universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master‘s Thesis

„Economics in the Laboratory. A Case Study on Epistemic Practices and Valuations in Experimental Economics“

verfasst von / submitted by

Helene Sorgner BA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien, 2017 / Vienna 2017

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 906
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Science-Technology-Society
Betreut von / Supervisor:	Ass.-Prof. Mag. Dr. Maximilian Fochler

Acknowledgements

I am greatly indebted to all the VCEE researchers who generously agreed to share their work and their personal thoughts on the methodology and discipline of experimental economics. My special thanks go to Jean-Robert Tyran, who enabled me to do this work by granting me access to the VCEE and its laboratory, and to whom I owe my fascination with experimental economics.

I would also like to thank Pablo Torija and all the lab assistants who instructed me how to run experiments and provided many valuable insights about the daily work at the laboratory. Working at the lab was a truly exciting experience, and I hope that I was not too much of a burden to them after all.

Arthur Schram and Marco Piovesan, while visiting the VCEE, also found time to talk to me about their work and experimental economics in general, for which I am grateful.

My supervisor, Max Fochler, patiently guided me through the ups and downs of this project over the course of more than a year. Our discussions, whether on Skype or in person, were always helpful, supportive, and motivating, and I was very fortunate to find a supervisor who is ready to invest so much time in reading and discussing a master's thesis in all its details.

I would also like to thank all my colleagues at the STS department who commented on this project and helped to improve it while it was still in its early stages.

Bogusz Lewandowski kindly provided detailed comments on a particularly difficult paragraph during the final days, which was more helpful than he is probably aware.

My wonderful colleagues Sophie Ritson and Barbara Grimpe thankfully agreed to do some last minute proofreading and language editing, and I hope that I will soon be able to return the favour.

Finally, I am immensely grateful for all the support I have received during the last year and throughout my studies from my friends and family who had to endure my frequent mood swings and/or mental absence.

Joona Nikinmaa, apart from being the best, helped me enormously not only by making food almost every day, but also by diligently proofreading this thesis and earlier drafts, which improved my writing considerably.

My mother, Barbara Sorgner, has always supported and encouraged me in so many ways, and never once questioned how I used the excellent education she provided for me despite the many challenges of being a single parent. Franz Hochgerner, far from being absent, also supported me throughout my studies. Thank you both for helping me find my way.

I admire economists for their rigour and precision, and I am clearly not one of them. All the remaining mistakes are entirely my own.

TABLE OF CONTENTS

1 Introduction

Economics, with its sophisticated mathematical methods and its wide-ranging institutional influence ranging from providing explicit policy mechanisms and advice to governments, to educating the personnel filling the higher ranks in both public administration and industry, is considered to be the most prestigious of the social sciences. The method of laboratory experiments, identified with testing hypotheses by creating and observing phenomena under meticulously controlled conditions, is often referred to as the hallmark of modern science. Its introduction to the study of nature marks a shift in the history of Western thought and the corresponding conception of reliable knowledge that is to this day referred to as the “scientific revolution”. Until recently, it was assumed that economics could not employ this method, a deficiency that made the knowledge it produces seem somewhat inferior to that of the “exact” natural sciences. From John Stuart Mill to Milton Friedman, all influential theorists of economic methodology were convinced that economics as a science must be largely content to observe (cf. Guala, 2005).

This view was rendered inadequate at the latest in 2002, when Vernon L. Smith and Daniel Kahnemann were awarded the Nobel Memorial Prize in Economic Sciences "for having integrated insights from psychological research into economic science, especially concerning human judgment and decision-making under uncertainty" (Kahnemann’s contribution) and "for having established laboratory experiments as a tool in empirical economic analysis, especially in the study of alternative market mechanisms" (Smith’s contribution).[1] Kahnemann together with the late Amos Tversky is considered one of the forerunners of what is today known as behavioural economics, while Smith is often attributed with having established a methodological foundation for experimental economics. Both research programmes make extensive use of controlled laboratory experiments, in which participants, engage in bargaining and decision-making through computer networks and are afterwards compensated with monetary rewards. The discipline and method of laboratory experimentation in economics is still developing and not without its internal controversies, but its institutionalisation through the establishment of dedicated journals, research groups, laboratories both material and virtual, textbooks, university curricula, and an increasing amount of publications in major journals has made considerable progress during the last three decades.

How laboratory experiments are used in a science that has traditionally been non-experimental is the focus of this thesis. Laboratory research, and more specifically, the practices involved in producing scientific facts, have long been at the centre of analysis in social studies of science and technology (STS). The first laboratory studies conducted during the 1970s marked a move within sociology of

---

[1] Nobel Media AB. (2014) All Prizes in Economic Sciences. *Nobelprize.org.* Retrieved from https://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/ , 30.04.2017.

science from studying representations of scientific work to studying scientific practices in real-time, as embedded in specific research cultures with their distinct social and material environments. These studies laid the groundwork, both methodologically and conceptually, that characterises STS and its approach to the study of science and epistemic communities to this day. Building on this background, it is only a small step of applying the approach of laboratory studies to the recently established practices of experimental economics, albeit a step that so far has not been taken. In fact, the social sciences in general and their various research practices have received very little attention from science studies scholars so far. By interviewing researchers working at the *Vienna Center for Experimental Economics* (VCEE), a research unit established at the University of Vienna in 2011, and by collecting ethnographic observations while assisting in laboratory experiments, my aim was to take this first step towards a more profound engagement from STS with the research practices of contemporary economics.

There are several reasons why a study on experimental economics is a worthwhile project for a master's thesis, beyond the fact that the novel use of laboratories makes economics available for a laboratory ethnography. The high prestige of economics as an academic discipline has suffered during the latest financial crisis, which many commentators attributed to the detachment of academic economists from real-world problems. Indeed, especially financial economics and macroeconomics were accused of faring poorly with regard to real-world developments. Paul Krugman, a notorious critic of his own discipline, went so far as to claim that „much of the past 30 years of macroeconomics was ‚spectacularly useless at best, and positively harmful at worst'" ("What went wrong with economics," 2009). Other commentators from within the discipline identified the narrow specialisation and disregard of wider political and societal questions within economics (Shiller & Shiller, 2011), or the reliance on the predictions of highly idealised models over empirical research in financial economics (Lux & Westerhoff, 2009) as contributing to the crisis.

One of the criticisms levelled against economics models in this later paper is that they only assume a single, perfectly rational agent (representing a household or firm), often also referred to as "homo economicus". These models therefore not only abstract from interactions between agents, they also assume that the lonely representative agent "has unlimited insightfulness and capability of deliberation" and „manages his financial affairs as a side-aspect of his utility maximisation problem, taking into account all potential future happenings with the correct probabilities" (Lux & Westerhoff, 2009: 2). The authors point out that research in behavioural and experimental economics has already shown the limits of human rationality in many ways, and call upon theoreticians to take these insights and the implications of human interaction more seriously. This proposal illustrates the idea of experiments in economics as a „repair shop" for standard theory, and the hope that the behavioural regularities established in laboratory experiments can be used productively as an addition to theoretical models in microeconomics and beyond. The use of laboratory experiments in economics on such accounts is presented as a remedy for the discipline's notorious detachment from empirical reality.

However, the experimental investigation of human behaviour is not without its difficulties either. A large-scale study (Open Science Collaboration, 2015) recently investigated how well the results of laboratory experiments in social psychology could be reproduced in faithful replications. The outcomes were alarming, as only 36 out of the 100 replicated experiments still had statistically significant results (while this was the case for 97 of the original studies) and effect sizes decreased notably. Some of these outcomes can be attributed to a publication bias that favours positive over negative results. Accordingly, the phenomenon that the original effect size decreases in replication studies can be observed in many other experimental sciences as well (cf. Lehrer, 2010). A recent survey reports that experiences of failing to reproduce a colleague's or even one's own experiment are quite common across the sciences (Baker, 2016). Depending on the subject matter, it is actually very likely that "most published results are false", at least according to a widely cited article by John Ioannidis (2005). Using a Bayesian model, Ioannidis shows that the likelihood of finding a true relationship depends on how many true relationships there are among all of those tested in the first place. This means that when many different hypotheses are tested, but only very few of those are actually true, the outcome of the experiment is likely to be a false positive.

Other explanations for the poor reproducibility of experimental psychology research focus on the particular difficulty of the research subject, especially when the participants are infants (Peterson, 2016). In addition to human subjects not being standardisable and the exact research design therefore being hard to replicate in the first place, the research practices of psychology have also been criticised for allowing too many "researcher degrees of freedom" (Simmons, Nelson, & Simonsohn, 2011). This is especially problematic when only those treatments and control variables that yield significant results are reported in the final publication, and when data collection is stopped as soon as significant results are achieved.

Experimental economics, however, has so far circumvented this discussion. Indeed, as an answer to the debate on reproducibility, a group of experimental economists conducted their own replications, concluding that the 16 recent studies investigated show much better replication rates than those reported for psychology (Camerer et al., 2016). There are some caveats to this result, because the sample is much smaller than the psychology sample, and some "wearing off" of the effect size can also be observed here. In general, however, experimental economists seem confident that should there be any problematic aspects of their method (such as the common use of the "convenience sample" of undergraduate students or the question of how well results from laboratory experiments generalise to other settings), this is a motivation for doing more, rather than less, experimental research (Falk & Heckman, 2009).

What these examples of methodological discussion exemplify is one of the main insights of laboratory studies in STS, namely that scientific practices are diverse, that methods and the approach to the empirical are specific to individual research cultures, and that the production of a scientific fact is always a local accomplishment embedded in a particular material, social and institutional context. My question will therefore not be *why* economics experiments work, but *how*. In the tradition of laboratory studies in STS, I will look at the practices involved in designing and conducting an experiment, and into the strategies used to make experiments successful. The underlying assumption is that the *epistemic practices* of different disciplines vary according to their subject matter, their social organisation, their institutionalisation and other factors: experimentation in high energy physics is different from experimentation in molecular biology, and it is to be expected that experimentation in economics presents yet another kind. The question is then**, how have economists appropriated the method of laboratory experiments** to suit their questions, theories and disciplinary criteria? In other words, my first focus in this study will be on the *epistemic culture* of experimental economists, observed through the lens of experimental practices. Here I will aim to describe the practices and strategies that I learned about when interviewing researchers and assisting in laboratory experiments, and carve out the main characteristics of the approaches I have discussed and observed.

However, while there seems to be a very stable idea of what the "textbook approach" in experimental economics is, I have found that there are also views on experimentation that dissent from this approach. It seemed to me that beyond the common practices of making laboratory experiments in economics work, there are different ideas about the purpose of this method, which correspond to different accounts in the secondary literature on experimental economics. My second question is therefore, **how do experimental economists evaluate the method of laboratory experiments**? This question signifies a shift in focus to a more person-centred approach, analysing individual accounts of experimentation and the evaluative principles that are enacted in these accounts. This focus builds on a growing literature in STS that deals with valuation and evaluation practices in science, both on an institutionalised and on a personal level, and more concretely with the question of where institutional and personal accounts of the value and worth of scientific work come apart. My research therefore not only aims to describe the epistemic culture of experimental economists, but also how experimental economists locate themselves within this culture and the wider landscape of academic knowledge-production in general.

Asking what is valuable about experimental economics results is not a purely academic question. "Whispering in the ears of princes", that is, giving concrete policy advice, is after all one of the three "promises" of experimental economics that Alvin Roth lists in the introduction to a recent methodology handbook (Roth, 2015). Prominent whisperers include Cass Sunstein and Richard Thaler (2008), advisors to US president Barack Obama and bestseller authors, who base their policy programme of "nudging" on the insights of behavioural economics. On a different scene, the promise and success of experiments in economics also seems to carry over to the discipline of political science, where experiments are seen by some as the method of most cutting-edge research.[2] Studying how economists conduct and evaluate experiments therefore gives us insight into the expectations surrounding knowledge production in the social sciences today, and how the import and appropriation of epistemic practices from other research traditions serves to fulfil these expectations.

In order to answer my research questions and embed my analysis in the existing scholarship on research cultures, the conditions for academic knowledge production and experimental economics in particular, the text will proceed in the following way. First, I will review the state of the art in the STS literature informing my research. In a next step, a number of theoretical concepts that are particularly prominent in my analysis are discussed in greater detail. I will then describe my case study, including a short overview of the methodology of experimental economics and a brief introduction to the location of my fieldwork, the Vienna Center for Experimental Economics. In the same chapter, I will also describe my research and analysis methods and reflect on how my study developed, how I negotiated access to the field, what role I took on as a researcher, and how this influenced my observations. The second part of the thesis is comprised of the analysis, which proceeds in two parts: the first part focuses on epistemic practices in developing and running laboratory experiments, and the second part focuses on the discursive practices of valuation I encountered during my interviews. In the conclusions, I take up some themes from the state of the art and relate them to my analytical observations and the insights that can be drawn on the purpose and outlook of economics as a laboratory science.

---

2 Consider that at least two handbooks on the topic have appeared in recent years:
Kittel, B. (Ed.). (2012). *Experimental political science: principles and practices*. Basingstoke, Hampshire: Palgrave Macmillan.
Druckman, J. N., Green, D. P., Kuklinski, J. H., & Lupia, A. (Eds.). (2011*). Cambridge handbook of experimental political science*. Cambridge: Cambridge Univ. Press.

2 State of the Art

This chapter describes the current body of research that informs my project in terms of the methodological approach chosen, the epistemic interest, and the analytical conclusions I make. The first part is dedicated to a tradition of research in STS referred to as "laboratory studies", which has inspired and informed my approach to studying experimental economics. The second part describes more recent studies on the effects that an increasing demand for the accountability of research and its measurement has on the research practices of communities and individuals. The third part engages with the state of the art in research on experimental economics, and extensively discusses the three major contributions on the methodology, epistemology and history of this discipline. I will end this chapter with a brief suggestion of the contributions my research makes to this body of literature.

2.1 Studying Laboratory Practices

Much of the conceptual, theoretical and methodological content of what today forms the discipline of *Science and Technology Studies* (STS) has its origin in a research programme that developed during the 1970s and was later subsumed under the name of *laboratory studies*. At this time, several researchers independently of each other began to investigate the practices involved in producing scientific knowledge in their original context, scientific laboratories (Lynch, 1985; Latour & Woolgar, 1986; Knorr Cetina, 1981; Traweek, 1992; Collins, 1992). Their novel approach to studying science signified a "move towards studying scientific practice, what scientists actually do, and the associated move towards scientific culture, meaning the field of resources that practice operates in and on." (Pickering, 1992: 2). On a conceptual level, laboratory studies were a step towards dissolving the boundary between what was perceived as the technical and social content of science. A common insight of all these studies was that in the daily practices of researchers, the social and the technical cannot be clearly distinguished. More specifically, they showed that this distinction is the result of a gradual de-contextualisation and elimination of qualifications and "modalities" from statements and claims produced in the laboratory as they leave their immediate context of origin (Latour & Woolgar, 1986, chapter 2) and are presented in the end product of research, a scientific paper (Knorr Cetina, 1981, chapter 5).

Laboratory studies also marked a methodological shift in science studies. Before the first laboratory studies, sociologists of science had typically conducted historical case studies and relied on scientists' own accounts of their practice. Adopting the methods of ethnography and participant observation from cultural anthropology for the study of science meant that knowledge production could be studied in real-time to find out "how it is that the realities of scientific practice become transformed into statements about how science has been done" (Latour & Woolgar, 1986: 29). To answer this question, "the prolonged immersion of an outside observer in the daily activities of scientists" (ibid.) was considered an appropriate method. This immersion, however also required a reflexive attitude on the part of the observers. On the one hand, they needed to distance themselves from taken-for-granted beliefs and shared understandings about science in order to approach and study researchers as practitioners of a specific, alien *culture* (akin to the foreign "tribes" studied in classical anthropology). On the other hand, this approach also required the observers to consciously reflect on their own methods of producing facts and claims, since they are not so different from those of the scientists they observe. An early proponent of the approach therefore pointed out that ethnographies of science should not be understood as providing an image of science "as it really is", since that would amount to claiming for the account of the anthropologist the same objectivist stance that laboratory studies seek to problematize about natural science (Woolgar, 1988).

The methodological approach of laboratory studies is mainly characterised by participant observation in the community to be studied, which can last for several years. In the course of their research, observers often take on minor tasks in the laboratories, which allows them to acquire routine and familiarity with the environment. Observations are recorded in detailed field notes. Along with documents, photographs, interviews and sometimes also video recordings, they form the empirical material that is gathered in the field and subsequently analysed (cf. Woolgar, 1988). Apart from these methodological communalities and a general attention to the constructive nature of knowledge-making, there was and is however a considerable diversity within laboratory studies. How exactly the observing, recording, transcribing and analysing of the empirical material proceeds depends on the disciplinary background and epistemic interests of individual researchers. For example, Knorr Cetina's early work (1981) was influenced by theories of social and radical constructivism foregrounding an interest in epistemological questions, whereas Lynch's (1985) laboratory study presents the application of the ethnomethodological approach in sociology, from its detailed analysis of mundane practices to scientific work. Knorr Cetina in her review article (1995) states that the introduction of the method of ethnography to studies of science and technology is one of the lasting influences that laboratory studies have had on the field. As Latour and Woolgar note in the post-script to the second edition of *Laboratory Life* (1986), there are however widely differing understandings of what an "ethnography" should be like, and according to some, their own account fell short of these criteria. This becomes obvious when comparing *Laboratory Life* to the work of anthropologist Sharon Traweek (1992). Around the same time as Latour conducted his participant observation, Traweek, who as a graduate student in anthropology had worked as a tour guide at the facilities of the *Stanford Linear Accelerator*, embarked on a several year-long ethnographic study of high-energy physics communities in the US and Japan. The result is a thick description of high-energy physicists' work environments and machines, career trajectories, norms of interaction, hierarchies within and between communities, and the cultural differences between Japan and the US. Traweek describes her approach as that of classical anthropological fieldwork, the outcome of which should give an account of four domains of a community's life:

> The first is ecology: the group's means of subsistence, the environment that supports it, the tools and other artifacts used in getting a living from the environment. The second is social organization: how the group structures itself, formally and informally, in order to do work, to form factions, to maintain and resolve conflicts, and to exchange goods and information. The third is the developmental cycle: how the group transmits to novices the skills, values, and knowledge that constitute a sensible, competent person; the stages of a life and the characteristic attributes of a person at each of those stages. The fourth is cosmology: the group's system of knowledge, skills, and beliefs, what is valued and what is denigrated. (Traweek, 1992: 7).

In other words, Traweek's aim was to reconstruct the commonly shared worldview of the high-energy physics community, how it is transmitted and sustained, and how this worldview is reflected in and shapes their social reality. For Traweek, physicists' theories of space and time – describing both the universal, timeless laws of nature that they aim to discover and the short and precarious spans of lifetime each of them can contribute to this task – do indeed illuminate their social practices. High-energy physics, on her account, is a culture of striving for "extreme objectivity". The products of material practice – detectors – are understood simultaneously as the incorporation of the history and skills of a specific research group and as an instrument to read the eternal book of nature without the interference of human agency.

Traweek's study stands out as a comprehensive anthropological account of a research community, since the other laboratory studies conducted at the same time had a more explicit interest in "the way in which the daily activities of working scientists lead to the construction of facts" (Latour & Woolgar, 1986: 40). As Doing (2008) observes, laboratory studies in general may be divided into those that focus on the specific processes and practices involved in producing knowledge claims, and those that more comprehensively describe the social organisation and material culture of research communities. Knorr Cetina's later work in this sense marks a shift towards a more global analysis of research practices. With her seminal work *Epistemic Cultures* (1999), Knorr Cetina moved from the study of knowledge construction practices in a single laboratory to the comparative study of the construction of "machineries of knowing" in different fields. By comparing the "practices of creating and warranting knowledge" (Knorr Cetina, 1999: 246) in high-energy physics at CERN and in molecular biology in several German research laboratories, she highlights the "disunity" of the sciences rather than their continuity. Knorr Cetina therefore also framed this study as an investigation into the diversity of the communities that make up contemporary "knowledge societies". On this account, molecular biology and high-energy physics, although both at the frontier of research, are characterised by markedly different approaches to the empirical, which correspond to different solutions to social organisation. In high-energy physics, a preoccupation with the vast apparatus, its functioning and its limits corresponds to the research collaboration as the unit and main agent of research. In molecular biology, the importance of embodied skills in the direct engagement with the research material in contrast brings individual researchers and their accomplishments much more to the fore. Knorr Cetina's overall argument for moving towards studying the "machineries" of knowledge production is that a comparative analysis can show "that the meaning of the empirical changes as these machines are implemented in different fields" (Knorr Cetina, 1995: 158). This observation underwrites the related argument that the construction of a scientific fact is always a local accomplishment; the power of laboratories resides in their ability to produce local reconfigurations of natural and social orders (more on this in section 2.2). Knorr Cetina takes this as contrasting the traditional image of sciences as being based on a unified methodology and points out that consequently, models of scientific practice "that generalize across all fields and disciplines are premature at best" (ibid. 159).

If the categories used to describe research practices and the processes involved in constructing knowledge do not generally apply to all epistemic cultures, as Knorr Cetina suggests, this raises interesting questions for the study of disciplines that were modelled on existing types of research, such as experimental economics. Experimental economics has appropriated the method of laboratory experiments in an effort to emulate sciences that are characterised by a stronger integration of theorising and empirical data than economics has traditionally been. If Knorr Cetina's argument about the diversity of local practices and of the machineries of knowledge production that make up *epistemic cultures* is correct, then it is to be assumed that the practices that are involved in laboratory experimentation in economics are unique to that field. In other words, it is to be expected that what experimentation in economics *means* in practice cannot be understood by extrapolating from experimental and laboratory practices in other sciences. Rather, the existing epistemic culture of the academic discipline of economics into which the experimental method was introduced and gradually embedded might have shaped experimental practices just as much as the new method is said to have changed economics.

Another reason to pay attention to the appropriation of the experimental method by a science that was not traditionally experimental is that laboratory research is associated with a high degree of control and reliability, an expectation that actual research practices cannot always fulfil. In a very recent contribution, Peterson (2016) describes the intricacies of developmental psychology experiments where the research subjects are toddlers and infants, and the strategies that psychologists use to achieve statistically significant results despite the inherent difficulty of their research object. This study is particularly interesting considering recent debates on the (lack of) reproducibility of results from experimental psychology (Open Science Collaboration, 2015). The failure to reproduce most of the effects reported in published psychological studies calls into question what was often considered one of the hallmarks of reliable science, namely the replicability, and therefore, the context-independence of its results. This debate – and Peterson's study - also underwrites an argument of the classic laboratory studies outlined above, that the production of a knowledge claim is always embedded in a specific context (Knorr Cetina 1981, 1995), and that the success of replication even of a very established fact can therefore not be taken for granted (cf. Collins, 1992). Another important insight to be gained from these studies is that replication (in the sense of doing exactly the same experiment, i.e. the same design under similar conditions, again) is actually very rarely done in science and that the value of the experimental method must therefore depend on other aspects besides the purported replicability of its results.

When considering how the field of laboratory studies developed, it seems that the "cultural approach" to the study of science (Knorr Cetina, 1995) has been used predominantly to map out the diversity of research practices in contemporary science. Particularly in the work of graduate students, laboratory ethnographies were conducted in a variety of new research contexts, while also mobilising new bodies of theoretical work to interpret the observed practices. Cyrus Mody (2001), for example, studied how the notions of purity and contamination are constructed and mobilised in material science research during an ethnographic study of the use of microscopes in a university research laboratory. Morana Alac's (2008) research on how the interpretation of digital images of brain scans takes on a form of embodied action, involving gestures and a sense of practical manipulation, builds on Michael Lynch's ethnomethodological laboratory studies. Natasha Myers (2008) in her ethnographic account of how crystallographers develop and make us of "embodied models" in their search of the structure of protein molecules likewise brings the scientists' body and tacit sensual skills back into the focus of analysis. All of these studies thus are concerned with describing and reconstructing epistemic practices as such. This can be interpreted as a sign for the ongoing interest in the diversity of the local accomplishments that make up the practice of science, particularly where these depart from the prevailing image of science as cognitive, rational and theory-driven work.

The study of how these different epistemic practices of research cultures correspond to different types of social organisation that can be understood as the second branch of laboratory studies has not been continued to the same extent. The reason for this seems to be a general move in science studies towards studying the (external) conditions for doing research and how those are implicated in local research practices. Although this type of research, insofar as it uses participant observation, could be understood as exemplifying the epistemic cultures approach, in contrast to Knorr Cetina's work it instigates a move back towards individual researchers, their experience of disciplinary environments, and the considerations that shape their research practices as the focus of analysis. On a conceptual level, this shift could be referred to as that from the study of epistemic cultures to the study of "epistemic living spaces" (Felt, 2009). In the context of the general development of STS as a discipline, it reflects an interest in the changing governance and conditions of research that has been growing steadily since the mid-1990s, calling for analyses of "how societal imaginaries and policy framings as well as changes in research assessment and monitoring exercises are experienced by individual researchers and how they affect their practices." (Fochler, Felt, & Müller, 2016: 181). The literature concerned with these developments and how they affect research practices in contemporary sciences is the topic of the following section.

2.2 Research Evaluation and its Effects on Research Practices

After *Epistemic Cultures* (1999) the focus of what could still be referred to as "laboratory studies" shifted away from research practices as such to how different communities and individual researchers adapt to the institutional transformations marking a new governance of science during the last three decades (e.g. Felt, 2009).[3] Developments in science policy and governance emphasising a need for accountability, transparency and efficiency are often described as being part of a broader trend towards the institutionalisation of "audit cultures" (Power, 1997, 2008; Strathern, 2000) in public and professional life. Transformations in the conditions for academic research are attributed particularly to the implementation of research evaluation systems on a national and institutional level (Whitley, 2007) and the increasing importance of measures such as public rankings of institutions (Espeland & Sauder, 2007) and citation-based impact factors of journals and publications (Rijcke, Wouters, Rushforth, Franssen, & Hammarfelt, 2016). While evaluation systems assessing the quality of a research proposal or the use of public funding are typically introduced by public administrators and funding bodies, metrics such as university rankings and citation counts are provided by private companies (newspapers and scientific publishing houses) mainly for commercial reasons. Whether their explicit purpose is to enable financial "auditing" and quality control or to provide information to the interested public, measures and indicators increasingly shape valuation processes within scientific practice. Taken together, these measures and evaluation criteria result in a complex "metric assemblage" (Burrows, 2012) to which individual researcher find themselves and their research inevitably subjected.

Using the example of the magazine U.S. News' rankings of US-American law schools (first introduced in 1987), Espeland and Sauder (2007) show how public measures affect the objects measured and instigate them to comply with the measurement criteria. The authors argue that this "reactivity" of public measures is the result of two different mechanisms. The mechanism of *self-fulfilling prophecies,* for example, operates on the level of external audiences, who will react to a "bad" ranking of an institution by withdrawing support (either by not applying to be admitted to this school or by ceasing to fund it), which then results in a decrease in quality of education and reputation. Likewise, earlier rankings may affect the reputation of lesser known schools, funding may be allocated on the basis of rankings, and schools in general may aspire to conform more to the criteria of the rankings, which then confirms the validity of the measures. The mechanism of *commensuration* operates on the cognitive level by directing attention away from the manifold qualitative differences between entities towards those differences that can be represented quantitatively on the scale of a common metric. The hierarchies that are established by comparing and ranking schools according to a small set of predefined criteria often depend on very small differences that do not reflect actual quality differences between schools, and do not take regional differences and specific missions (such as supporting minorities) into account. Nevertheless, "keeping up the numbers" is seen as necessary due to the im-

[3] This shift in the epistemic interest of laboratory studies is also visible in the latest edition of the *Handbook of Science and Technology Studies* (Felt, Fouché, Miller, & Smith-Doerr, 2017). Whereas the two former editions featured designated entries discussing the scope and purpose of laboratory studies as such (Knorr Cetina, 1995; Doing, 2008), the sections discussing studies on laboratory science in the latest handbook do so from the perspective of changes in the organisation, governance and regulation of scientific research.

portance that these rankings have acquired for prospective applicants and funding bodies. Espeland and Sauder report that as a consequence, school administrators have shifted priorities in admissions, career advice and budget spending within their schools to better conform with the USN rankings criteria, and often resort to "gaming strategies" when reporting information, which ironically have led to increasing distrust between administrators and a perceived loss of transparency.

Another metric that has achieved overarching importance in academic life is the journal impact factor (JIF). This number, first proposed by Eugene Garfield in 1955 (Garfield, 2006) and calculated by the company Thomson Reuters based on the publications listed in its *Web of Science*-database, represents a measure of the average number of citations that articles published in a scientific journal receive within a given timespan.[4] The impact factor itself and its importance for the evaluation of research vary dramatically across scientific disciplines. Particularly in the life sciences, this metric has taken on a life of its own and figures prominently in researcher's epistemic and authorship practices (Rushforth & Rijcke, 2015). The prominence of the journal impact factor shows that beyond the general imperative to "publish or perish", researchers also need to take into account *where* they publish. Senior figures in the life sciences have warned that their field's pre-occupation with publishing in very few high-ranking journals negatively affects the quality of publications, because even a short letter in *Science* or *Nature* is seen as counting more than a substantial publication in a less prominent journal (Lawrence, 2003). Also, a preoccupation with citation counts leads both scientists and editors to go for „flashy" rather than sound research (Schekman, 2013). A common critique is that since the impact factor denotes an average, it is by no means indicative of the quality (or, for that matter, the number of citations) of any individual article and therefore should not be used as a measure for the quality of an individual's research (DORA, 2013). In contrast to this negative view of the significance of impact factors and citation counts, Dahler-Larsen (2014) has proposed to investigate the meaning that these indicators acquire through their use in scientific practice. Rushforth and de Rijcke (2015), following this suggestion, treat the widespread accounts about the use and influence of the JIF as scientists' "folk theories" and investigate how these ideas come to matter in biomedical researchers' day to day practices. They find that the journal impact factor is used as a device for assessing ongoing work and deciding whether a project is worth investing more effort in, given its likelihood to be published in a high-impact journal. The impact factor is seen as correlating with the novelty of research typically published in a given journal, and in this sense is used to estimate the likelihood of having a paper accepted or rejected by this journal. Also, it is generally assumed that journals with higher impact factors (i.e. journals that on average receive more citations) will be read more widely and therefore cited more often, another instance of the "reactivity" of public measures. Researchers who instead of publishing in high-impact general interest journals typically write for small audiences of specialists in

[4] To be precise: "A journal's impact factor is based on 2 elements: the numerator, which is the number of citations in the current year to items published in the previous 2 years, and the denominator, which is the number of substantive articles and reviews published in the same 2 years." (Garfield, 2006: 90).

turn find it difficult to account for the value of their work when confronted with funding bodies who evaluate researchers based on the ranking of their publications.

Arguably, the increased importance of publishing in highly ranked scientific journals (instead of other types of publications) has also affected how junior researchers are building careers. Müller (2012) in her study on researchers in the Austrian life sciences describes how post-doc researchers mediate the conflicting imperatives of engaging in collaborative work and being competitive as an individual. Since first-authored publications are considered the main currency on the academic job market in these disciplines, post-docs prefer individualised projects to collaborative work and need to weigh accepting or giving support to others against the need of having to share authorship, which is seen as decreasing the value of a publication. Based on the same interview material, Fochler, Felt & Müller (2016) argue that post-doc researchers in general seem to evaluate their own work and their research environment primarily in terms of whether it will allow them to publish productively and succeed on the academic job market.

In economics, similar trends are observable. The need to publish productively at an earlier stage, for example, is visible in the trend towards PhD theses that consist of several articles rather than a monograph (Stock & Siegfried, 2013). The average number of publications in high-ranking journals has significantly increased for economists being awarded tenure in the German-speaking area (Graber, Launov, & Wälde, 2008). While the JIF as a measure and as an indicator of research quality plays a lesser role in economics than in other sciences, public journal rankings devised specifically for economics take on this function, and they can be shown to influence a journal's reputation among economists (Haucap & Muck, 2015). Outspoken critics of these developments, in economics (Frey, 2003) and elsewhere (Burrows, 2012; Lawrence, 2003), condemn the powerlessness researchers experience when confronted with the need to publish in specific journals and to conform to a pre-defined set of external evaluative criteria that is constantly expanding. Not least because of the worry that the "publish or perish"-imperative will eventually harm the quality of research, alternative suggestions for evaluating research than the use of quantitative metrics have recently been put forward (DORA, 2013; Hicks, Wouters, Waltman, de Rijcke, & Rafols, 2015).

These developments also raise the more fundamental question of the value of scientific work and how it can be adequately assessed when taking into account the plurality and heterogeneity of institutional and cultural contexts for contemporary knowledge production. The emerging field of „valuation studies" (Lamont, 2012; Kjellberg & Mallard, 2013; Dussauge, Helgesson, & Lee, 2015) here signifies a reorientation towards these fundamental questions and investigates practices of ascribing and assessing worth as the processes in which values are enacted, negotiated and stabilised. In particular, these studies are interested in how different forms of value – epistemic values and economic values, for example – are differentiated and stabilised in practices of valuation, and what principles of ascribing worth are referred to in evaluative processes. An aspect often criticised about the introduction of evaluation systems and forms of governance in research, academia and higher education is that they, in an effort to make heterogeneous objects commensurable, reduce a plurality of values to the uniform logic of cost-benefit-analysis. Perhaps due to the advent of economic and managerial thinking in the realm of academia, economics as a discipline and its pervasive influence on contemporary life has also become the focus of a dedicated research programme in STS.

2.3 Studying Experimental Economics

Economics has come into the focus of sociological research in the tradition of science studies as a provider of theories, models and descriptions of the social world that are performative in nature and therefore shape the objects and relations they describe. The research programme investigating the "performativity" of economics was launched by Michel Callon (1998) in the introduction to a collection on the anthropology of markets. His controversial thesis is that "economics, in the broad sense of the term, performs, shapes and formats the economy, rather than observing how it functions." (Callon, 1998: 2). There is no such thing as the "economy" that exists independently of the activities of economists and that could be described in a correct or incorrect way. Rather, the "economy" is performed through the numerous ways in which economics as a discipline, but also economics in the broader sense of knowledge and technologies used in markets, relates to and acts upon its objects: "by observing them, by measuring them, by predicting them, by providing theories to explain them or instruments to regulate them, by spreading some functional technique about them (or just some suggestive vocabulary to deal with them), by designing them in a laboratory, by inventing them, and so on." (MacKenzie, Muniesa, & Siu, 2007: 6) What economics does, is to constitute economic things.

Laboratory experiments at a first glance look like a trivial example of "performed economies", and so do the policy interventions that were initially commissioned from and designed by academic economists (Guala, 2007). However, the performativity of laboratory experiments appears to be less trivial when its parallels with other activities that bring about economic things are considered. Muniesa (2014) describes, for example, the use of focus groups in marketing research as a "provocation" that brings about consumers and their preferences as a performative achievement in a highly artificial environment. On the outset, focus group testing has nothing to do with how consumers in "natural" markets buy products, but consumer testing does not serve as a proxy or a representation for a "real market" . Rather, it is a first step towards constituting a consumer market, since the preferences elicited here will, through a series of translations, be turned into products and marketing strategies. The description of financial products, the setting (or "discovery") of stock prices, the instruction of business school students, and of course, the introduction of performance indicators, rankings and evaluation systems already discussed above according to Muniesa are further examples of practices which provoke and constitute economic things. The difference to (laboratory) experiments is then that the latter quite explicitly aim to provoke what they refer to: "Economic experiments perform economic objects, in quite a general sense. What experimenters describe is indeed produced by them in the experimental setting. They account for what they provoke." (Muniesa & Callon, 2007: 163). In fact, Muniesa's description of economic things as provoked in the activities that constitute the economic owes much to the constructivist critique of science developed in laboratory studies, particularly to Pickering's (1995) suggestion to complement the "representational idiom" of science with a "performative idiom". The performative idiom is more in line with what was often observed in studies of scientific practice, that "objects turn objective because they are provoked (literally, created) in the laboratory." (Muniesa, 2014: 10-11). If, as Muniesa suggests, what is considered "economic" in general is constituted through active provocation, then laboratory experiments in economics only take advantage of the performed nature of markets, consumers, their preferences and their strategic behaviour, which allows these economic things to be created and elicited at will.

As Muniesa notes, the thesis of the performativity of economics is at odds with an idiom of naturalism that is characteristic of modern sciences and also ascribed to economics, where it amounts to the idea that "we may all have different cultures, opinions and beliefs, but we all share the same economic laws." (Muniesa, 2014: 37). He suggests that this idea should be abandoned; economists seem to use a naturalistic idiom on the surface, while (as a discipline) being quite aware and even proud of how well they "perform". This is an interesting suggestion in the context of experimental economics in particular, but in fact, experiments in economics have received attention in the performativity programme only insofar as they are used to design interventions and applications, i.e. insofar as they can serve as "demonstrations" (Muniesa & Callon, 2007). This focus on applicability and intervention resonates with a particular understanding of the purpose of economic laboratory experiments that has been argued for by Francesco Guala (2007), and that he also explicitly contributed within the discussion on the performativity of economics. Guala's *The Methodology of Experimental Economics* (2005) is the most established of three detailed studies on experimental economics from the perspective of philosophy and history of science. These three studies largely form the conceptual background for my own research, which is why I will discuss their claims and arguments in some detail now.

2.3.1 Guala's *Methodology of Experimental Economics*

Guala, a philosopher of science and of economics in particular, aims in his monograph (2005) to clarify and resolve some methodological problems involved in using laboratory experiments in economics. It is therefore predominantly a normative account of what good scientific practice in experimental economics can and should be, although with the often repeated caveat that what a good method is ultimately depends on the intended purpose.

Guala first sets out with describing what, from a philosophical perspective, is the most challenging aspect of experimentation, inductive inference. In contrast to deductive inference, which is a logical relationship between a set of statements and a conclusion, inductive inference is fallible. It is not straightforward that we can infer from any given observation that a statement (a hypothesis) meant to explain this observation is true or false; the evidence in favour of the hypothesis only makes the hypothesis more likely to be true. In particular, it is not clear that an experiment can be used to test any given prediction, since developing a prediction usually involves taking into account some initial conditions based on which the prediction is made, and some auxiliary assumptions about the conditions under which the prediction is expected to hold (this is known in philosophy of science as the *Duhem-Quine-Problem*). The experiment is then always a test of the prediction in combination with the auxiliary assumptions and initial conditions; if the experimental result refutes the prediction, it is not obvious whether that is because the prediction failed, or because the experimenter had the initial conditions or the auxiliary assumptions wrong.

To answer these worries, Guala develops an account of inductive inference for the practice of experimental economics via the model of a "perfectly controlled experiment". A perfectly controlled experiment is built on randomisation and comparison between two (or more) experimental groups. Both groups are subjected to the same experimental conditions, except for one single factor that is varied between the two. It is assumed that the two groups are otherwise completely uniform, and the most common strategy to ensure (statistical) uniformity is random assignment of participants to either group. It is further assumed that all the background conditions are held constant across both experimental conditions. If such an experiment could be performed, and the behaviour of the group subjected to the experimental variation would be significantly different than the behaviour in the other group, it would be straightforward to identify the parameter that was the only difference between the two groups as the cause of this changed behaviour. According to Guala, this model can explain how experimenters actually arrive at valid inferences in laboratory experiments. The validity of these inference is, most of all, an empirical question, and the context and circumstances of the research matter in each individual case. If a phenomenon that refutes a well-established theory has been identified in experimental research, experimental economists proceed by checking whether it is an artefact of the experimental procedure, whether it is robust to variations in the design and the background conditions, and whether the auxiliary assumptions about the conditions in which the theory is expected to hold have been adequately instantiated in the experiment. In other words, the validity of inductive inferences (about the truth or falsity of a hypothesis or prediction) is assessed by investigating the experimental "background" conditions through series of follow-up experiments, and by successively eliminating alternative explanations for the observed phenomenon.

This account of inductive inference therefore concerns the *internal validity* of claims about the relationship between experimental evidence and a hypothesis. Guala's project, however, also includes an account of the *external validity* of laboratory experiments in economics, i.e. how economists can use laboratory experiments to learn something about behaviour in real-world economies. As Guala argues, there is a trade-off between internal and external validity in laboratory experiments. The more tightly controlled and simple an experiment it is, the easier it is to derive valid inferences about the causal effects of any given factor from it, but the less likely it is that these inferences are also valid when applied to much messier real-world contexts in which many more unknown background factors interact. The problem then lies in the inherent difference between laboratory experiments and real-world contexts, a problem that affects all laboratory sciences to some degree. Guala's solution to this problem is similar to his solution for inductive inference: It is mostly an empirical question in what ways laboratory experiments differ from their "target" real-world contexts, and whether these differences are causally relevant. In particular, experiments can use the comparison of experimental results to field data to infer whether the causes they have identified in the lab are also likely to obtain in the real world. This is an important prerequisite of a specific type of experimental economics that engages in "economic engineering", i.e. the design of real-world institutions (such as specific types of markets) which are first tested in the laboratory. It is, however, also necessary when experiments are used to investigate phenomena that were first observed in real-life contexts. In any case, solving the problem of external validity is only possible when the experimental results are intended to apply to a very specific target system, such that the experimental system can be modified to resemble it until the remaining dissimilarities are causally irrelevant. In some cases, as in the design of auctions, it is also possible to make the target system resemble the experiment. Guala argues that experimental economists in general seem to downplay the question of external validity, due to a rhetoric that was used in the early years of establishing the new research programme mainly as a tool for theory-testing:

> Unlike some of its neighbor disciplines, such as experimental psychology, experimental economics grew within (and had to defend itself from) a scientific paradigm that attributes enormous importance to theory. This is probably why it was sometimes easier and more effective from a rhetorical viewpoint to present experimental economics as primarily devoted to theory testing. This view is mistaken, because the proper role of experimental economics is to mediate between abstract theory and concrete problem solving in the real world. […] Like models, experimental results must eventually teach us something about the real world. However, many experimental results in economics are never applied to real-world situations. (Guala, 2005: 229).

In his view, any research programme "should eventually produce applicable results (that's what scientists are paid for, after all)" but Guala allows that a specific experiment "may just highlight a phenomenon or mechanism, to be later exploited by applied scientists when they deal with specific cases" (ibid., 230).

In a later text (2007), in which Guala revisits some arguments of his book in the context of the performativity-debate mentioned above, his claim that experimental economics should ultimately be understood as application-oriented research is even more pronounced. Here, Guala describes what he sees as two different „logics of experimentation", the approach of (theory)-„testing" and the approach of (institution)-„building". The approach of the "testers", according to him, is ultimately misguided; it is not at all clear what the benefit or epistemic gain of testing game-theoretic models in laboratory experiments should be. Ultimately, what can be tested is only how well a model applies to a given situation. In contrast, the "builders" have successfully used experiments to intervene in the economy, and Guala therefore expects a shift towards more application-oriented research in experimental economics in general.

2.3.2 Santos' *Social Epistemology of Experimental Economics*

A much less normative philosophical account of how experiments in economics can produce reliable knowledge was provided by Ana Cordeiro dos Santos (2010). Similar to Guala, Santos devotes the first half of her project to developing a theory of scientific experiments in general, the "social epistemology of experiment". From Pickering (1989), Santos takes the idea that experiments end once a *three-way coherence* between the experimenter's conceptual understanding of the phenomenon, the conceptual understanding of the apparatus and the results of the experimental procedure has been achieved. Drawing on Pickering's (1993, 1995) conception of a *dialectic of resistance and accommodation* in experimental practice, Santos also emphasises that an important aspect of the epistemic value of experiments is the material participation of the object of interest, which can and does "resist" the experimenter's theoretical expectations. Beyond Pickering's account, Santos also highlights the importance of the "social dimension" of experimentation, i.e. that experimental procedures and their results are subject to critical assessment by the scientific community. Scientific institutions that facilitate critical discourse, especially where it involves the exchange of ideas between scientists coming from different backgrounds, can counteract the tendency of individual scientists to experimentally confirm their own expectations. The *material resistance* of the phenomenon of interest and the *social resistance* of the scientific community are therefore the two main sources of epistemic value for experimental results.

Applying this theoretical framework to experimental economics, Santos argues that the participation of human subjects is what makes experiment epistemically superior to mathematical models in economics. Santos disagrees with Guala's claim that experiments should strive for greater *external validity* so that experimental results can be applied to specific real-world situations, because this view does not do justice to most of the contemporary research in experimental economics. She instead argues for the view that the results of economics experiments enable us to draw *generic inferences* to a class of situations (rather than a specific real-world target) that are structurally similar. Most experiments, on Santos' account, are therefore inspired by theory or previous experimental results, rather than a specific real-world situation.

It is, however, important to assess in each individual case how far the experimental design constrains the agency of the participants and therefore limits the potential for *material resistance.* In order to produce unambiguous results that can be attributed to the effects of the experimental design rather than any other unknown factor, experimental economists need to exert a considerably high level of control over participants' behaviour:

> The key epistemic question of scientific experimentation concerns the trade-off between experimenter's actions and the agency of the material world. Control is necessary to produce valid results, i.e. results that scientists believe are no artefacts of the experimental procedures. But the more control is exercised, the more the results are the outcome of scientists' actions rather than the agency of the material world. (Santos, 2010: 181)

In particular, the standard methodology of experimental economics requires that experimenters induce participants to prefer certain outcomes over others, by providing monetary incentives and imposing a particular "reward structure" on the experimental system, which attributes actual gains and losses to the range of possible decisions. In some types of experiments, these constraints are so rigid that participants have very little option but to behave according to the experimenter's intentions. These are experiments that study how to bring about a desired outcome on the aggregate level, for example by designing a market mechanism that helps buyers and sellers to achieve the most efficient transactions. The advantage of these experiments is that they can be exported to real-life contexts where similarly rigid conditions can be imposed. Santos calls these types of experiments "technological experiments", because they provide insights on how human behaviour can be controlled by imposing certain institutions to achieve a desired outcome. Their object of study thus is the institution (e.g. the market design) itself, rather than the behaviour of the participants.

"Behavioural experiments", in contrast, study how individuals interact with strategic environments, and allow far greater input and variety of individual behaviour. Some behavioural experiments, or rather series of experiments, have led to the discovery and establishment of behavioural regularities that contradict the predictions of standard economic theory. The results of the so called *ultimatum game* (UG) experiments, for example, show that in a situation where two people need to divide a sum of money between themselves (one offers a distribution and the other can only accept or reject the offer, but in the case of rejection both earn nothing), social norms of fair sharing rather than rational motives of income-maximisation are guiding the majority of interactions. Santos also analyses the ensuing attempts to disprove the results of the ultimatum game by creating situations in which strategic considerations do override fairness, but shows that these experiments are too restrictive on the part of participants' agency. For a proper *behavioural experiment* to be in place, an experiment needs to allow participants to express their motives through choices that will result in outcomes that are in line with their motives. For example, in the ultimatum game experiments, participants who prefer a fair distribution can choose to either propose a fair distribution or reject an unfair distribution (and thus punish the proposer), which eventually (because the proposer can also anticipate the rejection, or learn from the punishment) results in the outcome of a fair distribution. Experiments in which such an alignment cannot be achieved, because the choice of equal contributions, for example, results in unequal outcomes, or because fair distributions cannot be enforced, do not disprove the results of the ultimatum game.

Behavioural experiments cannot be used as "test-beds" for policy mechanisms, but according to Santos, they nevertheless offer important insights on human behaviour in particularly interesting situations. For example, the *generic inference* drawn from the ultimatum game is that fairness plays a role in social contexts where individuals who prefer fair distributions can enforce this social norm on others. The results of behavioural experiments can therefore be used to develop new theories and models of human behaviour, and according to Santos, this is precisely the task of the discipline of *behavioural* *economics*. However, in order to draw valid inferences about the motives behind human behaviour, experimenters must allow for human agency to play a substantial role in experiments. Santos' analysis shows that through the experimental design, i.e. by either enforcing or relaxing constraints on choices and outcomes, economists are in principle equally able to produce "standard" rational income-maximising behaviour and "anomalous" behaviour. This results in the two types of experiments described above, but it also produces a trade-off situation between experimenters' control and participants' agency. Too little control jeopardizes the results of experiments, because they cannot be clearly attributed to one of the known features of the experimental system; and too much control diminishes the epistemic value of results, because they will be the outcome of experimenters' actions rather than participants' agency. How, in practice, experimenters handle this trade-off situation will be the topic of the first part of my analysis.

2.3.3 Svorencik's *Experimental Turn*

The third monograph on experimental economics is a recent doctoral thesis written by Andrej Svorencik (2015) at the University of Utrecht. It differs from the two studies discussed above not only in terms of its method and epistemic interest, but also in its characterisation of experimental economics as a method and discipline. In order to reconstruct what Svorencik calls the "experimental turn" in economics, he interviewed around fifty researchers who had been involved in the development of the discipline between the 1970s and early 1990s. These interviews were supplemented by archival research and document analysis of the private files of some prominent experimental economists, and the statements that a group of pioneering researchers had given during a "witness seminar" on the early years of experimental economics in 2010. The result is a very detailed analysis on the emergence and institutionalisation of experimental economics according to its protagonists. It covers the personal "conversions" of some early researchers towards the experimental method, the impact that the development of computerised laboratories had on the institutionalisation and proliferation of experimental research, the struggles within the community leading up to the incorporation of a dedicated research society - the *Economic Science Association* (and the controversies surrounding the choice of this name) – the struggles between experimentalists and journal editors for the acceptance of this kind of research in the economic mainstream, and finally, the inception of a dedicated journal (*Experimental Economics)* in 1998.

Svorencik's main argument is that laboratory experiments succeeded as a method in economics because they supply theory-building with rigorously produced high-quality data. He illustrates this argument by reconstructing four "driving forces" behind the "experimental turn":

a) A conceptual reconfiguration of what counts as data and evidence in economics (integrity),
b) Knowing how and by whom such data are created (rigorousness),
c) A realization that experimental research is most potent when it goes in tandem with economic theory (a virtuous circle),

d) Elevating experimental data to the same level as theory (symmetry).
(Svorencik, 2015: 6)

According to Svorencik, experiments under controlled conditions offered an enormous advantage over field data, because the latter is often collected in unknown circumstances or with unreliable methods (integrity). In experiments, the researchers themselves could oversee the data collection process and make sure that the methods were reliable (rigorousness). The insight that experimental data can illuminate theory (symmetry) and that experimental results can even be used to build new theories (the virtuous circle) is what enabled experimental economics to become a research programme, as experiments generated new questions that needed to be answered through further experiments and theory-building. The core of the "performative turn" therefore was not simply an agenda for introducing the experimental method to economics, but "a claim that is deeply normative and at the same time empirical; namely, that rigorously controlled data collected in the laboratory or the field is a better foundation for economic theorising than other types of data, evidence or observation." (ibid. 236). From this perspective, Svorencik also criticises the notion of experimental economics being mainly theory-driven. His argument is that experimentalists cared first and foremost about the quality of the data they produced. Only data produced in known and tightly controlled conditions could be used to test theories, since it was in such conditions that the theory was given its "best shot": "Experiments finally produced data that were fit to test theory, data that were rigorous enough. These were not haphazard data, but data that had been geared towards theory, so that theory could not escape." (ibid.)

The introduction of computerised laboratories in the 1980s greatly facilitated the collection of data under controlled and standardised conditions, and it also enabled the design of more complex experiments. Svorencik also points out that institutionalised laboratories helped small research communities to form around them, which gradually led to the formation of the new research identity of experimental economists. Further steps towards the institutionalisation of the field were annual meetings and conferences, which served as precursors to the founding of the *Economic Science Association* in 1986. At this point, Svorencik argues, the aim of the most influential group around the later Nobel laureate Vernon Smith was still to promote the value of observational methods for economics in general (thus making it more "scientific"), which is why this (somewhat pretentious-sounding) name was chosen in favour of a name referring explicitly to experiments. One of the main reasons why it took more than ten years to also found a dedicated journal for experimental economics was that experimenters did not want to become "ghettoised" and preferred to try and publish their research in high-ranking general interest journals such as the *American Economic Review.* This, however, was not always easy, and Svorencik dedicates a particularly detailed chapter to the analysis of the different rhetorical strategies that early experimental economists used to convince journal editors and referees that their research was worth publishing. Apart from pragmatic request for being reviewed by a referee with some knowledge in experimental methods and asking critical referees to provide concrete arguments of where and how the experiment had failed, these strategies also include the often-voiced argument that experiments presented real, albeit simple, economic situations, and that general theories could be expected to apply to these simple situations as well. Interestingly, this emphasis on the relevance of laboratory experiments for theories – which is in line with the "driving forces" described above – is less prominent in Svorencik's description of the strategies that were used to promote experimental research towards funding bodies. Since boards of funding bodies like the American *National Science Foundation* are often staffed primarily with natural scientists, highlighting the relevance of economics experiments to real-world applications (instead of theories) made it easier to argue for its importance in these contexts. What more, the funding of experimental research according to Svorencik helped the economic programme within the NSF to expand, because the use of the experimental method in economics "was also a means of increasing scientific credibility vis-à-vis other more mature sciences operating within the NSF." (Svorencik, 2015: 191).

As is probably notable already from this short summary, Svorencik's account of the institutionalisation of experimental economics is a highly appreciative historiography. The positive values exemplified in the "driving forces" are identified as the overall explanation for the considerably fast success of experimental economics both as a method and as a field. In contrast to Guala and Santos, Svorencik touches upon methodological discussions only in passing in his analysis of a particular controversy, which he uses mainly to illustrate how it helped to establish the credibility of experimental research towards the community of mainstream economists. The most striking difference in his account, compared to those of Santos and Guala, is that application-oriented research – what Guala calls the "building" approach and Santos refers to as "technological experiments" – does not play a role at all in explaining the success or the value of experimental economics. If Guala (2007) is right and the shift towards application-oriented research is a rather recent phenomenon, then this discrepancy could be explained with Svorencik's sampling of mainly first-generation experimenters and his research interest in the period when experimental economics still had to prove its worth within a very theory-driven discipline. However, Svorencik's notion that experiments are generating new theoretical questions and in this way serve as the starting points for entire research programmes conforms well to Santos' account of the use of experiments in economics, and her argument that experimental results can yield "generic inferences" even when they are not applicable to a specific real-world context. I highlight these tensions in the three studies of experimental economics discussed above because they foreshadow some topics that also emerged from my interviews. Questions concerning the respective roles of theory and observation, and the ideal of scientific experimentation that economics experiments should methodologically aspire to, seem to be still unresolved within the discipline, and this is likely what the differences in these three accounts signify as well.

2.4 My Contribution

Given the detail and analytical sophistication of Guala's and Santos accounts, my aim is certainly not to add to a discussion of the methodology and the epistemic value of economics experiments. Likewise, Svorencik's historical study is far more comprehensive than the very partial and context-specific perspective I am able to offer based on the interviews I conducted. My research project, however, also has a different orientation towards experimental economics, namely an approach grounded in the methodology and epistemic interest of laboratory studies in STS. In line with the initial idea behind these ethnographies of particular research communities, my focus will be on the epistemic practices involved in doing laboratory experiments in economics. Guided by the observation that epistemic cultures are diverse and bring about very different strategies of approaching the empirical world and organising research, my specific interest lies in the ways that economists have appropriated the experimental method and put it to use within a very established and unified academic discipline that was traditionally non-experimental.

The above outline of research in the tradition of *laboratory studies* shows that with very few exceptions, social sciences have not received much attention from STS research so far. This is even more obvious for experimental social or behavioural science. The only account of social science experimentation within STS that I had encountered when starting this research project was Knorr Cetina's typology of laboratories, which includes a brief characterisation of social science labs. Knorr Cetina (1992) here compares social science experimentation to simulations or sand-boxed war games, in which the action of interest is merely represented or "staged" in "mock behaviour" scenarios. This is a description that methodologists and experimental economists, who invest a lot of money and effort into making sure that the stakes and therefore the behaviour they observe in an experiment are "real", would certainly reject. At the time that this characterisation was written, experimental economics had only recently been institutionalised, and Knorr Cetina is certainly not to blame for not being aware of this type of research; also, social science laboratories, although part of her typology, are a site of knowledge production she has not studied herself. The seeming inadequacy of this solitary account of social science experimentation in STS, however, makes an investigation in the laboratory practices of economists all the more worthwhile. While the research practices of particle physics and molecular biology have not changed in dramatic ways in the twenty years since Knorr Cetina concluded her research on epistemic cultures, the incorporation of laboratory experiments in the mainstream of economics has certainly affected economists' self-understanding of their methodological possibilities – at least if Svorencik's thesis of the "experimental turn" is correct. Svorencik also explicitly mentions the lack of attention towards the practices of this considerably new discipline from sociologists of science.

The contribution my research project makes, however small it may be, can be understood as a first step towards filling this gap in the laboratory studies tradition.

There is, however, more to this research project than just exploiting the fact that economists now use laboratories and can therefore be an object for laboratory studies. Economics has come under scrutiny by scholars from other disciplines not least since the last financial crisis. The policy-relevance of economic research, including experimental studies, and the pervasive influence of economists not only as academics, but also as advisors and public administrators warrants a heightened interest in the practices on which their claims are based. The research programme on the performativity of economics is an example of this interest, but this kind of research has largely studied economic activities outside of academic contexts. Whether these accounts are critical or supportive of how economics is done, they do highlight the high expectations surrounding the power of economics to actually intervene in contemporary conditions of public life. Meanwhile, the transformations taking place in the government of science and academia have not left the academic discipline of economics unaffected. It is therefore worthwhile to ask what role the method of laboratory experimentation plays in a discipline that needs to demonstrate its accountability not only according to the criteria and metrics of academic research evaluation systems, but also towards a public that demands solutions for urgent problems. Due to the relative novelty of the discipline of experimental economics, it is likely that the answers to the question of what the value of experimental research is will differ between individual researchers. The diverging ideas, argued for by Guala, Santos and Svorencik, of what the epistemic gain and therefore the purpose of laboratory experiments is, additionally indicate that there might be an interesting heterogeneity to explore in experimenters' accounts of their own research.

3 Theories and Sensitising Concepts

This chapter clarifies the theoretical notions of "experiment" and "laboratory" by discussing various definitions and uses of these concepts in the STS literature. I will carve out the characteristics of experiments and laboratories that can be identified within these theoretical accounts and that are informative in the context of my own research project. The second part of this chapter is dedicated to the conceptual notions of principles and regimes of valuation, and describes the situations and contexts in which an analysis of valuation practices can provide insights that help to illuminate how (epistemic, economic, moral…) values are stabilised. The concept of epistemic cultures, proposed by Knorr Cetina (1999) serves as a bridge between my theoretical interest in epistemic practices and valuation practices. An engagement with valuation practices also opens up the perspective of my analysis towards the institutional context of governing and evaluating academic research that my respondents and their research practices are embedded in beyond the specificities of their particular epistemic culture.

3.1 What is an Experiment?

Before investigating how experimental economists appropriate the experimental method, it will be useful to clarify what "experimenting" means and how this practice has been characterised in previous studies. Experimentation as a method is often associated with the origin of modern science, so much so that "'experimental method' used to be just another name for scientific method" (Hacking, 1983: 149). In its briefest definition, experimentation is the "testing of a hypothesis under controlled conditions".[5] This means that experiments are considered as mainly theory-driven; only after a hypothesis is deduced from a theory can it be tested experimentally. If the experimental results are sufficiently similar to the results predicted by the hypothesis, the hypothesis and therefore the initial theory is confirmed; if not, the theory is refuted (see also Gooding, 1992 for a graphic model of this view). This "received view" of experimentation has been criticised for several reasons. The first is a theoretical argument known as the *Duhem-Quine-Thesis*, saying essentially that since no hypothesis can ever be tested in complete isolation, every hypothesis can be defended against a negative experimental result through the assumption of auxiliary hypotheses that explain the result while keeping the theory intact. "Falsifications" of theories as envisaged by Popper or instances of Bacon's "experimentum crucis", which enables a definitive decision between two competing theories, are therefore much rarer and much less important than traditional philosophers of science have assumed. The second reason why the received view has been criticised is that it is also descriptively false. As Ian Hacking (1983) points out, experiments need not necessarily start with a clearly defined hypothesis or even a theory; many

[5] This definition can be found in: *McGraw-Hill dictionary of scientific and technical terms*. (2003) (6. ed). New York, NY: McGraw-Hill.

important scientific discoveries were made in an explorative fashion, out of curiosity. Only retrospectively were these experiments then represented as testing a specific theory or conjecture. There must therefore be other common characteristics of experimentation that make it recognisable as a specific type of practice across different scientific disciplines and research cultures.

A more elaborate definition of experiments, found in the *Dictionary of the History of Science*, thus reads like this:

> An experiment, unlike an experience, is a designed practical intervention in Nature: its upshot is a socially contrived set of observations, carried out under artificially produced and deliberately controlled, reproducible conditions. At the experiment's core is the notion that the conditions for producing a given effect can be separated into independently variable factors, in such a way as to demonstrate how the factors behave in their natural (i.e. in the non-experimental) state. The crucial assumption is that the factors studied – and represented in experimental design as independent and dependent variables – retain their identities (and dispositional properties) whether or not other conditions are held constant, as in the laboratory, or freely vary, as in extra-experimental reality. (Bhaskar, 1985: 136).

This definition will be useful for my further analysis because it is sufficiently abstract to identify the beliefs, so to speak, which also underlie laboratory experiments in economics. An experiment is an active intervention, usually based on a design to create a specific effect under known and controlled conditions. The factors believed to create the effect are deliberately introduced by the experimenter, with the idea that they will affect the experimental situation in the same way as they would in a "natural environment". However, this is also a purely cognitive model of experiments, abstracting from the actual processes involved, and the definition given would not apply to field experiments (which are not necessarily done under artificial conditions) or thought experiments (which are not practical interventions in nature). Bhaskar therefore also qualifies the initial description by adding that experiments involve work on all four levels of theory, design and instrumentation, experimental production, and control and observation. To get "a single type of mechanism going (experimental production) in relative isolation (experimental control)" (ibid. 138) therefore requires both theoretical knowledge (including adequate auxiliary hypotheses) and practical skill.

3.1.1 Experimentation as a skilful practice

It is this level of practical skill, which includes the work on design, instrumentation, the control of the background conditions and the production and observation of effects, which has received the most attention within STS. Against the traditional view that reduces the function of experiments to the controlled production of observations to test theoretical hypotheses, historians, philosophers and sociologists of science since the early 1980s increasingly emphasised the practical dimension of experimentation. For some, this was due to epistemological reasons. According to Hacking (1983), turning to scientific practice is a more reliable guide of finding out what we can know. Rather than discussing the logical consistency of theories and therefore restricting themselves to discussion about representation, philosophers should look at scientists' practices to learn about reality. The fact that experimenters not only produce theoretically predicted entities such as electrons in the lab, but also use them as tools to investigate other phenomena, is for Hacking a strong support to the claim that these entities really exist.

Harry Collins (1992) also studied how changing knowledge claims arise from experimental practices, and highlighted the relevance of *tacit knowledge* (a term first introduced by Michael Polanyi) and practical skill for experimentation. He shows that the replication of scientific results is a delicate achievement even in those cases, where - as in the building of a laser - there are clear criteria for success, because local and procedural knowledge that often cannot be articulated plays an important role. Cases where neither the functioning of the apparatus nor the nature of the results to be obtained is known beforehand can therefore lead to a situation of *experimenters' regress:* to know whether the phenomenon exists, we need to build an instrument that can detect it, but we do not know whether the instrument actually can detect the phenomenon while the characteristics and the existence of the phenomenon are still uncertain; and we do not know these characteristics before we have detected it. In Collins' case, regarding the early debate about whether gravitational waves had been detected with a specific apparatus or not, closure was finally achieved through the invocation of various criteria for the credibility of experimental results, not all of which were purely scientific. Collins' argument is therefore that we should conceive science as consisting of "skilful communities" possessing the relevant expertise to assess knowledge claims within their field, rather than following an "algorithmic" logic of fact-production and infallibility.

Having accepted that experimental practice plays a vital part in scientific knowledge production, STS scholars also set out to describe what this practice consists of. For Gooding (1992), all forms of experimentation are characterised by the simultaneous manipulation of both material and conceptual objects, i.e. the continuous fine-tuning of both experimental technology and theory. Other common features of experiments are human agency in the form of continuous choices (which sounds more trivial than it is), the occurrence of unexpected events, and the absence of a linear, logical structure. Such a structure arises only when experimental practices are reconstructed into linear narratives for different purposes later on. Gooding emphasised the importance of skills and learning in experimentation, aspects that tend to be retrospectively concealed in verbal accounts of actual experiments (such as in publications or science textbooks). Contingent *choices* made during the procedure are then often "upgraded" to *decisions* with a specific goal. Experimental success, according to Gooding, is the gradual convergence of observations and expected outcomes; once the representations obtained in experimentation are sufficiently stable, they are perceived as something corresponding to things in the world rather than being the products of human agency. Reconstructing the actual experimental practices of scientists shows that experimental phenomena are always the products, and not the starting point, of human activity: "all natural phenomena are bounded by human activity whose products express the culture in which it occurs" (Gooding, 1992: 109).

A corresponding view of experimentation is that of Andrew Pickering, who like Gooding sees experimentation as the ongoing moulding of conceptual and material resources. To the convergence of observations and expected outcomes, Pickering adds the concept of an *instrumental model* that explains how the observations where achieved and why they correspond to the expected outcomes. According to Pickering (1989), an experimental fact is the result of a *three-way coherence* between the *phenomenal model* (the conceptual understanding of the phenomenon under investigation), the *instrumental model* (the conceptual understanding of how the experimental apparatus works), and the *material procedure* which is the actual experimental intervention performed with the help of the apparatus. Aligning these three elements in a state of *interactive stabilisation* is a precarious achievement of the experimenter, since new incoherencies are always likely to appear with further investigation. Pickering refers to these incoherencies as *resistances* of the material world and points out that they would not exist, were it not for the experimenter's expectations, which are usually derived from the phenomenal model. *Three-way coherence* is achieved by *accommodating* these resistances through adapting one's conceptual understanding or the material procedure.

In his later work, Pickering refers to this "dialectic of resistance and accommodation" with the metaphor of the "mangle of practice" (Pickering, 1993). This metaphor signifies Pickering's adoption of the "posthuman" approach of Actor-Network-Theory, which grants nonhuman actors agency to the same extent as human actors, and highlights nonhuman contributions to knowledge production. According to Pickering, both human and nonhuman agency is *temporally emergent*, and in experimental work, this emergence takes a reciprocal form of mutual restructuring. Material agency and the theoretical understanding thereof both evolve "in real time" during experimental practice. The same can be said, according to Pickering, of scientists' practical goals. Goals are formulated based on the models of future states and outcomes, and these models are subject to the cultural and material field in which they are developed. They also open up a variety of possible goals, and as the material field changes ("is mangled") in experimentation, as do the models and the goals. The modelling that underlies experimentation is therefore an open-ended process, and since there is no obvious end to the dialectic of accommodation and resistance, experimentation can be considered as open-ended too. Models goals, material agency, human agency and knowledge are all temporally emergent in experimentation.

Santos (2010) builds on Pickering's reconstruction of experimentation for her project of developing a "social epistemology of experimentation". She takes Pickering's concept of *three-way coherence* as describing the main epistemological principle of experimental practice, which can explain why experimenters believe that they have actually produced the phenomenon of interest (rather than an artefact) and why they trust the knowledge stemming from an open-ended process "in which both the means and the outcomes of that process are at stake" (Santos, 2010: 14). Coherence can be achieved by using different "coherence strategies", such as confirming the same result with different experiments, confirming predicted effects, eliminating alternative explanations, and using statistical arguments. However, since theories are in principle always *under-determined* by empirical results, which means that many phenomenal models can in principle be aligned with the outcomes of a material pro-

cedure, coherence alone does not account for the epistemic value of experiments. Santos here also mentions the constraints on researchers' ability of manipulating their conceptual and material resources as a decisive factor: The more established and well-understood an *instrumental model* and the more *rigid* an experimental system is, the fewer options an experimenter has for prematurely ending the experiment and simply confirming prior expectations through conceptual and material manoeuvring. Very *rigid* experimental systems offer the experimenter a high level of control over the experimental conditions, but they might in turn reduce the participation of the material world "to fairly confined and recognizable problem-situations" (Santos, 2010: 44). However, in Santos' view it is not only the participation of the material world in experimentation and the *resistances* it offers to the conceptualisations of the experimenter that accounts for the epistemic value of experiments. Another important dimension are the *social resistances* and the *social validation* offered by the criticism and appraisal of the scientific community. According to Santos, experiments are of the highest epistemic value when the participation of the material world is enabled to a high degree, such that the outcomes of the experiment are determined by the material agency rather than the experimenter's agency, when they are *socially robust* and consensually appraised by the scientific community, and when their results, once established, can be put to use in subsequent scientific (and even non-scientific) practice.
Going back to the very beginnings of modern science, it is notable that the social appraisal of knowledge was instrumental in establishing the credibility of the experimental method itself. One of the methods for establishing this credibility was to perform experiments in front of a selected (i.e., trustworthy) public, as Shapin and Schaffer (1985) show in their account of the debate between Robert Boyle and Thomas Hobbes in 17th century England. According to them, Boyle succeeded in introducing his experimental method of natural philosophy by creating "matters of fact". It was important that the insights gained through this new science were viewed as given by nature, not as created by man. Boyle achieved this through employing three different technologies: The *material technology* of the air-pump, the scientific apparatus used in public experiments, the *literary technology* of "virtual witnessing", through which Boyle ensured that his experiments could be replicated or at least asserted credibility by those who had not actively witnessed them, and a *social technology* which regulated the norms of witnessing, assenting and dissenting within the noble community of those considered intellectually and morally capable of giving testimony about experiments. All of these served as "objectifying resources", making it possible that the results of experimental science would be matters of fact, which could not only be assented to, but also believed with "moral certainty" (cf. Shapin & Schaffer, 1985, chapter II).

The public witnessing of experiments, but also the trustworthiness of the public, had an important function for establishing the credibility of this method. Given that these "public" experiments were usually replications of earlier successful experiments done in the laboratory (where the laboratory also derived some of its credibility from being installed in the house of a gentleman, see below), in our contemporary view they would rather be considered demonstrations than experiments. Today, what Shapin and Schaffer call “virtual witnessing” has become more important for establishing the credibility of experimental results: this is the *literary technique* of disclosing all the elements of an experimental design in the publication of the results, such that fellow scientists can understand exactly how the results were obtained, and would in principle be able to reproduce the procedure. Santos (2010) refers to this possibility of assessing the experimental procedure without actually needing to reproduce the experiment as an important element of *social validation.* The recent debate surrounding the reproducibility of results in experimental psychology (Open Science Collaboration, 2015) exemplifies how much the credibility of experimental science still rests on the idea that replication should in principle be possible and actively supported by complete and accurate reports of the experimental procedure.

Combining the definitions and views discussed above, I would like to retain the following aspects of experiments and keep them in mind for my own analysis:

- All aspects of experimentation (the building of an apparatus, the control of the experimental conditions, the creation of phenomena, the recording and interpretation of observations) involve *practical skill* and often *tacit knowledge*. This skill is acquired through immersion in a community of experimenters, through experience and failures.

- *Material intervention*: Experimentation consists of some practical intervention in the material world, and the feedback or “resistance” offered by the material is a source of knowledge for scientists. The participation of the material world is what makes experiments epistemologically useful, and this can be said of experimental economics as well:

> The direct participation of human subjects is the major source of epistemic value of economics experiments. Participants in economics experiments may ‘resist’ economists’ expectations and thereby prevent experimental results from being exclusively determined [by] economists’ material and conceptual interventions on the experimental microeconomic systems. (Santos, 2010: 182).

The epistemic value of an experiment consequently depends on how much material participation and resistance it allows. In other words, experimentation is always a *trade-off between experimental control and material agency*. Often the phenomena (whether intended or not) that are created through these material interventions would not exist otherwise, but as they are increasingly stabilised in experimental practice, they are identified as elements of reality.

- An experiment ends when the experimenter’s conceptual understanding (the *phenomenal model*), the *instrumental model* of how the experimental apparatus works, and the results of the *material procedure* of the experiment are aligned in a state of *three-way coherence.* Achieving this coherence is what gives experimenters confidence in their results and their method.

- Experiments have historically had different purposes and functions. While some of them certainly also are designed in order to test particular hypotheses, others had an *explorative* function; in other words, while some experimentation is explicitly theory-driven, experimentation as a practice within science also has “a life to its own” (cf. Hacking, 1983). Outside of the proper realm of real-time research, experiments and accounts of experiments can be used as public *demonstrations*, which legitimise and lend credibility to a specific type of research. The credibility of experimental results also rests on their potential for *replicability* and on the *assent* of the peer community. The *social validation* of the scientific community and the subsequent use of the results further confer epistemic value on an experiment.

3.2 What is a Laboratory?

In my analysis I will follow Knorr Cetina’s (1992, 1999) proposal to distinguish experiments and their characteristics from the laboratories in which they are taking place. Knorr Cetina called for such a distinction because in her view, laboratories had been underrepresented in theoretical accounts of knowledge production in comparison to experiments. In science studies, however, the laboratory was increasingly recognised as an “important agent of scientific development” and the “locus of mechanisms and processes which can be taken to account for the success of science” (Knorr Cetina, 1992: 116). Moreover, studying laboratories is instructive because even though their principal epistemic function remains the same across science, the technologies through which this function is achieved vary according to the specific *epistemic culture* of a field. The principal function of laboratories is thus that they achieve

> a reconfiguration of the system of ‘self-others-things’, of the ‘phenomenal field’ in which experience is made in science. As a consequence of these reconfigurations, the structure of symmetry relationships which obtains between the social order and the natural order, between actors and environments, is changed. (Knorr Cetina, 1992: 116)

The “enhanced environment” (ibid.) of the lab achieves a reconfiguration of the natural order by taking the objects of research out of their natural context. This means that objects that would not be accessible in their natural environment are brought into the lab through different means (for example, through the use of digital image processing technologies in astronomy), that natural time scales are subverted through the acceleration, slowing down or repetition of processes, and that objects are processed only in parts, after substantial re-engineering (consider transgenic mice), or in the form of models, visualisations or traces. Knorr Cetina also refers to this as an “enculturation of natural objects” (ibid. 118), a transition that is achieved through different processes of *laboratorisation* (Knorr Cetina, 1988), which make the objects of research accessible and manipulable at will.

The process of reconfiguration does not stop at the human agents. According to Knorr Cetina, especially in sciences which involve a high amount of direct contact with the “natural objects” (such as molecular biology, see Knorr Cetina 1999, chapter 4), scientists are enculturated and enhanced in various ways to fit the new emerging order of self-others-things: “In the laboratory, scientists are ‘methods’ of going about inquiry; they are part of a field’s research strategy and a technical device in the production of knowledge." (Knorr Cetina, 1992: 119). They are, for example, turned into measurement devices, capable of solving different problems and making "good guesses" by invoking embodied knowledge that is achieved during practical routines, training and experience. This observation underwrites Knorr Cetina's argument that laboratories, once the process of reconfiguration is complete, are characterised by an order that is "neither social nor natural" (ibid. 121). Both the natural objects and the researchers studying them are transformed, or, to use a more Foucauldian term, "disciplined" in and by the laboratory, to become more apt for the production of knowledge. In other words, laboratories work by achieving specific differences between what goes on within them and what goes on in everyday life, but also between themselves and other laboratories and scientific fields. These "cumulative processes of differentiation" (Knorr Cetina, 1999: 45) may explain why laboratory processes are also markedly different in different research fields.

Michael Guggenheim (2012) in his discussion of how the term "laboratory" was used in the history of sociology also stresses these differentiation processes, and defines laboratories by the separation they achieve between the relevant and stable "inside" and the irrelevant "outside":

> The laboratory is the result of a procedure that separates between an outside, an environment that is considered negligible for some epistemic claim or technological invention, and an inside, a (partly) controlled environment that is considered relevant for this claim or invention. The lab is not so much a closed space, but a procedure that often results in a space with the properties to separate controlled inside from uncontrolled outside. Control means not necessarily physical control but a procedure whereby data and objects are managed to behave in a way the scientist wishes. The separation between inside and outside allows for the two central features of the lab, placelessness and consequence-free research. (Guggenheim, 2012: 101).

It is notable that Guggenheim explicitly defines a laboratory in a procedural and praxeological sense rather than as a place with specific characteristics; a laboratory as a "placeless" place only exists because "scientists created boundaries that separate the lab from the world" and those boundaries are established by "scientific operations" (Guggenheim, 2012: 102). These operations are necessary to render the laboratory environment stable, so that the controlled environment can be clearly distinguished from the unstable research objects studied in it. If there is a difference in observations, it must be attributable to a difference in the object, not in the environment. Once such stability has been achieved, the laboratory can function as a "mechanism for generalisation", because "epistemic claims or objects derived from labs can be extended to other non-controlled environments" (ibid.). Note that this function of generalisation rests on the assumption, mentioned above, that the factors present in the laboratory have the same qualities they would have in uncontrolled environments.

3.2.1 The placelessness of laboratories

What Guggenheim's definition also highlights is that laboratories are mechanisms for generalisation because, as places, they do not have any relevant characteristics; on the contrary, their main characteristic is that they are in principle placeless and dispersible, detached from the contingencies of local cultural and natural conditions. As Robert Kohler (2008) points out, this notion of placelessness is a powerful social fiction, since laboratories in different countries, cultures and continents of course vary and have distinctive qualities (similar to other widely dispersible institutions like a lingua franca). Moreover, the idea of the "placeless" laboratory is a typically modern one. In early modern science, the credibility of the knowledge produced in laboratories was achieved because those laboratories were installed in the houses of gentlemen. Since noble houses were considered to be public places and marked by a social code of hospitality and openness, experimental knowledge produced there had an advantage in credibility over the obscure knowledge coming from the secluded alchemists' kitchens (Shapin, 1988). In other words, in early modern science, place mattered for epistemic authority. Placelessness, however, according to Kohler is the hallmark of *modern* institutions in general:

> Placelessness marks lab-made facts as true not just to their local makers but to everyone, anywhere. It marks the lab as a social form that travels and is easy to adopt, because it seems rooted in no particular cultural soil but, rather, in a universal modernity. Placeless means dispersible. (Kohler, 2008: 766).

The idea of placelessness attached to modern laboratories thus follows a greater logic of universality and dispersion that is compatible with other modern developments such as mass public education, managerial large-scale industrialism and the modern bureaucratic state.

Tom Gieryn (2002) offers a practical explanation for the placelessness of modern laboratories. Especially in the life sciences, laboratories have become increasingly standardised not only for epistemological, but also for financial reasons: It is simply more cost-efficient to re-use architectural elements from earlier laboratories that scientists were satisfied with, and to have labs fitted out by firms specialised in providing this particular equipment. Also, scientists generally assume that their peers in different laboratories are working under similar conditions as themselves. The actual and assumed standardisation of laboratory environments therefore renders the place of knowledge production irrelevant: "Place (as a unique spot in the universe) no longer adds credibility to scientific claims, but only because those places of inquiry have been carefully designed and built to be identical" (Gieryn, 2002: 127). Beyond the material outfit, scientists can also expect behavioural codes and, indeed, the people in other laboratories to be similar to themselves. This is also the case in the behavioural sciences where, due to their difficult research objects (Peterson, 2016), material arrangements are not sufficient to achieve a standardised environment. Not only are there rules in place, which keep those lacking sufficient training and credentials out, but laboratories also function as "normative landscapes" (Gieryn, 2002: 128) in which certain codes of conduct prevail. Very similar to the gentlemanly houses of early modern science, the principal openness to peers and the shared rules of behaviour help to underwrite the credibility of knowledge claims, although in contemporary science this remains more implicit. What seems to apply to the notion of a laboratory and the particular form of credibility it confers across these examples, however, is the idea that the circumstances of knowledge production can be generally *known*, considered to be common knowledge, and therefore be rendered inconsequential.

3.2.2 Laboratory and society

All of the theoretical accounts of laboratories mentioned so far emphasise the distinction between laboratories and the natural world, and the procedures necessary to achieve a boundary. This seems to run counter to many contemporary uses of the word that extend the notion of "laboratory" to refer to places of knowledge production in general, such as "the city as a laboratory" or "society as a laboratory". For Guggenheim (2012) this metaphorical use is somewhat misguided, because it insinuates that the places referred to as laboratories share some of the characteristics of laboratories as defined above, whereas most often they can neither be controlled nor contained. However, there is one very well-known account of laboratory sciences which argues that they are successful precisely because they manage to enrol society as a whole in their particular form of knowledge production. For Latour (1983), a laboratory achieves its epistemic benefits by a series of "displacements of actors", the dissolution of the boundaries between inside and outside, and the active reversal of differences of scale. His example for such displacements and reversals is Louis Pasteur's strategy in dealing with the historical anthrax epidemic. Pasteur first moved his laboratory to anthrax-contaminated farms, then brought back the cultivated microorganism to the lab where it could be made visible, controlled and manipulated, and finally brought a weaker breed of the microorganism back to the farms in form of a vaccine. Latour claims that this amounts to extending the lab first to French farms and subsequently to the French society as a whole: "In this succession of displacements, no one can say *where the laboratory is* and *where the society is*." (Latour, 1983: 265). The laboratory extends its power by dissolving its boundaries and enlisting a wide range of societal actors in the researcher's interests.

Despite (or because of) the great influence of this position, there are reasons to be sceptical of whether the power of laboratories actually resides in their capability to dissolve boundaries and expand to society as a whole. In a recent study referring to Latour's claims, Matthias Gross (2015) uses the example of geothermal research to show that while some experiments are necessarily performed outside of laboratories, the process of justification of the knowledge gained still takes place inside laboratories. These laboratories are either established as a clearly delineated entity within, or completely removed from the field site. Moreover,

> the real-world experiment does not seem to be able to exist without the laboratory experiment. Indeed, all the novel experimental processes taking place in the world of engineering are then downscaled back into the closed world of the laboratory. (Gross, 2015: 9).

Gross also uses the notion of laboratory for a second type of justification process, namely the invited participation of stakeholders in discussing the risks and benefits of "real-world experiments" such as drilling projects. The reason why these stakeholder invitations can be seen as laboratories is because they tend to take place in a setting clearly removed from the everyday life of participants, and the outcomes of these debates are often without any practical consequences. Gross argues that in both cases, those aspects of the experiment that are seen as problematic are taken to the laboratory in order to be investigated separately. Unsettled knowledge claims that arise during the drilling project are solved through careful investigation under contained laboratory conditions, and questions concerning social and ecological responsibility are addressed during the stakeholder events. The laboratory can then assume a legitimising function, precisely because it offers the possibility to try out riskier aspects of the experiment without grave consequences. This does not mean, however, that the knowledge, once legitimised in the laboratory, cannot travel back into society. Re-interpreting Latour's case study, this is also exactly how Pasteur worked: He certainly *experimented* on the real world by introducing his vaccine, but not before he had tested out its functioning in the *contained space* of his Parisian *laboratory*.

In the more colloquial use of the word, "experiment" still refers to an activity which is open-ended and the results of which cannot be anticipated beforehand. This notion is also what Krohn and Weyer (1994) emphasised when they warned of the emergence of an "experimental society" some twenty years ago. Trial and error, success and failure (i.e. resistance and accommodation) according to them are part and parcel of scientific research, but one of the characteristics of modern science is that the failure of an experiment does not have any moral consequences - as long as the research process itself is "contained" within the community of scientists and, more specifically, within the walls of a laboratory. However, certain kinds of knowledge cannot be gained within the laboratory because they involve large-scale technological or ecological systems. Examples are knowledge about the (mal)functioning of a complex technology (as in nuclear power plants or defence systems) or the long-term effects of chemical substances and ecological hazards on the environment. Some interventions and technologies simply involve many *unknown* risks, even when all *known* risks have been duly assessed and pre-empted beforehand. Through the introduction of new technologies on a large scale, society as a whole is therefore increasingly involved in "experimental research", if this latter notion is defined in the following way:

> - the drawing up of a theoretical design and the formulation of hypotheses concerning the occurrence of (future) events;
>
> - the setting up of an experimental situation that can serve as a means of testing theoretical assumptions; this includes knowledge of the relevant boundary conditions, the controlled variation of parameters as well as the making of arrangements for the observation of effects;
>
> - the establishment of an organised research process embedded in a network of scientific institutions. (Krohn & Weyer, 1994: 175)

An experimental situation is present when all of the three factors are in place. Krohn and Weyer offer example cases for research that is experimental in nature, but clearly expands beyond the limits of the laboratory, such as missile testing, or the deliberate contamination of pastries with radioactive substances. Although irresponsible interventions of the latter sort are normally prohibited by law, some risks inherent in introducing new technologies can never be avoided even if they are legal and publicly known. This means that wherever experimental research oversteps the boundaries of the laboratory, scientists will probably find themselves confronted with new responsibilities and a moral accountability that is not usually a feature of experimental research in laboratories.

It is therefore useful to distinguish between laboratories and experiments also when describing processes on the societal level. While it makes sense to speak of "real-life experiments" or of an "experimental society" when referring to large-scale interventions with open outcomes that are nevertheless in some way designed to produce knowledge, the existence of such interventions does not make the places and societies affected a "laboratory". Based on the characteristics of laboratories discussed here, it would be more apt to simply refer to the location of such an intervention as a "field", as opposed to a laboratory.

To sum up, the characteristics of laboratories I would like to keep in mind are:

- *Placelessness*: Through a certain level of standardisation both of material and social conditions, laboratories (at least within the same discipline) achieve a level of sameness that renders the exact location of knowledge production irrelevant. Indeed, if idiosyncrasies of a laboratory are reported in a scientific publication, this actually raises doubts about the credibility of the results produced there (Gieryn, 2002). Only when background conditions are irrelevant can results be *generalised*. Placelessness is therefore a function of:

- A *known and controlled environment.* The laboratory works by introducing a systematic difference between *inside and outside*, between the stable, controlled conditions of the laboratory and the uncontrolled conditions of the field. By raising this boundary, the laboratory also offers the promise of *containment* and enables riskier research to be carried out without consequences.

- Most importantly, laboratories achieve epistemic advantages through a *reconfiguration* of both the natural and the social order, and by enculturating and disciplining both the research objects and the researchers.

3.3 Epistemic Cultures

As mentioned above, Knorr Cetina emphasises that the processes of differentiation and reconfiguration that make up a laboratory differ across research fields. Speaking of these differences, Knorr Cetina invokes the idea of the diversity of *epistemic cultures* as opposed to the idea that all (scientific) knowledge production is necessarily similar. On this account, high-energy physics is characterised by the processing of signs and traces rather than of the natural objects themselves. The collider experiments at CERN also exhibit great pre-occupation with the experimental apparatus and its functioning, the investigation and assurance of which makes up most of the epistemic labour in high-energy physics. In Knorr Cetina's words, this epistemic culture is therefore occupied with "the experiment itself, *with observing, controlling, improving, and understanding its components and processes*" (Knorr Cetina, 1999: 56, original emphasis). Molecular biology, in contrast, is marked by a direct engagement with the natural world and a movement away from signs and representations, towards the natural objects. Where particle physicists try to understand every aspect of their experiment and their apparatus, even if they are not of immediate relevance to experimental success, biologists rather solve the problems occurring during experimentation with the strategy of "blind variation and natural selection" (ibid. 91), meaning that several workarounds are tried out until one provides the desired outcome. The epistemic strategies of molecular biologists involve using the researcher's body as a sensory instrument and a repository for embodied knowledge, which enables the scientists to get a "feeling" for their objects and for workable solutions. High-energy physicists, on the contrary, employ epistemic strategies that focus on understanding the experiment and also understanding what *cannot* be known with the means at hand.

The differences in the "epistemic machineries" of the two research cultures are mirrored by differences in their "social machineries": The collider experiments at CERN are large organisations with several hundred to several thousand contributors. The individual scientist vanishes within the social organisation of the experiment, which is also expressed by the field's authorship conventions, where all contributors are named in alphabetical order on any publication resulting from the experiment. Molecular biology, in contrast, is still and necessarily a small-scale science, with individual laboratories (headed by a laboratory leader) always further divided into several individual projects. . The epistemic strategy of "blind variation" that is characteristic for this type of research requires the experimenter to become an instrument of research, an enhanced repository for embodied skill. In turn, the credit for scientific accomplishments is also given to the individual researcher. The authorship conventions of this field feature clearly defined rules of how to rank the authors of a collaborative publication in relation to their individual contribution to the project. This individualised reward system can result in struggles for authorship recognition and finds its continuation on the level of individual lab leaders and their competition for reputation, resources and personnel.

Epistemic cultures are therefore characterised and differentiated not only by how they engage with the natural world, but also by how they organise social structures around their specific epistemic goals and strategies. The specific form of how laboratories achieve a reconfiguration of both the natural and the social order can be described as one particular feature that contributes to and expresses an epistemic culture. In my description and analysis of laboratory experiments in experimental economics, I will consider them as *epistemic practices*, characteristic of an epistemic culture that also comprises institutions, organisations, technologies, conventions, attitudes and ways of relating to the empirical.

Yet epistemic cultures do not exist in isolation. Knorr Cetina (1999) herself refers to the study of *knowledge societies* as a motivation for studying *epistemic cultures*; understanding how (scientific) epistemic cultures work, on her view, will also help us to understand better how societies work that run on knowledge. Due to its focus on the internal characteristics of epistemic cultures, this approach cannot account for how external societal and institutional logics shape and influence scientific knowledge production. As contributions on research organisation and the interplay between science, science policy, and societal demands have shown, the rise of new types of management and the establishment of an "audit culture" also had profound impact on research communities and scientists' career paths (Power, 2011; Rushforth & de Rijcke, 2015; Müller, 2012). Auditing and evaluation of research increasingly use metrics such as the number and impact of journal publications a given research group or individual researcher has achieved in order to assess accountability and a transparent use of funding budget. At the same time, societal expectations of science – formulated in public funding plans and research policies, and nourished through various activities of science communication – are that research will yield useful applications or contribute to the common good in the long term. There are thus several conceptions of what makes "good science" and good research, which, when they become institutionally entrenched, might cause epistemic cultures to reflexively adopt the respective evaluative principles. An attention to how these evaluations are invoked when experimental economists describe their work thus opens up the perspective of epistemic cultures and points towards the overarching regimes of valuation in which individual researchers situate their work and their careers. It does, however, also open up room for reconstructing individual conceptions of what is worth doing and knowing in experimental economics that may be in conflict with more widely shared institutional logics. How things, people, and knowledge acquire value and the sites where different principles of assigning worth conflict is the focus of the recently evolved field of *valuation studies*. It is from this field that I gather the analytical concepts guiding the answer to my second research question.

3.4 Values in Practice

The idea that science should be "value-free" in order to be immune to the influence of changing political and moral orders (or abused for other, less universal and honourable ends than the search for truth) is traditionally implied with the notion that scientists should be "objective". This is not the place to discuss how the concept of objectivity has evolved and changed over time. Suffice to say, much research in sociology and history of science has shown that what we consider as the right values to guide scientific inquiry and produce reliable knowledge is contingent on social and historical formations, but this does not imply that the knowledge produced is not objective (Shapin & Schaffer, 1985; Daston & Galison, 2010). The "value-free" ideal of science has, for example, come under attack from feminist scholars who argue that instead of masking their moral and social commitments, scientists should actively communicate the values guiding their research and make sure that they are socially responsible (e.g. Anderson, 2004; Kourany, 2010). These discussions, however, mostly take place in philosophy of science and treat values as stable commitments that are guiding practice (such as the selection of methods and theory choice).[6] When values are understood as antecedents of practice, exploring and describing them becomes a purely conceptual, and therefore, philosophical endeavour.

In STS, the recent development of *valuation studies* (Kjellberg & Mallard, 2013; Lamont, 2012) has brought the question of values and evaluation into the focus of social research, not only, but also in sites of scientific knowledge production. Where STS scholars study values, they in contrast to philosophers do not conceptualise them as predefined and antecedent, but as enacted in practice. In other words, it is not so much the values that are studied, but the *practices of valuation* that bring them about. Considering values as the upshots of practices means that they cannot be invoked as explanations, but rather need to be explored. The interesting questions concerning values in science from the perspective of valuation studies then are how, in a given field, objectives, methods, facts and knowledge are ascribed worth, and how it is decided what is worth knowing in the first place (Dussauge, Helgesson, Lee, & Woolgar, 2015).

3.4.1 Negotiating values in situations of conflict

The recent interest in valuation is concerned with a much wider range of practices than those immediately involved in scientific knowledge production. Particularly interesting from this perspective are situations where different practices of valuation come into conflict, or where there is uncertainty over what is to be valued, and by whom. The editors of a recent collection on values in the life sciences refer to the latter as situations of "stakemaking", as they occurred after the nuclear disasters of Chernobyl and Fukushima:

> What is at stake is not a settled matter. People struggle to make sense of knowledge, companies struggle to survive, the government struggles to provide guidance and aid those who need it. […] Valuations of life, knowledge, and money become matters of concern. Whose assessments of radioactive fallout are valid? Whose evaluations of lives, quality of life, and livestock? Assessments of different values are intertwined: economic value, risk values, health values, quality of life. (Dussauge, Helgesson, Lee, & Woolgar, 2015: 11)

In such situations, it is neither clear what the concern is nor how it should be adequately assessed (Dussauge, Helgesson, & Lee, 2015). As a consequence, valuation practices are unsettled and need to be renegotiated simultaneously as expertise and authority is ascribed and established. Other sites for observing values and valuations in-the-making are highly "economised" sites such as different types of markets and accounting practices, but also regulatory agencies which shape value systems, and sites of social change (Kjellberg & Mallard, 2013). Particularly the relations and conflicts between what is usually described as economic value (i.e. value that translates into prices) and the plurality of other values is of interest to valuation studies. In the life sciences, the resolution of conflicts between

[6] For a recent overview on this debate, see: Douglas, Heather (2016). Values in Science. In *The Oxford Handbook of Philosophy of Science*, edited by Paul Humphreys. New York, NY: Oxford University Press. DOI:10.1093/oxfordhb/9780199368815.013.28

economic and "purely" scientific interests in medical trials through the adherence to "objective procedures" is an example, but also the construction of market devices for allocating scarce health care resources. Solutions to such conflicts present sites of value-making: what is fair, just, efficient, morally right, worth knowing and economic is not pre-defined and only established in the process (Dussauge, Helgesson, Lee, & Woolgar, 2015).

Within science, valuation practices can be routinely observed in the processes of *evaluating* scientific work in peer review. Lamont (2012) suggests that evaluative practices in peer review should be analysed particularly with regards to how they are shaped by the methods of comparison, by criteria, conventions, non-human supports and the self-formation of individuals, and whether these practises allow for a plurality or a rather narrow range of evaluative principles. She takes some inspiration from Knorr Cetina's study of epistemic cultures and their machineries of knowledge production, which according to Lamont "refer to the social and cultural structures that channel, constrain, define, and enable the production and evaluation of knowledge—indeed, such structures are both preconditions and constraints for the latter." (Lamont, 2012: 211). This rightly suggests that valuation practices are an integral part of much earlier phases in the "fact-making" process of knowledge production. By choosing one research topic over another, using one method rather than another, selecting or discarding data and presenting results, scientists just as routinely decide and establish what is worth knowing. As Helgesson, Lee and Lindén (2016) show, the development of a specific research design already involves a series of valuations. In their case study on textbook and journal articles about the use of randomised controlled trials and the comparatively new method of biomarker experiments in biomedical research, they reconstruct how different versions of what is considered "ethical", "economic" and "epistemically valuable" are articulated in descriptions of the two methods. Whereas for randomised controlled trials, genuine uncertainty about effectiveness of a treatment is a precondition for "ethical" research, biomarker experiments actively mobilise knowledge about how treatments may interact with individual characteristics in order to arrive at an "ethical" sorting of participants. What is seen as an economic use of resources and "clinical relevance" of outcomes similarly differs within these two types of experimental designs.

Different interests concerning the goal of a research project will also result in different "yardsticks" for evaluating research methodologies. Lee (2015) demonstrates that this is particularly the case in application-oriented research, where questions concerning the aim of the research and the right methodological approach cannot be disentangled from questions of science-industry-relations. Ultimately, these questions are therefore sorted out in negotiations settling funding and ownership, since industry-related and public funding bodies may pursue very different interests with the same line of research. What is at stake in these negotiations is the definition of "valuable" knowledge itself, how it should be produced and how it should be assessed – in short, "the question at stake is the negotiation, delineation, and coordination of several versions of 'Good Science'" (Lee, 2015: 210).

Yet also branches of science that are less application-oriented and therefore have little engagement with private industries are seen as increasingly subjected to a regime of evaluation that assesses people and products primarily in terms of their market performance (Burrows, 2012). As contemporary academia is undergoing a transformation process towards a more market-oriented logic of evaluation, the question of how worth is given to products and individuals, and how different orders of evaluation can be sustained has become ever more relevant and unsettled. In planning and conducting research, principles of valuation concerning research questions and methods become intertwined with what is perceived as the dominant principles guiding research evaluation (Rushforth & de Rijcke, 2015). Some observers fear that scientists may tend to choose research questions that are likely to attract greater funding, prefer less risky methods and procedures in order to have a greater chance of obtaining positive results, and write up their research in a way that is perceived as more appealing to potential peer reviewers (Lawrence, 2003). The valuation practices that scientists draw upon may be considered, as Lamont (2012) suggests, as an upshot of a specific epistemic culture and its machineries of knowledge production. Valuation practices that articulate epistemic values may also be described as potentially entering into conflicts with the institutional logics that ascribe worth to epistemic work based on performance indicators, applicability or even commercialisation prospects. It is these conflicts that the study of values as enacted and articulated in (epistemic) practices aims to make visible.

3.4.2 From "economies of worth" to "registers of valuing"

The recent interest in studying "orders of worth" was sparked by the work of Boltanski and Thèvenot (2006). In their elaborate study of the conventions that help resolve uncertainties and justify actions, they identify six different "common worlds", or economies of worth, each unified by a different principle of assigning worth and reaching agreement. Boltanski and Thèvenot show how these common worlds are invoked in handbooks of business management, and how they correspond to different conceptions of the common good and the well-ordered polity in political philosophy.

Other scholars took up their pragmatic approach of studying how different orders of worth are constituted in practices of evaluation. Economic sociologist David Stark (2009) in his analysis of three different firms came to the conclusion that a *heterarchy* of evaluative principles creates a type of uncertainty that is beneficial for innovation and entrepreneurism. Heterarchies are non-hierarchical orders of evaluative principles. Rather than enforcing "a single principle of evaluation as the only legitimate framework, they recognise that it is legitimate to articulate alternative conceptions of what is valuable, what is worthy, what counts" (Stark, 2009: 5). Heterarchies are not necessarily frictionless, but produce a "resourceful dissonance" (ibid. 6) that opens up new possibilities for action. Since heterarchical organisations tend to actively produce situations of uncertainty rather than suppressing them, evaluative principles, according to Stark, can be studied best by not only looking at the practice of valuations in specific cultural settings, but also by shifting the level of analysis from institutions to indeterminate *situations*.

STS scholars taking up the pragmatist orientation towards values subsequently shifted their focus away from identifying universal "grammars" of worth to the analysis of valuation practices in specific situations. An instructive example of this type of research is provided by Heuts and Mol (2013). In their study on how different actors involved in the production, selling, and consumption of tomatoes describe what a "good tomato" is, they consciously performed a re-orientation towards activities and the flexibility of assessments:

> "[We] shifted from talking about 'worth' (a quality) to foregrounding 'valuing' (an activity) and from 'economies' (that come with a single gradient each) to 'registers' (that indicate a shared relevance, while what is or isn't *good* in relation to this relevance may differ from one situation to another)." (Heuts & Mol, 2013: 129)

These *registers of valuing* are distinguished by a central concern – monetary value, for example, or handling tomatoes, or sensual appeal – in relation to which it is established what is good or bad in a certain situation. For example, tomatoes can be good to handle if they are juicy (for salads), but also bad to handle if they are juicy (for making sandwiches). Heuts and Mol find that the registers of valuing they identify are not neatly separated. Instead, they overlap, and often come into conflict or show internal tensions. Conflicts can be resolved by finding compromises, or by one register simply trumping another (customers might go for cheaper tomatoes even if the more expensive breeds taste better). Sometimes, tensions are apparently also irredeemable, such as the tension between a good tomato in terms of naturalness, and a good tomato in terms of taste. Good tomatoes, however, are not simply given, they are made: "valuing does not just have to do with the question how to appreciate reality as it is, but also with the question what is appropriate to do to improve things." (Heuts & Mol, 2013: 137). Assessing and improving – practices that are sometimes distinguished as evaluating and valorising – therefore are intertwined in situations where ascribing value is not a matter of distant judgements, but of hands-on work.

3.4.3 Evaluative principles and regimes of valuation

How, then, can valuation practices be studied in the context of scientific knowledge production? My own approach to investigate how values are made in science is taken from two recent studies on laboratory researchers in the life sciences. Fochler, Felt, and Müller (2016) study how PhD students and post-docs in the Austrian life-sciences ascribe worth to their own and other's work as they speak about their current situation and their future. In the context of life sciences (but this observation might easily be extended to academia in general), warnings of an over-reliance on quantitative methods of valuation that favour journal impact factors and numbers of citation over the quality of publications have recently been voiced by senior figures of the field. The authors therefore investigate whether such a narrowing down of evaluative repertoires is indeed observable in the accounts of life science researchers, and find that this is the case primarily for post-docs. Whereas PhD students enact a variety of *evaluative principles,* ascribing worth also in terms of how much others and themselves contribute to and care about the community of researchers, post-docs seemingly are restricted to a regime of valuation that favours individual productivity above everything else. Fochler, Felt and Müller introduce the notion of *regimes of valuation* to stress that certain practices of valuation are institutionalised and build on existing infrastructures, for example the bibliometric technologies used to assess the importance and "impact" of a publication:

> *[E]valuative principles* denote how worth is ascribed and argued for in a concrete situation, and *regimes of valuation* point to the broader discursive, material and institutional background this concrete evaluation draws on. We assume that regimes of valuation are comprised not only of institutionalized discourses, practices and material and digital infrastructures, but also of people living in, complying with and resisting these very regimes. (Fochler, Felt, & Müller, 2016: 180; original emphasis)

The authors aimed to reconstruct the different *evaluative principles* that their participants invoked when reflecting about what they value in working together, in making epistemic decisions and in planning their personal career. PhD students were found to value people and workplaces in terms of their sociability, honouring a helpful climate of mutual support, and following mainly their interests in choosing research topics and future career perspectives. Post-docs, on the other hand, saw themselves as having already chosen an exclusively scientific career, and therefore evaluated their workplaces, their co-workers and their own research mostly in terms of how well they could contribute to the success of this career. Their restriction to invoking one dominant *regime of valuation* can therefore be seen as a result of choosing a clearly defined, but nevertheless very risky path ahead, in contrast to PhD students whose future prospects are more open and diverse. That those pursuing an academic career find themselves restricted to a single regime of ascribing worth is also an expression of an institutionalised academic culture that assesses an individual researcher's worth (as in employability and credibility) almost exclusively in terms of their productivity, where the assessment of productivity is mainly based on quantitative techniques. The authors conclude that this narrow regime of valuation that junior researchers in the life sciences grow into may have potentially harmful long-term consequences, both for individual researchers and the scientific community. Researchers who subject all biographical choices to the single goal of individual productivity are left without alternatives to following an academic career, even though not all of them are likely to succeed. This might discourage younger researchers who are not willing to sacrifice other aspects of life for their careers, and may result in a scientific workforce that is even less diverse and less socially cohesive than it is now. Researchers' need to focus on publications may also result in choosing research topics that are seen as more likely to succeed, thus negatively affecting the quality and innovativeness of research, and also increasingly detaching the notion of successful science from that of social relevance and responsibility.

Fochler (2016) refers to the current cultural framework in which researchers accumulate worth through their productivity as *epistemic capitalism*. More specifically, the term denotes “the accumulation of capital, as worth made durable, through the act of doing research, both in and beyond academia.” (Fochler, 2016: 924; original emphasis). Central to the notion is the *accumulation* of worth (beyond the *production* and *ascription* of worth as in regimes of valuation), which is then reinvested in a “cycle of credit” to produce more worth, where “worth” need not automatically translate into monetary value. The notion of epistemic capitalism ties into the broader one of an emerging or existing *knowledge economy*, in which knowledge is increasingly commodified, but also seen as an asset to produce more value, and as a driving force of economic activities. While sociological research often fouses on the effects this transformation towards knowledge-based economies has on industries, working conditions and inequality (e.g. Powell & Snellman, 2004), researchers in STS have pointed out that it is also affecting the institutions and conditions of knowledge production itself. One of the transformations is the institutionalisation of a specific, productivity-oriented regime of valuation as discussed above, but also the development of regimes of accumulation as described by Fochler (2016). An example of such a regime is the employability of researchers, which, when seen as a function of their individual output in terms of publications, will lead them to avoid risks and thus forego research projects that are considered as more interesting, but less publishable, or avoid questions and methods where success is uncertain. Within a framework where knowledge production is evaluated in terms of how well it contributes to the accumulation of a specific form of (symbolic or monetary) capital, chances then are that the epistemic value of research is diminished in favour of other forms of worth.

Fochler (2016) and Fochler, Felt and Müller (2016) based their analysis on biographical interviews in which participants where prompted to reflect on specific instances and interrelations in their professional development. My own approach in interviewing was less biographical than problem-centred, but I likewise prompted my respondents to reflect on specific decisions and choices they made in the process of developing and conducting a laboratory experiment. Building on the literature cited above, I will aim to reconstruct the evaluative principles that experimental economists invoke in specific situations during this process, but also in terms of how they ascribe worth to their own and other’s work, and to the discipline as a whole.

4 My Case Study

This chapter is meant to give an overview of my field and the methods I used to approach it. Since the methodological principles of experimental economics will be referred to at different points throughout my analysis, I begin with a short overview of the most commonly discussed points. I will also try to clarify the relation between experimental and behavioural economics. The rest of the chapter is a reflective account of my fieldwork and analysis, and of how the observations I discuss in the empirical part gradually took shape.

4.1 The Methodology of Experimental Economics

The methodological principles that contemporary research in experimental economics builds on, including all the laboratory experiments discussed in my interviews, were formulated by Vernon Smith (1976, 1982). Smith himself had been inspired to build experimental markets by the classroom experiments of his teacher Edward Chamberlin at Harvard University in the 1940s. As an assistant professor, Smith conducted a series of market experiments to examine the predictions of competitive equilibrium theory. In these experiments, Smith modified Chamberlin's approach by using the design of a double auction in which bids and offers are immediately made public, and he repeated the interaction for several rounds (Friedman & Sunder, 1994, chapter 9). In contrast to Chamberlin's classroom experiments, Smith's results confirmed the predictions of competitive market theory: over time, the prices and final allocations converged to the competitive equilibrium. These experiments were published in 1962 and are commonly referred to as the beginning of experimental economics as a research programme. It took, however, the influence of research in experimental psychology and the development of new welfare economics with its focus on the design of resource allocation mechanisms for the methodology of experimental economics to take shape. From the bargaining experiments pioneered by the psychologist Sidney Siegel, Smith adopted the idea that participants' choices in the experiment should be connected to monetary rewards (Friedman & Sunder, 1994, chapter 9). From theories of mechanism design in welfare economics, Smith adopted a view of microeconomic systems as consisting of environments (the agents, their preferences and a set of commodities) and institutions (the rules according to which commodities can be exchanged and owned) that can also be realised in laboratory experiments (Santos, 2010, chapter 7). Smith's laboratory experiments according to Santos could even be presented as complementary to theories of mechanism design, in that they tested institutions for the allocation of commodities (i.e., market mechanisms) for their performance and efficiency in laboratory conditions. This alignment with a well-established research programme in turn lent some credibility to the new method of laboratory experiments: "The point was simply that experimenters are only doing what theoreticians do by different means." (Santos, 2010: 87).

Smith's early market experiments had not used monetary incentives, but this later became a standard, if not mandatory practice in experimental economics. The reasons for this practice were first outlined in Smith's "induced value theory" (1976):

> Control is the essence of experimental methodology, and in experimental exchange studies it is important that one be able to state that, as between two experiments, individual values (e.g., demand or supply) either do or do not differ in a specified way. Such control can be achieved by using a reward structure to induce prescribed monetary value on actions." (Smith, 1976: 275).

By distinguishing two otherwise identical choices through the reward that each of them yields, the experimenter can be sure that subjects will prefer the choice with the higher reward. This simple mechanism of control depends on some important "precepts" that Smith discussed in greater detail in a later publication (1982):

- *Nonsatiation* requires that subjects are not indifferent to the reward medium and always prefer more of it over less of it.
- *Saliency* describes subjects' awareness that their rewards depend on their own and others' actions in the experiment according to the specified rules, i.e. that the experimental outcomes will be translated into a corresponding amount of the reward medium.
- *Dominance* means that the reward structure is sufficiently strong to crowd out other motivations that might influence subjects' decision-making, including the subjective costs of participating in the experiment (i.e. the reward structure motivates subjects to make an effort despite getting tired or bored).
- *Privacy* ensures that subjects are not influenced by considerations concerning the outcomes of other subjects, because they only know about their own preferences and payoffs.

(cf. Friedman & Sunder, 1994, chapter 2; Santos, 2010, chapter 3; Guala, 2005, chapter 11)

These four precepts, according to Smith, ensure that the experimenter controls the preferences of her subjects through the reward structure specified in the rules of the experiment. In principle, the reward medium can be anything that satisfies the precepts, but the most straightforward medium to use is the local currency. For the purpose of testing (or falsifying) the predictions of economic theories, satisfying these four precepts should be enough. They provide for a controlled experiment that "consists of a far richer and more complex set of circumstances than is parameterized in our theories" (Smith, 1982: 936) because the decisions are made by real human agents.

Santos (2010) adds that what she refers to as the "instrumental model" of experiments in economics consists of these precepts along with a set of principles such as *simplicity* (in order to make the experiment easier to understand for subjects), the abstract and neutral formulation of the experimental task (to keep subjects' individual interpretations to a minimum), and preserving *anonymity*. Also, there is a strong conviction that deceiving subjects about the actual purpose of an experimental task violates the standards of good methodological practice.[7] Along with the use of monetary incentives, this latter requirement is what experimental economists themselves often cite as distinguishing their experiments from those of psychologists. One of my participants, for example, explained that the danger of deceiving subjects lies in losing control immediately (because deception makes it impossible to know how subjects interpret the instructions) and on the long run (because subjects may distrust and second-guess experimenters' instructions in the future). These methodological requirements have become fairly entrenched through textbooks (e.g. Friedman & Sunder, 1994), but also because they were increasingly used as criteria for publishing experimental studies, as this excerpt from the editorial to the first edition of the journal *Experimental Economics* exemplifies:

> Papers must meet certain high standards in terms of methodology. For most economic issues, it is important to provide subjects with real financial incentives to make careful decisions. Also, any deception should be carefully explained. Procedures should be reported in a manner that permits a replication that would be accepted by the authors as being valid. (Schram & Holt, 1998: 5)

Indeed, the provision of detailed accounts of the experimental procedure in publications, often along with the original experimental instructions as an appendix, is another characteristic of experimental economics research.

Entrenched as they may be, the methodological principles described above are not undisputed. According to the authors of a detailed examination of these principles, there actually are "ongoing controversies within experimental economics about what its methodology should be" (Bardsley et al., 2010: 332). Particularly the use of monetary incentives to control for subjects' preferences has been disputed for (behavioural) experiments which aim to study exactly those preferences (cf. Guala, 2005, 2007). A different question concerns the validity of results that were generated in artificial laboratory settings with student populations for contexts outside of the laboratory. While some experimental economists are strongly advocating for the value of field experiments with "natural" populations to produce results that are more "generalizable" (Levitt & List, 2015), the tenor seems to be that field experiments, too, have their shortcomings and that ideally, different methods should be combined to arrive at valid insights (Camerer, 2015; Falk & Heckman, 2009).

---

[7] Deception was mentioned to me on several occasions as a common characteristic of experiments in social psychology, and indeed, a number of particularly famous experiments, such as Milgram's experiments on obedience to authorities, are based on participants' ignorance about what the actual purpose of the experimental task is. There are some examples of experimental economics research which uses deception, such as the first experiment I assisted in that was presented as an opportunity for students to earn a little money by helping to prepare a mass mailing, but actually measured their prejudice against working with someone from a different ethnic background. Field experiments, which are meant to study whether behaviour in the lab corresponds to behaviour in the field (for example, whether participants who behave cooperatively in the lab are also more likely to return a "misdirected letter") also sometimes use deception, e.g. Stoop, J. (2014). From the lab to the field: envelopes, dictators and manners. *Experimental Economics*, *17*(2), 304–313.

4.1.1 Purposes and types of experimental studies

The heterogeneity of views on methodological principles within experimental economics may, as Guala (2005) also suggests, very well be the result of the heterogeneity of purposes for experimental research. A common distinction of experimental research in economics classifies experiments according to the theoretical tradition they are based on. According to this distinction, experiments typically either explore market mechanisms, strategic interaction between two or more individuals (based on game-theory), or individual decision-making. Another influential categorisation was initially proposed and recently revisited by Alvin Roth (2015) and sorts experimental research according to its aims into theory-testing ("speaking to theorists"), finding and describing unexplained empirical regularities ("searching for facts") and developing and testing policy interventions such as markets or auctions for the allocation of scarce goods ("whispering in the ears of princes"). Guala (2007) likewise distinguishes experiments according to their purposes, but collapses the testing of theory and the building of new theories based on observed regularities into the category of a "testing" logic of experiments. The "building" logic, in contrast, uses experiments to design interventions with real-world applicability. For Guala, this second approach, which allows economists to "do things with experiments" should be the main motivation for doing laboratory experiments. The most famous example of this "building" approach of experimental economics are the auctions for wireless personal communication spectrum licences commissioned by the US Congress in the early 1990s. The market mechanism was designed by game theorists, but many details were worked out by using laboratory experiments as "test beds" for the auctions, which were generally considered highly successful due to the high revenues they eventually produced for the federal administration.[8] Because such experiments are testing a "technology" in order to find out under which conditions a certain market outcome can be achieved, Santos (2010) also refers to experiments in market design as "technological experiments". Both their purpose and their experimental design, which restricts participants' decisions in important ways, distinguish them from "behavioural experiments" in which decision-making and strategic interaction among individuals are explored.

The distinction between experimental economics and behavioural economics in general is probably more familiar, but not always easy to draw due to institutional and methodological overlaps. Behavioural economics has its origin in psychology, and started out as an investigation of the systematic biases human beings exhibit in decision-making. In other words, behavioural economists study the situations in which actual human behaviour departs from the predictions of economic theory, and often proposes alternative theories to describe these phenomena. The research of Tversky and Kahnemann

8 Ever since, these auctions have also become a focal point for philosophical debate on experimental economics. More specifically, the role of experimental economists and game theorists in the process has been questioned, as has the perceived success of the auctions, since many of the political goals set by the US Congress – for example, allowing smaller businesses to buy some of the licences – were actually not met. See Reiss, 2008, and the contribution by Edward Nik-Khah and Philip Mirowski in MacKenzie, Muniesa, & Siu, 2007.

(1979), for example, showed that the predictions of expected utility theory were consistently violated by subjects choosing between two options (one being a lottery and one a certain gain). This led them to develop an alternative theory of decision-making under risk that focuses on perceived gains and losses rather than on the ordering of probabilities ("prospect theory"). These results were based on hypothetical questions, since subjects did not actually receive the monetary rewards corresponding to their choices. Since then, behaviour that systematically deviates from the predictions of standard economic theory has also been investigated extensively in incentivised laboratory experiments following the methodological principles above.

To illustrate this kind of research, research on *public goods games* may serve as an example (cf. Ledyard, 1995). Public goods games are situations in which a group of people are individually presented with an endowment and the choice to either contribute some or all of this endowment to a public good, or keep all of it for themselves. The sum of all the contributions in the group is then doubled up and divided between the group members. In other words, in groups of four, each member earns one half of the total sum of the group's contributions. This means that the strategy maximising the overall profit would be for each member of the group to contribute everything: If there are four members in the group who each contribute a total of 20 tokens and the total contributions are multiplied by 0.5, then each individual earns 40 tokens. The optimal strategy to maximise one's individual income, however, is to keep one's endowment and additionally profit from the contributions of other members: if all but one member of the group contribute 20 tokens, then the free-rider will earn 50, and the others will earn 30. Due to this strong incentive to free-ride on other's contributions, game theory predicts that none of the group members will contribute anything to the public good. For every token they contribute, an individual only gets half a token back. Since economic agents are expected to always maximise their own income, not contributing would be the rational thing to do in order not to get ripped off. In game-theoretic terms, not contributing is the "dominant strategy". Experimental research on public goods games, in contrast, shows that in laboratory situations, people consistently contribute some of their endowment to the public good, that contributions decrease when the game is repeated over time (usually with a sharp decline in the last round), that the possibility to punish free-riders helps to keep contributions on a high level, and that people are willing to punish others even when this comes at a cost. These persistent results are known as the "stylised facts" of public goods games, and they are an example of the empirical regularities contradicting theoretical predictions that have been observed in many game-theory based experiments. What these results show is that the preferences of experimental subjects are not perfectly controlled in these situations, and that their decisions must be based on other motivations than the incentive structure imposed by the experimenter. One influential attempt to formalise the "social" preferences that play a role here is the theory of "inequity aversion" (Fehr & Schmidt, 1999), which proposes to model subjects' preferences based on the assumption that some people strongly dislike being worse or better off than others; they are therefore motivated more by fairness considerations than by monetary incentives.[9]

This type of research – laboratory experiments on phenomena that run counter to the predictions of economic theory, and the development of new theoretical models to explain and accommodate these phenomena – is often done by economists, who work at economics departments, and publish in economics journals, even though they might be often willing to collaborate with psychologists. The research programme of behavioural economics spans across the disciplines of economics and psychology and still also involves other methods than laboratory experiments. For this reason, I agree with the suggestion "to define all forms of experimental research in economics as 'experimental economics' and to use the term 'behavioral economics' to refer to work, whether experimental or not, that uses psychological hypotheses to explain economic behaviour." (Bardsley et al., 2010: 2). Behavioural economics research has acquired considerable influence in policy-making in recent years, based on the insight that small differences in the way choices are presented have a significant impact on people's decision-making. Knowing these differences can therefore be used to develop "choice architectures" that "nudge" citizens towards choosing a healthier diet or saving up more for retirement without coercing them to do so – a policy orientation that its inventors have dubbed "liberal paternalism" (Thaler & Sunstein, 2008). A recent report by the newly created *Foresight and Behavioural Insight Unit* of the European Commission (Lourenço, Ciriolo, Almeida, & Troussard, 2016) shows that the use of „behavioural insights" in policy-making is actively encouraged on a European Level, as was the case in the US during the Obama administration. Some of the insights gained from laboratory experiments as those described here are therefore increasingly being used as resources for „evidence-based policy" in government and public administration.

4.2 Approaching My Field: The *Vienna Center for Experimental Economics*

The *Vienna Center for Experimental Economics* (VCEE) was established in 2011 as a research unit at the University of Vienna's Faculty of Business, Economics and Statistics. Its two directors, Wieland Müller and Jean-Robert Tyran, had shortly before been appointed as professors in the department of economics. The two directors come from rather different traditions within experimental economics, with Tyran, a former student of Ernst Fehr, having a background in behaviourally oriented research, and Müller in game theory and industrial organisation. The VCEE as a whole consequently unites a

[9] As always, the whole story is a bit more complicated than this very simplified account. As Ledyard (1995) describes, what makes the results of public goods games so puzzling is that a majority of participants acts neither completely pro-social nor completely selfish. Also, the particular behaviour observed seems to be influenced by many different aspects – for example, the gender of participants, whether there is communication, the amount of the endowment and the value of other experimental parameters, and so forth. Guala (2005, chapter 2) describes the intricacies of running a public goods experiment based on his own experience. That the research programme on public goods games is ongoing probably shows that all of these intricacies still have not been completely understood and explored.

variety of different approaches to experimental economics that were also reflected in my interviews. At the time that my fieldwork was conducted, during the summer term of 2016, the team of the VCEE apart from the professors consisted of 6 post-doctoral researchers, one of them being the lab manager, and 8 PhD students.

The founding of the VCEE, along with the furnishing of a laboratory at the university campus, was enabled by an initial grant by the University of Vienna. The VCEE has since also acquired a considerable amount of third party funding through different research grants (Vienna Center for Experimental Economics, 2016). While there is no dedicated degree programme in experimental economics, VCEE members teach courses on topics and methods in experimental research for undergraduate and graduate students in economics. It is, for example, possible to do an experimental master's thesis in economics, and the different consecutive courses offered are designed to prepare students who want to do so. Additionally, the VCEE regularly invites guest lecturers and hosts workshops and events. Since the autumn of 2016, this is also done through the newly established *Vienna Behavioural Economics Network*, which explicitly aims at promoting "the dialogue between researchers in behavioral economics and stakeholders in the economy, the public administration, nonprofit organisations and interest groups, and society at large."[10] The participation in this network indicates not only that the VCEE sees its mission in going beyond the academic realm, it also highlights the strong interest of some of its faculty in behavioural research. As was mentioned before, the line between behavioural and experimental economics is often very difficult to draw, particularly on the level of academic institutions where both research programmes meet. In the context of my analysis, it is perhaps most important to note that the VCEE is a research unit clearly affiliated with economics (not psychology), and that all the researchers I interviewed were trained economists who sought to publish in economics journals and oriented their career path along the requirements of the academic discipline of economics.

4.2.1 Preparing and adjusting a fieldwork strategy

I first encountered experimental economics in a course offered at the department of philosophy at the University of Vienna, which was co-taught by a professor of philosophy and a professor of experimental economics. The course reflected on the methodology of experimental economics with viewpoints both from experimental economists and philosophers engaging with the topic. At the time, I had already acquired some background in the philosophy of economics, and my interest in this general topic was the reason why I decided to attend the course. At the end of the semester, I started thinking whether experimental economics might actually be a suitable topic for a master's thesis in STS. I had engaged with the literature on economics in STS before, but found very little on the actual research practices of economists; in general, I felt that the social sciences were poorly covered by STS research. Experimental economics, however, was not only a social science, but a science that had only

---

10 Vienna Center for Experimental Economics. (2017) Vienna Behavioral Economics Network. Retrieved from http://vcee.univie.ac.at/vben/ , 15.04.2017.

recently moved to the laboratory. Studying experimental economics therefore seemed like a promising topic because it presented a recently established research discipline, and because it lent itself to a particular approach that had been one of the building blocks of the discipline of science studies, namely a laboratory ethnography. My first idea for the fieldwork was to try to follow a research project from its development stage to the laboratory sessions and the data analysis.

In order to prepare for my research project, I attended the introductory lecture on experimental and behavioural economics for master's level students in economics. This was an intense, four-hours-a-week course with a mid-term and an end-term exam. During that semester, I approached the lecturer and head of the VCEE, Jean-Robert Tyran, to ask whether and how I could do a case study on his research department for my master's thesis. He suggested that I apply for the position of a lab assistant, which I did. To my first inquiry, the lab manager kindly responded that they had enough personnel at the moment, but that I should send a CV and he would keep it on file in case there were any openings. This made my plan of doing a laboratory ethnography with participant observation seem much more unrealistic of a sudden, since it was unclear whether I would ever be able to actually access the lab. Additionally, Professor Tyran on one occasion introduced me to one of the younger researchers as a potential interviewee, who was rather sceptical of my idea of observing the day-to-day work of the researchers or attending their meetings. The reason for this was that he did not think that there was much to observe for me; for example, even in research collaborations, many discussions were taking place in a more informal and spontaneous manner than in specifically prepared meetings. Also, I realised that the preparatory phases for experiments can take very long, too long for me to observe the development and implementation of one particular project. At this point, I also discovered that an ethnography might not be the best approach, given that the practices involved in the planning and design of experiments that I was interested in would most likely not even be observable. Fortunately, I was not the first one to run into these problems.

The difficulties of applying the ethnographic approach of laboratory studies in STS to social science research were described by Lisa Garforth (2012). According to her, many of those difficulties may be attributed to the predominance of a rhetoric of "witnessing" in science studies, both in terms of content (the focus of the classic laboratory studies was often scientists' *observational practices*) and approach (laboratory studies would typically describe the *visible* practices and components of laboratory work). When research practices are solitary and do not involve any particular apparatus or materials, they in turn become invisible and thus unobservable to the traditional ethnographer. Much social science work is of this sort; it is often carried out by the researcher being on their own, at their desk or computer. When Garforth attempted to observe what social scientists were doing in their offices, she “experienced [her] presence there as intrusive and disruptive” (Garforth, 2012: 274). In other words, the cognitive labour of social scientists does not lend itself very well to participant observation.[11]

There were some attempts to study this solitary cognitive labour in earlier STS research. For example, Knorr Cetina and Merz (1997) applied the ethnographic approach to what they call the “thinking science” of theoretical physicists. The sociologists Yonay and Breslau (2006) referred to this idea of a “thinking science” in their own study on mathematical economists and their modelling practices. The two authors followed an ethnographic approach, but supplemented their observations of research colloquia and other public activities at economics departments with in-depth interviews on their participants’ work. These interviews would particularly focused on a piece of recently concluded research which:

> We asked our subjects to give us drafts of recently completed papers and read them very closely before interviewing the authors about the process of writing those drafts. The interviews included questions about the academic experiences of the interviewees, their careers, their work, and the social features of their fields and subfields. (Yonay & Breslau, 2006: 353).

The bulk of the interviews was formed by conversations about the research papers that the respondents had provided. Yonay and Breslau would ask them how they became interested in the topic and what they had known about it before they started the research project, and then ask questions about the considerations guiding the process of theoretical modelling, for example, which actors and variables were included in the model and why. In this way, the authors hoped to arrive at a detailed reconstruction of the reasoning behind theoretical economists’ conceptual practices. As they explain in a footnote, Yonay and Breslau are aware that retrospective accounts may provide only an idealised or even distorted image of scientific practices, but this does not run counter to their epistemic interest, as their goal was “to understand a way of thinking, not to reconstruct how actual projects have evolved.” (ibid.). They saw interviews as a method to elucidate the style of reasoning and the conceptual practices that are embedded in the epistemic culture of mathematical economics, and therefore as an apt way to arrive at an understanding of this epistemic culture. Since I was motivated by a similar research interest, only with regard to experimental economists, I found this approach immediately appealing.

4.2.2 Interviewing

Relieved that there seemed to be a viable alternative to doing a classic laboratory ethnography that actually might be more suitable for my own research interests, I decided to follow the approach described by Yonay & Breslau (2006). I asked my respondents to provide me with a description of one of their recent research projects (a draft paper, a presentation or a proposal, for example), on the basis of which I would prepare the interview questions. The interview guideline I developed was loosely

[11] Garforth actually goes one step further and argues that the insistence on making science observable that is implicit in laboratory studies is not so different from the insistence on measuring and assessing the „output“ of research in contemporary evaluation systems – in both cases, invisible work does not really count.

based on the recommendations given by Bogner, Littig & Menz (2014) for preparing and conducting expert interviews. Where expert interviews, as in my case, aim for the interpretative (as opposed to technical or procedural) knowledge of the expert and thus address her as representative member of a certain expert group, and when those interviews form the basis of the research rather than a first exploration of the field, Bogner, Littig and Menz (2014) speak of "theory-generating interviews". These kinds of interviews do not require a standardised guideline that needs to be strictly followed in every case, since the questions of the interviewer are meant to induce the respondent to freely narrate and reflect. The function of the guideline is therefore mostly to prepare the interviewer for the interview situation and to serve as a memory aid. For my own approach, I combined the outline for theory-generating expert interviews provided by Bogner, Littig & Menz (2014) with the different questions types defined for ethnographic interviews by Spradley (1979). More specifically, I prepared a guideline with several thematic clusters or "theme blocks" and related sub-questions (as recommended by Bogner, Littig & Menz), and made sure to include some questions that are more conducive to narration and detailed explanations, such as "grand-tour questions" and "example questions" (as recommended by Spradley). This guideline was once formulated very generally, to lay out the themes I wanted to address, and then adapted to each interview and participant individually, based on the preparatory material I received. An example of an adapted interview guideline can be found in the appendix. Through the formulation of my questions and the introduction of my research project, I presented myself as a layperson interested in the methodology of experiments. Despite similar preparations, the interview situations differed greatly and I was lucky that the first interviews I conducted were also the easiest ones. All of the interviews with postdoctoral researchers took place at their offices (except for the first one, which we did in the common room of the department), but the interviews with the two PhD students were conducted outside the department during extended lunch breaks. Perhaps unsurprisingly, the interviews became more challenging, the older and more established my interview partners were. My self-awareness as a young, female, social researcher made it certainly more difficult to present myself as a serious researcher in those situations, and the initial scepticism of some of my interview partners towards my methodological approach ("how is a study based on a few interviews going to be representative?") additionally highlighted this asymmetrical relationship.

My strategy for finding interview partners was to start from the junior researcher that I had already been introduced to, and ask him as well as Professor Tyran for recommendations whom to approach next. I would contact potential interviewees with an email including a short one-page description of my research project and how I planned to conduct interviews. The considerations guiding my sample selections were to allow for diversity in terms of career stages and approaches. In those cases where an interview could be arranged, I asked for some written material on a research project the researcher wanted to discuss (in most cases, the researcher let me choose the project myself). I prepared a sheet informing my respondents that I would record and transcribe the interview, that it would only be used in the context of my master's thesis and that their responses would be anonymised. It was also mentioned that full anonymisation might not be attained, given the small number of participants and their shared institutional background. Also, there was only a very small number of female researchers to approach as potential interview respondents. As in many other academic disciplines, the number of women in economics decreases dramatically from the undergraduate level to the ranks of professorship, and this is reflected in the number of women working at the VCEE. I did interview both male and female researchers, and the gender ratio of my interview sample very roughly corresponded the gender ratio of the VCEE staff. Since identifying the gender of a respondent along with their career stage would have consequently made it very easy to identify particularly the female researchers, I decided to use gender-neutral pronouns in the text of my thesis. Throughout the text, participants are only identified by a number (P1, P3, P4…) that corresponds to the chronological order in which the interviews were conducted. My respondents were asked to sign the informed consent sheet before the interview started, and one copy of the sheet remained with them.

All in all, I conducted seven interviews between March and June 2016. The interviews lasted between 45 minutes and two hours. One of these interviews (P2) was not transcribed and completely omitted from the analysis, because the respondent was a PhD student in sociology, not economics. It was, however, a useful exercise and also helped to clarify several questions that came up during my fieldwork. Another interview was very informative, but I was not allowed to record it and had to take notes instead (apparently, I had not communicated my need to record the interview very clearly when arranging the meeting). All of the remaining interviews were recorded and transcribed in verbatim transcription style. They were then uploaded to the Atlas.ti software and coded inductively following the principles of *Grounded Theory* (Charmaz, 2006). During the term of my fieldwork, I did not attend any classes at the department of economics, but I attended a research seminar, a workshop on experimental sociology where two of my interviewees presented their work, and I interviewed two guest lecturers, Marco Piovesan and Arthur Schram. These latter interviews for me were again useful resources for contextualising the research I discussed with my respondents. I also received some additional recommendations for literature on theory and methodology during these interviews, but they were not used as material for my analysis.

4.2.3 Participant observation

Around the same time as I started preparing for my first interview, I received an email from the lab manager that they actually needed another assistant. My first assignment was to help out in an experiment that was conducted during the Easter break. This experiment was very different from the research usually conducted at the VCEE, because it involved a considerable amount of deception. Students were hired via the VCEE database, but asked to help out with preparing a mass mailing of informational flyers to the students of the faculty. They would be paid according to the amount of letters they were able to prepare within an hour. The experiment itself took place at the department, where two classrooms had been rearranged for that purpose. Several months after helping to run this experiment, I found out that it actually studied how willing students were to agree to work with somebody who had a foreign-sounding name. The experiment was a follow-up session to a collaborative project that one of the researchers had started earlier at a different university, and as he later ensured me, it was not very representative neither of his own work nor of the research normally done at the VCEE.

As I started working in actual laboratory experiments, my methodological approach to my research project changed yet again. All of a sudden, I had been granted much more field access than I had hoped for, and the conversations with other lab assistants and the lab manager proved to be very insightful. I therefore became much more sensitised to the practical aspects of running experiments, the importance of having smooth and routine procedures and clear guidelines, and of performing rather menial, but necessary tasks (such as counting money) with diligence and care. In total, I assisted in laboratory experiments seven times between the end of March and the end of June, sometimes only for one or two sessions, sometimes for an entire workday. In all of the sessions, there was at least another experienced lab assistant present. On some occasions, there would be two of us along with the experimenter himself, or two of us along with the lab manager. For one of the experiments, there was a separate training session, because the experiment required participants to "bring a friend" and involved an unusual and elaborate matching process that had to be explained beforehand. I was paid for every hour I spent at the laboratory just like the other laboratory assistants, but I soon noticed that I was not considered to be a laboratory assistant like every other. The obvious reason for this was that I had told everyone involved that I was doing research on the laboratory, and this was also visible since I was constantly taking notes. The other reason that I presume played a role in how I was approached by the staff of the laboratory was that I did not have a substantive background in economics and for that reason could not be completely trusted with running an experiment and also explaining it to participants if there were questions. In hindsight, this latter concern was probably quite justified. I remember one experiment during which I tried to answer some questions at the beginning, but in the end had to ask my colleague to answer them. I was assigned responsibility for an experiment only once, when I was asked to run an experiment together with the researcher, which meant that I would need to take care of all the administrative procedures and the bookkeeping. Even in this case, another lab assistant was present, although she was just casually spending time at the laboratory and doing her homework there (this was not very unusual; since some of the assistants are good friends, they would sometimes visit each other at the lab and spend time there even when they were not "on duty"). I ended up making several mistakes on this occasion; I came a little late, was very stressed during the preparations and managed to mix up the seating cards for the extras with those of the other participants. In the end, everything went well, of course, also because the researcher helped me with some of the standard procedures, but I was not assigned sole responsibility for an experiment again.

The lab manager also in some cases assigned me to experiments that he thought would be "interesting" to me and consequently spared me those that were considered less interesting. This was probably the reason why I never participated in the pre-testing of an experiment, since both the lab manager and the assistants considered this to be particularly boring and tedious work.[12] Ironically, helping in a pre-test might actually have been very interesting for me, and I unfortunately only figured out that I was deliberately left out of this common practice very late in my field work. This aspect exemplifies the ambiguous position I took on while working as a lab assistant: Officially, I was treated like the other lab assistants. Unofficially, I received a special treatment. I was assigned to experiments were I was not actually needed because the experiments were thought to be interesting for me, and was then also compensated for these hours. Due to my lack of experience, I was not assigned to very difficult or tedious tasks and at every single occasion, there would be another experienced lab assistant around in case I needed help. It was obvious that I lacked the background in economics that all the other lab assistants had, and my own research and approach were met with curiosity and sometimes even created astonishment or irritation (for example, when I wrote down a comment about the phenomenon of vanishing pens after very boring experiments). At the time, I did not mind so much, but I worried of course that my being at the lab was actually more of a burden than a help.[13]

4.2.4 Analysis

The analysis of my material then took several months. Given that my focus developed and changed during the fieldwork and that the first interviews I conducted influenced the questions I would ask in the later ones, my fieldwork and analysis process could be described as following an "iterative-inductive" logic (O'Reilly, 2005). For example, the idea that an experiment should ideally be based on a theoretical model was something I had not encountered before (transcribing and coding) my first interview, and it then became one of the aspects I would pay particular attention to in all the following interviews. Over the course of my fieldwork and my analysis, my research questions also changed several times, as new interesting aspects and related questions arose from my engagement with the data. Before starting my fieldwork, I had been particularly interested in what could be described as "translation processes": how do experimental economists "translate" a real-world problem into a laboratory experiment? And how do they translate the results back into an insight about economic behaviour? In other words, I was curious about the relationships between laboratory experiments and real-world contexts, and how experimental economists might defend their research against the accusations that it is "unrealistic". Through the interviews, I gradually understood that laboratory experiments are rarely intended to map real-world problems directly, and I became interested in the reasons

12 Garforth (2012) reports the similar experience that bioscience researchers were reluctant to involve her, as a participant observer, in work that they considered to be routine and „boring".
13 The fact that I was not included officially as a student assistant in the annual report of the VCEE (Vienna Center for Experimental Economics, 2016) seems to underwrite this concern. However, I was officially credited as an assistant in the paper publishing the results of one of the experiments I assisted in.

and explanations why this is so. The final research questions that are discussed below emerged after I had left the field and started concentrating on coding the interview transcripts. Having worked through all the interviews, I realised that there was considerable heterogeneity in how individual researchers describe the properties of a "good experiment", and I decided to make the evaluation of experiments the second focus of my analysis. After a first round of instance-to-instance initial coding, I began to group my codes according to different phases in the experiment and the strategies recurrently described. During the following rounds of focused coding I developed the analytical categories to describe my observations more theoretically (Charmaz, 2006). In this stage, I focused on instances of valuations, on epistemic practices, particularly those that involved the reduction of complexity, on practices that were described as establishing or losing control, on motivations and interests, and on descriptions of how participants and their behaviour are managed in experiments. The final research questions and the results of this analysis are described below.

The empirical research I did for this thesis, all in all, certainly fell short of what would normally be considered an ethnography in anthropology or sociology (Delamont, 2007). One reason is that my fieldwork was conducted over a relatively short period of time, and my rather irregular working schedules at the laboratory did not allow for proper "immersion". Even if I would have gotten the chance to be there regularly several times a week for a year, what can be observed at the laboratory is only one aspect of the epistemic work involved in experimental economics. Instead, I used episodes of participant observation at the laboratory and interviews, during which I could also clarify some of the questions that had come up when running experiments, as complementary strategies to reconstruct the epistemic practices involved in developing and running experiments. Since "the data from each can be used to illuminate the other" (Hammersley & Atkinson, 2007: 102), using both interviews and participant observation certainly helped me to develop a more comprehensive reconstruction of these practices, particularly of those that are not observable, than if I had only used one of these two methods. Naturally, this does not mean that my reconstruction applies to the practices of experimental economists in general; it is the result of a very specific and situated investigation into a particular context at a particular time, guided by a particular research interest. The following analysis should be read with these qualifications in mind.

4.3 Research Questions

The systematic use of laboratory experiments is a relatively new method in economics. Experimentation as a method is closely connected to the development of modern science and the epistemic benefit of controlled observation. From its context of origin in the physical sciences, it was increasingly imported to other research fields, including those studying human behaviour. Its application to study economic questions, however, is not straightforward, as economics was traditionally considered to be a non-experimental science. While "standard" economic theory has recently come under critical scrutiny for its inability to solve pressing real-world problems, experimental economics on the one hand seems to promise an approach to make abstract models more applicable. On the other hand, experimental economists take care to distinguish their methods from those of other behavioural sciences with a longer tradition of experimentation. While much experimental research in the behavioural sciences suffers from a replication crisis, due not at least to the impossibility of meeting the same levels of standardisation and control as in the natural sciences, experimental economics is still expanding its methodology and scope. It is therefore worthwhile to ask whether and how economists have found a specific approach to experiments, which allows them to benefit from this method despite the difficulties and risks inherent in studying human behaviour, and the abstract and simplified character of the theoretical models it applies.

Also, as Knorr Cetina (1992, 1999) shows, the characteristics of experiments differ across sciences, in accordance with different technologies of enculturating natural objects that make up various types of laboratories. Thus, while the experimental method in a general sense is applied across the sciences, the actual practices and meanings encompassed in experimenting depend upon the specific epistemic culture of each field. Vice versa, looking at the practice of experimentation can also tell us something about the epistemic culture in question.

My first principal research question is therefore, *how do experimental economists appropriate the method of laboratory experiments for the purposes and demands of their own discipline?* This question points to the specific epistemic culture of experimental economics, and the epistemic practices available to members of this culture. It can be answered by looking for the strategies involved in designing and conducting laboratory experiments, and by explaining which specific theoretical demands and practical challenges they address:

- What epistemic practices are involved in transforming problems of interest to economists into phenomena that are tractable in an experimental setting? What does this tell about economists' understanding of how laboratory experiments relate to the world?
- How is the relation between theory and experiment described? What theoretical interests do experiments serve, and what do economists perceive as the advantage of experiments over theoretical models?
- What standards for producing reliable results are implicitly and explicitly addressed in experimental practice?

Experiments with human subjects always pose specific challenges to the researchers. While standardisation is difficult (and might be impossible) to achieve in other experimental sciences as well, human behaviour as an object of research can be expected to be particularly unpredictable (cf. Peterson, 2016). The fact that human behaviour systematically diverges from the predictions of economic theory is also the reason why economics experiments are done in the first place. It is therefore worth investigating how experimental economists themselves reflect on the challenges and benefits involved in doing experiments with human subjects.

- How do experimental economists conceptualize experimental subjects, their agency and contribution to knowledge production, and how is this conceptualisation articulated when reflecting on the practices involved in experimental design?

Finally, the laboratory as a place for building controlled environments arguably plays an important role in producing reliable and comparable results. Shifting attention from the epistemic practices of experimental design to the practices and material realities of conducting experiments, I also ask:

- How do the processes, routines and material arrangements that make up an economics laboratory support the epistemic strategies and practices involved in experimentation? Using Knorr Cetina's terms, how does the laboratory of experimental economics help researchers to reconfigure natural orders and achieve an epistemic advantage vis-à-vis their object of study?
- What do the material arrangements and codes of conduct in the laboratory tell us about the methodological demands and priorities of experimental economics, and how do they help to enrol both research subjects and researchers into a specific form of knowledge production?
- In particular, how are participants disciplined in order to achieve a reconfiguration that caters to the epistemic interests and conceptions of experimental economists?

It is not self-evident that the overall idea of what good experimental practice is and what kind of knowledge experiments should aim to provide is shared by all the individuals that are part of a recently established epistemic community such as experimental economics. Shifting the focus from the level of the epistemic culture to the level of individual researchers, then, it will be important to ask what tensions arise between the methodological demands of the discipline and the individual researchers' view of why experiments are valuable and how they should be conducted:
How experimental economists evaluate the method of laboratory experiments is therefore my second principal research question. It signifies a methodological shift in perspective, from the collective level of the epistemic community to a more person-centred approach focused on individuals' evaluative principles. To answer this question, I will analyse how my respondents reason about the purposes of experiments and explain specific choices and problems in view of these purposes. Referring to the work of Fochler et al. (2016), I will look for the evaluative principles and regimes of valuation that are addressed and enacted by experimental economists.

- What epistemic qualities and values do economists ascribe to the method of laboratory experiments?
- How do individual researchers justify and evaluate specific design choices?
- What are experimenters' motivations for selecting particular problems and questions as worth investigating?
- How do researchers characterise their own approach with reference to what they perceive as

the standard methodology of their discipline?

Choosing certain questions, designs and approaches over others, researchers are enacting the evaluative principles that also guide their epistemic practices. Approaching these practices from two perspectives, that of the epistemic community and that of the individual researchers, will therefore allow me to pay attention both to those unifying aspects characteristic of a specific epistemic culture, and the heterogeneity of individual approaches it can accommodate.

5 Analysis, Part 1: Epistemic Practices in Developing and Conducting Experiments

This first part of my analysis aims to reconstruct the epistemic practices and strategies involved in developing, designing and running a laboratory experiment in experimental economics. I trace these epistemic practices along four stages in the process of experimentation: developing the experiment from an initial idea, developing a theoretical model and making predictions, developing the experimental design, and conducting the actual laboratory experiment. Differentiating these four stages is a deliberate choice that I have made in order to organize my analytical observations; in practice, they often mesh or overlap and may not be so easily distinguished. To answer my first principal research question, *how economists appropriate the method of laboratory experiments,* I focus on the one hand on the relation between economic theory and experiment, and on the other hand on the conception of the role of participants in the specific type of experimentation that my respondents describe and practise. Consequently, I will first investigate how researchers develop an experiment from an initial idea and what the role of theoretical models is in that process. Describing how they prepare the implementation of an experiment, what considerations enter particular design choices, and how the laboratory and the rules governing its use help in producing reliable results, I then reconstruct how participants are conceptualised. I also describe what resources are mobilised in order to enrol participants into the epistemic machinery of laboratory experiments, while still allowing for a trade-off between experimental control and participants' agency.

My argument proceeds in the following way. I first clarify the conceptualisation of the relationship between the world, the theories, concepts and models economists use, and the experiments. Having established that there seems to be a tension between conceptions of experiments giving more weight to theory, and those giving more weight to observations, I reconstruct a similar tension in experimental economists' conception of the role of their human subjects. This conception of human subjects is most clearly articulated when economists speak about how they design their experiments, but it also informs and guides laboratory practices. I argue that the laboratory achieves a reconfiguration of the natural objects economists study by allowing economists to study "real behaviour" (i.e. behaviour that is not merely modelled or simulated) in conditions that are very close to those described by theoretical models. Economists, on this account, appropriate the method of laboratory experiments to create environments that are sufficiently similar to theoretical descriptions that the insights derived from observing behaviour in such situations can inform theory-building in economics.

5.1 Relating Experiments to Real-World Contexts and Theory

Where do experiments start from? The following sections will try to answer this question for the sample of experiments that I was able to study and discuss with their authors. While all of the experiments I talked about were presented as having some relation and relevance for real-world problems, two of these were directly inspired by a current debate that attracted researchers' interest. To be suitable for an experiment, a problem needs to lend itself to a certain kind of interpretation, which reconceptualises the actors and actions involved in terms of a situation that can be described by economic theory and practically implemented in a laboratory. In particular, researchers need to identify a single dimension that might affect behaviour and can be isolated and varied in a controlled way in the laboratory. In a second step, researchers then derive hypotheses and predictions from their conceptualisations. The most established way to do so is by developing a formal mathematical model, but there are also other approaches to experiments, which give more weight to observation and make predictions based on previously established results.

5.1.1 How do experiments relate to the world?

Of the experimental studies I discussed with their authors, two were explicitly presented as investigating a real-life problem that had recently received media coverage or was a topic of academic discussion. One of them was an experiment on consumers' reaction to corporate tax avoidance. The inspiration for this experiment had been the public debate, particularly in the UK, on how tax avoidance by multinational corporations could be curbed. Given that tax avoidance usually takes advantage of a legal grey zone, it is not necessarily unlawful. Consumers, however, might understandably find it unfair that large corporations are paying a lower percentage of tax (or even none at all) than the average citizen. The researcher told me that they had been particularly puzzled about how differently two multi-national corporations – Google and Starbucks, respectively – reacted to the public outcry over their tax avoidance practices. While Starbucks apologised and paid some amount of the avoided tax back, a spokesperson for Google pointed out that their liability was primarily with their shareholders' interests in maximising profits and that they were only "playing by the rules". The researcher's first idea was therefore to experimentally investigate why these companies could avoid paying taxes and why some would get through with it more easily than others. Eventually, the experiment then developed into a test for a policy solution that would empower consumers through a "quality rating" of companies, showing how much taxes they had paid. Such a rating would give consumers the chance to decide for themselves whether they still wanted to buy products from a tax avoiding company or boycott it in favour of companies with a better rating.

To arrive at an experimental setup from the initial inspiration of the tax avoidance cases of Google and Starbucks, several steps of *reconceptualising* had to take place. I have chosen the term "reconceptualising" for this process (rather than simply "conceptualising"), because understanding the problem as a case of tax avoidance is already the result of a specific conceptualisation. Perceiving a case of tax avoidance as a situation that can be investigated experimentally, therefore, requires an additional step of reconceptualisation from the point of view of an experimental economist: Analysing this situation, the researcher told me, they were looking for the "economic fundamentals" behind the status quo.

They began by looking at the *market structure*: In this case, what kind of products the two different companies were selling and how much competition there was on their respective markets. Through this lens, one company could be easily identified as a monopolist. The researcher then decided that consumers should also play a role in the experiment, because consumers' demands for stopping tax evasion had been a driving force in the public debate. Including consumers meant that the experiment would consist of a two-sided market with firms and consumers. While the real-world market consists of a number of firms and millions of potential consumers, the experiment would feature a "downscaled" version of a market, consisting of a group with a small number of participants (two buyers/consumers and one or two sellers/firms, respectively). To think of the initial problem in terms of a two-sided market situation with either two firms or one firm (a "duopoly" or a "monopoly") therefore was the first step of developing the experiment. The "economic fundamentals" that had been identified by the researcher were the structures of two different markets, one of which was a monopoly, and a number of consumers in both markets who can decide to buy or not buy the products of these firms. That means that the actors involved in the situation were reduced to firms and consumers, and their range of activity and interaction was reduced to (not) buying and selling a (nondescript and homogeneous) product.

Given that the real-world situation – the public discussion over tax evasion - that inspired this research not only involved buyers and sellers, but also legislators and news media, and took place in a complicated legal framework where tax evasion is somehow lawful albeit morally problematic, it is not obvious that the problem of interest should have been reconceptualised primarily as a market, i.e., as a situation of exchanging goods. The reason why this particular reconceptualisation took place here might be that the researcher knew they wanted to do an experimental study from the start, so they had to think about "bringing [the situation] to the lab" (Transcript P1). Markets are the kind of concepts that lend themselves to a laboratory experiment, because they are very easy to model and implement in the lab; in this case, a group of four participants is enough to create a duopoly market.[14] However, the researcher gave up on the first idea for a baseline setup because the solution for an experiment investigating the market power of the two different firms seemed too obvious: It was to be expected that the monopolist would sell more products despite avoiding tax, compared to firms operating in a non-monopolistic market.

> (P1) So of course I could have modelled it somehow like that, so just have one firm or something like that, and in another treatment if you want, many firms, and then see what happens, but this would be the obvious thing. So, ahm, theory would predict that Google, or that the monopolist shouldn't care, and I would have highly doubted to find anything else in the lab. (Transcript P1)

---

14 Note that for all intents and purposes, the laboratory experiment is not understood as a „simulation" of a market situation, but as a „real" market. The participants that take on the role of sellers and buyers trade a (nondescript and homogeneous) virtual good, but they do trade it „for real", since the profit they make through trading during the experiment is converted into actual monetary earnings at the end of the session.

A treatment is a particular variation of the experimental design. In most cases, experimenters design and run a "baseline" condition and then one or several "treatment" conditions, in order to study the effects that changing a specific "treatment" variable or factor has on behaviour.[15] In this case, designing an experiment with only a "monopoly"-treatment and a treatment with several firms was expected to produce a very predictable and therefore uninteresting outcome. What was still missing was a relevant factor that could be expected to make a difference for the decisions of consumers and firms, and that could also be varied in a controlled way. This factor was identified as whether consumers were or were not informed about the tax avoidance of firms, which signifies another step of narrowing down and reconceptualising the initial situation. The researcher moved on to a different idea that focused on the role of *information* on consumer behaviour, rather than the structure of the markets. After all, there would not have been a public outcry if nobody had been informed about the companies' tax evasion practices:

> (P1) [T]hen I thought of well, I didn't even know that they, ah that they avoided taxes so much. I wasn't so interested in that at the time, so it was the information that actually got it all, got it all started. So in the sense, I was not informed […] and I thought ok, information, how, how can information actually change the behaviour or the market interaction the consumers and and firms? I mean from this whole CSR perspective, ahm, and that's where Starbucks again comes in, […] they voluntarily paid this say 20 million, and why should they do so? If they didn't care about these accusations and how consumers might react, they shouldn't have done that, I mean that wouldn't make sense. So, I was interested in following up on this idea, so why do they do that? And what's ahm what dimensions could potentially play a role, ok? That I could, that I could change and play around with in the lab. And there was this information. Because if there was no information, they shouldn't care, but if there was an information they might care. And then it's not so clear how much they should care. (Transcript P1)

The baseline setup for the experiment then consisted of a market with two buyers and two sellers. Buyers could buy goods, which for the sellers had a certain production cost and for the buyers had a certain consumption value. The only way for the participants to make profit was to trade (and sell at a higher price than their production cost or buy at a lower price than their consumption value). Prices were denoted in "experimental currency units", which would be converted into Euros at a known rate at the end of the session. There was a fixed exchange fee (i.e. a tax) on each unit sold by the sellers, but sellers could "withdraw" some or all of the units sold at a lower cost than the exchange fee (i.e. they could avoid taxes).[16] The exchange fee would then not be paid by the seller, but distributed across all other members of the group. In the baseline treatment, the buyers had no way of finding out whether sellers were avoiding the exchange fee. There would be two more treatments: one in which

---

[15] The term "treatment" itself seems to be taken from medical research, as Friedman & Sunder (1994) notice. In the second part of this analysis, I will identify medical trials as an ideal type of scientific experimentation that was invoked by one of my respondents. Guala's (2005) notion of the "perfectly controlled design" is also modelled after medical trials and their principles of randomisation and treatment variation.

[16] Using the term "exchange fee" instead of "tax" is an example of experimenter's concern to avoid suggestive language. I will discuss this requirement for "neutral language" in instructions below. In this case, using the word "tax" was probably avoided because it would have framed the experiment in a particular way and produced overly strong reactions on the consumer's side because of the moral connotation of "tax avoidance". In order to avoid the criticism that effects are only observed because of such moral connotations, experimenters try to use neutral formulations, since they are aiming to only measure the effectiveness of the intervention they are testing (rather than the controversial nature of the topic).

sellers could decide to disclose the percentage of taxes they paid, and one in which the percentage of taxes paid was automatically disclosed after each round. The question the experiment sought to answer was thus: If buyers see how much tax firms decided to pay (rather than shifting the burden of tax on all the other market participants), will they prefer the firms paying higher taxes (but in turn asking higher prices for their products) and possibly manage to coordinate a boycott, in order to make the tax evaders change their practices? More specifically, is a voluntary disclosure of the taxes paid by firms enough to make consumers better off?

Identifying the "crucial dimension" that can cause a difference across treatments is the basic building block of an experiment. Information is a factor that can easily be controlled and varied in the lab. Thinking about the public debate on how tax evasion of multinationals could and should be tackled in terms of a market where buyers might be informed about the tax discipline of prospective sellers, the researcher conceptually arrived at a situation that could also be practically implemented in the lab. In the course of this conceptual procedure, however, the initial situation was transformed in quite a remarkable way: from the observation that national governments cannot force multinational corporations to pay taxes, but consumers might have some kind of aggregate power, to an experimental market in which two buyers could decide to buy or not buy the products of two sellers.

It is notable that reconceptualisation also helps to reduce a particular problem, which served as the inspiration for the research, to a more general problem that is interesting from the point of view of the discipline of economics. The example above and the different steps of developing the experiment illustrate this process of reduction: Looking at a problem that they find interesting, researchers interpret it from the perspective of their theoretical and methodological background, and reformulate it through applying concepts available in this theoretical repository. These concepts then work like a filter, removing the particularities of the initial situation and leaving only a more abstract and generic situation. Some of the metaphors the researcher used when describing this development also clearly invoke a process of reduction:

> (P1) [Y]ou need to downscale everything and you you want to make sure […] that you mirror the important dimensions that you think the real environment has, and just boil them down to the, to the essentials.[…] So all the unnecessary stuff needs to go, because you have, don't have space for that, ah, and no distractions and what not, but you need to boil it down to the crede- ah to the essentials. Ahm, yeah. That's how I came up with the baseline kind of setup. (Transcript P1)

Describing the process in this way, the researcher indirectly offers an explanation for how experiments relate to the world: An experiment shares with a real environment some important dimensions, namely those that are seen as *essential* from the economists' point of view. In the example above, the essential features were those that could be reconceptualised in terms of economic theory ("markets", "monopoly"). The epistemic process of developing an experiment then consists of finding these dimensions through a process of reduction – "boiling down" – that removes all the non-essential features of the real-world environment. Conversely, it can be assumed that economic fundamentals can be found "underneath" the flesh of everyday life; they provide the skeleton of our mundane interactions, and understanding the skeleton will help us to understand how the whole thing moves.[17] What is abstracted away in the process of developing an experiment are all the particularities that make the original situation unique, such that it can be identified as being only one case of a more general situation, an example of a more generic concept (a market, for example).[18] This corresponds to Santos' claim that the results of laboratory experiments allow us to draw *generic* inferences. Rather than applying to one specific real-world context, the insights gained from experiments apply to the class of situations that are structurally similar. To see whether the conclusions obtained from the experiment also hold in a real-life context, further studies such as field experiments would be necessary; and this was also what the researcher suggested when presenting the results.

Not all experiments, however, start from the inspiration of a real-life case.[19] In some cases, the research project starts with the aim to investigate a particular theoretical concept experimentally, and the insights gained from the experiment are then interpreted as corresponding to a certain class of real-life situations. The experiments I encountered in my interviews were all either of the first or the second kind; they would all either start from an theoretical concept or a real-life inspiration that was then developed into an experiment by finding a relevant treatment dimension that could be expected to affect behaviour and be subjected to controlled variation. My participants also mentioned other ways of developing an experiment, which are more methodologically motivated, such as developing a new kind of game, or a new tool (the use of MRT-scanners in the recently developing field of neuroeconomics would be an example of the latter). One of my participants told me about their first experiment, which was developed by making use of a particular feature of the software, which allowed them to impose time-lags on decisions and investigate the effect of these time-lags on participants' behaviour. Another one wanted to investigate unstructured bargaining, and had developed a graphic interface to enable this particular kind of interaction between three players. The practices of reconceptualisation and reduction that I have described above, in my view, are exemplary of a certain way of seeing and interpreting the world ("in terms of") that is characteristic of the epistemic culture of experimental economists, but they might not strictly apply to all experiments and their histories. Where they do, they sometimes culminate in the most reductive reconceptualisation one can give of a strategic interaction, a game-theoretic model.

---

17 This idea that there are fundamental economic principles that apply everywhere is what Muniesa (2014, 36-38) refers to as the „naturalistic style" in economic reasoning.
18 Nancy Cartwright refers to this process as "climbing up the ladder of abstraction", arguing that the higher the level of abstraction, the easier it is to draw valid inferences (cf. Cartwright, 2010).
19 In fact, most experimental studies probably are not inspired by a particular debate, but by previous studies or theoretical interests. That I had two studies in my sample, which referred to very specific real-life problems, is mostly a consequence of my initial research interest of how real-life situations are translated into laboratory experiments.

5.1.2 How do experiments relate to theory? The *textbook approach*

On several occasions during my fieldwork, I encountered what could be named the "textbook approach" of developing an experiment. This approach would be referred to in two different ways. Either the stories researchers told me about how they developed their experiment were exemplary of this approach, or they invoked it – implicitly or explicitly – as a contrast to their actual practice. The following two accounts of the "textbook approach" reflect both kinds:

> H Ok. But what, what would you say are the cha- the characteristics of a paper that is very, that has very good chances of getting published?
>
> P3 This one.
>
> H This one, ok.
>
> P3 You have a non-cooperative model, which you implement in the lab, yeah, clear predictions that you implement in the lab. This is classic
>
> H Ok
>
> P3 experimental economics methodology. Non-cooperative model, game-theoretic model, have predictions, run it in the lab, see if those predictions hold. (Transcript P3)

> P4 Hm, ok. I mean there's the, the textbook ahm approach, I guess, where you first think about your research question, then you go to the literature, you real-, you identify some research gap, you think how to model this, ah you set up a design, you develop some hypothesis, ideally a theoretical model to predict what you expect to happen, and then you go to the lab
>
> H mhm
>
> P4 that would be the textbook approach. (Transcript P4)

What characterises the "textbook approach" according to these two statements is a systematic step-by-step procedure, and the use of the experiment to test theoretical predictions. In the ideal case, these predictions are made with the use of a theoretical model, and more specifically, a model following the conventions of non-cooperative game theory. There are two reasons, according to my interviewees, why researchers do not always follow this "textbook approach": One the one hand, there usually are more "loops" between the different steps of developing the experiment, when researchers go back and forth between their ideas and possible designs; on the other hand, not all experiments start with a fully formulated theoretical model.

What is the kind of theoretical model my respondents refer to? A game-theoretic model (a "game") is a formalisation of strategic situation. It specifies the agents involved (players), the rules according to which they can make decisions (who moves first, available actions, etc.) with available information specified at each decision point, and the payoffs for each outcome of a game. The outcomes are ordered numerically according to the payoffs; this order reflects an individual player's preferences over the outcomes. Models are typically solved through searching for an equilibrium (or several equilibria, depending on the game), meaning, by finding a "stable" state in which all of the agents act in such a way that they would not benefit from changing their strategy, given what the other players are doing. The assumption guiding this method is that the target systems represented by the model will, all other things being equal, converge towards one of the equilibria identified. From finding an equilibrium (or several equilibria), it is therefore possible to make theoretical predictions on how each of the actors is ideally going to behave in the long run, given the structure of the game, their assumed preferences, and the absence of outside influences. Furthermore, it can be studied theoretically how their behaviour should change when some parameters are modified.[20]

Incidentally, only one of the experiments I discussed with my interviewees was based on such a fully developed game-theoretic model. This experiment was motivated by the researchers' interest to study corruption and lobbying, a general interest that they had narrowed down to the question of how people interpret information that they know may have been paid for by a third party. Again, transparency, i.e. being informed whether the advisor had taken the payment, was identified as the controllable dimension. The draft paper framed the study as a contribution to current debates about whether scientists giving advice to policy-makers should be obliged to disclose their sources of funding. Some argue that decision-makers need to know whether their expert advisers suffer from a conflict of interest (the household example being a medical scientist working for the tobacco industry). Others point out that transparency about third party funding might in turn give rise to the different bias that the advice of scientists who benefit from third party funding is generally disregarded, even though it might be truthful. The authors then set out to investigate both theoretically and experimentally whether transparency can result in better decision-making. However, both authors explained to me that they considered their model to be very general, in that the so-called advisor could be thought of as a lobbyist or a scientist, but also simply as a doctor who may or may not be paid by the pharmaceutical industry to recommend a certain medication. The researchers specifically wanted to study a situation in which being paid would not necessarily force the advisor to lie, if the situation were such that the advice preferred by the third party would also be truthful. What the model then described was a situation in which one of two possible states of the world obtained, an advisor who would know which state it was, a decision-maker who wanted to make one out of two possible decisions in line with the actual state, and a third party who preferred one of the two decisions, regardless of the actual state of the world. The advisor could receive a side-payment if she advised the decision-maker to chose the option preferred by the third party, but she could also decline this payment. In the final version of the

[20] Cf. Don Ross' discussion of the notion of „equilibria“: „When we say that a physical system is in equilibrium, we mean that it is in a *stable* state, one in which all the causal forces internal to the system balance each other out and so leave it 'at rest' until and unless it is perturbed by the intervention of some exogenous (that is, 'external') force. This is what economists have traditionally meant in talking about 'equilibria'; they read economic systems as being networks of mutually constraining (often causal) relations, just like physical systems, and the equilibria of such systems are then their endogenously stable states. (Note that, in both physical and economic systems, endogenously stable states might never be directly observed because the systems in question are never isolated from exogenous influences that move and destabilize them. In both classical mechanics and in economics, equilibrium concepts are tools for analysis, not predictions of what we expect to observe.)"
Ross, D. (2016) Game Theory. In E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Retrieved from: https://plato.stanford.edu/archives/win2016/entries/game-theory , 15.02.2017

experiment, this side-payment was higher in the state where the advisor would have to lie to give the advice preferred by the third party.

It makes sense, in my view, to think of modelling as a special, more formalised and more conventional case of the processes of reconceptualisation and reduction described above. When developing a model, researchers likewise look for a theoretical interpretation that is part of their conceptual repository. In this particular case, the model was a variation (or rather extension) of the so-called *signalling game*, a well-studied type of game in which one player, who has an informational advantage, sends a message to the second player. The second player, who knows only the likelihood of the first player's message being true, will then have to decide whether she believes this message or not. What the researchers did to arrive at a model was therefore to think of their problem - advice on decision-making involving a possible conflict of interest - *in terms of* a signalling game, with the extension of a third party offering a side payment. The most important criterion for the researchers was to create a model that was as simple as possible, since they would later need to implement it as an experiment:

> P3 So, yeah, it's, it's a model where the person who is giving you information may or may not have taken payment and you may, may or may not be aware whether or not this payment has been taken, and you have to make a decision. And this is, you know, the simplest model that we could think of that, that captures this.
>
> H mhm
>
> P3 Ahm, you know, of course in reality there would have been some other stuff going on, ahm, but in the experiment you try to ask like one particular fact that you're interested in.
>
> (Transcript P3)

The process of modelling here is described as being part of developing an experiment, and as helping with finding the simplest possible setup and identifying one single causal factor that can affect behaviour.[21] The theoretical modelling should also be sufficiently simple to be implemented as a laboratory experiment, and the simpler the model is, the easier it is to implement. In this case, this meant that the third party offering the side-payment to the advisor was modelled as "inactive", so that she did not have any other options than offering the payment of a pre-defined size. In the experiment, this third party was "played" by the computer. The researchers explained to me that they had made this choice, because a three-player game would have been much more difficult to understand for the participants, possibly resulting in more mistakes and less useful data:

> P3 The questions in the seminar were, if you modelled the special interest group, are these the offers that would be made, so, ahm, for a theoretical paper, we should have also modelled the decisions of the special interest group, to look at what is the, what kind of payments would they offer. And, for an experimental paper, as I said, you want to keep it simple. Cause if you have the three player game, ahm,

21 There is a rich and diverse literature in philosophy of economics, which discusses the process of economic model building and the nature of models' relationship with reality. This is not the place to engage with this debate. Let me just add that what I have observed in my participants' accounts of model-building corresponds with Morgan's characterisation of „idealisation", which „suggests that models are arrived at by processes of *abstracting* to the level of ideas or concepts; of *simplifying* the case or system treated by omitting irrelevant or negligible influences; of *isolating* the elements that are really thought to be important by *ceteris paribus* clauses" (Morgan, 2008: 659).

it's really complicated. Learning gets much more difficult, and we really wanted to focus on the central action between the second and third player in the game. So it was really because it's an experimental, ahm, it's already difficult for subjects with these two, just the two roles, if we added a third role as well, ahm, it'd be, you know, take a lo-, it's, yeah, it's very, it's very hard for subjects to understand these things, and learning takes a lot longer, because with the, the two players, things start to settle down after a bit, but if you have a third player, who is also making mistakes and slowly learning, ah, yeah, so that, as I said before, in an experimental study, you really want to make it as simple as possible, and, with as few people as possible, so. (Transcript P3)

According to this statement, the complexity of theoretical modelling needs to be traded off against the practical feasibility of the experiment. The limits of feasibility are presented by participants' ability to understand the structure of the game in the time available, and probably also by the costs of having more roles and therefore, more participants to be paid. Note that the researcher here anticipates that participants will need to "learn" how to play the game during the experiment, i.e. that they will gradually understand the structure and consequences of their decisions through experience. Due to these practical constraints, modelling for a laboratory experiment is different from modelling as a purely theoretical work. Given that what they had modelled was only a *sub-game* of a bigger three-player game, the researchers anticipated that their model was insufficient from a theoretician's point of view. Indeed, they received one review that criticised their model for being "too simplistic". Yet the model was not intended to provide a fully developed theoretical account for the situation, or a theory of lobbying. What it served to do was to show theoretically whether transparency could improve decision-making in a simplified situation like the one implemented in the laboratory, and what to expect from the experimental results. One might also say that since this model would have been "too simplistic" to serve as a theoretical account, it was primarily developed towards corroborating its predictions with experimental data. I was told that it is also possible and acceptable to take already existing theoretical models and simplify them for an experimental study. The practical function that the model has can then, in my view, be interpreted as that of structuring the experiment and narrowing down the range of possible explanations for the results. Developing a model or simplifying an existing one helps with the "hard choices" of finding an experimental setting and abstracting away everything "unnecessary", because the practice of modelling is guided by conventions and builds on existing conceptual resources. Additionally, a theoretical model helps with the experimental design, because it enables the researcher to calculate which parameter values yield the most interesting predictions for the experiment, and it situates the research project within the existing literature.

Having developed the model and analysed all the equilibria under transparency and non-transparency, the researchers found that their model predicted a positive effect of transparency on decision-making. The then set out to test whether this outcome could also be observed in the laboratory:

> P3 Right, so we start with this theoretical model, we get some predictions, according to the model, in equilibrium, transparency can be better than non-transparency. And then we implemented this model exactly in the laboratory to see if it worked when we had real human subjects. (Transcript P3)

The function of the laboratory experiment, according to this reasoning, is to test the theoretical predictions of the model when some assumptions about the economic agents (that they behave perfectly rational, for example) are relaxed, since “real human subjects” will typically not behave like idealised economic agents. In other words, the experiment is a test of the robustness of the theoretical account that was presented in the theoretical model. The epistemic value it has over the model is that it enriches the results of the research with some of the particularities - the idiosyncratic preferences, interpretations and cognitive limitations of individual human beings - that were originally abstracted away.

5.1.3 Experiments “without” theory

Whether experimental economists follow the textbook approach outlined above seems to depend on their own background, i.e. the training they received and the influences other scholars had on them, which results in different preferences and ideas about the role of theory in experiments. Especially researchers identifying themselves as coming from a more “behavioural” tradition seem to follow a less formalised approach. Whether researchers find it necessary to provide a model and theoretical predictions before the experiment starts also depends on the type of experiment, i.e., whether it is a more “standard” experiment or a more “behavioural” experiment.[22] “Standard” experiments are experiments which can be modelled, and the outcomes of which can be predicted using standard theory and the typical assumptions of risk-neutrality, rationality, and so on. “Behavioural” experiments are experiments in which established behavioural regularities, for example social preferences, are expected to play a significant role, to the effect that standard (game) theory is known to fare poorly with predicting the results.

As an example, consider a “behavioural” experiment that investigated a situation in which a “client” could consult an advisor to receive either a “severe” (i.e. expensive) or a “mild” (i.e. cheap) treatment, under the risk that the advisor would sell the expensive treatment (and “overtreat” the client) when this was not necessary. In the publication, the authors present the underlying game in the extensive form, i.e. as a decision-tree diagram, which specifies the choices and outcomes for each player at every stage, and solve this game with *backward induction*[23] to arrive at “standard” game theory predictions (under the assumption of rationality, self-interestedness, and so forth). However, it was explicitly anticipated that these “straw-man predictions” would be refuted by the experimental results:

---

22 This distinction between “standard” and “behavioural” experiments is based on how my participants used these terms, and it is not meant to provide a clear-cut classification. Santos (2010, chapter 10) distinguishes between „technological“ and „behavioural“ experiments, where technological experiments are typically those investigating market mechanisms and other „technologies“ which, if successful in the laboratory, can be directly implemented as policy instruments. According to Santos, technological experiments usually are based on theories from the field of market design and industrial organisation, whereas behavioural experiments are based on game theory and study strategic interaction. In my sample, all the experiments were based on game theory, but still some were apparently considered to be more “behavioural” than others.

23 *Backward induction* refers to the method of solving a game by starting from the possible outcomes that each action and expected counteraction will have, to choose the action that yields the most desirable outcome. It often comes down to deciding the first move of the first player, given the possible outcomes. Cf. this slightly more technical, but also more elegant textbook definition: “Whenever a player has to move, she deduces, for each of her possible actions, the actions that the players (including herself) will subsequently rationally take, and chooses the action that yields the terminal history she most prefers.” (Osborne, 2004: 158).

H Mhm. And the predictions that you have, you derived them from certain assumptions? Rationality assumption and so on?

P6 Exactly, that's what we always have in these papers, we always have the straw-man prediction, but which is not useful.

H Mhm.

P6 The straw-man prediction says, essentially says, nobody should trust, nobody should buy, yeah, it's gonna, the market is gonna break down. But that's, we know that this is not gonna happen.

H Hm

P6 Because we have a lot of experience with these type of markets, we know that this is not gonna happen. So, but it's still, kinda for the, for the un-, uninitiated economist it's still interesting to see, wow, you know, here they have this setting where this, this prediction doesn't happen, wow. To us it's not a surprise, that, the level is different. (Transcript P6)

This statement is in line with a type of research that considers behavioural economics as the "repair shop of economics" (Transcript P1). What behavioural economics does, on this account, is to study exactly where the predictions of standard economic theory (in this case, game theory) fail, and then systematically describe the reasons for this failure, such that they can be taken into account in future research. As another researcher explained to me, it is possible to integrate well-known behavioural regularities, which counteract the predictions of standard game theory, into one's modelling to arrive at more empirically adequate and possibly novel and unique predictions. The improved empirical adequacy of such "behavioural" models comes at the cost of their generality. While standard game-theoretic models are sufficiently abstract to be very generally applicable to a range of situations (as we have seen in the case of the lobbying model, which was based on the even more general model of a signalling game), behavioural models are considered to work "in niches".

(P1) [In] terms of these behavioural models, well, the nice thing about the standard models is, they are very, very ah, very very clean and parsimonious, or? They don't assume much, they just say you want more, basically always, and you want variety, you know, these kind of standard preferences. They are very applicably, well, very generally applicable. But ah these, these empiri-, these behavioural models, they work in niches, so they have been shown to work in niches.

H mhm

P1 So if you have a particular problem, then this model can explain it well. So there is no grand behavioural model that explains all kind of, say deviations from the standard model in all possible cases. (Transcript P1)

Standard models here are referred to as a kind of minimal case, because they only assume "standard preferences". Interestingly, "standard preferences" are considered to be the minimum assumptions about the expected behaviour of agents. Behavioural assumptions, which build in social preferences and other non-rational types of behaviour, are seen as an addition to these standard assumptions. The deviations from standard preferences that behavioural models describe are consequently expected to only apply to particular contexts. This seems to highlight a strong and prevailing assumption that the economic behaviour described by standard theory is in some way more fundamental than the "social" behaviour described by behavioural models. Behavioural models need to make additional assumptions in order to predict and explain phenomena. Conversely, these phenomena are theoretically expected to exist when an "additional" factor (such as a preference for equality or fairness, or a social norm) is present, otherwise, "standard" economic behaviour should be observed.

In the experimental practice of my interviewees, the function of behavioural models – in contrast to standard models – seems to be more to provide a theoretical explanation of the behaviour observed in the lab, than an a priori prediction of what is going to happen in the lab. One explanation I received for this pointed towards the different purposes experiments can serve: If an experiment is considered to be more "exploratory", because it does not test an already existing theory or model, it is not seen as necessary to start from very precise theoretical predictions. Behavioural models then take the form of a post-hoc formalisation of the experimenters' initial hypotheses. These hypotheses are not developed by building and analysing a model, but rather by drawing on previous experience and established results. Since formalised models are not needed to arrive at hypotheses and predictions, they are sometimes omitted completely. This, however, is a risky move, because not providing a theoretical explanation for one's results in the form of a model is seen as limiting publication chances:

> (H) You didn't have the theoretical modelling for this situation.
>
> P6 Exactly. That was a weakness in this sample.
>
> H Oh?
>
> P6 That was a bit, a bit a problem in, ah, we were lucky that we published this well. Because the theory we have is ah, somewhat hand-waving. It's not, you know it's not a precise point prediction we're testing here. We're doing some qualitative things, where are, we say well, more competition should discipline the suppliers, and there, and because it disciplines them, they should provide more high quality, and that should induce trust. But we can't say, you know, 17.3 is the prediction, and then we test 17.3, it's qualitative prediction, or "comparative static", as we say.
>
> H Ok?
>
> P6 So we say, when we compare two treatments, then we think, in this dimension it should go up, in that dimension it should go down, this dimension should stay the same. This is comparative static. And then we could go this way, then this should go up, and this, or everything should go up. This is called comparative static.
>
> H Ok.
>
> P6 It's a weird expression, but ah it means "qualitative". It's, it's, you don't have any precise point prediction. (Transcript P6)

Formal models, also behavioural ones, allow experimenters to make rather precise predictions, for example in terms of quantities and probabilities. In the absence of such a formal model, researchers can only formulate roughly what they expect to see when they change a particular factor. These "qualitative predictions" turned out to be accurate enough in the present case. Even though the authors did not provide a theoretical explanation for the behaviour they observed in the laboratory, they were also right in anticipating to what degree the "straw-man predictions" would be false.

Whether it is possible to make predictions (be they qualitative or "standard") about the outcome of an experiment without a model depends on the underlying game. Especially one-shot games with "perfect information" – i.e. games that are only repeated once, and in which all players know about the available moves and payoffs for all other players – can normally be solved with backward induction.[24] Once the game has been solved and the predictions have been made in this way, taking into account behavioural assumptions, it does not seem necessary to provide a formal theoretical explanation immediately either.[25] Rather, it is apparently common practice to develop the behavioural model at a later stage:

> (H) You didn't have a model, right? At, at first, like, because in the textbook approach, as you, as you called it, you should have a model first and then test, find the parameters?
>
> P4 Mhm, yeah
>
> H And implement those in the lab. So, how did you derive your hypothesis and your parameters?
>
> P4 Ah, based on, well since it's a one-shot game, it's quite easy
>
> H ok
>
> P4 So with backward induction, it's clear what's going to happen if you assume rational, self-interested and risk-neutral agents, it's straightforward what's going to happen with backward induction
>
> H mhm
>
> P4 So in that sense, it wasn't necessary to implement the model here, and then, next I bia-, I was assuming that if some, a share of people is, has social preference, how could this change, ah, behaviour? But you're right, ah, I should still ah do the modelling, with Fehr-Schmidt preferences, for example
>
> H mhm
>
> P4 Ahm, but somehow, like everybody I was talking to said, yeah, you can do that afterwards. So, in that particular case, it's seemed, since this the, ahm, the standard game theoretic approach with this risk-neutral etcetera assumptions, is so clear what's going to happen, it's apparently, it wasn't that necessary that I write the model first. (Transcript P4)

To arrive at predictions and find the right parameters, this researcher had used what could be referred to as a "proto-model" specifying the actors and their available choices. This was basically identical with the experimental design, but it was not a "formal" theory or model. Developing models and theorising "ex post" seems to be an acceptable practice that is not considered to affect the reliability of experimental studies. In the study on tax avoidance described above, the experimenters asked a theorist to help develop a model that captured and explained what they observed in their experiment, namely that the information whether a firm has paid taxes or not can distinguish an otherwise homogeneous product from that of other firms. One author told me that s/he was quite happy that this model had not just served as a "fig leaf model" (Transcript P1) to increase their chances for publication in a top journal, but that its predictions had also inspired some parts of the data analysis, which eventually provided an unexpected result. This indicates that such retrospectively added "fig leaf models" in other cases do not necessarily contribute much to the epistemic gain of an

---

[24] In contrast, the signalling game that the model on transparency in decision-making was based on is a game with imperfect information, because only one of the two players knows the actual state of the world, whereas the other one needs to form a belief based on the known probability of each state and the message she receives. This type of game needs to be solved by a formal equilibrium analysis.

[25] The terms „theory" and „model" were often used interchangeably by my respondents, which I interpreted as signifying that formal models are considered to provide theories (i.e. theoretical explanations) for specific situations.

experimental study, since the predictions or hypotheses tested in the experiments were made without the model anyway.

If formal models are not needed to develop experiments, hypotheses and predictions, why are they still considered as so necessary that not providing a model might result in not getting one's experiment published? Why does standard methodology require experimenters to work with predictions derived from a model? My tentative answer draws on Santos' (2010) and Pickering's (1989) account of the epistemic value of experiments. A fully developed mathematical model is a far more *rigid* system, to use Santos' (2010) terminology, than a proto-model, which is still open to conceptual manoeuvring. Having a fully developed model before starting the experimental sessions therefore limits experimenters' ability for conceptual manipulation in an important way, because it provides them with a stable *phenomenal model* of the process they are studying. Given that the *instrumental model* of how the experimental "apparatus" works is also fairly stable – all the experiments in my sample used the same technologies and adhered to the same methodological principles, and this instrumental uniformity is fairly institutionalised in experimental economics – the only thing that is left to manipulate, in Pickering's (1989) and Santos' terms, is the *material procedure*, i.e. the actual experimental design and implementation. From this perspective, it is easier to understand why the "textbook approach" favours formal models and theoretical predictions. As one senior researcher told me, formal models and theoretical predictions reduce "researcher's degrees of freedom" (Transcript P7). If the predictions of a formally correct model and the results of an experiment using an established methodology do not match, then the flaw must either be with the experimental design, or the researchers have found a genuinely unexpected result. In both cases, the solution will be to do follow-up experiments either to improve the procedure (by changing parameter values or other details of the design) or to better understand the unexpected result and its cause.

According to the "textbook approach", a theoretical model thus serves epistemic as well as practical purposes. From a practical perspective, the function of a model can be described as reducing complexity and "degrees of freedom". It provides a precise description of a strategic situation that can be explored both formally and through direct implementation in a laboratory experiment with "real human subjects". From an epistemic perspective, a model provides an account of the experimental results that is either in line with existing theory or proposes a new theoretical explanation. The prominent position that theoretical reasoning assumes in the "textbook approach" is somewhat counteracted by the much less theory-driven practice of "more behavioural" research. Although researchers doing "behavioural experiments" acknowledge the requirement of eventually providing a theoretical account of the experimental results, they seem to grant experimental observation itself much more epistemic weight. The reason for this different approach is clearly that behavioural experiments study phenomena that standard theory cannot adequately describe or predict. Hypotheses and predictions then must be developed by taking into account what has been observed before, and modelling such behaviour always requires making additional assumptions beyond those describing the behaviour of ideal economic agents. The fact that there consequently seem to be two different conceptions of scientific experimentation in play is only described as problematic when it comes to publishing papers in economic journals. At least according to my respondents, theory-driven experiments or experiments that at least come with a fully developed theoretical model are clearly preferred in the current state of the discipline over experiments that start from and focus on observation.

5.2 Designing the Laboratory Experiment

Having developed a conceptual understanding of the experimental situation in the form of a proto-model or a model, experimenters set out to develop the experimental design and plan its implementation in the laboratory. Designing involves decisions about the sequence of different steps and phases of the experiment, the number of participants, the number of rounds and interactions, and the value of parameters, but also more subtle decisions such as how to formulate instructions and facilitate understanding and a smooth procedure. Using Pickering's and Santos' terminology, designing the experiment amounts to creating the *material procedure* that is expected to best answer the research question. This means that through the experimental design and the interactive situation that it creates, the experimenters aim to produce the behaviour they want to study. The assumption guiding the design process, and the practice of experimentation in general, is that on the one hand, the behaviour of experimental subjects can be induced to some extent, while on the other hand, the independent agency of subjects is what makes experimental observations insightful.

At the VCEE, almost all experiments are run on a software called z-Tree that was developed specifically for the purpose of running economics experiments (Fischbacher, 2007). Z-Tree was described to me as the "industry standard" (Transcript P6), it is free to use and frequently updated to accommodate new requirements (for example, new languages), which has contributed to its use in economics laboratories around the world. An important part of the actual implementation process, the programming of the experiment in the z-Tree software, will not be covered here. The reason for this omission is that I was not granted access to observe experimenters' more intimate work (such as the programming), and that I also lack the technical understanding and skills to analyse computer code. It would certainly be an interesting question whether and how the software affords particular solutions to design problems and precludes others, but this goes beyond the scope of my research project. However, much of what experimenters do when designing the features and structure of experiments is also conceptual work that can be reconstructed in interviews. What became clear from my interviews is that questions of experimental design are inseparable from the prevailing conception of participants, their capabilities and their role in the experiment. Planning and preparing an experiment, which includes writing the instructions for participants, is guided by considerations of how to best convey the structure of the experiment and the task at hand, without influencing participants' behaviour in the sense of creating “experimenter demand effects”. Ideally, participants should understand exactly what decisions they can make and what consequences those decisions will have, but they should not be able to guess the researcher’s hypothesis or feel like they are expected to behave in a certain way. While influence in the sense of conveying the expectations of experimenters should therefore be avoided, researchers also try to avoid producing inconclusive results by anticipating participants’ interpretations of the experimental situation and considering possible complications and problems. The practices involved in designing an experiment are characterised by an inherent trade-off between experimenters’ control and participants’ agency: Participants are expected to decide according to their “true preferences”, but experimenters feel that they need to “elicit” these preferences by actively excluding all other possible influences on behaviour.

Experimentation in economics in general rests on the methodological assumption that it is possible to *induce* behaviour through providing a particular *incentive structure* and by adhering to some precepts, which guarantee that this incentive structure dominates all other motivations that participants might probably have. In behavioural experiments, some of these precepts might be relaxed (e.g., participants may not only know about their own, but also about others’ payoffs, see Santos, 2010, chapter 10), but the underlying assumption that a particular incentive structure can be imposed, and that the choices participants make express their preferences when subjected to this incentive structure, still holds. Santos (2010) identifies Smith’s (1982) methodological principles as the *instrumental model* of experimental economics. Smith’s methodology treats laboratory experiments as “microeconomic systems” and is based on the assumption that subject’ preferences can be controlled by using monetary incentives to “induce value” and therefore preferences over different available choices, because some choices will result in higher earnings. Since these principles have been laid out in the 1980s and the use of monetary incentives is mentioned as one of the criteria for accepting papers to the journal *Experimental Economics* (Schram & Holt, 1998) the *instrumental model* of experimental economics can be considered to be fairly rigid and allow little room for individual adjustments. Especially in experimental studies where experimenters develop a theoretical model and therefore also a very rigid *phenomenal model* beforehand, the only possibility of handling problems and unexpected results is by making changes in the *material procedure*, i.e., by adjusting the experimental design.

Designing an experiment is probably the part of the experimentation process where skills and experience matter most. This is the case because the conceptual understanding of a situation – the *phenomenal model* - does not directly translate into an experimental set-up. Even when experimenters have developed a theoretical model and clear-cut predictions, there are many different ways of implementing the model as a procedure in the laboratory.[26] Which concrete implementation to choose can be a matter of great concern for experimenters, because even small details of the design might have an

[26] This also, but not only, concerns how to program the software. As one of the experimenters put it, there is “not one way to program a public goods game, but a hundred” (Transcript Field Notes).

effect on how participants interpret the situation, or make the experiment as a whole more difficult to understand. Some of the younger researchers therefore mentioned that it was very important to have good supervision when beginning to do experiments, because more experienced researchers can help one to make better design choices and let go of ideas that seem less promising. (One of the researchers who did not have such supervision told me that it had cost them a lot of time to figure out the best design by trial-and-error on their own.) When asked about design choices, a senior researcher told me that they had by now acquired a "good feeling" for parameterisation and that design choices such as the number of periods were also "a matter of hunch" (Transcript P6). The implication is that some of the practical knowledge involved in experimentation can only be acquired in "learning by doing".

Some informal rules of thumb for design choices do exist, such as the rule that an experiment should ideally be between one and a half and two hours long. Researchers therefore need to plan carefully how to best obtain a sufficient amount of information without making experiments overly complicated or long. There are also formal methods for calculating how many observations (and therefore, how many participants and sessions) one will need to arrive at statistically significant results, which is particularly important for researchers who only have a limited budget. The method of "power analysis" used to calculate the number of observations needed is directly related to the type of statistical tests used in data analysis. One researcher additionally mentioned that Reinhard Selten, who established the first experimental economics laboratory in Europe, introduced the general rule of having at least five independent observations. Many design choices are also contingent on the type of underlying model. For example, if the situation is modelled as a one-shot interaction, this means that participants in the laboratory will never play twice in a row with the same person, even though the same interaction might be repeated for many rounds during the experiment. So-called "partner matching", where participants play with the same person or group for several periods, in contrast can be used to study repeated interactions in which the formation of trust or reputation over time is expected to play a role.

5.2.1 Anticipating behaviour and facilitating *learning*

For many other design choices, no defined rules apply. These choices are then made based on assumptions of how participants will interpret the experiment and its instructions, and what the researchers could do to help participants understand the structure of the game more easily. How experimenters reflect on the process of designing experiments and writing instructions consequently sheds some light on their conception of the role and agency of experimental subjects. For example, a choice that was described to me as potentially having serious implications for the success of the experiment is whether participants should remain in the same role throughout the experiment or change the role at some point. Whether roles should be changed, at what point during the experiment, and how often, is a decision that does not make a difference from a theoretical point of view, but experimenters expect it to affect the behaviour they will observe. In the experiment on transparency and lobbying, the researchers used a single role-switch after 20 rounds as an instrument to facilitate participants' understanding of the game. The reason for introducing this role-switch was that they had observed "some very strange behaviour" in a pilot study without role switching:

> (P3) So in the pilot we didn't have role-switching, and we got some very strange behaviour.
>
> H Mhm
>
> P3 And we thought that it was due to people not understanding the experiment. That's why we introduced role-switching, because if they'd experience the previous role, then they would understand the incentives of the other role better. And perhaps, behave differently, if they fully understood. (Transcript P3)

The "strange behaviour" was that decision-makers followed the advice given when the extra payment had been taken (i.e. when the advisor took the "bribe" and was compelled to send a specific message no matter what the actual state of the world was), but not when the payment had been rejected, which was both counter-intuitive (since an advisor who had rejected the side payment is more likely to give honest advice) and against the theoretical predictions. However, it turned out that the "strange behaviour" only disappeared after the parameters were changed so that the side payment the advisor would receive was bigger in the state where the advisor had to lie than in the state where the advisor could both accept the payment and send a true message. The experimenters changed these parameters for purely theoretical reasons, although they later explained that the new parameter values were also more intuitive, because an expert in real life might actually receive a higher "bribe" if she is expected to lie in return.[27]

In hindsight, both authors agreed that the role switching by itself probably did not have a substantive effect on behaviour, since participants were learning (as was indicated by their more efficient decision-making) very fast already in the 20 rounds before the switch. That participants in general need to "learn" how to make decisions in accordance with their preferences during the experiment is a commonly held assumption. This assumption might explain why interactions that are actually based on one-shot games are repeated up to 40 times during an experimental session.[28] Over the course of these repetitions, the behaviour of participants who gradually come to understand the structure of the game is expected to converge towards the equilibrium predictions, or at least settle on a distinct strategy. For this reason, it is also a common practice to exclude data from the first rounds of an experiment, i.e. the "learning" phase:

---

[27] With the original payoffs, the equilibria the authors had found under the condition of transparency were not robust to so-called "equilibrium refinements", which are formal criteria of selecting out of several possible equilibria in a game those that are more likely to occur (because they provide better outcomes for at least one of the players, for example). Apparently, making the payoffs robust to these theoretical criteria also helped participants to make better decisions in the experiment (i.e. decisions that were more efficient and thus profitable), a result that had not been anticipated by the experimenters.

[28] Repetitions of course also produce more data, but this does not result in more observations, because the behaviour of each group of participants over the experimental session as a whole only counts as one independent observation. The main reason why one-short interactions are repeated so often therefore really seems to be to observe how participants' strategies evolve over time and whether behaviour converges to an equilibrium.

P5 Something very standard is leaving out early stage data. That, that is standard. Because, many people know that, people also learn over time, so they don't just learn from the instructions, but they also learn over time, so it's more or less standard to leave out the first stages of the experiment, but one can also discuss them and say, well, so we do it in our joint work, because we know that they behave much more consistent with equilibrium over time.

H Mhm

P5 Because they are learning, by playing. That type of leaving out is standard, because of learning. (Transcript P5)

This statement implies a notion of participants' behaviour as something that can be and should be optimised in response to the experimental environment. Usually, optimisation means that participants gradually learn to make decisions that are in their own best interest in terms of making profit. The behaviour that the experimenters described as "very strange" for example was disadvantageous for participants in the role of decision-makers and was therefore interpreted as being rooted in a misunderstanding that could probably be solved by making participants switch to the other role.

Role-switching is methodologically possible because of the assumption that an agent's behaviour is a best response, given their preferences, to the strategic environment they are presented with. In other words, a participant does not need to "identify" with being a decision-maker or an advisor, but should be able to find a good strategy in each of the roles. The reason why the experimenters had decided to make participants change roles only once and not continuously in this particular case was that participants who know that they will change roles every other round might take advantage of this:

(P6) [If the] advisor accepts the payment in some state, he must give a dishonest advice. But he gets an additional payoff for this.

H yes

P5 If we continuously changed the roles, people may collude, it might become a norm to always accept the payment, because you know that you will be in that role very soon, and you accept the payment. So there can be collusion among people, which is not there in the theory, the theory is one-shot. And it doesn't say, oh, the decision-maker will become an advisor in the next period. But if your design is not like that, if the roles are continuously changing, people may just collude and it may be a norm to accept the payment and it may be a norm very easily.

H mhm

P5 So we didn't want to change the roles continuously, we didn't want to switch them too many times. So we switched only one time. Ahm, yeah, that was also not great (laughs), but that was better than not switching at all. (Transcript P5)

In other words, a continuous role-switch might have allowed participants to come to the implicit agreement that it is acceptable to always take the side-payment and give dishonest advice, because everyone will equally profit from this behaviour. Collusion would be detrimental for the experiment, because participants' decisions would be motivated by something else than the factors that were deliberately introduced as the incentive structure of the game. Collusion would also be problematic because it cannot be accounted for theoretically. The researcher described the risk of participants starting to "collude", if that was allowed by the experimental design, as one of the "behavioural issues" that make designing experiments a challenge. Because of these "behavioural issues", also theory-driven experimentation in economics requires specialist knowledge that goes beyond what can be modelled and predicted theoretically. This specialist knowledge must be acquired either through experience or supervision. Conventions such as leaving out the early stages data in analysis show that anticipating and handling "behavioural issues" is part of the standard practice of experimental economics.

5.2.2 Writing instructions

Much that is involved in designing an experiment comes down to presenting complex information in the most efficient and simple way. The instructions for the experiment on transparency in decision-making exemplify several strategies that experimenters use to make the experiment easier to understand: First of all, they introduce the task at hand as "making a series of decisions". Participants are then reminded that they can earn "a considerable amount of money" if they follow the instructions carefully, that the amount they earn depends on their own decisions and those of other participants, and that they will be paid in cash at the end of the experiment. This first paragraph therefore amounts to pointing out the monetary incentives that are considered as the main methodological tool to elicit "true preferences". In particular, it establishes Smith's (1976, 1982) precepts of *dominance* (earning money is the main motivation) and *saliency* (participants' decisions during the experiment will determine how much they get paid). Participants are then informed that they should remain silent and raise a hand if they need help. After these general remarks, the experiment itself is described. The description covers the two available roles ("sender" or "receiver") and the number of times that each sender will interact with various randomly chosen receivers (20 rounds). The receiver must choose between two options ("A" and "B") and a "spinner" randomly selects one out of two states ("L" or "R") that determines which option is better for the receiver. The probability for each of the states to be selected is then given in frequencies, presumably because these are easier to interpret than probabilities: "On average, the Spinner selects L 2 out of 5 times, and R 3 out of 5 times." It is pointed out that only the sender will know which of the two states was selected, that they then need to decide whether they will accept the side-payment and recommend one of the two options to the receiver. The note that the sender needs to recommend one specific option if he/she accepts the payment, and that the amount of the extra payment depends on the state chosen by the spinner is *italicised*. What follows is a recap of the structure of the game in bullet points, and an overview over the earnings that each player receives in every possible outcome. These earnings are expressed in points (between 1 point and 10 points) and the participants are informed that two out of all the rounds they played will be randomly selected at the end of the experiments. The points they have earned in those two rounds will then be converted into Euros at a rate of 3:1 and "privately paid to you in cash". This first part of the instructions takes up two A4 sheets in a large font with double spacing. They end with the request, in bold letters, to type a certain number code into the box on the computer screen and click OK, which means that after this first part of the instructions, the participants will need to answer some control questions. There is also a second part of further instructions (another A4 sheet), which explains how each player can enter their decisions during the experiment. This part is necessary, because the authors decided to use the "strategy method" to get more information about the strategic responses of receivers. The strategy method, introduced by Reinhard Selten, consists of asking a player what their response would be to each of the possible actions of the other player. In this case, the receivers, before seeing the recommendation of the senders, were asked to decide which option they will chose if the sender recommends A, and which option they will choose if the sender recommends B. The decision that matched the actual recommendation given was then implemented and determined the earnings for this round. This third page of instructions ends with the announcement that there will be a few practice rounds before the experiment starts. At the bottom of the page, there is again a sentence printed in bold letters telling participants which code to type in on the screen once they are ready.

Compared to other instructions I have seen, these are particularly easy to read, not only in terms of the language and wording used, but also in terms of the layout and the generous use of the space available. In fact, there is very little text compared to other experiments, and the researchers told me that they took great care not to make the instructions too long. For example, one reason why they would not use the strategy method with both players was that this would have been too complicated to explain in the instructions:

> (P5) [We] just didn't want to make the instructions longer. So you always run a risk. It would prob-, it might have been better to do that, but to do that, you also write more, and you might distract the subjects more. So there is always a trade-off, there is always a more detailed design, which you might like, but that would also make the instructions lengthier, and it may distract your subjects, which would in the end maybe not work at all. (Transcript P5)

The use of points or some kind of experimental currency unit is also common practice in experiments, and was explained as a means of producing "rounder numbers" that are easier to calculate with. As these instructions show, communicating that participants will be paid in cash according to the decisions they make, and that their earnings are private, is seen as paramount to ensure that the precepts for "inducing value" (Smith, 1976) and therefore controlling subjects' preferences are met. Additionally, it is necessary to ensure that participants read the instructions carefully. This is done here both by including an instruction for the next step on the bottom of the page (such that participants only know how to continue once they have read the entire page) and by making participants go through a tutorial before the actual experiment starts. All the experiments I discussed and helped to run included some form of control questions, and participants could only proceed with the experiment when they had answered all of them correctly.

In fact, participants who do not understand the experiment – either because of language difficulties or because it is too complicated – and therefore make "random" decisions, were described as a more or less inescapable threat by all of the researchers I interviewed. For once, it is almost always the case that some of the participants do not understand the instructions, and even by doing test rounds and questionnaires at the beginning of the experiment, this cannot be completely avoided. Although early-stage data can be left out, it is not permissible to exclude any individual observation from the analysis when it seems to be the result of “random” behaviour: “You don’t know whether it’s a mistake or there’s some strange reasoning behind it, that we don’t get. So it’s very dubious to like pick out things that don’t fit the theory and say ‚these were mistakes’, so, you just have to leave them in there, typically.“ (Transcript P3). That participants understand what they can do and how it will affect them is a precondition for making conclusions based on the experimental results. This explains why some of my respondents described writing the instructions as such an important and also challenging task.

One feature of all the experiments I discussed and observed that I found to be in contrast with making experimental instructions easily comprehensible was the emphasis researchers placed on formulating instructions for participants as neutrally as possible. For example, there is no mention of “decision-maker” and “adviser” in the instructions for the experiment above, but the two roles were referred to as “receiver” and “sender”. Likewise, in a different study on norms of solidarity, two options referred to by the author as “helping out” and “passing” would be named “Option A” and “Option B” in the experiment. Initially, this did not make any sense to me: especially when the research question itself was formulated in terms of clearly value-laden concepts, such as “solidarity” or “conflict of interest”, why should participants not know that these were the ideas underlying their interactions in the laboratory? In other words, why would the real-life problem that inspired the researchers in the first place be abstracted away in the laboratory?

This requirement for neutral language seems to be a result of the importance of aligning theoretical models and the experimental design. It is also another illustration of an expected discrepancy between theoretical predictions and actual behaviour. When an interaction is analysed from the perspective of game theory, i.e. finding out what strategies are available to the participants given their goals and the expected outcomes, then it should not make a difference what these goals and strategies are called, or how the players are referred to. Due to the structure and the specifications of the game, some strategies are simply preferable to other strategies (they are “dominant”). To use a well-known example, it obviously does not matter for the solution of the so-called “Prisoner’s dilemma” whether the situation is framed in terms of two prisoners’ being confronted with the possibility to confess, or whether we only speak of two “players” who are given the options to “cooperate” and “defect”. The story of the prisoners only serves as an illustration. In the words of my interviewee, “from a game-theoretic perspective, labels should not matter” (Transcript Participant 3).[29]

The fact that how a situation is framed does not make a difference in theory results in a strong conviction that if an experiment is meant to produce the same results as a theoretical model, it should do so without any framing. On the other hand, economists of course know that it makes a difference in prac-

[29] The same interviewee mentioned that other aspects that also should not matter from a game-theoretic perspective are indeed considered in experimental design. For example, the difference between payoffs should be big enough to be perceived as a motivation by participants, even though the size of the difference in payoffs is insignificant for the theoretical predictions.

tice whether they call an earning a "payment" or a "bribe". Since the experiment should implement the model as exactly as possible to confirm or falsify its predictions, using language that has strong connotations is seen as an illegitimate way of producing stronger results:

> (P5) So, in economics it is perceived as if, if you get the result in the lab, with that, with framing, it weakens your theory part. There is this thought: you obtained your results because you framed it. So it's more, it's assumed to be more successful if you obtain the result without the help of the frame. So I think it's to make the results more robust. If you get the results with framing, people are like, yeah, we get it, cause you frame, you just call it what, bribe, third party, and you got the result, not because of the theory. So people think you're not very successful, I think that was one reason why people started using neutral framing more and more, to show the power of the theory.
>
> H Ok.
>
> P5 So that was one thing. We also thought, oh why don't we use framing, and so on, but then theorists say, well then, what is the success of the theory? You just get the result because of framing. (Transcript P5)

When using value-laden terms, researchers are also seen as giving up control, because they cannot be sure that the behaviour they are observing is a response to the incentive structure they impose, or whether it is the result of participants' interpretations of these terms that have nothing to do with what the experiment is designed to measure. For this reason, the researchers were certain that their chances of publishing the paper would have been very low if they had labelled the experiment from the start. In this particular case, using neutral language was explained to me as a concession to current academic standards, intended to avoid unfavourable reviews; the researchers themselves would have actually been interested in using framing for their experiment. There are some additional reasons why framing the experiment in a particular way was not an option here. The only possibility they saw for doing so would have been to use framing as a treatment condition and compare the effects of different framings on behaviour, which was not feasible because running additional sessions would have exceeded their budget. This is considered a permissible use of suggestive language, but it also corresponds to a different epistemic interest in how situational cues affect behaviour. In other words, studying the effects of framing is a suitable question for behavioural research, but it is of less importance in theory-driven experimentation. The aim of this particular study was to study theoretically and experimentally whether transparency can improve decision-making in general. Framing the experiment as a particular situation, the researchers explained, would therefore have also lowered the generality of their results.

5.2.3 Inducing behaviour

As was mentioned earlier, experimental methodology is characterised by the tension between actively producing the behaviour that is of interest to the experimenters and at the same time making sure that participants are not "influenced" so much that they have no choice but to conform to the experimenters' expectations. Santos (2010) refers to this tension as the trade-off between experimenters' control and participants' agency: Through the experimental design, researchers try to take control over participants' behaviour just so much that their agency is still a valuable input to the results, but enough that their behaviour can be attributed clearly to one of the factors that are deliberately introduced by the researchers, rather than anything else. Anticipating how subjects are going to behave plays an important role in the designing process, and the accumulated experience within the scientific community of experimental economists is particularly valuable to this end. For example, it is generally known that student populations behave very cooperatively in laboratory experiments. This is easy to explain, since students, even if they do not know the identity of other participants, know that in general they will play against their peers. If an experiment investigates whether a certain change in design will enhance cooperation in comparison to the control treatment, it is therefore important that there is room for improvement in the baseline rates of cooperation. Consequently, experimenters "stack the incentives against cooperation" in the control treatment, which means that they choose a design and payoff values that make cooperation a very risky choice (for example, if cooperation means relying on reciprocity, and the "favour" you are doing your partner is not returned, you will not earn anything in this round or even make a loss).

In one of the experiments I discussed, this strategy worked out, and the change introduced in the treatment condition resulted in higher rates of cooperation than in the control treatment (i.e., participants made mutually beneficial rather than selfish choices). In another experiment that one of the researchers told me about during the course of the interview, the strategy failed and the persistently high rates of cooperation in one of the control treatments turned into a serious problem for the researchers. This experiment was about trust in a lender-borrower relationship. To study how different rules that protected the lender would affect trust (i.e. cooperation) rates, there had to be a treatment in which trusting the borrower to return the money was very risky, because there was no protection. To the researchers' astonishment, participants' behaviour only counteracted the theoretical predictions in this "no trust" treatment:

> P5 In the, in the lab, people continuously cooperate in this game with no trust and high trust.
>
> H Ok.
>
> P5 That was, it's too risky in theory, they shouldn't, but they just always cooperate. In the other game, everything is similar to the equilibrium predictions. And consistent, and this control treatment is, is not consistent. People just send money all the time, and return money all the time. They are very nice. And I found it very interesting. In reality I'm not sure people would care, what I'm trying to say is the borrower should default, so we wrote parameters in a way that you should just default in the first round. Stop the relationship. But they don't do it.
>
> H Mhm
>
> P5 Yeah, so.
>
> H So, here you don't really know if that is going to be a problem, yet? Or, it's just
>
> P5 It's in such a stark contrast with what we predict!
>
> H Yeah
>
> P5 We don't know what to do, we don't know how to justify. That's why we did the follow-up experiments, we changed the parameters, it's the same
>
> H It doesn't go down.

> P5 No, it just doesn't go away. So, we thought, what can we do, we'll write it like this and, and we write possible explanations, send it, and if referees like it and say, well do this and that follow-up experiment, we'll do it. Otherwise, I don't know. (Transcript P5)

The researcher had no satisfactory explanation why the cooperative behaviour persisted, since it would be very unlikely for lenders to give money and borrowers to return money, instead of defaulting, if the lenders have no securities and no way of enforcing cooperation. In the interview excerpt above, this phenomenon is therefore identified as an artefact of the laboratory situation; "in reality", people would not be so trusting and cooperative. The number of sessions they had run made it unlikely that this behaviour was due to the participants knowing each other personally, and the fact that they even had to do some of the follow-up sessions in a different laboratory (because they had "consumed" the subject pool in Vienna) ruled out that it was a local phenomenon. Their general explanation was just that students tend to be "very nice" in the laboratory, because the laboratory as such has "high social capital".

It appears that the fact that participants did not behave according to predictions was less of a problem than the fact the researchers could neither come up with a satisfying explanation for the deviations, nor reduce them by changing the parameters or running more sessions. After all, from years of behavioural studies it is known that student populations tend to be cooperative and that in general, people do not behave as rationally and selfishly as is assumed in theoretical predictions. In the experiment on transparency and lobbying, for example, the researchers found that even in the treatment where there was no transparency, a certain percentage of the advisors would always reject the side-payment and tell the truth, even though this was not in their best interest. The researchers reported this result in their draft paper and explained that it was most likely caused by an inherent *lying aversion* in some of the participants; this deviation from predictions was therefore not entirely unexpected. Where an unexpected deviation can neither be explained nor reduced with additional sessions and adjustments, this on the contrary indicates a lack of control on the part of the experimenters. Control, after all, means that all the conditions and factors influencing behaviour can be at least known and, ideally, manipulated at will. Lying aversion, for example, is a phenomenon that has been established and systematically investigated in earlier experiments, and the same is true for other behavioural regularities that result in cooperative behaviour. These factors, which are inherent in participants, cannot be fully controlled, but they are known and can serve as explanations. The high tendency for cooperation in student populations is also known, but so are the different design strategies (e.g. making cooperation a very risky choice) that normally serve to significantly reduce cooperation rates such that they are more in line with theoretical predictions. Where this does not succeed, as in the example above, it results in a breakdown of the experimental *three-way coherence* between conceptual understanding, instrumental model and material procedure described by Pickering (1989) as the epistemic goal of experimentation. The material resistance presented by overly cooperative behaviour causes this breakdown. Since the *phenomenal model* (the experiment was based on a theoretical paper and game-theoretic model) and the *instrumental model* (the methodological tenets of experimental economics) are fairly rigid, the only way of accommodating this resistance is by changing something in the *material procedure.* Consequently, the flaw is first suspected to be an effect of insufficient randomisation of participants (which is why additional sessions should help, as they would statistically even out this effect), or of using parameter values that motivate participants in an unintended way. Having excluded both of these possibilities, the researchers apparently were at loss for an explanation, and progress on the project had momentarily stalled. Since a considerable amount of money had already been invested in the sessions, simply giving up on the project was not a viable option; the best strategy, at least at the time of the interview, was then to submit the paper and wait for anonymous referees to suggest new adjustments.

The observations described above open up some further questions concerning the relationship between theory and experiment, and the role that participants play. The requirement to use neutral language, for example, again points to a primacy of theoretical reasoning. It also invokes the view that the behaviour described by economic theory is more fundamental and general than any kind of behaviour that is a response to social or moral factors. If experimenters want to study "general" phenomena, or develop and test generally applicable models, they need to make sure that context-specific interpretations of the situation are excluded and that what they observe is only a response to the experimental procedure they designed. The common concern for enabling participants to "learn" how to behave strategically additionally indicates that the "economic" behaviour that can be described in theoretical models needs to be carefully elicited and encouraged. A majority of participants, apparently, will not behave in best response to the experimental environment right from the start. Since it is known that student populations tend to be very cooperative in the lab, some types of behaviour even need to be actively discouraged. Experimenters in general seemed to be quite confident that they are able to deliberately induce and preclude certain types of behaviour, as long as they can anticipate the possible reactions of participants in the design process. Yet there are always limits to how successful these strategies are, partly because of the cognitive limitations of participants, and partly because some types of non-rational behaviour are apparently very persistent. All of this indicates that experimenters' conception of experimental subjects is that of a malleable, but not fully controllable resource. The malleability of behaviour, the fact that participants can be trained to adequately respond to the experimental situation and pursue monetary interests, is what allows experimental economists to study behaviour in the laboratory in the first place. Referring back to Muniesa's (2014) claim, experimental economics takes advantage of the fact that economic behaviour can be actively "provoked" by exposing ordinary people to the right conditions. That this resource is not perfectly controllable, that participants will sometimes behave according to other preferences than those assumed by the experimenter, is what presents the epistemic advantage of experiments over theoretical models. This ambiguous conception of participants' role reflects the tension, described earlier, between the aim to evaluate and build theories with the help of experimental results, and the aim to study behaviour that is not adequately described by economic theory. Experimenters ultimately aim to be able to predict behaviour with the means of economic theory, but if they were actually able to fully predict it, there would be nothing to be gained from doing experiments. A consequence of the epistemic value of unpredictability is the risk that the assumptions that experimenters make about their subjects' preferences are inadequate, and that subjects consequently react to the experimental setup in a way that was not anticipated beforehand and cannot be explained. This is the case, for example, when other preferences that were not taken into consideration override the incentive structure of the experiment, or when participants interpret the situation in a completely different way than was intended by the researcher. Ultimately, such unpredictable aspects are inherent in the nature of research on human behaviour and therefore irreducible, but much of what experimenters do can be interpreted as strategies of accommodating the risk they entail. Along with ensuring participants' comprehension, accommodating the risk of alternative preferences and interpretations is precisely the concern that the processes and material arrangements that make up the laboratory seek to address.

5.3 In the Laboratory

The laboratory, as a theoretical notion, can be described as processes through which a reconfiguration of natural objects (Knorr Cetina, 1992, 1999), a difference between inside and outside (Guggenheim, 2012), and in general, enhanced control over both the material and the background conditions are achieved. In experimental economics, the laboratory achieves this reconfiguration not only through spatial and material arrangements, but also because of the processes of recruiting and managing participants. For the experimental economics laboratory to function, it is important that a steady flow of new subjects can be easily attracted and recruited. The laboratory therefore has to be easily accessible, and it has to be advertised in such a way that the reward structure, i.e. the monetary incentives normally used to elicit "true preferences", is present from the very beginning. Consequently flyers, posters and the website advertising the experiments to the local student population do that with the slogan "Earn Money in Economics Experiments"[30]. Also, the lab is centrally located on the university campus so that participants can easily fit experiments into their daily schedules and take part in between regular university courses. The advertising of experiments and the location of the laboratory on university premises finally makes sure that the participant pool will consist of the typical "convenience sample" of the local student population. Since this is the case with all experimental economics laboratories, it enables a certain amount of comparability between experiments conducted in different geographic locations (although that comparability is sometimes questioned by experimenters themselves).

---

30 For example, on the registration website where new participants can sign up for the database: Vienna Center for Experimental Economics. (2017) Earn money in economics experiments. Retrieved from http://www.univie.ac.at/orsee/orsee-3.0.3/public/, 06.02.2017.

Although it is embedded in the university campus, the laboratory as a place is also very distinct from its surrounding environments. It achieves this difference between the outside – which, apart from being a university environment, also serves as a recreational space both for students and the general public – and the inside, a place with a dedicated function and use, by addressing the participants in their function as experimental subjects before they even enter. The door leading to the lab from a small hall on the ground floor of the building always sports a sign before and during experiments, which either reads “Please wait here for experiment” or “Experiment in progress, do not knock”. Such a sign, along with the creation of three makeshift workstations that were as far apart as possible, also sufficed to turn a normal lecture room into a laboratory for the first experiment that I observed; upon seeing this sign, participants take on the role they had signed up for when registering for the experiment, and wait for further instructions.

The lab itself is a spacious light room with a row of windows on the right wall. All surfaces are either light grey or made from glass. Right after the entrance, there are toilets and a small kitchen to the left and the control room to the right. The kitchen is sometimes used by lab assistants to store drinks and make tea. As I noticed later on, neither the kitchen nor the restrooms were renovated when the lab was installed, and especially the kitchen surfaces show their age, presenting a noticeable contrast to the shiny and clean appearance of the laboratory space. The door to the control room has a window with an opening above a little windowsill, very similar to a counter at the railway station, for example. This is where participants show their ID to one of the assistants when they arrive, and where they get paid after the experiment. The control room hosts all the materials necessary for experiments (mostly paper, instructions, pens, and other office material - there are desks, cupboards and shelves for that), including a safe where the money is stored, two servers (Windows) with monitors, a Linux server with a monitor, and a printer. Lab assistants leave their stuff in the control room and try to hide it away (especially when it is food), in order to keep up tidy appearances. On the front side, the control room is separated from the lab with a white wooden partition, similar to a shop counter. In between this counter and the glass partition where the cubicles begin, there is some room where extra subjects are asked to wait for their turn to either still participate in the experiment or to get paid the show-up fee. There is a glass door in the glass partition, which is usually left open. Participants queue at this glass door for their turn to get paid after the experiment. The glass partition makes it possible for the experimenters to observe the laboratory from the control room, at the same time as reducing noise from the control room. Behind the glass partition are the 28 computer cubicles where participants are seated. These cubicles are separated by tall grey panels; there is a desktop computer with a monitor, a keyboard and a mouse, an office chair and a pen in each cubicle. All in all, the laboratory has a cool, elegant and professional atmosphere. It is clearly designed to serve the purpose of conducting experiments only, and does not invite for spending more time in it than necessary.

Once inside the laboratory, participants’ actions are reduced to a clearly defined range of possible tasks, which consists of: showing their ID at the window to the control room and receiving a seating card; sitting down at a work station or waiting to be selected as an extra in the designated waiting area in front of the control room; performing the tasks specified in the instructions for the experiment; raising their hand and waiting for an assistant to answer if they have a question; queuing for getting paid after the experiment has ended; and using the toilet (which they should ideally do before the experiment starts, or after it has ended). The lab rules specifying what is and is not permissible to do are on display in every cubicle. Participants may drink something during the experiment, of course, and some of them even bring something to read with them, which was tolerated at least in one case I observed. Since there is a pen and paper in every cubicle, participants often start drawing when they get bored, and there are some nice products of particularly elaborate doodling on display in the control room. In general, the laboratory as a space affords only a limited set of activities, and the prevailing codes of conduct that are communicated explicitly and implicitly also discourage any activity that does not serve the purpose of conducting experiments. This is even true for the lab assistants. Although the assistants are typically in control of the procedure and possess greater knowledge and familiarity with the place, they are themselves subject to a set of rules and norms that corresponds directly with the epistemic function that the laboratory serves in the practice of experimentation.

5.3.1 Codes of conduct in running experiments

The laboratory of the VCEE is administrated by a designated lab manager, a post-doc researcher who is employed full-time by the university and does not have teaching duties. Jean-Robert Tyran, one of the two heads of the VCEE around whom the research centre was established in 2011, told me that having a full-time lab manager had been one of his main demands when negotiating his transition to Vienna with the university. The lab manager schedules experimental sessions, assigns participants and lab assistants, and deals with the overall financial, administrative and technical issues concerning the lab. The experimental sessions themselves are run by lab assistants, sometimes together with the researcher who designed the experiment. Lab assistants are typically undergraduate economics students who are employed for working a maximum of 40 hours per month. If the experimenter is not running the experiment him/herself, there will always be two lab assistants at each experiment. After I had been accepted as a lab assistant, I first helped out with an experiment that did not take place at the lab (which is rather unusual for VCEE research). My next time was a training session, in which the experimenter (a PhD student), another lab assistant, and the lab manager would run the experiment and instruct me at the same time, after the lab manager had given me an introduction to the laboratory. I did several experiments together with more experienced lab assistants, until I was assigned to be "responsible" for an experiment, together with the researcher (it was the same PhD student who had already run my first laboratory experiment). However, another lab assistant was still present at this session in case I would need help.

On my first visit to the laboratory, the lab manager handed me a copy of the VCEE *Guide for Running a Session*. These guidelines are 15 pages long and consist of a checklist along with detailed description of each task that is involved in running an experiment, and when it should be carried out. It starts with a reminder to hang a sign on the lab door which tells participants to wait for the experiment, then describes exactly all the different steps of starting the computers, the computer programme, preparing the forms necessary for accounting and counting the cash for paying participants, welcoming participants to the lab, paying them, collecting data and tidying up afterwards (including shutting down the computers and turning off the lights). It also contains guidelines for handling unexpected events and more common problems, such as dealing with latecomers or participants who do not have an ID. As is clearly stated on the first page, this guide serves the aim of standardising the practical procedures involved in running an experiment as much as possible:

> We have two goals when we run experiments: 1) to run all sessions as similarly as possible, to minimize any potential "experimenter bias", and 2) to treat our subjects in a fair and consistent manner. The following guidelines are written with these goals in mind, so please try to follow them as closely as possible.[31]

Avoiding experimenter's bias means that it should not make a difference for participants' behaviour who runs the experiment (e.g., the experimenter should not influence the participants by being particularly friendly or unfriendly). The guidelines, apart from being a useful checklist that prevents lab assistants from making mistakes or forgetting one of the necessary tasks, therefore also are an instrument for establishing control over the background conditions of the experiment. This is most explicit in the section that contains a photo of how every cubicle should look like before the experiment starts; the lab assistants are also reminded here that they are responsible for leaving the lab in this condition after each session. Providing a tidy and clean environment was mentioned on several occasions as being an important part of the lab assistants' work. As the lab manager formulated it on my first day: "But the more clean it is, the more standardised it is! If it is very aseptic, you have a boring environment that is always the same" (Transcript Field Notes). Interestingly, the lab manager here borrowed a notion of "contamination" that is typical for wet labs and medical environments. To keep the environment *aseptic*, it is necessary to remove all kinds of distractions, and reduce details that cannot be held constant across sessions to a minimum. Otherwise, just like germs in a wet lab might contaminate biological data, unknown and uncontrolled background conditions might contaminate the data of economics experiments.

Apart from the general guidelines for running a session, lab assistants are usually provided with specific instructions, i.e. a protocol to follow for each individual experiment. These stated, for example, how many participants there should be, if there are any special requirements (e.g. recording participants' conversations or matching them into groups in a specific way), and how participants should be instructed. Usually, participants also receive written instructions at the beginning of the experiment. In order to minimize the potential for experimenter demand effects, lab assistants do not read out the instructions aloud (and one of the assistants was rather unhappy when a researcher did so while run-

[31] Vienna Center for Experimental Economics. (2016, March) *Guide for Running a Session*,.

ning his own experiment, particularly since they had told him not to do so before). This is an example of the informal codes of conduct that are not written down, but rather communicated when prompted. Another example is the requirement to monitor participants' progress without making them feel observed. On one occasion, when I was walking down the aisles between the workstations to see if any of the participants needed help, the other lab assistant chastised me for looking at a girl for too long: "Don't stare at them. We have the principle that if you know you are being watched, it changes your behaviour." (Transcript Field Notes). Since apart from the standardisation of background conditions and avoiding experimenter demand effects, randomisation and anonymity were often explicitly mentioned to me as paramount for conducting experiments, I will discuss both in more detail now.

5.3.2 Randomisation

Random assignment of participants to different treatment conditions is a basic principle of any research involving human subjects, be it behavioural or medical. Without randomisation, researchers can never be sure whether the treatment effects they are observing are not actually selection effects, caused by the specific characteristics of the group of people that were (self-)selected into one of the treatments. Randomisation also serves as an insurance against all the intrinsic individual factors that cannot be controlled by experimenters:

> P3 I mean there clearly are things you can't, can't control for, like different preferences, what a person had for breakfast, right? I mean, these are still kinds of things you can't control for, what you just hope is that these things are randomly distributed between your two treatments, so you'll not falsely find a treatment effect. […] [Whenever] you're looking at an experiment, there are things that maybe you think, oh, we need to control for this, and maybe they do influence behaviour, but as long as they influence behaviour in both treatments in the same way, that doesn't matter. (Transcript P3)

Although it is such an important principle for ensuring the credibility of experimental results, random assignment of participants appears to be more of a presupposition than a prerequisite that is actively enforced in practice. The lab manager usually selects a large number of registered subjects from the VCEE database and invites them per email to participate in experimental sessions; subjects then register themselves for the sessions they want to participate in. There are therefore limits to randomisation, particularly when the different treatments conditions are not implemented within the same session. Participants who normally sign up later might consequently sign up for other treatments than those normally signing up early, which could result in a selection effect (if the intrinsic characteristics that make them sing up early somehow also affect the behaviour that is studied in the experiment). This became apparent to me in an experiment that required groups with a particular gender ratio. While running the experiment, one of the lab assistants started worrying that the requirements could not be met, because overall, more females than males had signed up for the sessions. The lab manager, who was also present that day, answered that the gender ratio might even out later on, since males tend answer more slowly to invitations and therefore participate in the later sessions. The lab manager's tentative explanation for this phenomenon was that female students use their smartphones more often and are therefore quicker to respond to invitations. "Ideal methodology", I was told, would require that all treatment conditions be implemented in the same sessions, so that self-selection into a session does not entail self-selection into a treatment.

While these more subtle selection effects are not always taken into consideration, researchers do take care not to run the sessions with different treatments under obviously different conditions. For example, since most participants are students, the subject pool might differ substantially during the winter and summer breaks, and some researchers take care to run all of their sessions during term time:

> P3 We can get subjects during the holidays in Vienna, but it's gonna be a different, a different subject pool than if we do it during term time. Because we won't have Erasmus students, ahm, yeah, maybe richer people more like to be on holiday, poorer students more like to stay in Vienna and be working, ahm, and so we were careful to have the third treatment during term time as well. (Transcript P3)

What season of the year it is could also make a difference, and even running all the sessions of one treatment in the morning and all the other sessions in the afternoon might result in false treatment effects and should be avoided. Also, if additional sessions are needed (e.g. because they are a condition for publishing the research) it is considered good practice to do them in the same laboratory as the original sessions, which in some cases might require researchers to return to a university they have left in the meantime. The reason for this, I was explained, is that both the composition of the subject pool (in terms of whether there are more economics or more humanities students, for example) and the economic situation of the student participants might be different across universities. Sessions for the same experiment could not be performed both in Copenhagen and in Eastern Germany, for example.[32] Potential differences in the subject pool, which are caused by the long time-lags between original and follow-up sessions, cannot be avoided in these cases.

In the laboratory itself, random assignment is established with the use of seating cards. Before the experiment starts, lab assistants prepare as many of these small plastic cards as participants are needed for the experiment, plus a few cards for "extras". The cards for participants have a number between 1-28 that corresponds to one of the workstations that will be used for the experiments; the cards for "extras" have a number higher than 28. These cards are then shuffled and given out to participants when they enter the laboratory and show their ID to the assistants. Participants who receive a number that corresponds to a workstation can take their seat right away, while participants who receive an "extra" number need to wait whether they can take the place of another registered participant who failed to show up. "Extras" who are not selected in the end do not participate in the experiment, but receive a flat show-up fee as compensation. The reason why the seating cards are shuffled is stated in the VCEE guide: "We don't want the people who show up earliest to always be selected".[33] On the one occasion that I witnessed when an "extra" was getting impatient about having to wait, an emphasis on randomness was also the answer he received from the lab assistant: "It is random, everything is

---

32 There are also experiments, which explicitly aim to study "cross-cultural differences" by running the same treatments in different countries.

33 Vienna Center for Experimental Economics. (2016, March) *Guide for Running a Session*.

chosen randomly." (Transcript Field Notes). This was an accurate response, but it still surprised me, since I personally would have explained the situation with our need to have a number of participants that is a multiple of four (and at that point, only 19 participants had arrived, so we would have needed to send 3 extras home if there had not been one latecomer still). The lab assistant, who was much more experienced than me, in contrast responded with an explanation why this particular person was selected to wait, rather than providing information about the experiment itself. This answer was meant to point out the neutrality of the selection procedure, and to avoid disclosing any information about the experiment, in case the extras would really need to be sent home and could then participate in another session of the same study. Having additional information about an experiment (or prior knowledge of the type of game that it is based on) can "spoil" a participant for further sessions, but the same is the case if participants are distrustful about the procedures and suspicious about the reasoning behind the experiment. Just as important as actually providing random selection is therefore making sure that participants *believe* that they are randomly selected.

5.3.3 Anonymity and privacy

The use of seating cards serves the additional function of assuring participants' anonymity during the experiment. Since participants are assigned to workstations with a randomly drawn number, they are only identifiable via this number during the experiment. After the experiment, they receive the payment that corresponds to their earnings during the experiment by returning the seating card to the lab assistants. The payment of each number is automatically calculated by the computer programme; the link between the number and the name of the participant is only established when they show their ID again, so, that the amount they are about to receive can be entered next to their name in the receipt that they then need to sign. This receipt, which is simply a list of the registered participants, along with the amounts they earned and their signatures, is the most critical breach of anonymity happening during the experiment. In particular, participants can see the earnings of other participants and know that others will also see how much they have earned. One of the PhD researchers running their own experiments therefore worried that this knowledge might influence future behaviour in experiments, because participants might not want to be known as the one "ripping others off".[34] To avoid this, they subsequently used a cover sheet for the receipts, through which participants could only see the line with their own earnings when signing. The downside of using this cover sheet was that it slowed down the procedure of paying participants considerably, which is probably the reason why none of the other experimenters used such precautions.

Anonymity is further created through the grey partitions, which separate individual workstations into cubicles, so that participants cannot be observed during the experiment, except by somebody standing directly behind them. They are therefore required to raise a hand when they have a question, and the

[34] In many behavioural experiments, e.g. public goods games, particularly high earnings are usually the result of free-riding, i.e. taking advantage of others' contributions while not contributing oneself.

lab assistants need to walk up and down the aisles to see if anyone looks like they are unsure what to do. There is a more indirect way of observing participants' progress through the experimenters' computer in the monitor room on which the experiment is run. The z-Tree programme shows how long each of the "z-leafs", i.e. each of the participants, takes to complete one step in the experiment. If one of them takes particularly long, a lab assistant will walk up to this workstation and see if the participant has merely forgotten to press the "continue" button after finishing their task, or is otherwise confused. This indirect monitoring has the considerable advantage that there is no need to "stare" at participants and make them feel observed, which is exactly what the creation of anonymity should prevent. Creating a feeling of anonymity, after all, was described to me as probably the main function that an experimental economics laboratory should serve:

> P4 But what's important I think is that people have the impression that there's anonymity, so there's some labs which actually have curtains, and people enter one after the other, so they never actually meet before. And I think that increases anonymity a lot, ahm, especially if your subject pool is not that big. If your subject pool is very large, the chances that you meet someone you know is not that high, but the subject pool is very small, and then in, before entering the lab you already meet someone you know, might affect results, probably not that much, but you never know, you don't have the control. […]
>
> (H)Why is anony-anonymity so important?
>
> P4 Ahm, because ahm we don't like, we know that social factors have a huge influence on behaviour, so if you, if you have like social norms, if you have the feeling that you are observed, that you're playing with a friend or someone you know, that might change behaviour, and when you're anonymous, if you have the feeling I'm not going to be observed, the experimenter doesn't know that it's me who's, ah, behaving in that way, I'm not playing with a friend, you are playing like very, like you are more playing the way your true preferences are, and which are not affected by other outside factors in that way. (Transcript P4)

The VCEE laboratory, which does not have curtains, but a large subject pool, was considered as being very "standard" in the level of anonymity that it provided, and certainly better than other labs which only use paper walls as partitions between workstations.

The effects of various levels of anonymity are also sometimes studied in experiments. I worked on the sessions of one experiment in which participants were asked to bring a friend, and would meet and briefly get to know a third person before the actual experiment started. The game they would then play was a public goods game, i.e. a type of game in which social preferences are known to affect outcomes. The implication of such research is that the less anonymous a player and the more "social" the interaction with other participants is, the more other considerations than the incentive structure of the game will play a role in decision-making. This experiment aimed to measure just how much of a difference social interaction and friendship made. In experiments that do not explicitly study such factors, social interaction is seen as detrimental. As the interview excerpt above suggests, the actions participants take are then considered to be a result of "social factors" rather than exhibiting their "true preferences". According to the same researcher, this is also the reason why surveys and interviews should not be expected to fully elicit participants' preferences:

> (P4) Many people will say "yeah, of course", like, because they are in an interview situation, there's an interviewer sitting in front of them, and they want to seem, ahm appear like a nice person with social

> preferences and so on, comply with some social norm. So in that, you, you don't have the control here. It might, might be that some people answer really honestly, and other people are driven by the social norms. (Transcript P4)

Conversely, where experimenters want to avoid the effects of "social norms", they need to make sure that personal exchange and communication between participants and the personal identification of participants are kept to a minimum during the experiment, and ideally, also before and after. Establishing conditions in which the influence of social norms can be reduced to a minimum on this account also amounts to establishing experimental control.

Anonymity in the laboratory is closely related to the precept of *privacy* that undergirds traditional experimental economics methodology (Smith, 1982). Privacy means that participants only know about their own outcomes. As Santos (2010) mentions, this precept is often violated in behavioural experiments, for example, when participants need to decide how to split a sum between themselves and another person. According to my respondent, such experiments therefore present participants with a trade-off between maximising their income and complying with social norms:

> (P4) And in the lab you have the advantage that you can incentivize behaviour, so it's not a hypothetical question, but a question which actually affects the payoff of the one taking that decision, and you have this anonymity, so it's not the social norms, of course can still affect behaviour, but not affect that someone, they are talking directly to an interviewer, for example, or that they think, well it doesn't cost me anything to comply with the social norm, so what. In this case, they, they have this trade-off between complying with the social norm, or increasing their own payoff, and that's a much better way to isolate the, their real preferences in that case. (Transcript P4)

The main assumption seems to be here that people in general comply with social norms when this is either cheap or really reflects their preferences, for example, because they care more about equality of incomes than the size of their own income. In environments where being social comes at a cost, participants are more likely to act according to their true preferences. Experimenters still need to take care that social norms do not become too dominant in the experiment, except if the influence of social norms on behaviour is precisely what they want to study. Since experimenters cannot know how exactly "social factors" are going to be interpreted, relaxing privacy and anonymity is seen as carrying the risk of losing experimental control. Reducing social connotations and social interactions is therefore a standard requirement for conducting experiments, which both the experimental design (for example, by avoiding suggestive language) and the laboratory procedures need to fulfil.

The different practices of establishing anonymity and reducing social connotations can also be interpreted with regards to Knorr Cetina's (1992, 1999) argument that laboratories can be described as processes of reconfiguration. Their epistemic function is to make the objects of study approachable and manageable for researchers by taking these objects out of where, how and when they naturally occur and subjecting them to a "social overhaul", whereby "the processes are 'brought home' and made subject only to the local conditions of the social order. The power of the laboratory (but of course also its restrictions) resides precisely in its enculturation of natural objects" (Knorr Cetina, 1992: 118). Knorr Cetina argues that in the course of this enculturation and the reconfiguration of both natural objects and researchers, a new order emerges in the laboratory that is neither purely natural nor purely social. In experimental economics, behaviour is produced in an artificial setting that is structurally similar to the situations in which it naturally occurs, but better observable, simpler and easier to control. In order to produce economic behaviour, the economic agents themselves need to be *laboratorised*. Since participants in laboratory experiments (as well as the experimenters) are social agents belonging to a variety of social orders themselves, laboratorising them essentially requires a certain amount of de-socialisation and de-personalisation. This is achieved through implementing the principles of anonymity and random (i.e., impersonal) selection above, through maintaining a standard and neutral environment, and a reduction of interpersonal communication to standardised (again, impersonal) interactions with the lab assistants.

5.3.4 Managing participants

A tool that further helps experimenters to reduce personal contact with participants and reduce individual students to experimental subjects is the ORSEE software package (Greiner, 2015), which is also used by the VCEE. ORSEE provides a database and a web interface both for participants and experimenters. At the VCEE, it is used by participants (who sign up for the database and enrol to individual sessions themselves), the lab manager (who creates experiments and sessions, and assigns participants to be invited), and the lab assistants (who need to fill in information about which of the registered participants actually participated after each session). Anyone can in principle sign up for the database, but in practice, the subject pool consists mostly of students. When signing up for the database, new prospective participants need to provide the following information: their first name, last name, an email address, what language they would like to receive emails in (German or English), whether they want to participate in experiments in English or German, their gender, their main field of studies and the year they began to study. Once they start enrolling in experiments, additional information is collected, which concerns the number and type of experiments they have participated in (for example, was it a public good experiment or a voting experiment), and the number of times they registered, but failed to show up for a session. The lab manager can use this information when assigning participants for a new experiment. Sometimes, as mentioned above, experimenters require a certain gender ratio; sometimes it is required that subjects have not participated in the same type of experiment before. There is also the option of excluding participants based on their field of study or if they have been to the lab more than a specified number of times. One of the PhD researchers told me that they had started to exclude very experienced subjects, because they would visibly behave differently from other participants due to their familiarity with the experimental procedures. Some of the other researchers told me that psychology students are often excluded, for the reason that they routinely participate in psychological experiments using deception, and might therefore second-guess the instructions they receive in economics experiments. One of the experiments I assisted in required that subjects did not know anything about public goods games; this resulted in a veritable recruitment campaign, where lab assistants promoted the experiments particularly among humanities students.

In general, very little is known about individual participants in an experiment, and information beyond the few features above, which determine participants' eligibility for specific experiments, is clearly not considered as relevant. In fact, the only piece of information that is routinely gathered during experiments is whether participants do take time to think about their decisions, instead of responding with an intuitive reaction. This is measured with a standard "cognitive reflection test" (Frederick, 2005) at the end of the experiment. Participants who perform poorly in this test are not excluded from the experimental data, but the cognitive reflection levels can be used as control variables in the data analysis. My interviewees told me that it is not particularly common to collect and control for demographic information in laboratory experiments, except when the purpose of the experiment is explicitly to study a correlation with demographic variables. One researcher, who had reported demographics in their draft paper, told me that the referees actually recommended them to cut out this part.

All of this suggests that the transformation that turns individual students into experimental subjects has a similar effect as the transformation that real-life problems undergo when they are reconceptualised as phenomena that can be studied in laboratory experiments. By abstracting away from particularities and individual characteristics and reducing them to a few salient features, students are transformed into members of the generic subject pool of experimental economics.

The inferences that can be drawn from experiments with this "convenience sample" of the student population are limited to situations with a similar group and similar conditions; in practice, this means that they are limited to other experiments. To apply them to any particular situation or group, further studies (e.g. field experiments) and additional qualifications are needed. Experimenters do not necessarily consider this to be a problem, for two reasons: First, it is part of the research culture of experimental economics to study its own limitations, which means that some researchers do actively conduct studies on how experimental results differ with different groups of participants (e.g. Engel, 2011). Second, experimental studies are frequently described as presenting a "lower bound": "if you can't even find an effect in the lab, for something were you think there should be an effect, […] there might not be one" (Transcript P1). That the subject pool used in most experiments is not representative of the general population is therefore not seen as detrimental for the validity of experimental results. Given that "the majority of experimental economics (and theoretical work inspired by it) revolves around the phenomena created in the laboratory" (Santos 2010: 114), it is more important that the subject pool used in different laboratories is sufficiently similar to allow comparisons between different studies. In this light, it is somewhat surprising that so little information about participants is actually collected in experiments. For example, participants are not required to state their age when enrolling in the database (although the minimum age can be inferred from the year they began their studies). One researcher mentioned that the subject pool in Vienna appears to be more heterogeneous than is commonly assumed, especially by researchers from the United States:

P5 Heterogenic is good, yeah. But we don't write it in the paper, so it would be good, so, in the US, for example, when people participate in the experiment, it's almost always the case that they are below 24, 25. So, there it's good, it would be better if this was more of a common knowledge that in Europe, students tend to be older, for example.

H Yeah.

P5 It would be good if this was also known. So it's not a, not I, I don't really, I don't really care that much apart from the English thing, but it might also be that people alway-, always read papers thinking this is a student population below 24, whereas it's not. So it might be good if maybe we'd write this. (Transcript P5)

Having encountered participants who did not look like the "typical student", this researcher also noted that English language skills tend to be a problem in Vienna, especially with older participants. Even though participants are asked for their language preferences when signing up for the database, and even though there are always control questions, which ensure that participants have understood the instructions before beginning the experiment, there might be a participant every now and then who does not sufficiently understand the language. This is problematic for experimenters, because participants with comprehension problems may slow down the overall progress of the session or at worst make random choices that also have an influence on other participants' behaviour. On the other hand, conducting experiments in English is more convenient for the researchers, because it is usually required that the instructions are included in publications, and English instructions make the experiment easier to reproduce for other researchers. Also, not all of the researchers at the VCEE do speak German themselves. Yet when researchers do want to make absolutely sure that there are no language difficulties, and that the participants also do not suspect that other participants have language difficulties, they write the experiment in German.

Overall, a considerable amount of the practical work involved in conducting experiments goes into managing participants: recruiting them, selecting and assigning them to sessions, treating them consistently, ensuring random assignment to groups and conditions, preserving their anonymity, making sure that they understand the experiment and, when it becomes apparent that they do not understand, report this back to the lab manager so that they can be excluded from further participation. The *reconfiguration* that the object of experimental economics is subjected to in the lab is achieved through all these practices of categorising and selecting participants according to very few criteria, accommodating and controlling for their cognitive abilities and comprehension levels, and most importantly, taking them out of their everyday social context as a university student and subjecting them to a range of standardised procedures, in which personal communication and the relevance of individual traits are reduced to a necessary minimum. That experimental subjects through these processes are disciplined into behaving more like economic agents becomes understandable when considering the dominant role that theoretical reasoning plays in experimental economics. Even behavioural models are in principle based on the assumptions and conventions of standard theory, although they extend these assumptions by adding selected factors (e.g. an aversion against inequality) that can explain "social behaviour" under some circumstances. In order to be able to theoretically describe the behaviour they observe, experimenters therefore need to make sure that it does not deviate too much from that of ideal economic agents, except, of course, in those domains where a deviation from the theoretical predictions is expected.

5.4 Answering My Research Questions

My analysis so far aimed to describe the epistemic practices involved in designing and running economic laboratory experiments, and answer the question of how economists have appropriated this method for the purposes of economic research. Based on my observations, I would suggest that the appropriation of laboratory experiments as a method for economics is characterised by the promise of observing "real behaviour" that can nevertheless be described and explained from a theoretical perspective. For economists, laboratory experiments offer a means of creating conditions, which are simplified enough that the predictions of economic theories and models can be directly tested. Consequently, experiments either investigate an existing theoretical concept, or a real-life situation that can be reconceptualised and reduced to some "economic fundamentals". Much of the conceptual and practical work involved in experimentation aims at identifying such fundamental structures, and then ensures that the experimental system is reduced to exactly those structures and the causal factors that can be controlled and varied in the lab. Even though they are thus fairly artificial and constrained, experiments are perceived as having an epistemic advantage over theoretical models, because they allow researchers to study behaviour and decision-making under specified and known conditions without assuming perfect rationality. Models and predictions can be tested in a situation where, due to the participation of real human agents, some of the standard assumptions needed to theoretically predict the behaviour of ideal economic agents are relaxed. When the behaviour that is produced in the laboratory confirms the predictions, this shows that the theory is robust to the behaviour of human agents who are not perfectly rational; when the predictions are persistently contradicted, it presents an opportunity of conceptualising the deviations as behavioural regularities and accommodating them in more specific models.

In either case, experimentalists need to be able to explain what dimension in the experimental system has caused the behaviour they are observing. To do so, they need to tightly align the experimental procedure with their conceptual understanding of the phenomenon to make it theoretically tractable, ideally by building a formal model that is then implemented in the lab. Economic theory therefore provides experimenters with direction and a clear structure that is based on familiar conventions and concepts. The dominance of theoretical reasoning, however, sometimes requires researchers to trade off practical considerations with the methodological requirement to be faithful to their model. Although experiments can easily be done without developing a formal model first, it is considered as good practice to at least provide a theoretical account of the observed behaviour afterwards; where experimenters cannot provide a formal theory, this is perceived as a weakness that lowers publication chances.

This preference for models and theoretical explanations, but also the methodological assumptions about eliciting preferences through incentivising choices in my view reflect the belief that what is understood as ideal (rational, self-interested, …) "economic" behaviour is more general and fundamental than the context-specific "social" behaviour. Nevertheless, there is a vital interest in studying why and how such "social" behaviour occurs, and this is incorporated by the research programme of behavioural economics. Even in "behavioural experiments", the epistemic requirement of isolating causal factors seems to make it necessary to consciously produce conditions for observing "economic behaviour" (by discouraging cooperation, for example), such that the factors provoking social behaviour can be identified more clearly. The notion of experimental control that my respondents invoked therefore seems to be closely related to what they perceive as the conditions necessary to facilitate "economic behaviour". Factors that distract participants from the incentive structure of the experiment and therefore discourage economic behaviour, such as a lack of anonymity, a lack of monetary incentives, or using deception, are consequently what were most often mentioned as carrying the risk of losing experimental control.

The laboratory provides material arrangements and codes of conduct that make it possible to control the experimental environment. This is done by reducing the range of activities, communication and interaction that are available and encouraged within the laboratory to standardised procedures. Upon entering the laboratory, participants are taken out of their familiar social contexts and subjected to the specific social context of the laboratory, where their role is not determined by their individual characteristics (at least not beyond their gender and area of studies), but by their function in an experimental system. They are only addressed as experimental subjects, and permanently instructed what to do next, from the moment that they enter to the moment that they leave. The material arrangements in the laboratory are standardised and "aseptic"; they provide participants with an impression of anonymity, but still make it possible to monitor their progress in the experiment.

Making sure that participants understand and interpret the experimental instructions as they are supposed to do is an important condition for ensuring experimental control and reliability, but experimenters acknowledge that there are limits to how completely human beings can be laboratorised. However, this irreducible element of human agency is what makes experiments epistemically valuable over theoretical models in the first place. As one researcher put it, even the fact that errors and misunderstandings cannot be avoided might be an advantage for experiments: "And of course in reality, people make mistakes as well, and people make decisions that I don't understand, so that, I guess adds the realism in a way" (Transcript P3). For others, the promise of controlling *almost* everything (but not everything) is what they call "the beauty of the lab" (Transcript P5). What presents the greatest risk to economics experiments – people's capacity to resist the expectations of others – is their main epistemic resource. That experimenters sometimes fail to produce the behaviour they intended, even when they take behavioural regularities into account, shows that despite several phases of reducing experimental systems to a controllable structure with known properties, participants' agency still presents an irreducible element that is genuinely unknown. Even though some behavioural regularities might be well understood by now, some selection biases can be avoided through exclusion and randomisation, and some comprehension problems can be solved in the laboratory, participants cannot be fully controlled.

6 Analysis, Part 2: Evaluating Experiments

One of the interests guiding my research project from the beginning was to find out why economists are doing laboratory experiments in the first place. What do economists perceive as the epistemic value of experiments, and what is their conception of a good experiment? This chapter aims to answer the second question guiding my analysis*, how economists evaluate the method of laboratory experiments*. Keeping in mind that values are not fixed, but open to negotiation, I will describe discursive valuation practices that I encountered in the interviews with experimental economists. To do so, I first describe the epistemic values and ideal types of scientific experimentation referred to by my respondents and then their individual motivations for pursuing a particular research project. In the final section, I will trace the valuation practices involved in experimental design. This will allow me to reconstruct several evaluative principles that experimental economists refer to when speaking about their own and others' work, identify which principles are invoked in which specific situations, and describe how some of these principles relate to a dominant regime of valuation in contemporary academia.

6.1 Epistemic Virtues of Laboratory Experiments

The general consensus amongst my respondents was that laboratory experiments are a valuable method because they allow economists to study phenomena that are difficult or even impossible to study in the field. I encountered various reasons why particular phenomena cannot be studied "in real life". Corruption, for example, was mentioned as a phenomenon that is not *observable*, because reliable data about naturally occurring corruption are hard to collect. Also, field experiments comparing corrupt and non-corrupt decision-making in a natural setting would be *ethically questionable*. In the case of the experiment on transparency in decision-making, the fact that transparency about third party funding and lobbying is not implemented yet was seen as not only motivating the research project itself, but also the method of a lab experiment: "Yeah, so, of course you can actually only do an experiment, if lobbying is not transparent, right? Cause if it's not transparent, you have no data to work with." (Transcript P3). The obscurity of some types of economic behaviour serves as an argument for doing laboratory experiments.

In the case of tax avoidance, the difficulty to observe the phenomenon in the field consisted not so much in the lack of naturally occurring data, but in a lack of criteria to unambiguously *identify* tax avoidance:

> Think of the ratings in in in in real life, how should that work? You know, that's very difficult, ahm it's not clear even what tax avoidance means, you know, I mean lawyers and economists are struggling all the time to define tax avoidance even, ahm, what does it mean, what's excessive tax avoidance then, you know, this is a very qualitative word, so what does it actually mean, it's very hard. In the lab it's very easy. We know you sold ten units, you, you ought to pay taxes for ten units, you only pay it for nine, we know you avoided one unit, taxes for one unit. Well that's very straightforward, ok, we know all that, we, we measured that, we observed that. In real life, we don't. (Transcript P1)

On this view, the epistemic advantage of the laboratory experiments consists of the researchers' ability to recreate a messy real-life phenomenon with a small number of observable and measurable processes. In the context of the laboratory experiment, economic agents only have a limited range of possible actions, and some of them can be identified as tax avoidance by definition. The ability to define what the phenomenon consists of, at least within the context of the laboratory experiment, also renders the phenomenon observable. In this particular case, the context of the laboratory experiment made it possible to translate a vague and "qualitative" notion into quantifiable operations that can easily be measured, and therefore, observed.

A different type of phenomena is described as observable in the field, but not *manipulable*. For example, the equality of opportunity can be measured within and across different societies, but it cannot be varied in a controlled way to see how the change in conditions affects behaviour. Several respondents mentioned the controlled manipulation of an individual factor as the hallmark of the method of laboratory experiments. This corresponds to Guala's (2005) model of the "perfectly controlled experiment" as a methodological ideal for generating reliable inductive inferences. As one respondent put it, laboratory experiments are extremely useful when researchers want to "turn a screw" (Transcript P7) and compare the baseline condition to what happens when a single factor is varied. In order to "turn a screw", researchers need to be able to single out this specific factor. Consequently, the difficulty to *isolate* a certain factor in the field was also mentioned as a reason for doing laboratory experiments. Isolation may be difficult to achieve in the field from the perspective of the researcher, because it may not be possible to only turn one "screw" at a time. It was also mentioned, however, that isolation in the field is impossible because there are too many distractions from the perspective of the subjects. In the reduced environment of the laboratory, this researcher expected subjects to be less distracted and easily focused on those dimensions of the environment that are salient from the experimenter's perspective:

> So people are, can actually very much focus on the treatment dimensions you have, they don't know that there are treatment dimensions, but they they know, they they see some numbers and they focus on these numbers, only on these numbers. In real world, they focus on everything, you know, or anything. And this, this would make them pay very little attention to the dimension that we're interested in, meaning that the treatment effect might be much lower. Probably not even existent. (Transcript P1).

Behind all of these considerations – how to observe, identify, manipulate and isolate particular phenomena – is the desire to study and describe relations of *cause and effect*. The opportunity to identify causal relationships with a high degree of probability was described as the main attraction of laboratory experiments in comparison to other methods. Indeed, whether an experimental design could be expected to identify a single causal factor also serves as an evaluative criterion:

> H Ok, what could be an apparent flaw?
>
> P4 Ahm, well maybe that it's not able to disentangle the, really the effect, that there are too many factors interacting and that in the end, which is like the big advantage of experimental, lab experiments, is that you can disentangle really one effect by controlling for the other effects. And if that's not possible in the design for whatever reason, then, well, it's lost in a way. (Transcript P4)

The epistemic value of identifying causal effects in the laboratory is at times slightly overshadowed by the threat that these causal effects might be identified in the laboratory *only*. Making conclusions from the laboratory setting to real-life settings is described as difficult and risky, precisely because the dimensions studied in the laboratory do not exist in the same state of observability and isolation in the social world:

> P3 Which is, which is the strength and weakness of experimental studies, it's the strength because you can really identify causality, you know exactly what's going on, what's making the change in outcomes. But the problem with this is obviously, you don't know if this other stuff in reality that may be going on is more important. (Transcript P3).

As mentioned by Bhaskar (1985), laboratory experimentation always rests on the assumption that the factors studied in the laboratory do behave in the same way as they would in natural conditions; i.e. that the reconfiguration of natural objects does not alter their properties (especially concerning causal relations) in a significant way. In economics experiments, particularly the interaction with many other factors that normally play a role in real life is stripped away. As the quote above shows, this is a matter of concern for experimentalists, but since reduction and isolation are also seen as necessary to identify causal factors, a certain tension between the internal validity of the experimenters' conclusions and their validity for application to other contexts cannot be avoided (cf. Guala, 2005). All of my respondents pointed out that the validity of experimental results is limited to the laboratory context. Making generalisations to other contexts or even direct applications to real-life situations was often described as a task that is separate from doing laboratory experiments, trying to answer an entirely different question. Yet that experimental results do apply to real-life contexts, at least on a general level by describing observable patterns of interaction, was also mentioned as an important epistemic quality of experiments:

> (P4) So I think these, in that way, it should be generalizable, that the patterns I observe in the lab have also some information for the field, for the real world, ahm, but it's maybe too much to claim that those who free-ride in the lab also free-ride in the field, like on that individual level, because there are so many other factors in the field which play a role, but in general, that you have these patterns, that some are free-riders, some are conditional co-operators, some are altruists, and that you also find these types then in the real world. I think in that sense, it's important. (Transcript P4)

This statement evokes a conception of laboratory experiments as allowing "generic inferences" from patterns observed in the lab to patterns in similar real-world contexts that corresponds to Santos' (2010) account. Along with the simplified setting, also the particular subject pool used makes it easier to compare experiments to other experiments, but less likely that experimental results are representative of the "general population". It is important to notice that experimental economists are aware of these limitations and therefore do not describe the method of laboratory experiments as giving them insights on real-world problems and real-world behaviour *directly,* despite the fact that experimental studies are sometimes inspired by such problems or even presented as offering contributions to their solution. For example, in the case of the tax avoidance study, the researcher suggested that a field experiment should be the next step for testing the effectiveness of the policy solution they proposed;

an experimental study by itself is not enough to establish how well a policy will work in real-life conditions.[35] For this reason, experimental studies were also described as a "lower bound" when it comes to finding effects that might also play a role in real life. The epistemic virtue of experiments by themselves is then seen rather in their being a method to test and challenge economists' established practices and assumptions, as the following quote illustrates:

> P5 So this is a certain population, it's not fully representative of the general population, of course. That's one problem I could think of. But I still think we cannot do without it**,** we cannot do without lab experiments, I think it's still a great way to show us direction
>
> H yeah
>
> P5 about the way we write our models, the way we think about the world. And, yeah, I, I think they're indispensable. I think they're helpful for us to understand more about the world, so to some extent, external validity is there, but it's true that we use a certain population. It's the undergraduate population.
>
> H And also in very simplified situations.
>
> P5 Exactly, also simplified settings. And, which still has its own advantages, that means we still have control over the setting, we still have control over the topic that we're trying to understand. But yeah, I mean, there are problems, but still I, I think we will, we can't do without lab experiments. They're useful. (Transcript P5)

Simplicity here is invoked as a virtue that is tightly linked to the ideal of experimental control. The problems that simplified settings and a homogeneous subject pool generate are considered negligible as long as experimental results can be used to build better models and theories. The use of laboratory experiments in economics as theory-testing devices is a dominant conception that I encountered frequently, and that has also been discussed in detail by other authors writing on the topic. However, testing and developing theories is not the only use of laboratory experiments. This conception of the use of experimentation consequently seems to be related to a specific ideal type of scientific practice that guides researchers' reasoning about their own methods. Yet there are also other (ideal) types of scientific experimentation around, each of them carrying with it different assumptions about the use of experiments and the relation between theory and data in producing valuable knowledge.

6.1.1 Ideal types of scientific experimentation

During my interviews, I came across three characterisations of experimentation that all invoked the model of a different natural science. I found this interesting, because the three different ideal types mentioned carry with them very different conceptualisations of good scientific practice. Consequently, invoking these ideal types suggests that individual researchers' ideas about the use and value of experiments in economics do differ in some respects. On a general note, that experimental economists refer to natural sciences when reasoning about their own method indicates how much experimental economics is indeed an effort to emulate these more empirically oriented types of research.

---

[35] Cf. Guala's (2005) suggestion that external validity is an empirical question that needs to be answered in each individual case by gradually assimilating the experimental conditions to the real-world situation that the experimental insights will be applied to.

The first characterisation that used a comparison to a different experimental science referred to the practice of using "control treatments" and varying only a single factor across different treatments. The researcher described this common strategy as similar to the practices of medical trials:

> P1 Well yeah, that comes to one big advantage to, to experiments, particular lab experiments, where you have this very clean environment,
>
> H mhm
>
> P1 so you can rule out plenty of explanations, ahm, ah ex ante, because they are still, they are just not there, […] and then you can, ahm, without any sample selection issues, or, what, say, say students in a lab is not a sample selection in that sense, but without any, ahm, any unobservables that might play a role, you can very, very ah tightly narrow down the conditions that you're checking. So you have one setting, we call, let's call it the baseline, this is, this is one setting, so here's a certain set of parameters, ok, that's it. And then the, the advantage of experiments is that you have different treatments that you compare to each other, so you have if you want the the doctors would say a treatment group and a control group.
>
> H mhm
>
> P1 Ok, there is one getting the placebo and the other one getting the real pill, ok? And that's the same what we do here, so, by keeping everything constant but the pill, basically, ahm we know if there is a difference between the two, it must have been the pill, because there was nothing else, ok? (Transcript P1)

In this comparison, the researcher not only likened experimental economics to medical trials, but also described the conditions that are necessary for the ideal type of medical experiments to succeed. The notion of a "clean environment" is invoked, as well as the need to avoid selection biases, and the idea that the conditions of the experimental setting can be known and controlled. The "clean environment" of the laboratory makes it possible to compare treatments, because it allows the presupposition that the background conditions remain constant. Together with the ability to manipulate a single factor in isolation, this setup allows researchers to identify causal effects. This ideal type is also what Guala (2005) describes as the model of a "perfectly controlled experiment", from which valid inferences about causal relations can be drawn. In this sense, it can be understood as the general model for all experiments that aim to identify and study relations of cause and effect. Note that there is no reference to theory or predictions in this particular excerpt, and neither is there in Guala's account. The ideal type of experimentation mentioned here consists of carefully establishing favourable conditions for observing causal effects, with the aim of learning how to actively produce these effects for a certain goal.

The second characterisation (Transcript P7) of experimental economics referring to another science used a comparison to experimentation in physics. To be precise, this researcher also compared the "learning by doing"-aspect of experimental economics to other skilful practices that require hands-on training, such as shoemaking and performing heart surgeries. The process of developing an experimental design was then compared to that of an experimental physicist gradually adjusting her experimental setup by identifying and accommodating outside influences such as room temperature and light. This however, should not imply that experimenters only "try around" long enough until they get the results they intended; rather, they are step-by-step taking into account all the possible factors influencing the results in order to arrive at a "correct" experimental design. According to this researcher, there are in general two approaches to experimentation: The first would be to "just experiment" and then induce regularities from what is observed, much like the legend describes Galilei's discovery of the laws of falling bodies. This approach first produces data and then builds theories inductively. The second approach, in contrast, is theory-driven experimentation: through introspection, the researcher arrives at a hypothesis for what result should follow from a given constellation, and then sets out to test this theoretical hypothesis with an experiment that is designed accordingly. According to this researcher, the first approach has little standing in today's "mathematically driven" economics. This was interpreted as beneficial for the practice of experimentation, since predictions and "benchmarks" are needed for good experiments, and theory would also provide "discipline". Theory-driven experimentation, in which the correct design to test a hypothesis is carefully developed, was seen as more productive than just producing facts, for example, by trying out different variations of already established games. With the latter approach, there might be some interesting new results, but in general, there would be too many "degrees of freedom" (Transcript P7).

The third characterisation of experimentation diametrically opposed the second one. Likewise focusing on the relationship between theory and experiment, this researcher bemoaned that the commonly known criteria for experimental economics research – the "textbook approach" – would not allow for unstructured and exploratory studies. Against the view that an experimental design should always be developed based on a theoretical model, this researcher invoked the idea that the experiment was itself a model, pointing to the material aspect of experimental intervention:

> P3 For an economist, a model is a non-cooperative game theoretic model, that's what it is. And, I think this view really limits what people can do in experimental economics, because, ahm, because an experiment in itself is a model, right? So a model is a simplification of reality that allows you to, focus on, you know, particular, particular elements for which you have controls, you can understand things more easily, because the environment is simplified.
>
> H mhm
>
> P3 Like for a biologist, a mouse is a model. A nematode worm is a, is a model. What they're really interested in, is learning things about human health, but it's just, the human body is complicated and so forth, so you take a very simple animal like a nematode worm, and this is, they call this, they call this a model.
>
> H mhm
>
> P3 So it's much easier to understand what's going on in a simple organism. (Transcript P3)

In this excerpt, biology is described as a natural science that engages its objects directly, albeit in a simplified form, an account that parallels some of what Knorr Cetina (1999) writes about the epistemic culture of molecular biology. The comparison is motivated by the assumption that laboratory experiments would allow economists to do something similar; just like mice and worms function as model organisms, microeconomic systems in the laboratory function as model economies. Such a model can be directly manipulated and studied, to explore its properties, and this idea of "pure experimentation" is something that the researcher is missing in contemporary experimental economics. Even though developing and testing hypotheses is the "ultimate goal", there should also be room for doing experiments without much theory, in order to just produce data and search for novel facts:

> P3 Ah, but also I don't think you need to make ahm, you don't, I don't think you always need to make exactly, predictions. I think it's, it's valid to just study what people do, in order to learn about what's going on, you know, if you are dissecting a nematode worm
>
> H Mhm
>
> P3 you don't necessarily hypothesize that, (unclear), you're just experimenting, like literally, literally just experimenting. Ahm, to to make, to come up with conclusions, about different policies, then you need a hypothesis in advance in order to apply a proper statistical test and this kind of stuff. But I think we should also have room for pure experimentation without necessarily having a hypothesis from the start. (Transcript P3)

The ideal type of experimentation invoked here is that of a direct material engagement with the object under study that is not guided by theoretical conceptions and predictions. In this researchers' opinion, such an approach is not only complementary to the prevailing theory-driven approach, but also necessary to study phenomena that cannot be modelled in a very definitive and structured way. A focus on non-cooperative game theory models as a criterion for valid experimental research would preclude experimental engagement with such phenomena. The researcher also expressed the hope that "somebody big" would start doing more unstructured experiments, so that this type of research would become more acceptable in the experimental economics community. Interestingly, in the second quote the need for having theoretical hypotheses is associated with application-oriented research, i.e. experiments that compare the effects of different policies in the lab. To arrive at a valid comparison, the results of such experiments need to be statistically analysed, and in order to do so, experimenters need to have a more or less precise idea of what effect it is they want to measure in the first place. Research that aims to simply generate new insights, on the contrary, should not be required to start from a hypothesis of what is going to be observed.

It is probably not a coincidence that the three ideal types mentioned here roughly correspond to the three purposes of experimental economics that were defined by Alvin Roth (e.g. 2015). According to Roth, experimental economists' tasks are "speaking to theorists", "searching for facts", and "whispering in the ears of princes", where the last one refers to informing policy-making, for example by designing and testing markets for goods that are difficult to allocate efficiently (a task that Roth himself has won a Nobel Memorial Prize for). These three tasks imply different aims of experimentation and different criteria for evaluating experimental results that mirror those invoked by the three natural science ideal types. Physics experiments, with their careful adjustments of apparatus and their introspectively developed hypotheses, were described as the ideal form of theory-testing that should be emulated by experimental economics. In contrast, the "pure experimentation" on model organisms that biologists are capable of was invoked as an ideal strategy for discovering new and unpredicted facts. Both of these two diametrically opposed conceptions can be instantiations of the more general ideal type of (perfectly) controlled trials. What this ideal type highlights is that to identify causal effects, theory is less central than being able to control the background conditions. Understanding how to practically bring about causal effects is more important than theoretical explanations in experimentation that has a clear policy-orientation and aims to design applications; however, this type of research clearly needs to be guided by hypotheses about what causal relations are likely to be observed, so that they can unambiguously identified in statistical analysis. Both exploratory and theory-driven research can probably supply such hypotheses and therefore inform policy-oriented research.

There might be a deeper analytical point to be made about these three different ideal types, since they also imply different conceptions of the relation of experimental practice to the empirical. "Just observing" or playing around with material arrangements involves a very different stance towards the empirical than carefully constructing theoretical predictions and testing them. While the "pure experimentation" ideal is based on the assumption that meaningful insights can be gained by just intervening in and closely observing nature, the theory-driven approach holds that this is an inefficient strategy of knowledge production. Observation and intervention, on this view, should be guided by clearly formulated expectations, and theoretical hypotheses about causal relations should be developed before it is tested whether they obtain. The ideal type of medical trials is to some extent an agnostic approach, due to its pragmatic outlook. Theory here helps to come up with hypotheses about what needs to be studied. It might then not be always possible to provide a theoretical explanation for the observed effect, but this is less important, as long as its cause can be clearly identified and controlled.

6.2 Motivations for Doing Experiments

One of the topics I explicitly discussed in each interview was the researcher's interest and motivation for doing a particular experimental study. How did they get interested in the topic, and why did they think that it should be investigated experimentally? While these two questions are distinct, the answers given were not necessarily distinct too, since all the projects I discussed were developed from the perspective of doing a laboratory experiment. An important criterion for choosing a topic therefore was presumably whether it can be studied experimentally. For researchers who describe themselves as putting more emphasis on theory and modelling, this means that they do experiments primarily to test theoretical models, but that they also develop models with the idea of doing an experiment later: "I always have the aim to use experimental methodology to evaluate the theory that I have." (Transcript P5). This particular researcher described developing models and testing them experimentally simply as the approach he/she personally preferred, given his/her strengths and educational background. It exemplifies the "virtuous circle" of theory-development and experimentation that Svorencik (2015) identifies as a driving force in the accounts of first-generation experimental economists. I was told that researchers coming from other traditions might prefer less theory-driven approaches: "First the model, because first you need to have a model and model predictions so that you think of an experiment.

Well, that's my approach, there are many people who do many behavioural experiments without much theory." (Transcript P5). This, of course, means that "speaking to theorists" is indeed a motivation for doing laboratory experiments in economics, but it might not be the main motivation for everyone.

What I did encounter in almost all my interviews was the researchers' desire for their topic to be *relevant.* Relevance can be established in different ways, the most straightforward being to refer to a current debate or problem that the experimental study is meant to illuminate. Presentations of experimental work in research workshops, or the introductions to draft papers, often feature explicit references to such unresolved "real-life" problems. In the case of the study on transparency in decision-making, both authors remembered a particular magazine article that described the debate about whether transparency about sources of research funding would improve or damage scientists' reputation. This article was also mentioned as a source of inspiration in the draft paper on the experimental results. The reference to this debate seemed to be more than just a rhetorical strategy for establishing the importance of their research. One of the authors even expressed a very clear political opinion on the matter whether transparency about third party funding should be mandatory:

> (P5) We already had an idea that transparency may not work, of course, because we have seen previous experiments about transparency that shows it difficult to work, but I was more hopeful. Yeah, it's not easy, it's not easy. But I still have a firm belief that it should be implemented. I just think it's, it's a dishonest world if it's not implemented. And I think it's still important to see it doesn't harm. And yeah, I, I, I really think it's a timely experiment. [...] I think it's an unfair world if things are not transparent. I understand there's a stigma to accepting a payment from a company that is related to the research, but in the long run, I think that's a good thing. Because, with, my, my ultimate opinion is that, in the companies, pharmaceuticals and this company and that company should not be able to give their money to whatever they want. (Transcript P5)

The motivation for doing the experiment, finding out whether transparency would improve or harm decision-making processes, and the results – it could at least be shown, both theoretically and experimentally, that transparency does not harm the processes – were both interpreted in the light of an ongoing debate. Apart from the researcher's personal interest in the topic, *timeliness,* i.e. a connection to what is perceived as current and topical, was therefore clearly invoked as a virtue of this research project:

> P3 I mean, it's big in the news, and it's one of the, you know lobbying in general is one of the most important things for determining what policies get implemented
>
> H mhm
>
> P3 So I think it's a very important thing to understand, ahm, yeah, so I think the topic is very important, and there are, yeah there are some things that you can look at, ah. (Transcript P3)

This was similar in the study on tax avoidance and transparency, where the researcher also expressed a hope that the topicality of the research would make it easier to publish the paper in a high-profile journal (at least if it could be finished in time):

> (P1) [It] will be some kind of hot topic always, I guess, tax avoidance, but now is still you know, EU is, is thinking of legislation, several nation states see it as an important topic, so now is the time to get it out. Ahm, because it's still something that is, that is new, that there is no paper like that, so I think we

> would have a good chance to even go to such a top journal, but we need to do it now, not in two years, because, you know. (Transcript P1)

Here, timeliness is additionally related to singularity and priority; with regards to publication chances, timeliness is seen as an advantage when the paper can also claim to be the first addressing this topic within the discipline. This statement was embedded in a longer reflection on how to coordinate with the co-authors to finish the paper, in which the researcher expressed the need to publish the paper rather sooner than later. This shows that linking one's research to ongoing debates also implies a certain risk. Especially with policy-oriented research that aims to provide an active contribution to the solution of a current problem, the risk is that the problem will have been solved by the time that the paper is actually ready to be published. If this happens to be the case, the research will be crucially devalued: it will lose the quality of timeliness, and this will also lower the chances of publishing well.

This is not the case for research projects that aim for relevance in a more timeless sense, by addressing problems that are perceived as being of *general interest*. The PhD student I interviewed, for example, clearly stated that his/her choice of topic was rooted in a desire to study something that also non-experts could find interesting and important:

> (P4) And so, the topic wasn't there right from the beginning, but what was there right from the beginning was the wish to do research in an area which is kind of a broader interest, of social, ah, interest of society also, and like not something which is really like only interes-, interesting to a very small scientific community, but to also the people outside, who, which I can tell my mother about and she will understand what I'm, like, why I'm interested in that topic, in a way. […] So that's how I ended up with redistribution, because I think we, ah, equality of opportunity, equality of incomes is really a topic which is in the media every day, and it's interesting to do research in that area. (Transcript P4)

The reference to the media here seems to indicate less the momentary and more the continuous public interest in questions of economic equality and redistribution. That the topic is "in the media every day" is an illustration of its general relevance, rather than a reference to a specific public debate. What valorises an experimental study, in the view of this researcher, is whether the topic it addresses manages to capture the interest of non-experts.

In this last case, once the particular research topic (of inequality and redistribution) had been identified, the researcher set out to develop a research question and an experimental design. In this stage, consulting the existing literature and finding a research gap played a decisive role. Once the research gap had been found, the idea for the experimental setup "came kind of naturally" (Transcript P4), although several different designs were developed before one was finally implemented in the laboratory. Producing *novelty* was invoked as a motivation for pursuing certain lines of research by each of my respondents. What novelty means depends on the individual case; being the first to introduce a new type of game or a new tool (e.g. FMRI scanners, which led to the new field of neuroeconomics) is a rare form of novelty. Finding a (timely) topic that has not been investigated experimentally, or finding a particular aspect or approach that has not been covered by the existing research, are more common. In any case, novelty in the sense of doing something for the first time is a criterion for selecting experimental designs; in my respondents' accounts, novelty is also mentioned as an indicator for good publishing chances. Since it is common in experimental economics to do follow-up research, there is a chance that an innovative design may become a "workhorse" and serve as a template for further experiments, which also increases citations. There are, however, some qualifications to be made: As mentioned earlier, using an unorthodox type of modelling or otherwise designing the experiment in a way that is not in accordance with established literature may be criticised by referees and consequently lower publication chances. As one researcher said, it is important to know the existing literature very well, in order to be able to defend one's approach against criticism by peer reviewers. Also, experiments that technically provide novel results by reproducing an existing study under slightly altered conditions were rather dismissively referred to as "epsilon contributions" (Transcript P1) by one of my participants. Novelty in the sense of simply doing something for the first time on this view is not enough to justify an experiment. The insight provided by a new study should be more substantial than showing the same thing again "with just one decimal place different [...] in the parameters" (Transcript P1).

What then is an experimental study that is worth doing? When I asked one of the senior researchers this question, I got an answer that is worth quoting in length, because it highlights the interplay of very different motivations and valuations:

> P6 Ha… Well the classic thing in science is what is good science, you're asking what is good science? It should be, *it should be new, it should be true, it should be important.* These are the three things, ah, so you have to find, ah, some field where you can make a contribution to the field, so, something that, ah, hasn't, hasn't been said before, or not in this way, at least. Ah, it should be true in the sense the analysis should be competent and correctly done, you should isolate the causal effect, so this is a matter of, of craftsmanship. And it should be important, and that's the most difficult part really, that, ah, the addition that you make should somehow be relevant to your audience, and your audience is typically professional economists, so, and that's the most difficult part, to get a sense for ah, what, what would people find interesting. And ah, I think a good paper is one where you give a short summary to someone who is not an expert in, in the field, and they will say, oh, that's kind of interesting. That's an intriguing thing. So for example this paper here, I presented it in the Rotary Club, here in Vienna last, last year, so a few months ago, and this one here, I will present in a health conference in Bern. So, these are all non-experts. (Transcript P6)

Again, novelty features prominently as a criterion for "good science", and so does the epistemic value of identifying causal effects that was discussed above. In this particular case, isolating the causal effect is not only a matter of the experimental design, but also of the researcher's data analysis skills. The biggest challenge, however, for this researcher is to do important and relevant research. On this account, it is hard to get a professional audience interested, but this task is even complicated by the personal desire to make contributions that will also be of general interest to non-experts. An ideal experiment – "good science" – therefore should combine the epistemic qualities associated with laboratory experiments with a sense of importance both to the academic peer community and the general public. That this is the "most difficult part" implies that the methodological aspect of "good science" can be mastered by training and is relatively unproblematic, since there are clear criteria for what is new and true. Importance, on the contrary, is a value that is open to negotiation and needs to be established in interaction with the scientific community and lay audiences in each individual case.

6.3 Valuations in Experimental Practice

Since the judgement and interest of academic peers is seen as so decisive, what are the types of papers that a professional audience is expected to find interesting? Novelty as a criterion for research worth doing was already mentioned several times, but some of my respondents also pointed towards other qualities that they think are valued by their professional peers. A concern for these qualities does not so much influence the selection of topics and methods, but it affects more intricate decisions concerning the experimental design and the presentation of results. Consider this interview excerpt, where the researcher describes what a journal paper should be like in order to be appealing to other economists:

> (P1) And in the end in the paper you have only one line of reasoning, and the paper, once it's written up, it it should read, you know, oh, that's a very interesting idea, that's exactly the way how to investigate that, oh the results even show what they expected, brilliant. You know, [...] it reads very, and it, this only has twelve pages or so. [...] So, it's it's only, a small percentage of, of the results that you present, and everything is very streamlined. Ah, yeah. Probably you could tell a different st- a what, you for sure can tell different stories with the same experiment.
>
> H Ok?
>
> P1 But you decide on one story, and then highlight the data that corroborates your story and and and checks your hypothesis and so on. [...] Cause also other researchers want to have an easy read.

The same researcher emphasised throughout the interview that an experiment should be designed in such a way that it could confirm one out of two alternative hypotheses; i.e. ideally, an experiment is interesting no matter what the results are. In order to achieve such an outcome, experimenters need to take into account all the possible explanations for the phenomenon they want to study, so that all but the hypothetical one can be ruled out by design:

> (P1) So I just, I just change an efficiency factor, ok, from public goods game, and just change it sufficiently, so that some explanations, just don't work anymore, they ha-, they don't have any predictive power theoretically. Ahm, and that's great, because I want to have as few explanations that poss-, as possible, because in the end I want to say it's actually that. (Transcript P1)

If an experiment cannot rule out alternative explanations for the observed effects, this would be interpreted as a weak and uninformative result, giving the researcher "a hard time selling this paper" (Transcript P1). A related issue is the requirement – which is not unique to experimental economics – that an experiment should produce significant results in order to be considered "interesting" and worth publishing. To illustrate this requirement, the researcher again used the example of story-telling:

> (P1) Well, significance, well, it's always easier to publish a study if you find some significant results, whatever you do.
>
> H Yeah, of course.
>
> P1 If you find, your idea does not work out, that's actually quite, a, quite valuable, but it's not sexy. So, it's not, it's not a, you think, you just think of any story, really. Somebody tells you a story and says, we find nothing, ok? Ah, we walked in the wood and it was so nice, but in the end there was nothing. So, you know, why did you go there? If you walked in the wood and then you found this brilliant cave, well, that's a story, you know? But, we walked to the wood, that's it, you know. We did it, because that was, but we didn't find anything particular, and not so interesting. (Transcript P1)

Providing a new insight in terms of finding out that a certain experiment does *not* produce any significant effects then by itself does not entail that the experiment will be considered "interesting". As this researcher explained, non-results would only be interesting if they contradicted a positive earlier result. For "pioneering" research, which builds on a hypothesis of finding something surprising or unprecedented, not finding anything significant would be a "non-starter" in terms of publication prospects.

These two requirements on the "stories" that experimental economists can tell with their research papers – that they should be cogent and easy to read, and that they should present a significant result – are taken into consideration when designing an experiment. In particular, researchers will have to decide which of the hypotheses they have in mind are the most "promising" to test with an experiment, i.e., which ones will yield the most interesting and presumably, also the most significant results. To some extent, this is also a practical necessity. Since doing experiments costs money (because subjects need to be paid), and a certain number of independent observations are required for every treatment, the number of treatments and consequently, the number of hypotheses (if they are tied to specific variations) researchers can test will depend on the budget they have available. Using "power calculations", researchers can calculate how many independent observations they will need to achieve a significant result, given that the effect they want to produce is as strong as they hypothesize. This is seen as a useful method to find out what would be the most promising line of research, given the budget available:

> P6 The number of groups, we think of how we wanna test things, what kind of statistical, statistical tests we need, and then we think, then we have a hunch what's gonna happen, and then we think well, if this is gonna happen, it will be significant with so and so many groups. And then this is the number of groups we have, so this is, we have to do these so-called "power calculations", we don't always do them very precisely, but you know, approximately. Ah, but these are important to estimate the budget. (Transcript P6)

Achieving an "interesting" result is not only a matter of the experimental design, but also a matter of the statistical tests used to analyse the data. Researchers can take their analysis techniques into account and choose the sample size – the number of independent observations, which, depending on the number of treatments and the size of groups, translates into the number of sessions and participants needed – to be able to produce statistically significant results with these techniques. Power analysis enables researchers to calculate roughly how much an experiment will cost. Conversely, if they know beforehand how many sessions and participants they will be able to afford, power analysis allows them to select the questions that they consider most likely to give them significant results with the sample size available. This was described as paramount especially for junior researchers whose funding is typically more limited:

> (P6) Because we were junior people, we didn't have a lot of money. So we were very careful and thinking about, you know, what we really wanna test, what are the most promising hypotheses to test, so typically, at the outset, we had lots of different ideas, and then we would go for one where we think, well here's, we gonna, we will, are likely to find an effect here. The most promising, so as, because we had limited budget.

H mhm

P6 So, then given that budget, we think ok, how many observations do we need to, to show this effect, if it's there, and ah, does it fit into the budget. And if not, we have to go for a different question. So now, my situation is much more comfortable, I'm not so constrained by the money, so, but for young researchers, it's a problem, do, you have to do these calculations to see well, is it gonna work out or not. (Transcript P6)

On this account, a limited budget in combination with the requirement that the experimental result should be a significant effect led the researchers to choose a research question that they hoped to answer with the amount of experimental observations they could afford at the time. Using power analysis therefore can be described as a strategy of mitigating epistemic risk, namely, the risk that experimental results will be inconclusive with the amount of data produced. As the interview excerpt illustrates, junior researchers are seen as particularly affected by this type of epistemic risk, because their funds are more limited and they cannot simply do follow-up sessions if they do not observe significant effects at first. The epistemic risk itself, however, is not only grounded in the nature of experimental research, but also in the requirements that a paper should fulfil in order to be considered worthy of publication. The perceived preference for significant results and the practical constraint of limited funding together are likely to put junior researchers in a position where they might forgo riskier lines of research in favour of topics and designs which are likely to produce stronger and less ambiguous effects.

Even after the research questions and the hypotheses to be tested have been chosen, researchers do have strategies available to optimise the experimental design for producing more "interesting" results. In the study on transparency in decision-making, the researchers for example chose the parameter values in such a way that the theoretical model would produce the predictions they found most interesting. What "interesting" meant here corresponded with their initial motivation of studying situations in which transparency might or might not improve the quality of decision-making. When developing the model, they found that in some cases there was no difference in predictions between the model with transparency and the model without transparency. This was the case, for example, if the payoff for giving honest advice was too high. According to the model, the advisors would have always given honest advice then, because it would be in their interest simply due to the high reward of being honest. Both authors described implementing such a model as an example of an uninteresting experiment. Since they wanted to investigate whether transparency could make a difference for decision-making, they needed to implement a design that corresponded to a model in which a difference due to transparency was at least theoretically possible. This particular design choice was justified with the nature of such decision-making processes in real life:

(P3) So we, so we chose parameter values to give us, ahm, partly to give us interesting predictions where there is a difference between transparency and non-transparency. And I think that's, that's valid because, ahm, we needed a situation where, if someone takes the payment, it doesn't necessarily mean that they're gonna be lying.

H mhm

> P3 And I, I think that paramount because in reality, you know, if the doctor's getting gifts, you may still trust them, you may think that only (unclear) because they can still tell the truth. […] So we needed the parameter values that allowed for this kind of ambiguity. (Transcript P3)

The choice of the value of a different parameter (the probability of state A) in contrast was motivated by theoretical considerations and the desire to arrive at a "clean design" (Transcript P5). If this value was > ½, this would mean that the state of the world which the third party did not prefer was more likely to be the case, and an advisor taking the payment would need to lie more often. One of the authors explained to me that in this case, it could still be shown that transparency was beneficial in theory, but there would be many more equilibria, making the predictions less clear-cut:

> P5 There are still cases, you can still find parameter values such that transparency benefits, it's a bit uglier, especially in the non-transparency case. There are mixed-strategy equilibria, and a corrupt equilibrium, so it's less clean, as we say, it's not clean, it's not a clean design. […] Yeah, so we thought, we wanted to know at least what we should expect in the non-transparency case. We were just expecting everyone to take the bribe, in theory. That's a very simple, clean prediction.
>
> H Yeah
>
> P5 And it would be, it wouldn't be like this if p is greater than ½. As I said, still transparency should make a difference in theory, but even discussing the predictions is uglier. (Transcript P5)

This notion of "clean" and "ugly" designs and "clean predictions" harks back to the notion of "easy-reading stories" mentioned above, and the general consensus that models and experimental designs should be as simple as possible.

As was mentioned already when discussing epistemic practices earlier, what researchers perceive as the expectations of their peer community (and especially of those peers who are likely to act as referees for journal submissions) affects how they will plan, approach and present their research projects from the start. This becomes evident especially where individual ideas and preferences of doing research come into conflict with what are perceived as the preferences and standards of the peer community. Consider how the researchers reasoned about whether to include a payoff for being honest in the experimental design. This payoff was included in the theoretical model in order to capture the benefit that individuals receive from telling the truth, either because it builds a good reputation or because they have an intrinsic aversion to lying. Since this aversion occurs "naturally" in the subject population, it strictly speaking would not have been necessary to include the payoff in the experimental design as well, and participants might have actually been confused why there was such a payoff in the first place. If anything, this additional information made the instructions slightly longer and more complicated. Not including the payoff would have meant to deviate from the model that the experiment was supposed to implement:

> (P3) Ahm, so, you know, if I was doing this, if I was designing what I think would have been the best experiment, I would have not had that payment. Ahm, but this was in order, you know, if, if, if you submit to these top journals, they expect the implementation to be exactly the same as the theoretical model, so we had to put it in. […] Or at least that was, that was what we thought. (Transcript P3)

In this case, the practical demand of making the experiment as simple as possible was overruled by the strategic consideration of what the experiment should be like in order to have good chances for publication. Another question where such a conflict arose was whether the researchers should use neutral or suggestive language ("framing") in the instructions. Both questions where decided in view of how potential referees might evaluate the paper. Since they expected peer reviewers to be critical of an experimental design that did not implement the theoretical model exactly, and of an experiment that was framed in suggestive terms, the researchers decided to go with the additional payment (even though it made the experiment slightly more complicated) and use neutral language (even though they personally thought that contextual biases might play a role for decision-making and should be captured). What is considered as standard and acceptable within the peer community of experimental economics was also invoked in one case as an argument to defend their research. When during a research seminar presentation another researcher criticised the conclusions that the authors had made based on a statistical test, the author presenting the paper defended both the test used and the conclusions made by referring to common standards as well as their status as junior researchers who cannot be expected to act contrary to established conventions. When I came back to this issue during the interview, s/he again emphasised that their analysis was in line with common practice:

> (P5) Every single experimental paper is like this, and [...] there is nothing unethical there. [...] This is how papers published in *Nature*, *Science* are working, I mean there is nothing weird we're doing. […] I am no statistician. I'm not gonna further the science of statistics. I'm just gonna use the tools that I need the way it's being used. (Transcript P5).

This justification both invokes the (collective) peer group and the conventions established in existing literature as the ultimate arbiter of what is acceptable research methodology, as well as the notion that junior researchers should not be expected to aim at changing or improving the professional standards.

What junior researchers need to do, most of all, is get papers published, especially if they want to secure future funding and ultimately, an academic career based on experimental research. The post-doc researchers I interviewed were very outspoken about this aspect of their work: "This is how academia is designed, that's, that's not great, they call it 'publish or perish'. If you don't publish, you perish. […] It's limiting in one sense, but you have no chance. Especially as a junior, I have to have it in mind." (Transcript P5). Another researcher highlighted that not only getting publications, but getting published in "good journals" is paramount for one's career perspective: "I only know economics in that sense, but publishing in good Econ journals is basically what, you know makes, or or hinders your career." (Transcript P1). Consequently, what was discussed above as the "textbook approach" was also described as grounded in the standards set by peer reviewers who expected experimental results to come with a theoretical model explaining them: "So, you want to publish well, and in order to, to have a chance in in in very good journals, ahm, say you increase your chances by having a theoretical model, predicting something, you know, at least to create hypotheses." (Transcript P1). While in this particular case, developing a model was described as a precautionary measure for improving publication chances (that turned out to be beneficial for data analysis as well), other researchers had already experienced negative feedback when they had deviated from the "textbook approach":

> (P3) [The] early stuff that I did was really unstructured. And, I had a lot of diff-difficulty publishing it. Ah, so now actually I try this, I'm, most of the stuff, most of the time, so, so, now I have a few projects where I start off with a non-cooperative model which I implement in the lab, and do in a standard way, ah, which is fine. (Transcript P3)

> (P6) [We] were fortunate to have published this well, but we always got the comment 'Where is the theory?', you know, you should have a behavioural theory that somehow explains what's going on. (Transcript P6)

"Publish or perish" was less of an issue for the senior researchers I interviewed, but they also frequently mentioned the need to plan publications, to "sell" papers to high-ranking journals and the sense that getting an experimental study published can never be taken for granted. The only exception to this general impression was the PhD student I interviewed. This researcher also mentioned that s/he would like to publish his/her thesis and not just do it "for the drawer", but described his/her reasoning and decision-making in the research process as guided by pragmatic considerations rather than a need to distinguish themselves academically. The reason for this different orientation was mainly that s/he had decided not to pursue an academic career after the PhD and wanted to work "more applied". This outlook enabled him/her to forego a search for the "best design" in favour of finishing the thesis in time, and was also mentioned as a justification for developing a theoretical model for his/her experiment only at the end of the research process:

> (P4) If you wanna test a theory which is out there, which is used to make predictions about what's going to happen in real life, then it's important to have your theoretical model, calculate with it your parameters etcetera, but sometimes if you have a more explorative approach, I guess it's less important. […] So, in that case, I think it's a bit less important, ahm, yeah, and it's also like I'm (laughs), I'm more applied, in a way, I'm not so, that much into theory, that's why I'm kind of procrastinating in that respect. (Transcript P4)

While other criteria for doing a good experiment – the novelty of the approach and the contribution to existing literature, the identification of a causal effect, and the relevance of the research to real life – informed the research process in important ways and were also explicitly mentioned during the interview, the criteria commonly invoked for getting a good publication did play a lesser role for this researcher.

It should be mentioned here that the laboratory assistants, who are eventually responsible for running the experiments in the lab and making sure that the data produced are valid, also articulated a clear opinion on what a good experiment is, especially in situations where their expectations were frustrated. For example, one of the first experiments that I helped to run was condemned as "shittily programmed" (Transcript Field Notes) by the lab assistants, because the first sessions had to be cancelled due to some problems with the software. After those had been fixed, this experiment still required lab assistants to manually insert data before starting a session, and among other shortcomings, it did not impose a time limit on the cognitive reflection test at the end, which meant that everyone had to wait a long time for the last participant to finish with this part. The fact that some participants left the experiment with very high earnings (up to 90 Euros) was also interpreted as a fault: "This is so high. It is the most expensive experiment ever, and the stupidest. I mean the guy is not stupid, he certainly has a

PhD, but this programming style really annoys me." (Transcript Field Notes). In this and another case where problems during the experiment were explained with a lack of proper preparation and testing, the assistants speculated that the responsible researchers might have considered themselves as experienced enough not to do a proper testing session. In general, laboratory assistants would complain about the bad design when an experiment seemed particularly long or boring for participants. They told me that the quality of an experiment could usually be inferred from participants' reactions and the atmosphere during the session. If participants got bored, they also got irritable and unfriendly; as one of the assistants told me, boring experiments can even be identified by the higher number of missing pens after a session. A good experiment, in contrast, would be easy to handle, ran smoothly, did not take too long, and left both participants and lab assistants in a content and satisfied state.

6.4 Answering My Research Questions

Having described the reasons and justifications given by experimental economists to do experiments, and to do them in a specific way, I can now set out to answer my second principal research question, *how experimental economists evaluate the method of laboratory experiments.* In particular, I will try to carve out the evaluative principles and regimes of valuation at work in the discursive and epistemic practices of my respondents. Evaluative principles "denote how worth is ascribed and argued for in a concrete situation" (Fochler, Felt, & Müller, 2016: 180). As the different descriptions above show, the worth of laboratory experiments as a method for economics is ascribed and argued for differently in different contexts:

When explicitly asked about why experiments are valuable (one of the questions I asked in interviews was, for example, what knowledge can be gained from running an experiment that goes beyond what can be known from developing a model), researchers answered in terms of epistemic qualities. These qualities were often illustrated by contrasting experiments to other methods (usually, collecting field data or constructing theoretical models) and showing how and in which ways laboratory experiments are superior to those methods. The notion of manipulability and control, and the possibility to identify and isolate causal effects are the main epistemic qualities associated with the method of laboratory experiments. Additionally, experiments have the advantage over models that they allow to test predictions with "real human agents", which is seen as a source of valuable insight. Laboratory experiments render complex causal relations observable in controlled conditions, and the assumption guiding this type of research is that there is an ontological correspondence to those causal relations as they happen "in the wild", because both the phenomena in the world and the phenomena in the lab are made from the same stuff: real human behaviour.

When researchers reflected on the merit of experimentation as a scientific method in general, however, they invoked conceptions of the use and value of scientific practice that go beyond the validity of claims about causal effects. In particular, it seemed that individual researchers were evaluating the use of experiments in economics with regards to very different ideal types of experimentation in other science disciplines. The identification of causality played a role in two of these ideal type conceptions, the physics experiment and the medical trial, but these two conceptions are distinguished by the importance that theory has in the process of experimental design and in the process of establishing and assessing knowledge claims. Moreover, the physics experiment, also in the descriptions that these researchers used (think of the Galilean experiments), has the primary aim of finding an empirical regularity for the sake of gaining knowledge. A medical trial is also about finding a stable cause-effect relationship, but for the sake of practical interventions and applications. Medical trials therefore have several stages, in which the experiment is gradually transferred from the laboratory to the field. Physics experiments (at least the ideal type invoked here) do not have a second stage beyond the lab; they are self-contained practices of knowledge production. The third ideal type of biology experimentation may or may not be targeted towards practical application, but it is marked by the absence of theoretical predictions to guide the practice, and by an idealised conception of "pure experimentation" that aims to "just observe".

I found that these ideal types do also serve as justifications when arguing for the value of a specific approach in experimentation. The researcher referring to the ideal type of physics experimentation expressed the opinion that trying to bring too much reality into the lab was a common "beginner's mistake" that would result in a loss of predictability and other epistemic qualities. An experiment should be like a good map; it should only contain what is of interest to answer the research question and only produce those effects that a researcher wants to measure. The only exceptions to this rule are experiments that function as "test-beds" for policy interventions, where it might be desirable that the experiment is more "true to reality". On this account, it is very clear that the desirable epistemic qualities of the approach to experimentation cannot be entangled from what is supposed to be an experiment's purpose. If experiments are seen primarily as instruments for testing or further developing economic theory, experiments that do not start out from theoretical predictions will easily be considered as flawed, and as allowing "too many degrees of freedom". If, on the contrary, experimentation is considered as a method to directly engage with and observe the material world in a simplified setting, the need to first derive theoretical models and predictions will be seen as precluding a wide range of research approaches that might also yield interesting and new knowledge.

The motivations for doing experiments should therefore be expected to make a difference in researcher's conceptions of good experimentation. While I have found this to be the case in some instances, the relation between motivations for doing experiments and choosing topics, and individual conceptions of good experimental practice is somewhat more complex. Also, my small sample of interviewees does not support sweeping claims. One researcher, as mentioned before, was very passionate about his/her research topics and their social and political relevance, yet his/her preferred research approach was very theory-driven and not primarily aimed at designing policy interventions. Those of my interviewees identifying themselves as coming from a tradition where theoretical modelling was less important expressed a desire for their research to be socially relevant, but they did not, at least in this academic setting, pursue research that was specifically aimed at practical applications.

What came out clearly in my interviews was how one particular situation, a researcher's career stage, matters for which evaluative principles are experienced as dominant and decisive. In line with Fochler, Felt and Müller (2016), I found that the PhD student I interviewed experienced him/herself as being able to refer to a wider variety of evaluative principles when assessing his/her work. S/he mentioned epistemic qualities, but also, as a driving force behind his/her research, the motivation of investigating a topic of general relevance and interest. S/he acknowledged that theory-building is an important aspect of experimentation and that publishing the thesis would be an ultimate goal, but felt that s/he could give these two demands a lesser priority because of his/her decision not to stay in academia and instead pursue a career in more "applied work". In contrast, junior researchers pursuing an academic career very explicitly described their need to conform to what is perceived as standards for high profile publications in (experimental) economics. Considerations of what potential referees might criticise, or of what referees had criticised in the past, therefore shaped their epistemic practices and informed concrete decisions concerning the experimental design.

Summing up the above observations, I can thus reconstruct several evaluative principles that experimental economists invoke when speaking about their own and others' work. One of them centres around *practicalities* and the *feasibility* of laboratory experiments. This principle is often referred to when reflecting on the designing and running of laboratory experiments. Experimental economists want to produce valid and useful data, and to do so, they need to take into account their participants' experience and reactions during the experiment. They need to find the right length for experimental sessions, the right level of complexity for experimental designs and instructions, and also, the right budget to be able to test the hypotheses they are interested in. From the perspective of the laboratory assistants, a good experiment is one that was tested beforehand to minimize bugs in the programme. Experiments that did not run smoothly because they had not been tested, or were too long, or too boring, or even experiments that used a lot of budget were what laboratory assistants considered to be bad experiments.

When asked directly about the *epistemic value* of the method of laboratory experiments, researchers in general referred to the ability to observe, identify and control causal effects. Achieving this epistemic value in practice demands a trade-off with what is commonly understood as "external validity", the applicability of insights gained from experimental studies to other contexts. Because of the simplifications and reductions that are seen as necessary to clearly study causal relations, insights from experiments apply to other contexts only on a rather general level. The invocation of different ideal types of scientific experimentation shows that epistemic values do not stand alone, but are articulated with regards to what is perceived as the purpose of experiments. Researchers' ideas of the type of knowledge that can be gained from experiments influences their view of what good experimental practice should and could be like. This is not an entirely trivial point, because it seems that the different purposes of experimentation imply different relations of experimental data to theory, and different conceptions of how the empirical should be approached. Also, the conception of good experimental practice that is endorsed by the most influential figures in the field is more likely to become institutionalised. This was the case with what was referred to as the “textbook approach” in experimental economics, a theory-driven conception of experimentation in which experimental data is seen primarily as a resource for testing and developing theoretical models. Although doing experiments “without much theory” seems to be a common practice as well, the “textbook approach” at least for the researchers I interviewed was generally understood as providing the yardstick for evaluating experimental papers.

The evaluative principle of *publishability* thus involved the textbook approach conception of theory-driven experimentation, along with more vague criteria such as telling a good and easy-to-read story. In practice, this principle seemed to come into conflict with considerations of practicality, but also with diverging epistemic conceptions of the purpose of experimentation. Other criteria mentioned by my respondents such as novelty, topicality and importance can also be interpreted in relation to the question whether a study is likely to get published. Presenting a study as a contribution to an ongoing debate or a topic of general interest is, of course, a discursive strategy for establishing its importance and raising the interest of one’s audience. Relevance is also a value that researchers aim to establish independently and beyond other qualities of experiments. This is particularly the case when researchers are in a more comfortable position to follow their personal interest, either because they have achieved senior status or because they are still doing their PhD. This is not to say that junior scholars cannot follow topics they are personally interested in, and in fact, all of the post-doc researchers I talked to seemed very invested in their research topics. Due to their more precarious career status, they need to take the interests and demands of their epistemic peers and potential referees into greater consideration.

All of these evaluative principles inform epistemic practices at various stages in the research process. Given that individual researchers have emphasised very different aspects about good experimentation during the interviews, there was a considerable heterogeneity in valuation practices visible in my small sample. The question of publishability, however, clearly emerged as a dominant principle. This is not surprising, given that the researchers rather explicitly situated themselves in a regime of valuation where academic achievements in the form of high-ranking publications are the main criterion for assessing individual researchers and their work. While this was seen as a more or less unchangeable state of affairs, the evaluative principles concerning the method of experimentation that are currently suggested by this regime were, all in all, described as dominant, but in principle malleable. The hope expressed by one of my respondents that “somebody big” might make unstructured experimentation more acceptable, and the frequently expressed notion that doing experiments with or without models was basically a matter of personal preference, show that questions about what is the right approach to experimental practice are considered as ultimately open to negotiation. This suggests that what ideal scientific experimentation is and how it should be done is not yet a settled issue within the experimental economics community.

7 Conclusions

This thesis is the product of my interest in the epistemic practices of experimental economists, and the evaluative principles I have observed that my respondents invoke when reflecting on their own and others' work. The questions that guided my analysis were how do experimental economists appropriate the method of laboratory experiments, and how they evaluate this method. In this final section, I will sum up once more how I answered these questions in the course of my analysis, reflect on some central themes in my argumentation, and relate my insights to the existing literature.

7.1 Producing and Observing "Economic Behaviour"

What I sought to answer by reconstructing experimental economists' epistemic practices was the question how economists have appropriated the method of laboratory experiments within an academic discipline, which is traditionally non-experimental. I followed these epistemic practices through different stages of conceptualising, designing, and running a laboratory experiment. The process of developing an experiment from real-life inspiration involves several steps of conceptualisation, reconceptualisation, and reduction that remove all factors and aspects of a situation except those that are considered relevant from an economist's point of view. A formal and conventional way of arriving at such a reconceptualisation is the construction of a theoretical model, which is then used to make predictions about the behaviour that should be observed in the lab (given that the agents are rational and self-interested), and implemented as the basis of the experimental design. Many experiments, however, are done without developing a "formal model" first, instead testing hypotheses that are based on experience and established knowledge about the behaviour of laboratory subjects. In these cases, I was told, it is still advisable to at least develop a formal explanation of the behaviour observed afterwards, because "having a model" improves publication chances. The epistemic advantage of laboratory experiments in comparison to models was described to me as the participation of "real human subjects" whose behaviour might deviate from those of the perfectly rational economic agents featured in formal models. Experiments therefore on the one hand serve as robustness tests for theoretical models; on the other hand, they provide insights for developing theory further and incorporating behavioural regularities to construct more empirically adequate models.

I have argued that the centrality of theoretical reasoning and the fact that such "behavioural models" are seen as less general each point toward an assumption that the behaviour described by standard economic theories – one might call it "economic behaviour" – is, methodologically speaking, the baseline phenomenon. Behaviour that is informed by "social factors" is seen as a deviation from this general case. One example is the use of neutral language and avoidance of value-laden terms in the instructions for experiments that aim to provide "general insights". I was told, for example, that framing an experiment that is based on a "general" model as an instance of a specific real-life situation by using value-laden terms would "weaken" the theory provided (Transcript P5).

Corresponding to this assumption, I found that the notion of experimental control my respondents invoked is associated with the ability to bring about "economic behaviour" or at least behaviour that can be described by theoretical concepts. The practices involved in experimental design and managing participants show that experimenters need to remove many obstacles in order to produce the conceptually tractable phenomena that interest them. Participants need to learn what their preferences are within the incentive structure that they are presented with, and what decisions correspond to those preferences. Even in the reduced and focused environment of the laboratory, strategic economic behaviour is not immediately observable but needs to be carefully elicited. Researchers need to write the right instructions, design the right kind of interactions (not too easy, not too difficult, not too few, not too many), and choose the right parameters and incentives. The laboratory is mobilised as an additional resource for producing economic behaviour through reconfiguring ordinary university students as economic agents, and providing conditions that resemble those represented by theoretical models. In the course of this laboratorisation of participants and their behaviour, the available courses of action are restricted to those that immediately matter for the success of the experiment. In particular, interaction and motivations are reduced to those factors that ideal economic agents would take into consideration. Social connotations are avoided, individual characteristics are rendered irrelevant, the material environment and the interactions with lab assistants are standardised, and indirect observation of participants' progress allows the experimenter to take precautionary measures if any one participant is about to jeopardize the success of the experimental session. The instructions that participants receive contribute to this reconfiguration by highlighting that the outcomes participants should prefer translate into higher monetary earnings at the end of the session.

The assumption that preferences over outcomes can be induced through monetary incentives and in this way, to some extent, are controlled by the experimenter (cf. Smith, 1976), is the methodological basis of experimental economics research. In Knorr Cetina's (1992, 1999) terms, the procedure of eliciting behaviour that is characteristic of experimental economics is a procedure that makes use of the "malleability" of its object. Economics experiments are possible because human agents can learn to behave in a way that is conceptualised as a best response to a strategic environment and laboratory conditions provide a context in which such learning is facilitated. Laboratory experiments in this sense are a straightforward example for the "performative" or "provocative" (Muniesa, 2014) nature of economics. In the laboratory, economists can build conditions favourable to producing the "economic behaviour" they are describing and studying. What we can learn from experiments is under what conditions this behaviour can (or cannot) be produced and observed. In response to my first research question, how economists appropriate the method of laboratory experiments, I have argued that laboratory experiments present a method that allows economists to produce and observe the phenomena that their theoretical models describe. The phenomena produced in the laboratory are reducible to a limited number of dimensions, and therefore still tractable by conceptual means. Economists have appropriated laboratory experiments as a method to produce "real" behaviour in controlled and simplified conditions, which allows them to either test standard theories or suggest the need for new theoretical concepts.

7.2 A Diversity of Conceptions Within One Dominant Regime

In accordance with this observation, experimental economists describe the epistemic value of their method as the ability to identify causal relations by observing, isolating, and manipulating factors in a way that would not be possible in the wild. Beyond these virtues, some of my respondents found value in other aspects of experimentation that they invoked indirectly through a comparison to the research practices of various natural sciences. With this comparison, they mobilised what I referred to as "ideal types" of experimental methodology to justify their individual views on the nature and the purpose of experimentation as a method: the theory-driven experiment of the physicist, the controlled medical trial and the "pure experimentation" a biologist performs on model organisms. From these characterisations, it becomes evident that there is no unified definition of the experimental method beyond a very general level, not even within the considerably young discipline of experimental economics. Rather, how experiments engage with the empirical, what is conceptualised as good and interesting results, and even the notions of relevance and novelty depend on the intended purpose of an experiment, i.e., on the answers to questions like: Is it an exploratory experiment that aims to "just observe" behaviour in controlled conditions in order to develop new insights and hypotheses; is it an experiment that is meant to test the causal efficacy of a policy or intervention; or is it an experiment that is guided by theoretical predictions and meant to provide inputs for theory-testing and theory development? The three purposes each entail considerably different approaches to the empirical, different levels and types of control, and different levels of integration with theoretical reasoning.

These different ideals of methodology corresponded to a range of motivations for doing experiments and several principles of evaluation concerning, for example, the practical feasibility of experimental studies and their relevance to professional and non-professional audiences. Still, what seemed to be the central concern of my interviewees, particularly of the junior researchers, is designing a study and writing it up in such a way that it has good chances to be published in a high-ranking journal. In line with earlier studies on evaluative practices in the life sciences (Müller, 2012; Fochler, Felt, & Müller, 2016), I argue that this reasoning corresponds to an institutionalised regime of valuation in which an individual researcher is predominantly assessed in terms of her journal publications. Accordingly, the junior researchers among my respondents described that their career perspective hinges on how well they can comply with the expectations of their peer community in terms of these publications. Considerations concerning the practical feasibility of experiments, for example, or the fruitfulness of unconventional approaches need to be traded off with conforming to what is perceived as the dominant view of proper experimental practice. The perceived criteria and standards of experimental practice that peer reviewers are expected to advocate inform a range of decisions in the experimental process, from the choice of the hypothesis to be tested, to the experimental design and choice of parameter values, the writing of instructions, and the choice of statistical techniques. Thus, while there are a variety of conceptions of good experimental practice, the approach to experimentation that translates into good publication chances is seen as providing the yardsticks for evaluating all experimental studies within a publication-oriented regime of valuation.

7.3 Theory-driven or Observation-Oriented Experiments?

What is this dominant approach, and how did it become so entrenched? A recurring theme in my observations was a tension between experimental approaches that focus on observation and empirical regularities, and approaches that focus on theory. This tension was manifested in many of the aspects I described in my analysis. It is very visible, for example, in the question of whether an experiment should start from formally derived theoretical predictions, or whether "qualitative predictions" based on previous knowledge and experience are sufficient. I referred to this as the question of whether or not to follow what was described as the "textbook approach" to economics experiments (namely building a formal model first and then implementing its elements exactly in the laboratory) so that the predictions of the model can be tested in similar conditions. The tension between this approach and the one that was dubbed as "more behavioural" comes from whether the experiment should start from theoretical predictions, but it also comes from how important it is to provide a formal theory to explain the observations. Experimenters who are interested in doing behavioural experiments or identify themselves as coming from a behavioural tradition give considerably more weight to the experimental observations themselves. Their approach is characterised by a practical understanding of what participants are likely to do in certain situations, and by exploring how behaviour changes when some aspects of the situation are changed. They do often provide theoretical explanations for the behavioural regularities they observe, but they see the production of such regularities as a viable epistemic practice in itself as well as being vital for enhancing the understanding of actual human behaviour.

This difference in approaches and in the value of experimental observations is reflected in how the role of participants is characterised. In general, the epistemic value of experiments resides in the capability of the experimental material – in this case, the human participants – to resist the expectations of the experimenter. Yet there is only so much resistance an experiment can afford in order to still provide useful results. The requirement, after all, is that experimenters will be able to interpret what they are observing as a response to one of the factors they deliberately introduced and varied in the experiment, in order to identify a causal relation. If experimenters are uncertain about what caused the behaviour they are observing – if there are too many alternative explanations – then the experiment, I was told, is “lost” (Transcript P4). It is therefore paramount to make sure that the actions of the participants in the laboratory is a response to the known characteristics of the laboratory environment and the experimental design, not something else. Furthermore, in the instances where idiosyncratic reasons for behaviour prevail, they are rendered insignificant through randomisation. Experiments that start from a theoretical prediction and implement a formal model in the laboratory environment have the advantage that the conceptual understanding of the expected behaviour is well-defined. If this understanding is mistaken, or if the experimental design was flawed, this will be shown directly when participants' behaviour is not what it was expected to be. Where this is the case, experimenters have a range of strategies available for adjusting the “material procedure” (Pickering, 1989) of the experiment. They can change parameter values, they can introduce a role-switch to enhance participants' understanding of the incentive structure of their partners, and they can, if they have a big enough budget, run more experimental sessions or even move to a different laboratory to see whether the deviant behaviour was a local or temporal phenomenon.

All of these strategies aim at aligning the observations more closely to the predictions and expectations of the experimenters. A failure to do so presents a serious problem for theory-driven experimentation, because the observed behaviour cannot be explained. In behavioural experiments, experimenters likewise aim to align participants' behaviour with their expectations, for example, by discouraging cooperation in a control treatment, such that an institution that is intended to enhance cooperation can be shown to work in other treatments. Yet these experiments start from the assumption that standard theory will fail to predict what is actually happening in the experiment, because it does not account for important driving forces of behaviour in certain situations. Behavioural experiments take phenomena that cannot be conceptualised by standard theory as their starting point and often produce them on purpose, without necessarily providing a formal explanation immediately. Since the factors driving behaviour that deviates from standard predictions are manifold and context-specific, the establishment of a behavioural regularity is more likely to spawn a series of new experiments than result in a unified theoretical account. In this sense, human agency itself provides the main source of new insights in behavioural experiments, whereas in theory-driven experiments, the aim is to arrive at a *three-way coherence* between a rigid phenomenal model, the established method of eliciting participants' preferences, and the results of the experimental procedure (Pickering, 1989; Santos, 2010). The tension between experimentation that focuses on theory and experimentation that focuses on observation, in this sense, corresponds to an ambiguous conception of participants' behaviour as being an epistemic resource and an epistemic risk at the same time.

Given that the theory-driven “textbook approach” which features theoretical predictions made with the help of a formal model is a more “rigid” experimental system (Santos, 2010) that allows less “conceptual manoeuvring” to accommodate unexpected results, there are methodological arguments explaining why this approach has become dominant in the discipline and widely associated with good publication chances. One of my respondents, for example, argued that experiments need “benchmarks” that can best be provided by theoretical predictions, and that experiments which pay less attention to theory might allow for “too many [researcher] degrees of freedom” (Transcript P7; cf. Simmons, Nelson, & Simonsohn, 2011). Beyond this methodological argument, there are historical reasons for the prominence of theory-centred approaches in experimental economics. According to Svorencik (2015), presenting the new method and discipline as a tool for testing and further developing the latest economic theories was necessary for experimental economics research to become accepted in the academic mainstream. More than three decades after the first experimental papers were published in prestigious general-interest economic journals, this alignment with theory is still very present in experimentalists’ epistemic practices, in their publication strategies and reflections and criticisms on their own discipline.

What, on Svorencik’s account, has been an early concern of experimental economists, namely to avoid becoming “ghettoised” by getting their research published in the leading general-interest journals, informs valuation practices in experimental economics to this day. In addition to satisfying the requirements of these journals, experimental economists have developed their own publication criteria by now, which were referred to on several occasions during my interviews. Through the general need to publish more and earlier, the criteria for journal publications has acquired an important epistemic role not only in experimental economics. This is also the case in other disciplines, as contemporary research on publication strategies in the life sciences (Rushforth & de Rijcke, 2015; Müller, 2012) shows. Such widely accepted criteria have helped to advance and unify the methodology of the new discipline of experimental economics. These criteria, in combination with the goal to align the novel type of research with the state of the art in economic theorising, also contributed to one particular type of laboratory experimentation becoming entrenched and known as the “textbook approach” that provides the yardsticks for evaluating experimental research in general. Asking experimental economists about how what they and others consider to be good experimental methodology informs their practical work of designing and running experiments therefore reveals that a discipline’s development is not only shaped by its internal “driving forces” (Svorencik, 2015), or the specific elements of its epistemic culture. It is also a matter of institutional logics and regimes of valuation that govern academic research in general, in correspondence to which disciplinary criteria and strategies of valuation are articulated.

7.4 The Advantages and Limitations of Economists’ Conceptual Tools

In the introduction to this thesis, I mentioned that experimental psychology has recently suffered from a “replication crisis” and that the behavioural sciences in general are suspected of operating with too many researcher degrees of freedom, resulting in unreliable and non-replicable experimental studies. I also mentioned that experimental and behavioural economists so far seem rather unimpressed by this threat, although direct replications of entire studies are considerably rare in this discipline as well.

When comparing the experiments I observed with those in developmental psychology described by Peterson (2016), it does appear that the reconfiguration of research objects is more straightforward in experimental economics, which gives the latter discipline an epistemic advantage over other behavioural sciences. The reason for this is that in their arsenal of theories and models, economists already have very established tools for conceptual reconfiguration at hand. The phenomena they then create and study in the lab are clearly delineated beforehand, theoretical predictions are precise, and deviations are considerably easy to identify. In other words, there are clear conventions available for arriving at a conceptual understanding - the "phenomenal model" in Pickering's (1989) terms - of many processes that are studied in the lab. The biggest methodological problem is that experimental economists do not know which of the auxiliary assumptions supporting their predictions are relaxed in the experimental environment, but by standardising the laboratory, minimising social connotations and inducing preferences via monetary rewards (by developing a standard "instrumental model"), they have found ways to approximate their notion of perfect control. Although economics was long considered to necessarily be a non-experimental science, it appears that it was only a small step from the idealised agents and institutions described in its theories and models to the laboratorised agents and institutions of experimental economics.

These powerful conceptual tools that laboratory experiments build on, whether they are theory-driven or "more behavioural", influence which phenomena experimenters can engage with and in what way. I mentioned earlier that even behavioural research, which is intended to be a "repair shop" and modify standard models to be more empirically adequate, is methodologically based on an assumption that the idealised "economic behaviour" described by standard theory is the general phenomenon. Whatever is observed in experiments needs to be conceptualised in relation to standard theory. In this sense, the conceptual tools provided by established economic theory also limit what experimenters can study, observe and describe, and what solutions they can provide to the problems they are investigating experimentally. Furthermore, the adherence to the methodological precepts that were defined by Smith (1976, 1982) for a certain type of experiments, those studying market institutions, to all types of experiments in economics might preclude a profound engagement with strategic behaviour that is driven by other motivations than monetary incentives. The latter is a point that Guala (2005, 2007) also raises in his critique of the theory-testing approach in economics. That experimental economics is so strongly grounded in and aligned with mainstream economic theory has had great advantages for the institutionalisation of the discipline (Svorencik, 2015). It has, however, also contributed to the import of some elements of the epistemic culture of academic economics that may not always be fruitful in the context of laboratory experimentation.

This raises questions for the recent import of the method of laboratory experiments to other social sciences, for example, political science and sociology. These sciences have a rich conceptual repository, but it is less unified than in economics, and they lack the technique of formal modelling that is such a central element of experimenting in economics. In addition, the concepts that experimental economists employ are of a specific kind. Experimental economists' definitions of notions that in real life carry many connotations, such as "advisor", "cooperation" or "tax avoidance", are such that they can be instantiated in a laboratory environment. This implies that the meaning attached to any notion of this kind is narrowed down to one particular understanding, which can be unambiguously described in theory and represented, for example, in the different choices an agent has available in an experiment. In my view, it is at least questionable whether political scientists and sociologists can and should strive to narrow down their theoretical concepts in a similar way, if this is a precondition, as my analysis suggests, for laboratorising their research objects. The experimental method might be most useful for social sciences that are not as strongly theory-driven as economics when it is used for designing interventions. As Guala (2007) argues, application-oriented research involves several steps of making sure that the laboratory and the target situation resemble each other in all respects that might be causally relevant. This type of experimental research is therefore ultimately less reductive than theory-centred experimentation, and since its insights are meant to be context-specific, not "generic" (Santos, 2010), it allows for a fuller range of meanings and connotations to be represented.

The narrow and decontextualized meaning of experimental economists' theoretical concepts also raises questions for the applicability of experimental insights to real-life situations. There is an inherent difficulty in extrapolating from any laboratory to real-life contexts, because real-life situations (for examples cases of corporate tax avoidance or lobbying) are always too rich to be fully captured either conceptually or in a laboratory experiment. When experimental economists want to study such situations, they reduce them to the elements that they consider as essential from an economists' point of view. In the tax avoidance experiment, these elements were identified as two buyers and two sellers interacting in a market with a transaction fee, as well as the information whether firms had paid an adequate amount of said fee. In the transparency experiment, the entire complex of lobbying, special interest groups, scientific advice and policymaking was reduced to a "signalling game" played by a less-informed decision-maker and a better-informed advisor who may have been paid for her advice. All the norms, moral connotations and political implications associated with both cases are carefully stripped away in the experiment. That economics experiments deal with highly reduced and abstracted notions of, say, "advisors" and "decision-makers" or "tax avoidance" is their strength, because it allows them to study situations and phenomena that cannot be studied otherwise. It becomes a problem, however, when the insights gained from laboratory experiments are presented as applicable to a wide range of real-life situations in which advising and decision-making or tax avoidance play a role. This is not to say that experimenters themselves would not be cautious with making such extrapolations. My respondents, for example, highlighted the obvious limitations of the insights gained from laboratory experiments stemming from the artificiality and simplicity of the experimental situation and the use of a "convenience sample" of undergraduate students. That the conclusions drawn from laboratory observations might be valid for the laboratory context only is not

necessarily a worry for experimental economists. On the one hand, they describe the correspondence between the lab and the world as being located on the level of very general patterns of behaviour, on the other hand they see the application of results to specific real-world situations as a separate task.
Still, some experimenters do aim to apply their insights to specific contexts in which behaviour is not as easily provoked as in laboratories. This came to the fore in one of my interviews, with the researcher's explicit motivation to be able to present his/her work to non-professional audiences. It is also expressed in the newly established *Vienna Behavioral Economics Network*[36] and its events, which are meant to engage stakeholders from industry and public administration in a dialogue with behavioural economics research. Studying how experimenters make such extrapolations in practice, when they present their results to non-academic audiences, would be an interesting topic for further research. What can be said, in the context of this thesis, is that wherever experimental insights are transferred beyond their context of origin we should be aware that they are the product of a research practice in which the concepts experimenters use take on a narrow and specific meaning that might be far removed from our everyday use.

7.5 Diversity in Practice

My research project started from the observation that there was a gap in STS literature where it comes to studying the research practices of social scientists. The existing secondary literature on experimental economics is concerned with epistemic (Santos, 2010), methodological (Guala, 2005) and historical (Svorencik, 2015) issues. My study is the first to explicitly focus on experimental economists' research practices, insofar as they can be observed during laboratory experiments and reconstructed from researchers' accounts. What, then, does my case study on practices, both epistemic and evaluative, in experimental economics reveal and contribute to the existing literature? In terms of the description of practices I have provided, I reported a great diversity in approaches and purposes even within my small sample of interview respondents and their research projects. Some of these approaches come into conflict in experimental practice and in the discursive valuation of experimental research.
The practice of designing an experiment is particularly multi-layered and involves a variety of skills both in terms of conceptual knowledge as well as experience and acquired skill. This is the case whether an experiment is what I have called "theory-driven" or "behavioural". Behavioural economists have a strong background in economic theory and know how to use the "standard models" their research seeks to overcome. Nevertheless, their experimental practice is less centred on these models and instead builds mainly on experiential knowledge about human behaviour in laboratory experiments, and on how to best design the "material procedure" of an experiment. Theory-driven

[36] Vienna Center for Experimental Economics. (2017) Vienna Behavioral Economics Network. Retrieved from http://vcee.univie.ac.at/vben/ , 15.04.2017.

experimenters, however, cannot do without this skill either, as they too need to anticipate participants' reactions and their interpretations of the experimental procedures they design in order to "induce behaviour" and enable "learning". In general, experimentation in economics involves the mobilisation of a range of resources. Examples I described in my analysis are conceptual tools (the games, models, and concepts that form the basis of what is studied in the lab), skilled experience and an empathic sense for participants' behaviour. Another important resource are the material conditions of a standardised laboratory environment, including assistants to perform all the menial tasks and follow the codes of conduct that are necessary for an experiment to run smoothly.

On the level of experimental purposes, I have observed greater diversity and a more nuanced register of motivations than it has been described in normative accounts of experimental economics methodology. In particular, the experimental studies I discussed with my respondents do not neatly fall into either the category of theory-testing or the institution-building approaches identified by Guala (2007). The only experiment in my sample that was clearly application-oriented was a test of a policy in the lab, namely, whether transparency about a firm's taxpaying would empower consumers through enabling them to boycott the firms that paid fewer taxes. In terms of its design, it was not inherently different from the other experiments I discussed, and it was clearly seen only a first step towards establishing that this policy could be efficient. The experiment on transparency in lobbying had a similar motivation. The main difference between these two experiments is that in the first one, developing a formal model was seen more as an obligatory act, whereas the second one aimed to make its point, that transparency is never harmful, first theoretically and then corroborate the insights gained from the modelling with experimental results. Incidentally, there was a third experiment in my sample, which started from a purely conceptual interest: the experimental exploration of markets of so-called credence goods; and was at a later stage reframed as providing insights about advice and decision-making between physicians and their patients. Here, the applicability to a specific real-world context was an outcome of the development of the research project, rather than motivating it from the beginning. What my analysis has shown is that there are more ways of relating to empirical reality and real-life problems than those criticising experimental economics as too theory-driven (Guala, 2007; Reiss, 2008) seem to allow for.

Considering Svorencik's historical study on the driving forces behind the "experimental turn" in economics, it seems that many of the motivations he identified amongst the first generation of experimental economists still ring true for those who are today looking to laboratory experiments as a means of integrating and improving economic theorising with empirical data. The motivations that my analysis identified beyond those described in his account characterise experimental economics as a contemporary academic discipline. Concerns of accountability, relevance and the topicality of research are amongst the concerns that come more to the fore once a new method has been established and accepted by the mainstream. This suggests the conclusion that the epistemic practices of a research community are only partly shaped by a distinct cultural form of producing "machineries of knowing"

(Knorr Cetina, 1999). Just as much as they are grounded in an epistemic culture, epistemic practices are responsive to changing competitive environments and the conditions in which academic research is produced. It will be interesting to observe how the discipline and the method of experimental economics will further develop, and in what ways its results will be put to use in real-world contexts. The future development of experimental economics will not only be a function of its intrinsic "driving forces" or of the purposes that its methodology is most apt to serve. It will also be shaped by future solutions to the organisation and governance of academic research, the distribution of knowledge, and the regimes of valuation that emerge with these solutions and provide the principles according to which individual researchers can assess their work and their personal capacities, and build academic careers.

If what is now the main regime of valuation in experimental economics and in academic research in general continues to be dominant in the future, I would not expect the research practices of experimental economics to change in a fundamental way. In particular, it can be expected that the theory-centred "textbook approach" continues to provide the model of an ideal methodology, and that there will not be a shift towards a more application oriented "building" approach any time soon. The reason for this is that there does not seem to be enough demand for the design of policy interventions, compared to the number of active experimental economists, and that research that does not neatly translate into (ideally, several) journal publications is not very attractive, particularly for younger researchers. If, on the contrary, new forms of valuation for academic research emerge, there is potential for a wide range of new approaches to experiments in economics as well. Approaches that at the time of writing were described as difficult to publish could become more accepted once they circulate through different outlets than high-profile journal articles and are found to be useful. The increasing use of online databases also provides an avenue for experimental research that is less focused on the computer laboratory as the main locus of knowledge production. These experiments give up some of the control that is characteristic of laboratory experiments for the advantage of collecting more "ecologically valid" data. Probably, controlled field experiments might also rise to greater prominence, for similar reasons. All of these developments would signify a shift towards studying behaviour in situations that are less like those represented by theoretical models, an increasing validity of observational data over theoretical reasoning, and, in this sense, an emancipation of experimental practice from the theory-driven paradigm of mainstream neoclassical economics.

8 References

Alac, M. (2008). Working with brain scans: Digital images and gestural interaction in fMRI laboratory. *Social Studies of Science*, *38*(4), 483–508. https://doi.org/10.1177/0306312708089715

Anderson, E. (2004). Uses of value judgments in science: A general argument, with lessons from a case study of feminist research on divorce. *Hypatia*, *19*(1), 1–24. https://doi.org/10.1111/j.1527-2001.2004.tb01266.x

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452–454. https://doi.org/10.1038/533452a

Bardsley, N., Cubitt, R., Loomes, G., Moffat, P., Starmer, C., & Sugden, R. (2010). *Experimental economics: Rethinking the rules*. Princeton: Princeton University Press.

Bhaskar, R. (1985). *Experiment*. In W. F. Bynum, E. J. Browne, & R. Porter (Eds.), *Dictionary of the history of science* (pp. 136–138). Princeton, NJ: Princeton University Press.

Bogner, A., Littig, B., & Menz, W. (2014). *Interviews mit Experten*. Wiesbaden: Springer Fachmedien Wiesbaden. Retrieved from http://link.springer.com/10.1007/978-3-531-19416-5

Burrows, R. (2012). Living with the h-index? Metric assemblages in the contemporary academy. *The Sociological Review*, *60*(2), 355–372. https://doi.org/10.1111/j.1467-954X.2012.02077.x

Callon, M. (Ed.). (1998). *The laws of the markets*. Oxford; Malden, MA: Blackwell Publishers/Sociological Review.

Camerer, C. F. (2015). The promise and success of lab-field generalizability in experimental economics: A critical reply to Levitt and List. In G. R. Fréchette & A. Schotter (Eds.), *Handbook of experimental economic methodology* (pp. 249–295). New York, NY: Oxford University Press.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., … Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*(6280), 1433–1436. https://doi.org/10.1126/science.aaf0918

Cartwright, N. (2010). Models: Parables v fables. In R. Frigg & M. Hunter (Eds.), *Beyond mimesis and convention* (Vol. 262, pp. 19–31). Dordrecht: Springer Netherlands. Retrieved from http://www.springerlink.com/index/10.1007/978-90-481-3851-7_2

Charmaz, K. (2006). *Constructing Grounded Theory: A practical guide through qualitative analysis.* London: Sage.

Collins, H. M. (1992). *Changing order: Replication and induction in scientific practice*. Chicago: University of Chicago Press.

Dahler-Larsen, P. (2014). Constitutive effects of performance indicators: Getting beyond unintended consequences. *Public Management Review*, *16*(7), 969–986. https://doi.org/10.1080/14719037.2013.770058

Daston, L., & Galison, P. (2010). *Objectivity* (Paperback ed). New York, NY: Zone Books.

Delamont, S. (2007). Ethnography and participant observation. In C. Seale, G. Gobo, J. F. Gubrium, & D. Silverman (Eds.), *Qualitative research practice* (pp. 205–217). London: Sage.

Doing, P. (2008). Give me a laboratory and I will raise a discipline: The past, present, and future politics of laboratory studies in STS. In E. J. Hackett, O. Amsterdamska, M. E. Lynch, & J. Wajcman (Eds.), *The handbook of science and technology studies* (pp. 279–295). Cambridge, Mass: MIT Press.

DORA. (2013). San Francisco declaration on research assessment. Retrieved from http://www.ascb.org/files/SFDeclarationFINAL.pdf?x30490

Dussauge, I., Helgesson, C.-F., & Lee, F. (2015). Valuography: Studying the making of values. In I. Dussauge, C.-F. Helgesson, & F. Lee (Eds.), *Value practices in the life sciences and medicine* (pp. 267–285). Oxford University Press.

Dussauge, I., Helgesson, C.-F., Lee, F., & Woolgar, S. (2015). On the omnipresence, diversity, and elusiveness of values in the life sciences and medicine. In I. Dussauge, C.-F. Helgesson, & F. Lee (Eds.), *Value practices in the life sciences and medicine* (pp. 1–28). Oxford University Press.

Engel, C. (2011). Dictator games: a meta study. *Experimental Economics*, *14*(4), 583–610. https://doi.org/10.1007/s10683-011-9283-7

Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, *113*(1), 1–40. https://doi.org/10.1086/517897

Falk, A., & Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, *326*(5952), 535–538. https://doi.org/10.1126/science.1168244

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868. https://doi.org/10.1162/003355399556151

Felt, U. (2009). Introduction: Knowing and living in academic research. In U. Felt (Ed.), *Knowing and living in academic research. Convergence and heterogeneity in research cultures in the European context* (pp. 17–39). Prague: Inst. of Sociology of the Acad. of Sciences of the Czech Republic.

Felt, U., Fouché, R., Miller, C. A., & Smith-Doerr, L. (Eds.). (2017). *The handbook of science and technology studies* (Fourth edition). Cambridge, Massachusetts: The MIT Press.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178. https://doi.org/10.1007/s10683-006-9159-4

Fochler, M. (2016). Variants of epistemic capitalism: Knowledge production and the accumulation of worth in commercial biotechnology and the academic life sciences. *Science, Technology & Human Values*, *41*(5), 922–948. https://doi.org/10.1177/0162243916652224

Fochler, M., Felt, U., & Müller, R. (2016). Unsustainable growth, hyper-competition, and worth in life science research: Narrowing evaluative repertoires in doctoral and postdoctoral scientists' work and lives. *Minerva*, *54*(2), 175–200. https://doi.org/10.1007/s11024-016-9292-y

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Frey, B. S. (2003). Publishing as prostitution? – Choosing between one's own ideas and academic success. *Public Choice*, *116*(1), 205–223. https://doi.org/10.1023/A:1024208701874

Garfield, E. (2006). The history and meaning of the Journal Impact Factor. *JAMA*, *295*(1), 90. https://doi.org/10.1001/jama.295.1.90

Garforth, L. (2012). In/Visibilities of research: Seeing and knowing in STS. *Science, Technology & Human Values*, *37*(2), 264–285. https://doi.org/10.1177/0162243911409248

Gieryn, T. F. (2002). Three truth-spots. *Journal of the History of the Behavioral Sciences*, *38*(2), 113–132. https://doi.org/10.1002/jhbs.10036

Gooding, D. (1992). Putting agency back into experiment. In A. Pickering (Ed.), *Science as practice and culture* (pp. 65–112). Chicago: University of Chicago Press.

Gooding, D., Pinch, T. J., & Schaffer, S. (Eds.). (1989). *The uses of experiment: Studies in the natural sciences*. Cambridge [England] ; New York: Cambridge University Press.

Graber, M., Launov, A., & Wälde, K. (2008). Publish or perish? The increasing importance of publications for prospective economics professors in Austria, Germany and Switzerland. *German Economic Review*, *9*(4), 457–472. https://doi.org/10.1111/j.1468-0475.2008.00449.x

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 114–125. https://doi.org/10.1007/s40881-015-0004-4

Gross, M. (2015). Give me an experiment and I will raise a laboratory. *Science, Technology & Human Values*. https://doi.org/10.1177/0162243915617005

Guala, F. (2005). *The methodology of experimental economics*. Cambridge; New York: Cambridge University Press.

Guala, F. (2007). How to do things with experimental economics. In D. MacKenzie, F. Muniesa, & L. Siu (Eds.), *Do economists make markets? On the performativity of economics.* (pp. 128–162). Princeton: Princeton University Press.

Guggenheim, M. (2012). Laboratizing and de-laboratizing the world: Changing sociological concepts for places of knowledge production. *History of the Human Sciences*, *25*(1), 99–118. https://doi.org/10.1177/0952695111422978

Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science*. Cambridge: Cambridge Univ. Press.

Hammersley, M., & Atkinson, P. (2007). Oral Accounts and the role of interviewing. In *Ethnography. Principles in Practice.* (3rd edition, pp. 97–120). London; New York: Routledge.

Haucap, J., & Muck, J. (2015). What drives the relevance and reputation of economics journals? An update from a survey among economists. *Scientometrics*, *103*(3), 849–877. https://doi.org/10.1007/s11192-015-1542-5

Helgesson, C.-F., Lee, F., & Lindén, L. (2016). Valuations of experimental designs in proteomic biomarker experiments and traditional randomised controlled trials. *Journal of Cultural Economy*, *9*(2), 157–172. https://doi.org/10.1080/17530350.2015.1108215

Heuts, F., & Mol, A. (2013). What is a good tomato? A case of valuing in practice. *Valuation Studies*, *1*(2), 125–146. https://doi.org/10.3384/vs.2001-5992.1312125

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature News*, *520*(7548), 429. https://doi.org/10.1038/520429a

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Kahnemann, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291.

Kjellberg, H., & Mallard, A. (2013). Valuation Studies? Our collective two cents. *Valuation Studies*, *1*(1), 11–30. https://doi.org/10.3384/vs.2001-5992.131111

Knorr Cetina, K. (1981). *The manufacture of knowledge: An essay on the constructivist and contextual nature of science*. Oxford; New York: Pergamon Press.

Knorr Cetina, K. (1988). Das naturwissenschaftliche Labor als Ort der “Verdichtung” von Gesellschaft. *Zeitschrift Für Soziologie*, *17*(2), 85–101.

Knorr Cetina, K. (1992). The Couch, the cathedral and the laboratory: On the relationship between experiment and laboratory in science. In A. Pickering (Ed.), *Science as practice and culture* (pp. 113–138). Chicago: University of Chicago Press.

Knorr Cetina, K. (1995). Laboratory studies. The cultural approach to the study of science. In S. Jasanoff, G. E. Markle, J. C. Peterson, & T. Pinch (Eds.), *Handbook of science and technology studies* (pp. 140–166). Thousand Oaks, Calif: Sage Publications.

Knorr Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, Mass: Harvard University Press.

Kohler, R. E. (2008). Lab history: Reflections. *Isis*, *99*(4), 761–768. https://doi.org/10.1086/595769

Kourany, J. A. (2010). *Philosophy of science after feminism*. Oxford; New York: Oxford University Press.

Krohn, W., & Weyer, J. (1994). Society as a laboratory: The social risks of experimental research. *Science & Public Policy*, *21*(3), 173–183.

Lamont, M. (2012). Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology*, *38*(1), 201–221. https://doi.org/10.1146/annurev-soc-070308-120022

Latour, B. (1983). Give me a laboratory and I will raise the world. In K. Knorr Cetina & M. Mulkay (Eds.), *Science observed: Perspectives on the social study of science* (pp. 141–170). London: Sage.

Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts*. Princeton, N.J: Princeton University Press.

Lawrence, P. A. (2003). The politics of publication. *Nature*, *422*(6929), 259–261. https://doi.org/10.1038/422259a

Ledyard, J. O. (1995). Public goods: A survey of experimental research. In J. H. Kagel & A. E. Roth (Eds.), *The handbook of experimental economics*. Princeton, NJ: Princeton University Press.

Lee, F. (2015). Purity and interest. On relational work and epistemic value in the biomedical sciences. In I. Dussauge, C.-F. Helgesson, & F. Lee (Eds.), *Value practices in the life sciences and medicine* (pp. 207–223). Oxford University Press.

Lehrer, J. (2010, December 13). The truth wears off. *The New Yorker*. Retrieved from http://www.newyorker.com/magazine/2010/12/13/the-truth-wears-off

Levitt, S. D., & List, J. A. (2015). What do laboratory experiments measuring social preferences reveal about the real world? In G. R. Fréchette & A. Schotter (Eds.), *Handbook of experimental economic methodology* (pp. 207–248). New York, NY: Oxford University Press.

Lourenço, J. S., Ciriolo, E., Almeida, S. R., & Troussard, X. (2016). *Behavioural insights applied to policy. European report 2016*. Publications Office of the European Union. Retrieved from http://publications.jrc.ec.europa.eu/repository/bitstream/JRC100146/kjna27726enn_new.pdf

Lux, T., & Westerhoff, F. (2009). Economics crisis. *Nature Physics*, *5*(1), 2–3. https://doi.org/10.1038/nphys1163

Lynch, M. (1985). *Art and artifact in laboratory science: A study of shop work and shop talk in a research laboratory*. London; Boston: Routledge & Kegan Paul.

MacKenzie, D. A., Muniesa, F., & Siu, L. (Eds.). (2007). *Do economists make markets? On the performativity of economics*. Princeton: Princeton University Press.

Merz, M., & Knorr Cetina, Karin. (1997). Deconstruction in a “thinking” science: Theoretical physicists at work. *Social Studies of Science*, *27*(1), 73–111.

Mody, C. C. M. (2001). A little dirt never hurt anyone: Knowledge-making and contamination in materials science. *Social Studies of Science*, *31*(1), 7–36. https://doi.org/10.1177/030631201031001002

Morgan, M. S. (2008). Models. In S. N. Durlauf & L. E. Blume (Eds.), *The new Palgrave dictionary of economics* (2nd ed., Vol. 5, pp. 654–663). Basingstoke: Palgrave Macmillan.

Müller, R. (2012). Collaborating in life science research groups: The question of authorship. *Higher Education Policy*, *25*(3), 289–311. https://doi.org/10.1057/hep.2012.11

Muniesa, F. (2014). *The provoked economy: Economic reality and the performative turn*. London; New York: Routledge.

Muniesa, F., & Callon, M. (2007). Economic experiments and the construction of markets. In D. MacKenzie, F. Muniesa, & L. Siu (Eds.), *Do economists make markets? On the performativity of economics* (pp. 163–189). Princeton: Princeton University Press.

Myers, N. (2008). Molecular embodiments and the body-work of modeling in protein crystallography. *Social Studies of Science*, *38*(2), 163–199. https://doi.org/10.1177/0306312707082969

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716–aac4716. https://doi.org/10.1126/science.aac4716

O’Reilly, K. (2005). *Ethnographic methods*. London ; New York: Routledge.

Osborne, M. J. (2004). *An introduction to game theory* (Repr.). New York, NY: Oxford Univ. Press.

Peterson, D. (2016). The baby factory: Difficult research objects, disciplinary standards, and the production of statistical significance. *Socius: Sociological Research for a Dynamic World*, *2*(0). https://doi.org/10.1177/2378023115625071

Pickering, A. (1989). Living in the material world: on realism and experimental practice. In D. Gooding, T. J. Pinch, & S. Schaffer (Eds.), *The uses of experiment* (pp. 275–297). Cambridge: Cambridge University Press.

Pickering, A. (1992). From science as knowledge to science as practice. In A. Pickering (Ed.), *Science as practice and culture* (pp. 1–26). Chicago: University of Chicago Press.

Pickering, A. (1993). The mangle of practice: Agency and emergence in the sociology of science. *American Journal of Sociology*, *99*(3), 559–589.

Pickering, A. (1995). *The mangle of practice: Time, agency, and science*. Chicago, Ill.: Univ. of Chicago Press.

Powell, W. W., & Snellman, K. (2004). The knowledge economy. *Annual Review of Sociology*, *30*(1), 199–220. https://doi.org/10.1146/annurev.soc.29.010202.100037

Power, M. (1997). *The audit society: Rituals of verification*. Oxford: Oxford Univ. Press.

Power, M. (2008). Research evaluation in the audit society. In H. Matthies & D. Simon (Eds.), *Wissenschaft unter Beobachtung. Effekte und Defekte von Evaluationen* (pp. 15–24). Wiesbaden: VS Verlag für Sozialwissenschaften.

Reiss, J. (2008). *Error in economics: Towards a more evidence-based methodology*. London; New York: Routledge.

Rijcke, S. de, Wouters, P. F., Rushforth, A. D., Franssen, T. P., & Hammarfelt, B. (2016). Evaluation practices and effects of indicator use—a literature review. *Research Evaluation*, *25*(2), 161–169. https://doi.org/10.1093/reseval/rvv038

Roth, A. E. (2015). Is experimental economics living up to its promise? In G. R. Fréchette & A. Schotter (Eds.), *Handbook of experimental economic methodology* (pp. 13–40). New York, NY: Oxford University Press.

Rushforth, A., & de Rijcke, S. (2015). Accounting for impact? The Journal Impact Factor and the making of biomedical research in the Netherlands. *Minerva*, *53*(2), 117–139. https://doi.org/10.1007/s11024-015-9274-5

Santos, A. C. dos. (2010). *The social epistemology of experimental economics*. London; New York: Routledge.

Schekman, R. (2013, September 12). How journals like Nature, Cell and Science are damaging science. *The Guardian*. Retrieved from https://www.theguardian.com/commentisfree/2013/dec/09/how-journals-nature-science-cell-damage-science

Schram, A. J. H. C., & Holt, C. A. (1998). Editor's preface. *Experimental Economics*, *1*, 5–6.

Shapin, S. (1988). The House of experiment in seventeenth-century England. *Isis*, *79*(3), 373–404. https://doi.org/10.1086/354773

Shapin, S., & Schaffer, S. (1985). *Leviathan and the air-pump: Hobbes, Boyle, and the experimental life*. Princeton, NJ: Princeton Univ. Press.

Shiller, R. J., & Shiller, V. M. (2011). Economists as worldly philosophers. *The American Economic Review*, *101*(3), 171–175.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Smith, V. L. (1962). An experimental study of competitive market behavior. *Journal of Political Economy*, *70*(2), 111–137.

Smith, V. L. (1976). Experimental economics: Induced value theory. *American Economic Review*, *66*(2), 274–279.

Smith, V. L. (1982). Microeconomic systems as an experimental science. *American Economic Review*, *72*(5), 923–955.

Spradley, J. P. (1979). *The ethnographic interview*. New York: Holt, Rinehart and Winston.

Stark, D. (2009). *The sense of dissonance: Accounts of worth in economic life*. Princeton: Princeton University Press.

Stock, W. A., & Siegfried, J. J. (2013). One essay on dissertation formats in economics. *American Economic Review*, *103*(3), 648–653. https://doi.org/10.1257/aer.103.3.648

Strathern, M. (Ed.). (2000). *Audit cultures: Anthropological studies in accountability, ethics, and the academy*. London New York: Routledge.

Svorencik, A. (2015, January 30). *The experimental turn in economics: A history of experimental economics* (Dissertation). Utrecht University, Utrecht. Retrieved from http://dspace.library.uu.nl/handle/1874/302983

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven: Yale University Press.

Traweek, S. (1992). *Beamtimes and lifetimes: The world of high energy physicists*. Cambridge, Mass: Harvard Univ. Press.

Vienna Center for Experimental Economics. (2016). Annual Report 2016. Retrieved from http://vcee.univie.ac.at/fileadmin/user_upload/proj_cent_experi_economics/Dokumente/Annual_Report_2016.pdf

What went wrong with economics. (2009, July 16). *The Economist*. Retrieved from http://www.economist.com.ezproxy.eui.eu/node/14031376/print

Whitley, R. (2007). Changing governance of the public sciences. The consequences of establishing research evaluation systems for knowledge production in different countries and scientific fields. In R. Whitley & J. Gläser (Eds.), *The changing governance of the sciences. The advent of research evaluation systems* (pp. 3–27). Dordrecht, the Netherlands: Springer.

Woolgar, S. (1988). Keeping inversion alive: Ethnography and reflexivity. In *Science: The very idea* (pp. 83–96). London; New York: Tavistock Publications.

Yonay, Y., & Breslau, D. (2006). Marketing models: The culture of mathematical economics. *Sociological Forum*, *21*(3), 345–386. https://doi.org/10.1007/s11206-006-9031-5

9 APPENDIX: Example of an Interview Guideline

BASIC INTERVIEW GUIDELINE

Adapted for Participant 5

- **Start with:** *introducing myself and my research topic. Ask for permission to record, assure anonymity, ask how much time the interviewee will have.*

- **Opening Question:**
  To begin with, can you tell me something about your personal motivation for doing experimental economics? Like, when did you first come in contact with this discipline, and why did you decide to do experimental studies yourself?
  - *Follow-up questions on interesting aspects, e.g.*
    - *Prior research specialisations*
    - *Specific research interests*
    - *Academic and personal influences*
    - *How long have you been doing experimental research?*
  - What did you have to learn in order to be able to conduct experimental studies? Where, and how did you learn it?
  - Have you been programming or running experiments yourself?

- **Theme Block 1:**
  How did you develop this particular study?

  <u>**The transparency study**</u>
  - In the paper, it is explained that this study is a contribution to the debate whether transparency can improve the accuracy of decision-making. Can you still recall why you became interested in investigating this problem with a theory-based experimental study?
  - How did you identify which aspects of this question would be suitable for theoretical modelling? To put the question differently, how do you ensure that your model captures the most important aspects and the underlying structure of this problem?
    - Did you encounter any difficulties when modelling the situation or when proving the model (your co-author said that s/he got the proofs wrong first time)?
    - Do you have the lab situation in mind already when you develop the model?

- I have read a few papers on experimental economics research, but I have never planned or conducted an experiment myself. Can you describe for me what the different intermediary steps were, from defining a research question to running the sessions? What were the different tasks involved and how did you distribute them between your co-author and yourself?
- Temporalities: When did you begin with planning and preparing the study, and when do you think that it might realistically be published?
- Did you encounter any problems when you were planning this study? Can you give me an example, and how you resolved it?

Questions about particular design choices:

- Do I understand correctly that you modelled the benefit of being honest – which in real-life cases might something like maintaining reputation or satisfying a preference not to lie – is modelled as a payoff? Is it a common strategy to quantify intrinsic values like this in order to incorporate them in theoretic models (and experimental designs)?
  [also referring to the question from the audience: how it can be modelled that beta > gamma if gamma is intrinsic?].
- In the presentation, you mentioned that payoffs were chosen such as to yield interesting predictions in the model, and to make a nice design for the experiment. Are parameters commonly chosen in this way?
- What were your expectations concerning the role switch after 20 periods? You said it was not necessary for learning and you would not do it again, why is that?
- Can you explain why you decided to use the strategy method?
- What did you ask participants in the questionnaire at the end?
- Did you exclude certain people (e.g. economics students or very experienced subjects) from participation? Why/not?
- Are there certain groups of participants you think are not so well suited for experiments? Why?
- Did you consider controlling for demographic characteristics in the analysis? Why/not?
- Concerning the **instructions**:
    - In the instructions, the two roles are referred to as “Sender” and “Receiver” rather than advisor and decision-maker. I have encountered the emphasis on neutral language already in other cases. Can you explain to me why neutral language is so important?
- Is there anything you would do differently if you were to do the study again? (In terms of modelling, design, analysis,…)

- There was quite some discussion at the end of the presentation whether your interpretation of the statistical tests was correct. Your co-author explained afterwards that the critique was justified, but that you would not be able to publish the paper in good journals otherwise. How much do the prospects of publication in general matter for your decisions when planning and writing up research?
    - What are the characteristics of an experimental study that has good chances of getting published?
    - What type of research would definitely be rejected?

- *What would you say were the concrete insights you got from the experimental results, which you could not have obtained from the model alone?*
- *What are the specific advantages of running an experiment in comparison to developing only the model?*
- *Could everything that you can develop a game-theoretic model of in principle also be implemented in a lab?*
- *Are all experiments based on such game-theoretic models or are there other ways of doing research in experimental economics that you are familiar with?*
- *What are the limits of experimental research? When does it get too complicated for a lab experiment?*

- **Can you maybe tell me something about other projects you are working on?**
- *What topics do you find most interesting? Why?*
- *Do you plan on staying in experimental economics? What would you like to focus on in the future?*

- **Theme Block 3:**
  What do you think that the limitations of experimental research are?
    - Do you share the concern of some critics of behavioural research that the subject pool – students and WEIRD populations – might not be very representative of the general population?
    - There has recently been some discussion about the **reproducibility** of experimental studies, particularly in the medical and behavioural sciences. Do you think that this is a relevant discussion for experimental economics, too?
    - How exactly do you **control the environment** for experiments? Can you give me an example of how you are standardising the procedure across different experiments?

        - Are there aspects which cannot potentially be controlled or standardised, or that are considered as irrelevant?
        - Are there any problems that you have frequently encountered in terms of standardising and controlling the environment, and ensuring the validity of experimental data?
    - *How important is **external validity**, in your opinion, for experiments in general?*

- **To conclude:**
    - How would you describe the relationship between experimental economics and more traditional mainstream economics?
    - How would you describe the research environment at the VCEE? Are there any noticeable differences to other places you have been working in? Is there something like a specific approach to this research department?
    - I felt that the atmosphere at the research seminar where you presented was almost hostile, at least a lot more agitated than I have seen in research seminars in my field. Is that something one would often encounter in experimental economics?

- **Is there anything you would like to add, that we have not yet talked about?**
- **Thank you very much for your time!**

10 Abstract Deutsch

Die vorliegende Arbeit beschäftigt sich aus der Perspektive der *Science and Technology Studies* (STS) mit Laborexperimenten in der Ökonomie. Traditionell waren die Wirtschafswissenschaften nicht-experimentell, doch seit den 1980er Jahren haben sich mit Behavioural Economics (Verhaltensökonomie) und Experimental Economics (Experimenteller Ökonomie) zwei Subdisziplinen herausgebildet, in denen Laborexperimente eine zentrale Rolle spielen. In diesen Experimenten agieren die TeilnehmerInnen zumeist über Computernetzwerke miteinander und treffen Entscheidungen, deren Resultate direkt in finanzielle Einnahmen für die TeilnehmerInnen übersetzt werden.

Während die epistemischen Praktiken, die soziale Organisation und die materielle Kultur der Natur- und Lebenswissenschaften in zahlreichen wissenschaftssoziologischen Studien detailliert beschrieben wurden, hat die wissenschaftliche Praxis der Sozialwissenschaften insgesamt noch sehr wenig Aufmerksamkeit erfahren. Gerade zu experimentellen Praktiken in den Sozialwissenschaften, einschließlich der oben genannten ökonomischen Subdisziplinen, gibt es aus wissenssoziologischer Perspektive nur vereinzelte Beiträge.

Die vorliegende Arbeit unternimmt einen ersten Schritt, diese Lücke zu schließen. In Bezug auf die Vielfalt „epistemischer Kulturen" (Knorr Cetina, 1999) und ihrer spezifischen Ansätze der Wissensproduktion stellt die Aneignung der experimentellen Methode in einer bisher nicht-experimentellen Disziplin einen interessanten Fall dar. Eine weitere Motivation ist der große institutionelle Einfluss der Ökonomie als akademischer Disziplin, sowie die zunehmende Bedeutung verhaltensökonomischer Maßnahmen in Politik und Verwaltung, die zum Teil auf Erkenntnissen aus Laborstudien beruhen.

In meiner Masterarbeit gehe ich den Fragen nach, wie sich ÖkonomInnen die Methode des Laborexperiments zu eigen machen, und nach welchen Prinzipien sie experimentelle Forschung bewerten. Diese Fragen beantworte ich anhand von sechs semi-strukturierten Interviews mit ForscherInnen am *Vienna Center for Experimental Economics* (VCEE) der Universität Wien, sowie durch Beobachtungen, die ich als Laborassistentin bei der Durchführung von Experimenten sammeln konnte. Sowohl Interviews als auch Beobachtungsnotizen wurden transkribiert und nach den Prinzipien der *Grounded Theory* (Charmaz, 2006) induktiv analysiert.

Ein zentrales Resultat meiner Analyse ist, dass Laborexperimente es möglich machen, Verhalten und Situationen zu produzieren, die sich an die Beschreibungen theoretischer Modelle annähern. Diese Annäherung wird durch mehrere Schritte der Reduktion erreicht: Die ForscherInnen identifizieren die aus ihrer Sicht fundamentalen ökonomischen Aspekte einer Situation, modellieren sie und benutzen diese Modelle als Grundlage eines experimentellen Designs. Dabei werden alle kontextspezifischen Elemente und Bedeutungen entfernt. In der Analyse der Gespräche zeigt sich, dass die epistemischen Praktiken der Forscherinnen durch eine Spannung zwischen theoriegeleiteten und

beobachtungsgeleiten Ansätzen gekennzeichnet sind. So haben experimentelle Studien einerseits den Anspruch, die Erkenntnisse und Vorhersagen theoretischer Modelle zu überprüfen, und müssen diese daher möglichst genau umsetzen. Zum anderen dienen Laborexperimente aber auch dazu, Phänomene zu untersuchen, die mit standardtheoretischen Modellen nicht beschrieben werden können. Diese Spannung setzt sich in der mehrdeutigen Rolle der TeilnehmerInnen fort. Einerseits bringen sie durch ihr – in Bezug auf theoretische Vorhersagen – „nicht-rationales" Verhalten einen epistemischen Mehrgewinn ein, und andererseits gefährden sie damit die Brauchbarkeit experimenteller Resultate. Dieses Risiko wird in Experimenten durch verschiedene Strategien vermindert, die den TeilnehmerInnen helfen sollen, strategisch günstiges Verhalten zu lernen.

In Bezug auf die Bewertungen experimenteller Forschung ergibt meine Analyse, dass die von mir Interviewten eine Vielzahl an Motivationen und unterschiedlichen Vorstellungen vom Nutzen experimenteller Studien und der idealen methodologischen Vorgehensweise haben. Viele dieser Vorstellungen werden aber letztlich der dringenderen Frage untergeordnet, ob eine Studie sich zur Publikation in hochrangingen Fachzeitschriften eignet. Was ForscherInnen als die Kriterien für solche Publikationen ansehen, beeinflusst daher ihre praktischen Entscheidungen während des gesamten Forschungsprozesses. In meiner Analyse zeigt sich, dass diese Kriterien einer bestimmten Konzeption von experimenteller Praxis entsprechen, wonach Experimente vor allem als Werkzeuge für die Evaluation und Weiterentwicklung aktueller ökonomischer Theorien anzusehen sind. Andere mögliche Funktionen von Experimenten, etwa die Suche nach bisher nicht theoretisierten Verhaltensformen, treten damit in den Hintergrund.

Historisch ist die Konzeption von Experimenten als (bloßem) Werkzeug der Theorieentwicklung in dem Bemühen begründet, experimentelle Forschung für den akademischen Mainstream interessant zu machen (vgl. Svorencik, 2015). Durch die zentrale Bedeutung von Publikationen in Fachzeitschriften ist dieser Theorie-fokussierte Ansatz bis heute so dominant, dass alternative Herangehensweisen von einigen meiner InterviewpartnerInnen als wenig aussichtsreich beschrieben wurden. Meine Analyse epistemischer und evaluativer Praktiken zeigt in diesem Fall die Einbettung zeitgenössischer Forschungspraktiken in institutionelle Logiken der Bewertung wissenschaftlicher Arbeit, und das mitunter konfliktreiche Verhältnis solcher Logiken und individueller Vorstellungen von guter wissenschaftlicher Praxis.

11 Abstract English

This thesis engages with the research practices of experimental economics from the perspective of science and technology studies (STS). Economics has traditionally been considered as a non-experimental science. During the last three decades, experimental and behavioural economics have emerged as two sub disciplines in which laboratory experiments play a central role. These experiments typically take place in computer laboratories at university campuses, where student participants bargain or make decisions and are paid according to their performance in the experimental task. Although STS has long had a strong interest in the role of laboratory practices in the production of scientific knowledge, this new discipline has not received attention from STS researchers as of yet. There is, in general, very little engagement with the research practices of social scientists from a STS perspective.

My thesis is intended as a first step towards filling this gap in the literature. The introduction of the experimental method to a non-experimental science presents an interesting case study in the diversity of "epistemic cultures" (Knorr Cetina, 1999) and a research community's context-specific solutions to knowledge production. While STS research on economics has so far focused on the application and influence of economics in various areas of public and professional life, a focus on the research practices of academic economists brings to the fore the background of the knowledge claims on which this influence builds.

My inquiry was guided by an interest in how economists appropriate the method of laboratory experiments, and how they evaluate the use of this method when reflecting on their work. In order to answer these questions, I conducted six semi-structured interviews with researchers in different career stages working at the *Vienna Center for Experimental Economics* (hereafter VCEE). Additionally, I assisted in running experiments at the VCEE and collected observational field notes at the laboratory. All my empirical material was transcribed and analysed inductively according to the principles of *Grounded Theory* (Charmaz, 2006).

The main finding of my analysis concerns the use of laboratory experiments in economics as a means of producing behaviour in conditions that closely resemble those described by theoretical models. This resemblance was achieved by the economists I observed with a serious of reductions, where the "economic fundamentals" of a situation are identified, modelled, and then implemented as a laboratory design by removing all context-specific elements and connotations. Experiments can be used to test the predictions of models, or to suggest new conceptual additions to existing theory, on the basis that all the factors influencing behaviour in the laboratory are in principle known and are theoretically tractable. In addition, my analysis points to a tension between approaches that focus on theory and the development of theoretical explanations, and approaches that focus more on observation and the production of behavioural phenomena that deviate from the predictions of standard theory. This tension is reflected in the ambiguous conception of participants and their capacity to frustrate experimenter's expectations as both an epistemic resource and an epistemic risk. In experiments, this

risk is accommodated by “inducing behaviour” and helping participants to learn which decisions reflect their strategic interests.

Analysing the principles of evaluation my respondents invoke when reflecting on their work, I found that experimenters articulate a range of motivations for doing experiments. Divergent conceptions of the purpose of experimentation correspond to different views on the methodology that can best serve these purposes. These individual conceptions of good methodology were, however, in practice often subordinated to the concern of getting one’s research published in high-ranking journals. This concern was so dominant amongst the economists I observed that it informed a wide range of practical decisions throughout their research processes. The evaluative criteria described as central for journal publications in my interviews correspond to an approach that has sought to align experimental practices with the state of the art in economic theorising. Historically, presenting laboratory experiments as tools for testing and developing economic theories was a strategy to make this new method acceptable for the academic mainstream (cf. Svorencik, 2015). The centrality of journal publications for the evaluation of contemporary academic research, in my view, is one reason why this theory-centred approach is still experienced as dominant by my respondents, and as precluding alternative conceptions of experimental methodology and applications. My focus on epistemic and evaluative practices in experimental economics therefore reveals how contemporary research practices are embedded in institutional logics of assessing and promoting research. Additionally this focus reveals where these logics come into conflict with researchers’ individual conceptions of good scientific practice.