Processing and Going-Concern Opinions: Experimental Evidence on IAS 17 vs. IFRS 16. 



Auditors' Information Processing and Going-Concern Opinions: Experimental Evidence on IAS 17 vs. IFRS 16



Thomas Hattenbach / Maik Lachmann / Ulrike Stefani
(ongoing project)

Motivation

Accounting treatment of lease contracts (lessee)

- IFRS (IAS 17) and US-GAAP follow the risk-and-rewards approach.
- **Finance leases** (“capital leases” in US-GAAP)
 - Both the **leased asset** and the corresponding **liability** are **recognized** in the balance sheet.
 - **Depreciation** of the asset and **interest expense** on the liability are reported in the income statement.
- **Operating leases**
 - **Disclosure** of the current and future lease payments in the notes.
 - **Rental expense** (so-called “straight-line lease expense”) (but no depreciation) is reported in the income statement.
- Lessors prefer off-balance operating leases.

Motivation

Accounting treatment of lease contracts (lessee)

- Joint project of the IASB (2013) and the FASB (2012): Property rights-theory perspective
- On 13 January 2016, the IASB issued IFRS 16 (Leases).
- All leases must be recognized in the lessee's balance sheet.
- Leased assets must be recognized by “capitalizing” the present value of the unavoidable future lease payments. Presentation of the asset in the balance sheet
 - as leased assets (“right-of-use assets”) or
 - together with “property, plant and equipment” (additional disclosure in the notes).
- Financial liabilities must be recognized.
- For all leases, both depreciation and interest expense must be reported in the income statement.
- Entities are required to apply IFRS 16 for annual periods beginning on or after 1 January 2019.

Motivation

Accounting treatment of lease contracts (lessee)

	IAS 17		IFRS 16
	Finance Leases	Operating Leases	All Leases
Assets	 	...	    
Liabilities	€€	..	€€ €€€€
Off-Balance Sheet Rights / Obligations	...	   €€€€	... ↑

Presenting lease arrangements under IAS 17 vs. IFRS 16 (following IASB (2016), p. 4)

Motivation

Effects of the revised accounting treatment of lease contracts on lessee's financial reports

- IFRS 16 will have a significant impact on companies with material off-balance sheet leases (e.g., in the sectors “Airlines”, “Retailers”, and “Travel and Leisure”).
 - Covenants
 - Key ratios derived from the balance sheet.
- The overall impact on profit or loss is expected to be relatively small.

Motivation

Effects of the revised accounting treatment of lease contracts on information processing

- Literature on “recognition vs. disclosure” (Aboody (1996), Davis-Friday et al. (1999), Davis-Friday et al. (2004), Ahmed et al. (2006)), focus on investors:
 - Recognized information is acquired and processed differently from disclosed information (Maines / McDaniel (2000)).
 - Recognizing information leads to an increase in value relevance (Davis-Friday et al. (1999)).
- Experiments on lease accounting, focus on external users (Munter / Radcliffe (1983), Wilkins / Zimmer (1983a, b), Wilkins (1984), Breton / Taffler (1995)): mixed results.
- Studies with a focus on auditors:
 - Auditors consider off-balance sheet liabilities when assessing client risk (Krishnan / Sengupta (2011)).
 - The location of information affects the auditor’s decision-making (Libby et al. (2006), Strand Norman et al. (2011)): Auditors believe that their misstatement-reduction responsibilities are lower for disclosed than for recognized amounts.
 - The presentation format affects auditors’ information processing (Lachmann et al. (2015)).

Research Question

Research question:

- Whether and how does IFRS 16 affect the propensity of auditors to issue a going-concern opinion on a lessee's financial report?
- Situation in which a company is in financial distress:
 - Long-term liabilities, including lease liabilities, exceed equity,
 - but long-term liabilities excluding lease liabilities do not suggest financial distress.
- The auditor's decision of whether to issue a going-concern opinion depends on
 - **Acquisition** (reading and storing) of information on lease contracts
 - **Evaluation** (interpretation) of information on lease contracts
 - **Weighting** of information on lease contracts to evaluate the financial situation
 - **Evaluation** of the financial situation (given lease contracts)

Research question:

- How does an additional signal indicating that the financial situation has worsened affect the auditor's decision to issue a going-concern opinion (signal versus no signal)?

Experimental Design

Design

- 3×2 between-subjects design
- Treatment variables:
 - Presentation format for lease contracts:
 - **Disclosure** of the lease arrangement (i.e., the right to use the leased asset and the obligation to make lease payments in the future) in the notes: DISC;
 - **Recognition** of the leased asset in the balance sheet as the separate item “**right-of-use assets**” (and recognition of the lease liability in the balance sheet): RUA_REC;
 - **Recognition** of the leased asset in the balance sheet under “**property, plant, and equipment**”, that is, together with other items actually owned by the company (and recognition of the lease liability in the balance sheet): PPE_REC.
 - Presence of a signal indicating financial problems:
 - **Presence of a signal indicating financial distress**: SFD;
 - **Absence of a signal indicating financial distress**: NO_SFD.
- Participants: **Auditors**.
- Random assignment of the participants to the six cells of the experiment.
- Computerized decision-experiment including **eye-tracking**.

Experimental Design

DISC: Balance Sheet

Excerpt from the Balance Sheet of Company XY
(all values in thousands of EUR)

Assets		<u>Current Year (t=1)</u>	<u>Previous Year (t=0)</u>
Current Assets			
Cash and cash equivalents		2.212	1.145
Trade and other receivables	(1)	8.172	8.179
Inventories	(2)	10.609	10.300
Available-for-sale financial assets		25.310	25.310
Non-current assets			
Property, plant and equipment		17.140	18.360
Goodwill		10.000	10.000
Research and development projects		9.600	10.800
Total assets		<u>83.043</u>	<u>84.094</u>
Liabilities and equity			
Trade payables	(3)	9.120	8.970
Long-term debt		24.100	24.100
Deferred tax liabilities		3.519	3.348
Total equity		46.304	47.676
Total liabilities and equity		<u>83.043</u>	<u>84.094</u>

Previous Screen

Notes

Next Screen

Experimental Design

DISC: Notes

Excerpt from the Notes of Company XY
(all values in thousands of EUR)

(1) Receivables: On Dec. 31, $t = 1$, trade receivables and other receivables included receivables from the sale of goods of EUR 7.942 ($t = 0$: EUR 7.949) and receivables from services of EUR 230 ($t = 0$: EUR 230). The average period of customer payment is 37 ($t = 0$: 37) days. Upon payment within 30 days of invoice, no interest is charged. Thereafter, 6% is applied. All receivables are faultless.

(2) Inventories: Inventories contain the following positions (all values in thousands of EUR):

	Current year ($t=1$)	Previous year ($t=0$)
Raw materials and manufacturing supplies	7.426	7.210
Work-in-progress	1.061	1.030
Finished goods, parts and products held for resale	2.122	2.060
Total	10.609	10.300

(3) Trade payables: Trade payables are financial obligations resulting from transactions with suppliers. The average period of payment is 45 ($t = 0$: 45) days.

Leases: At the end of $t = 1$, Company XY has leased a plant rather than purchasing it. The interest rate is 5%. The value of the leased asset is equal to the value of the unavoidable future lease payments for non-cancelable leases.

The net present value (in thousands of EUR) is:

54.681

Position is identical to the corresponding liability recognized
in PPE_REC and RUA_REC

Previous Screen

Next Screen

Experimental Design

RUA_REC: Balance Sheet

Excerpt from the Balance Sheet of Company XY
(all values in thousands of EUR)

		<u>Current Year (t=1)</u>	<u>Previous Year (t=0)</u>
Assets			
Current Assets			
Cash and cash equivalents		2.212	1.145
Trade and other receivables	(1)	8.172	8.179
Inventories	(2)	10.609	10.300
Available-for-sale financial assets		25.310	25.310
Non-current assets			
Property, plant and equipment		17.140	18.360
Goodwill		10.000	10.000
Research and development projects		9.600	10.800
Right-of-use assets	(4)	54.681	0
Total assets		137.724	84.094
Liabilities and equity			
Trade payables	(3)	9.120	8.970
Long-term debt		24.100	24.100
Lease liabilities	(4)	54.681	0
Deferred tax liabilities		3.510	3.348
Total			
Total			

Position is identical to the corresponding liability recognized in PPE_REC or disclosed in DISC

Previous Screen
Notes
Next Screen

Experimental Design

RUA_REC: Notes

Excerpt from the Notes of Company XY
(all values in thousands of EUR)

(1) Receivables: On Dec. 31, $t = 1$, trade receivables and other receivables included receivables from the sale of goods of EUR 7.942 ($t = 0$: EUR 7.949) and receivables from services of EUR 230 ($t = 0$: EUR 230). The average period of customer payment is 37 ($t = 0$: 37) days. Upon payment within 30 days of invoice, no interest is charged. Thereafter, 6% is applied. All receivables are faultless.

(2) Inventories: Inventories contain the following positions (all values in thousands of EUR):

	Current year ($t=1$)	Previous year ($t=0$)
Raw materials and manufacturing supplies	7.426	7.210
Work-in-progress	1.061	1.030
Finished goods, parts and products held for resale	2.122	2.060
Total	10.609	10.300

(3) Trade payables: Trade payables are financial obligations resulting from transactions with suppliers. The average period of payment is 45 ($t = 0$: 45) days.

(4) Leases: At the end of $t = 1$, Company XY has leased a plant rather than purchasing it. The interest rate is 5%. The leased asset, presented in the Balance Sheet as right-of-use asset, and the corresponding lease liability are measured at the present value of the unavoidable future lease payments for non-cancelable leases.

Previous Screen

Next Screen

Experimental Design



PPE_REC: Balance Sheet

Excerpt from the Balance Sheet of Company XY
(all values in thousands of EUR)

Assets		<u>Current Year (t=1)</u>	<u>Previous Year (t=0)</u>
Current Assets			
Cash and cash equivalents		2.212	1.145
Trade and other receivables	(1)	8.172	8.179
Inventories	(2)	10.609	10.300
Available-for-sale financial assets		25.310	25.310
Non-current assets			
Property, plant and equipment		17.140	18.360
Goodwill		10.000	10.000
Research and development projects		9.600	10.800
Right-of-use assets	(4)	54.681	0
Total assets		<u>137.724</u>	<u>84.094</u>
Liabilities and equity			
Trade payables	(3)	9.120	8.970
Long-term debt		24.100	24.100
Lease liabilities	(4)	54.681	0
Deferred tax liabilities		3.519	3.348
Total		<u>137.724</u>	<u>84.094</u>
Total		<u>137.724</u>	<u>84.094</u>

Position is identical to the corresponding liability recognized in RUA_REC or disclosed in DISC

Previous Screen
Notes
Next Screen

Experimental Design



PPE_REC: Notes

Excerpt from the Notes of Company XY
(all values in thousands of EUR)

(1) Receivables: On Dec. 31, $t = 1$, trade receivables and other receivables included receivables from the sale of goods of EUR 7.942 ($t = 0$: EUR 7.949) and receivables from services of EUR 230 ($t = 0$: EUR 230). The average period of customer payment is 37 ($t = 0$: 37) days. Upon payment within 30 days of invoice, no interest is charged. Thereafter, 6% is applied. All receivables are faultless.

(2) Inventories: Inventories contain the following positions (all values in thousands of EUR):

	Current year ($t=1$)	Previous year ($t=0$)
Raw materials and manufacturing supplies	7.426	7.210
Work-in-progress	1.061	1.030
Finished goods, parts and products held for resale	2.122	2.060
Total	10.609	10.300

(3) Trade payables: Trade payables are financial obligations resulting from transactions with suppliers. The average period of payment is 45 ($t = 0$: 45) days.

(4) Leases: At the end of $t = 1$, Company XY has leased a plant rather than purchasing it. The interest rate is 5%. The leased asset, presented in the Balance Sheet under property, plant and equipment, and the corresponding lease liability are measured at the present value of the unavoidable future lease payments for non-cancelable leases.

Previous Screen

Next Screen

Experimental Design

SFD

Additional information about Company XY

The following screens present the financial statements of Company XY for the current year $t = 1$. The financial statements also provide information on the previous year $t = 0$.

Directly after the publication of the financial statements for the current year $t = 1$, Company XY received a significant downgrade from two rating agencies because there was an indication that the company would face severe liquidity problems in the near future.

Please study the financial statements provided thoroughly.

Remember: **Your main task is to evaluate the financial position of Company XY.**

Next Screen

Experimental Design

NO_SFD

Additional information about Company XY

The following screens present the financial statements of Company XY for the current year $t = 1$. The financial statements also provide information on the previous year $t = 0$.

Company XY is rated by two rating agencies. After the publication of the financial statements for the current year, these rating agencies did not adjust their ratings.

Please study the financial statements provided thoroughly.

Remember: **Your main task is to evaluate the financial position of Company XY.**

Next Screen

Hypotheses: Acquisition of Information on Lease Contracts

Prominence:

- The **higher the prominence** of a single piece of information, the **lower the cognitive costs of acquiring** this information.
 - Information is more prominent for recognized than for disclosed information

H1a: Individuals are more likely to acquire the information on the lease liability (the leased asset) when the corresponding information is recognized (RUA_REC and PPE_REC) rather than disclosed (DISC).

- Information is more prominent for separate items than for summarized amounts.

H1b: Individuals are more likely to acquire the information on the leased asset when the corresponding information is presented as a separate item (RUA_REC) rather than subsumed under “property, plant, and equipment” (PPE_REC).

Signals indicating a downgrade are perceived as a “red flag”

- Individuals are more likely to read financial information more diligently to identify the reasons for the downgrade.

H1c: Differences in the likelihood of acquiring information on the lease liability (the leased asset) under disclosed versus recognized information are reduced when there is an additional signal indicating financial distress.

Testing

- Questionnaire; information on leased assets / lease liabilities as AOI's.

Hypotheses: Evaluation of Information on Lease Contracts

Labeling:

- Information about the lease liability is **labeled identically** in the treatments.

H2a: There are no differences in the likelihood that individuals will correctly evaluate previously acquired information on the lease liability (the leased asset) when the corresponding information is recognized (RUA_REC and PPE_REC) rather than disclosed (DISC).

- Labeling the leased asset as „right of use asset“ attracts participants' attention.

H2b: Individuals are more likely to correctly evaluate previously acquired information on the leased asset when the corresponding information is presented as a separate item (RUA_REC) rather than subsumed under “property, plant and equipment” (PPE_REC).

Testing

- Questionnaire; comparison of the search strategies across the treatments (saccades for deriving key ratios).

Hypotheses: Weighting of Information on Lease Contracts

Recognition vs. Disclosure:

- Auditors might believe that there is more **business risk** associated with recognized numbers than with disclosed amounts and that external users regard recognized information as more **relevant**.

H3a: Recognizing information on the lease liability (the leased asset) (RUA_REC and PPE_REC) rather than disclosing it (DISC) increases the weight individuals place on previously acquired information on the lease liability (the leased asset) when deciding whether to issue a going-concern opinion.

H3b: Presenting information on the leased asset as a separate item (RUA_REC) rather than subsuming it under “property, plant and equipment” (PPE_REC) increases the weight individuals place on previously acquired information on the leased asset when deciding whether to issue a going-concern opinion.

H3c: Differences in the weighting of previously acquired information on the lease liability (the leased asset) under disclosed versus recognized information are reduced when there is an additional signal indicating financial distress.

Testing

- Questionnaire; assuming that the length of a fixation is an indication of cognitive processes, eye-tracking data can be additionally used as a reference.

Hypotheses: Decision to Issue a Going-Concern-Opinion

H4a: Individuals are more likely to issue a going-concern opinion when the information on the lease liability (the leased asset) is recognized (RUA_REC and PPE_REC) rather than disclosed (DISC).

H4b: Individuals are more likely to issue a going-concern opinion when the leased asset is presented as a separate item (RUA_REC) rather than subsumed under “property, plant and equipment” (PPE_REC).

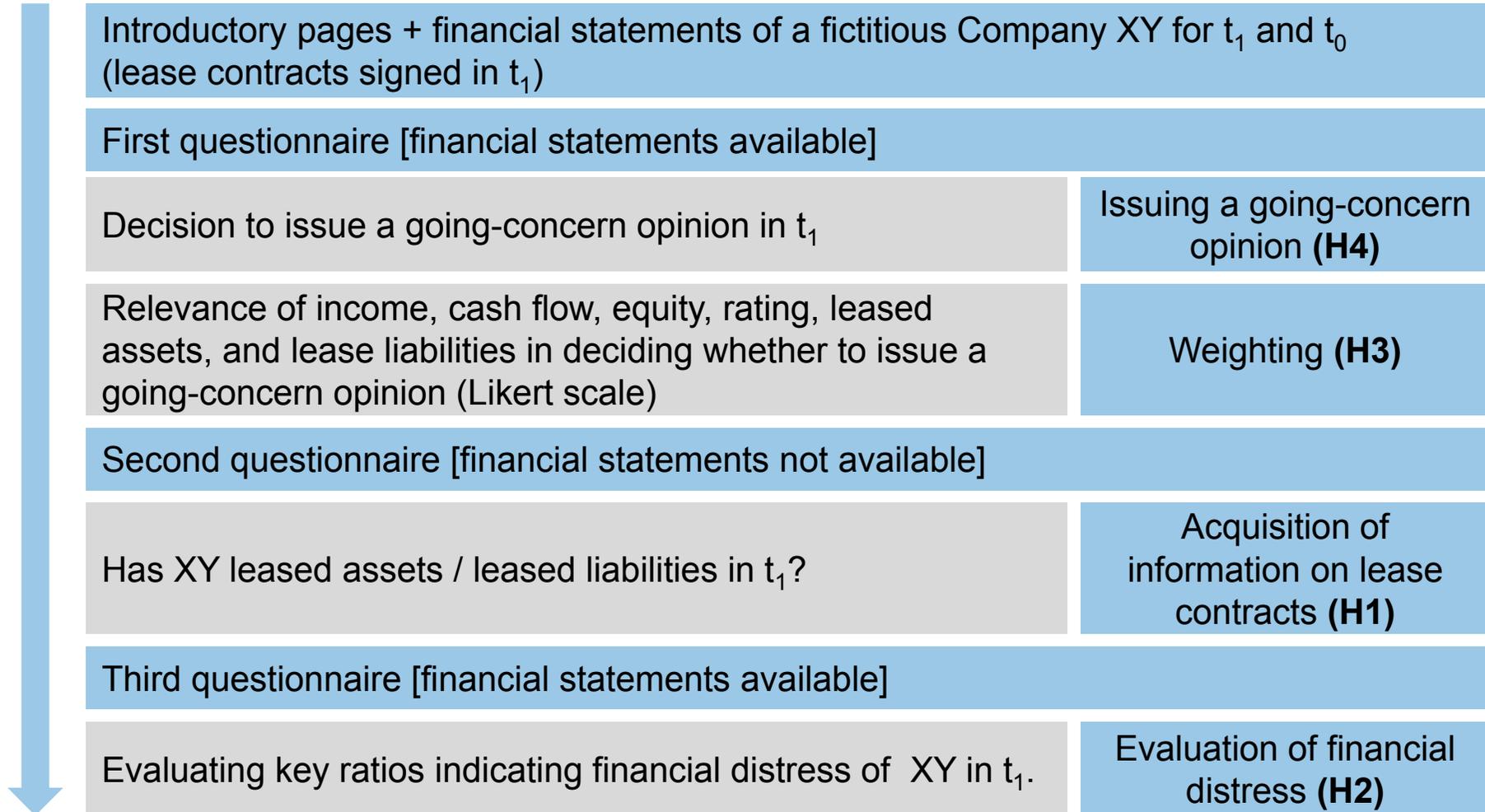
H4c: Differences in the likelihood of issuing a going-concern opinion on disclosed versus recognized information on lease contracts are reduced when there is an additional signal indicating financial distress.

Testing

- Questionnaire.

Experimental Task

Main decision task



Experimental Task

Additional tasks (control experiments), manipulation checks, and demographic data

Cognitive Reflection Test (Frederick (2005))

“A bat and a ball cost € 1.10. The bat costs € 1.00 more than the ball. How much does the ball cost?”

~~10 Cents~~

5 Cents

Disposition to think twice

Risk Lottery (Dohmen et al. (2011)) with 20 situations

	50/50-Lottery	Safe Payment
15	0/300	140
16	0/300	150
17	0/300	160
18	0/300	170

Risk preferences
(risk-averse,
risk-neutral,
risk-loving)

Locus of Control Test (Rotter (1966))

Extent to which individuals believe they can control events affecting them

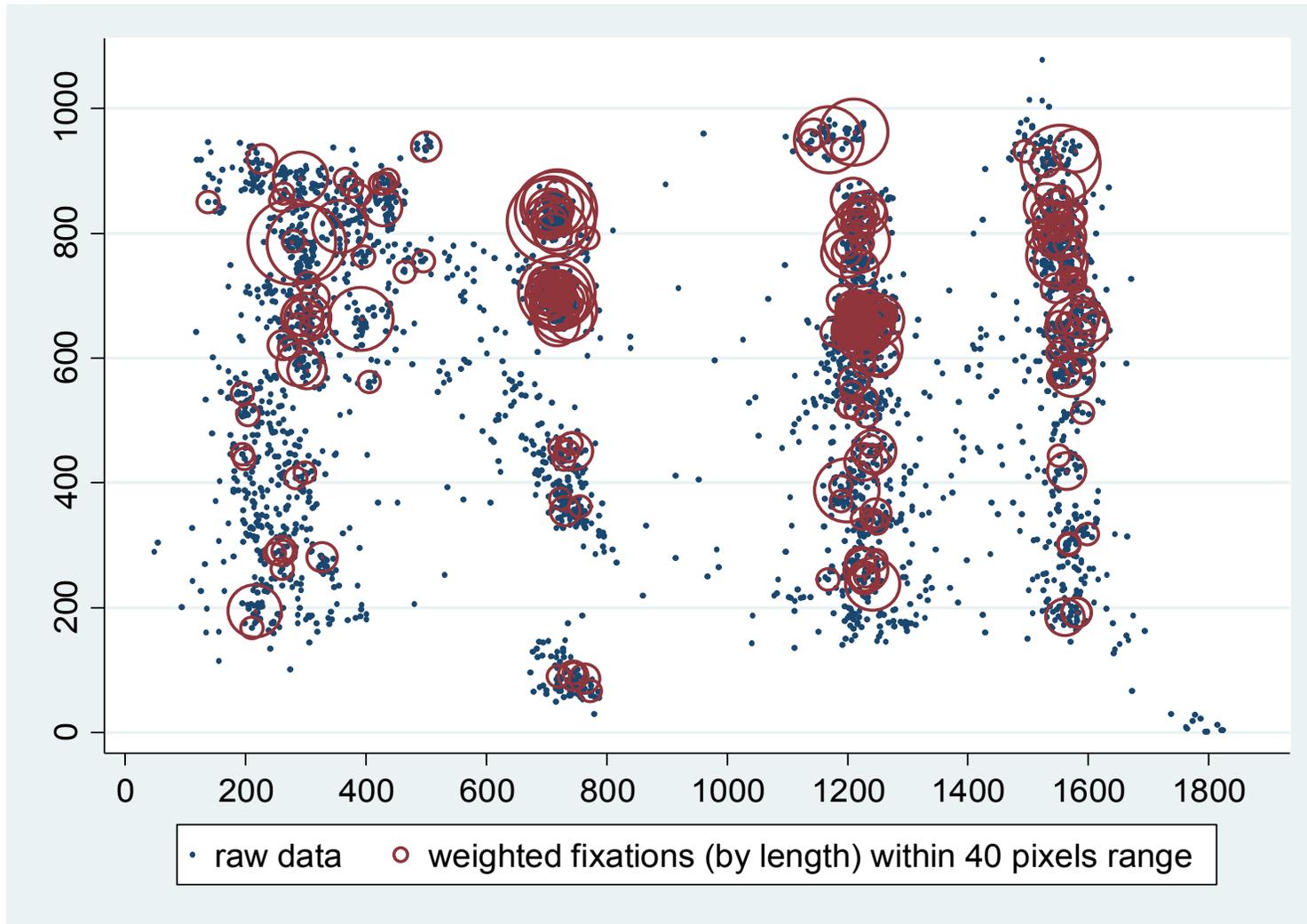
Manipulation checks

Demographic data

Faked Data



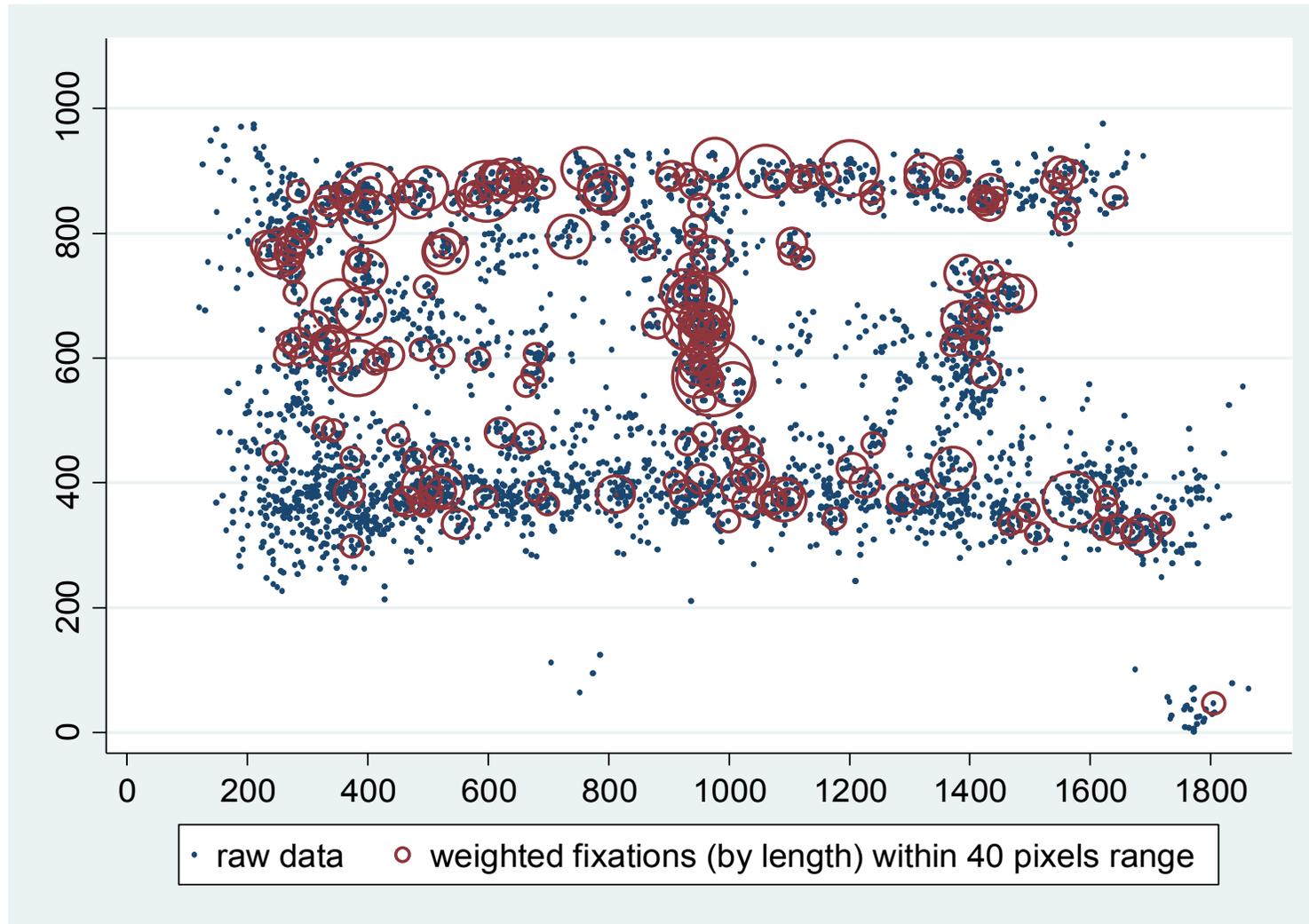
PPE_REC: Balance Sheet

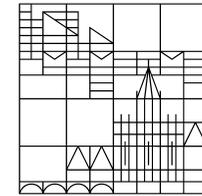


Faked Data



PPE_REC: Notes





**Thank you for your attention ...
... and for your comments!**

Prof. Dr. Ulrike Stefani
Department of Economics

Tel.: +49 (0) 75 31/88 - 5251

Fax: +49 (0) 75 31/88 - 5257

ulrike.stefani@uni-konstanz.de