In brains we trust: How neuroeconomists stylize trust, the brain, and the social world

Klaassen, P.

Citation for published version (APA):
Klaassen, P. (2014). In brains we trust: How neuroeconomists stylize trust, the brain, and the social world.
Chapter 1

Introduction

1.1 The new laboratory science of decision making and the discovery of trust

Vast changes are gathering from this new [neuro-] technology, propelling humanity toward a radical reshaping of our lives, families, societies, cultures, governments, economies, art, leisure, religion—absolutely everything that’s pivotal to humankind’s existence.

Lynch & Laursen (2009, p.7)

The scene: through the hospital doors, into the laboratory

In the center of Zürich, right across from the main building of the University of Zürich, one finds Switzerland’s oldest and largest hospital and its most important medical research center: the University Hospital of Zürich. With its staff of over 6000 people, working in more than 40 clinics, almost 200,000 patients a year receive care here. But not only patients and personnel populate this institution. For instance, some seven years ago, for days in a row early in the afternoon small groups of entirely healthy students, all males in their early twenties, entered the hospital’s fMRI facility. These students came to the hospital to be subjects in an experiment. They had agreed to participate in a project investigating the effects of the hormone and neuropeptide oxytocin on individual decision making. The results of this study were published in a much-cited scientific article by Thomas Baumgartner and colleagues (2008), and with hindsight we can therefore say that these students constituted
an essential linkage towards the establishment of novel facts regarding the mechanisms implicated in oxytocin’s effects on trust in decision making.

To participate in the study, the students had to be in good health and meet all sorts of requirements. For instance, participation in the experiment was disallowed to anyone suffering from claustrophobia or any medical condition that could impede one’s ability to lie still for some time. Also, subjects could not have certain medical devices, such as unremovable braces, a cardiac pacemaker or hearing implants.

Once approved for the study, each student, now a study subject, used a nose spray to administer either three puffs of placebo or three puffs of oxytocin in each nostril—neither the subjects nor the researchers on the floor knew which treatment was received. Approximately fifty minutes after receiving the placebo or oxytocin, a subject would insert ear plugs and lie supine on a table, which was then moved into a functional magnetic resonance imaging scanner—a huge cylindrical machine better known by its acronym fMRI. Inside the fMRI scanner, each subject played strategic games against a human or computer opponent.

At the beginning of each round, subjects were given twelve monetary units. They were told that they would be paid one Swiss franc (≈ € 0.80) for every five monetary units that they had at the end of the experiment; this would be in addition to the eighty CHFs (≈ € 65) they would receive for participating in the study. The moves in the game consisted of decisions to either transfer or not transfer monetary units to their opponents, who in turn could make similar choices. A crucial difference between the transfers of the subjects and those of their opponents was that the amount of the subject’s transfer was tripled by the experimenter before it reached the opponent, whereas the amount transferred by the opponent was not modified.

Lying as still as they could, the subjects viewed the game via a mirror that reflected the computer screen on which the game was played and made their choices for each turn with a four-button input device. All 49 subjects participated in 24 decision trials, which were divided into several categories depending on which game was played or which phase of the experiment the game was played in. Halfway through the experiment, subjects received feedback on how much they had earned or lost, depending on the moves they had made.

The meaning: trust unraveled

These healthy young men, 49 in all, were subjects in an experiment in what has become known as neuroeconomics. In this particular study the researchers aimed at assessing the effects of oxytocin on decision making and at exploring the neural circuitry involved. More specifically, the researchers were interested in the role of oxytocin in our decisions to trust.
The games used in the study were borrowed from behavioral game theory and were devised specifically for investigating trust.

At the time of this experiment, scientists already knew that the intranasal administration of oxytocin increased human feelings of trust. They also knew that several prefrontal cortical areas of the brain were implicated in decisions to trust. What was unclear, however, was just how oxytocin did what it did. Which specific brain areas were affected by this neuropeptide, and how did these different areas contribute to trust? And what about responses to breaches of trust? If trust is not returned, how does oxytocin influence one’s subsequent decision making? To help answer these questions, this experiment was designed to not only illuminate the role of oxytocin in human decisions involving trust of another but also to measure the neural activation patterns that occurred as those decisions were made.

The 49 students who participated in this experiment helped unravel the neural mechanisms involved in trust and in oxytocin’s implication therein. They helped establish that differences in trust levels due to oxytocin are associated with different activation patterns in several regions of the brain, most prominently including the amygdala, the midbrain regions and the dorsal striatum. These areas of the brain are associated with fear processing (amygdala and midbrain regions) and with learning (dorsal striatum). What we may surmise from this experiment, then, is that an increase in trust appears to be connected to a decrease in fear and a decreased capacity to adapt one’s behavior in response to feedback, all of which are influenced by oxytocin; in addition, trust is mediated by the amygdala, the midbrain regions and the dorsal striatum.

The technologies: oxytocin, fMRI, and the Trust Game

This experiment brought together two lines of research that each had an independent existence for some time: neuroimaging research into social decision making, and neuroendocrinological research concerning (human) sociality. In both lines of research, a prominent place had been reserved for research on trust, conventionally investigated using games similar to the one used here. Thus, besides the scientists who designed the study and the 49 subjects who participated, this event had three protagonists: oxytocin, fMRI, and the Trust Game.

The neuropeptide oxytocin is a phylogenetically old molecule that serves various functions in mammals. It is best known for its role in the release of milk in nursing mothers or lactating animals, but in the past few decades it has also been implicated in an ever growing set of social behaviors, including trust. However, though humans produce oxytocin endogenously, in order to test for its effect on the nervous system in a systematic and controllable fashion, exogenous oxytocin must be administered via nose sprays, as was done in
Baumgartner’s study. Indeed, Baumgartner’s design was similar to that used by Kosfeld and colleagues when they established oxytocin’s capacity to increase levels of trust in humans (Kosfeld et al. 2005).

Various technologies for investigating neural activation exist, all targeting different aspects of it. When it comes to neuroscientific research aimed at localizing psychological functions, no technology has been as successful as fMRI. Such fMRI scanners, as were used in Baumgartner et al.’s study, are basically huge magnets combined with fine-tunable radio transmitters and receivers. Together with the computers required to process the collected data, fMRI technology allows for a basically noninvasive way of localizing the neural activation patterns correlated with the cognitive process, (social) decision, or emotion that is taking place at any moment in time, usually while a test subject completes a particular task. For fMRI investigations no radioactive tracer substances have to be administered, as for instance is required in the case of positron emission tomography or PET scanning, and no electrodes have to be placed on the skull, as is required with electroencephalography or EEG. The most significant inconvenience of fMRI is the awful noise the scanners make: Sometimes this approaches 130 decibels, the noise level of a rock concert or a pneumatic drill. Because of this, subjects have to wear earplugs.

Technologically less sophisticated but no less crucial than oxytocin and fMRI is the so-called Trust Game as was used in this experiment. Trust games provide for a relatively straightforward way to understand the complex characteristics of trust in social interactions. Although many variations of this game exist, they all share one aspect—they are played by two players who are allowed to respond, only once, to the other player’s action. Both players begin the game with a specific number of monetary units. As described in the introduction, Players One and Two choose to transfer none, some, or all of their units to the other player; however, Player One always goes first and both players are aware that Player One’s transfer amount is always tripled by the researcher (figure 1.1). Thus, in theory both players can profit from the tripling of the monetary units sent by Player One. But this only transpires if Player One trusts Player Two to send back some of the monetary units, and if Player Two proves to be trustworthy and in fact returns an amount that is larger than the base amount that was tripled by the researcher. In this game, therefore, any move by Player One that constitutes a transferal of monetary units, is interpreted as an instance of trust.

The word “game” in Trust Game owes its place to the set of artfully constructed strategic situations that come from Game Theory, a formal and mathematical branch of the rational choice perspective in the study of human behavior. Since the 1970s more and more research has been devoted to behavioral experiments using Game Theory. It is assumed that observations of people’s behavior in such strategic games can teach us about real-life social
interactions, because many actual social situations supposedly have structural similarities to those modeled in these games. The idea is that by using a variety of games, different characteristics of social and strategic action can be investigated and weighed against the yardstick of rationality that Game Theory brings into view and helps articulate. In the last fifteen years, so-called Behavioral Game Theory has also been enrolled in neurobiological forms of experimentation. Research on the role of oxytocin in trust exemplifies this (e.g. Kosfeld et al. 2005), as does research on the neural pathways implicated in trust (e.g. McCabe et al. 2001). The experiment by Baumgartner et al. brings together both lines of inquiry.

1.2 Neuroeconomics, trust, and reality

In miniature form, almost all of the themes this dissertation touches on can be discerned in the section introducing the neuroeconomics of trust. To make things more explicit, however, I will disassemble the main concerns. These are at least the three following themes:

1. a new laboratory science of social decision making going by the name of “neuroeconomics,” i.e. a new science, which
2. takes on phenomena such as trust, and which,
3. in the process, might affect our views on the nature or reality of such social phenomena and humans qua social beings.
Before expounding on how these themes will be scrutinized philosophically in this dissertation, I will briefly introduce this triplet of science, trust, and (social) reality.

Experiments in **neuroeconomics**?

Although economists, neuroscientists, methodologists, and philosophers of science have been discussing neuroeconomics for some years now, and although it has even made some appearances in mainstream media, it deserves some more introductory and contextualizing words. As with most things *neuro-*-, neuroeconomics can be many things. As “neuroeconomics” is used here, the term denotes a particular way in which neuroscience and economics have been brought together during the last ten to fifteen years. Sometimes this is explained by saying that neuroscience gives economics the proper foundation it has been lacking for so long, or, more specifically, that neuroscience helps develop economic theories of decision making by adding (biological) constraints. This interpretation is sometimes called “behavioral economics in the scanner,” but, as we will see in chapters 4, 5, and 6, this description is somewhat too restricted.

Since the beginning of the new millennium, a lot of work has been done under this label of neuroeconomics. As the epithet behavioral economics in the scanner implies, neuroeconomics can be interpreted as the unsurprising next step in economics’ steady movement toward a conception of its most prominent subjects, viz. humans, as concrete, material, biological entities. Entities, moreover, that are empirically investigable. This movement, moreover, coincides with a movement away from the highly abstract and idealized former subject of economics, the **Homo economicus**. Whether this image indeed fits the neuroeconomic project, is one of the questions I will address in this dissertation.

There are, however, two other conceptions of neuroeconomics in circulation. One of the alternative meanings given to “neuroeconomics” refers to the scientific field in which a branch of economics is used to model and understand brain activity. The basic idea here is that economics is a substrate-independent tool, the Swiss Army Knife of intellectual labor, so to say. As such, economics can and should be used to model and understand not only objects such as markets, politics, ethics, partner choice, or education, but also brain activity.2 In this view, the brain is conceived of as a complex system in which different areas (or even different synapses) are in competition with one another in their (independent) struggle for resources such as oxygen and sources of energy.3

Furthermore, the term neuroeconomics is sometimes used to denote a market where neuro-products are being bought and sold. The designated market is huge, and the types of products on display are wide-ranging. Some products are complex, expensive, and labor and expertise intensive—think for instance of deep-brain stimulating “pacemakers” for the therapeutic use in patients with Parkinson’s disease4 and fMRI scanners for imaging brain
activity. Other products traded here are relatively easy to use and fairly cheap, such as bestsellers Seroxat and Ritalin for the treatment of depression and Attention deficit hyperactivity disorder (ADHD), respectively. Then there is also the variety of often simple and, equally often, scientifically suspicious products, ranging from fish-oil to “dream-machines” or “mind-machines” that help you relax (or learn better, or become more aware) by stimulating the optical nerve, to the variety of neurofeedback programs that can be used to train your brain.

For social scholars of science and society, this market has a plethora of riches on offer, as it constitutes a place where science and the public meet and interact. More specifically, neuroeconomics in this third sense embodies a place where the “neurobiologization” of subjectivity and society is enacted, questioned, and advertised. With the dimensions of this market, however, also comes a certain ethereality; it is difficult comprehend something as big and irregularly shaped as neuroeconomics understood in this way. As soon as one tries to focus on this neuroeconomy, one will find that the variety of products on sale and the variety of buyers, sellers, intermediaries, marketers, marketing channels, (corporate) stakeholders, and interested parties is simply stunning. Thus, in my this dissertation this meaning of the term is of subordinate importance, even though the topic of neurobiologization will surface occasionally and is present throughout in the background as a motivational theme.

Why neuroeconomics?

Besides the specific meaning of neuroeconomics at play in this dissertation, it is crucial to emphasize that neuroeconomics in itself constitutes a specific case of a much broader development, viz. the rise to prominence of all things “neuro.” The neuroeconomic market described above illustrates this, but so does, for instance, the fact that while in 1958 only 650 articles were published in the brain sciences, 2008 alone saw the publication of over 26,500 such articles (Rose & Abi-Rached 2013, p.5). Furthermore, consider the long list of novel academic fields employing the neuro prefix: neuroaesthetics, neuroethics, neurolaw, neurosociology, neurotheology, neuromarketing and so on. Today, the involvement of the brain in aesthetic judgment, in moral and lawful behavior, in social interactions, and in beliefs in or experiences of God is at the center of attention for many scholars throughout the world. According to many today, a significant part of the answer to the question of what kind of beings humans are, involves an explication of what the brain is and how it works; indeed, the hope invested in the brain is well expressed in the epigraph to this chapter.

There can of course be no doubt that brain-based accounts of who, how, and what we are exist next to many other such accounts. That some identify themselves and their peers in
(large) part in terms of their brains and their neurochemistries does not mean that psychological, developmental, socio-economic, gender, racial or other identities have been done away with; to claim otherwise would be utter nonsense. But this does not negate the fact that, from the 1960s onwards, the neurosciences have supplied us with a new way of thinking about ourselves and, potentially, for acting upon ourselves. And this new way of thinking deserves philosophical scrutiny.

To start with, then, we have to recognize that there are many dimensions to today’s upsurge in interest in the brain. The economic dimension has already been touched upon: there is a viable market for brain research and products aimed at the improvement of the human condition via actions on the brain. This dimension cannot easily be severed from what are probably best considered as cultural factors, however. That is to say, we are concerned with a phenomenon that has emerged most clearly in what was formerly known as “the West”—namely, that part of the world made up by advanced liberal democracies. It appears that neuroscience’s rise relates to and interacts with customary practices characteristic of such societies, such as the use of neuroscience in criminology and in developing strategies for crime control. In this regard Rose and Abi-Rached have convincingly made the case that a new neurobiologically informed “diagram of control” is taking shape (Rose & Abi-Rached 2013, p.166), central to which are policies of screening and intervention. Such policies, according to Rose and Abi-Rached,

are likely to contribute to a further widening of the net of the apparatus of control to the ‘precriminal’ or ‘predelinquent,’ and to play a part in the new ways in which subjects and subjectivities are governed in the name of freedom in an age of insecurity. (p.167)

Based on the idea that structural neuroimaging can tell us something about the susceptibility to violent, impulsive, dangerous, sexually predatory and similarly unwanted behaviors, neurobiology influenced the general publics’ beliefs on risk and its management.

Another obvious place to look for illustrations of the ways in which the rise to prominence of the neurosciences interacts with more general cultural factors is in psychiatry, which nowadays conceptualizes disorders almost always in terms of their underlying neurophysiology. Fidgety boys who cannot concentrate do not (merely) suffer from ADHD, they suffer from neurodevelopmental disorder ADHD. Non-communicative boys who show little empathy do not simply suffer from autism, but from neurodevelopmental disorder autism. In both these cases, which are only two examples of a wide-ranging phenomenon indeed, historically variable social norms and neurobiological research meet in highly complex ways and give rise to very particular practices of classification and management of types of behavior. There is a tendency at work here to conceptualize disordered behavior through an appeal to biology, and not by way of a symptom-based taxonomy, as the Diagnostic and
Statistical Manual of Mental Disorders (DSM) V—the presently used manual of psychiatric diagnoses—purports to do. This tendency is driven both by a scientific and technological push of ever more data and proven means of intervention, as well as a popular or cultural pull of (overblown) expectations and desires for clear-cut explanations in terms of something as purportedly real, stable, and universal as our biological constitution. Explanations in such terms, moreover, often function in order to release patients from personal responsibility for their abnormal behaviors and are therefore more than welcome: The oft-held belief seems to be that when one's disorderly behavior has its roots in a neurodevelopmental deficit of some sort, the behaviors accounted for by that deficit are beyond one's liability since they are a matter of fate, not of will.

Outside the criminological and psychiatric contexts, the economic and cultural aspects of the rise of neuroscience show in the huge nonacademic interest in neuroscience and its potential and professed implications. This is manifest in the great number of popular science books, television broadcasts, blogs, and the like dedicated to the topic and enthusiastically consumed by the public, as well as in the attention neuroscience receives in policy circles. This dimension at least partly depends on the academic one—the one that produces knowledge of the brain. The technological advances, particularly in the domains of neuroimaging, neurophysiology, and neuropharmacology, have made it possible to find out all sorts of things about the brain's involvement in human actions, (deviant) behaviors, beliefs, powers, experiences, desires, and the like, and that the possibilities of neuroscience have been realized with great enthusiasm. Thus, it appears that the combination of neuroscientific knowledge and cultural beliefs about the role of the brain in who we are and what we are disposed to do bears an obvious relation to all sorts of local societal demands.

All this, however, merely boils down to a restatement of the initial observation that there is an upsurge in interest in the brain. Pointing out ways in which this upsurge relates to all sorts of other characteristics of “late modern societies” does not by itself bring us closer to an explanation. One explanatory route simply relates the academic, economic, and cultural dimensions to each other: The fact that a lot of money can be made with anything from neuro-books to neuro-pharmaceuticals and other sorts of neuro-devices, follows from the knowledge divulged by the neuro-disciplines. Due to the latter we get to know more and more about brain-based disorders, capacities, dispositions and the like, about brain-affecting cures for these disorders, about tools that positively affect these capacities, and about means of governing these dispositions.

This explanation, however, bears all the signs of complacency. It is based on the premise that neuroscience flourishes because it gets things right, whereas it might well be the case that a lot of money can be made with products that do not spring from knowledge practices that get things right. The simple and attractive story above is that, because science gets
things right people are interested in it, disorders can be cured, unwanted behaviors and experiences mended—this is a story of how success reaps success. Unfortunately, it is much too simple. This account does not help us understand, for instance, why it is just now that so much hope, money, time, and work are invested in brain research and all that comes with it; nor does it help us understand anything about the nature of such investments or the rewards. It does not add, in other words, to an understanding of how neuroscience rose to such cultural, academic, and economic prominence as it has reached today but instead contents itself with the fact of the achievement.

This *veni, vidi, vici* explanation of the success of neuroscience arguably forestalls any critical assessment, whereas such assessment is surely needed. To give an example, it can be argued that (philosophical) beliefs about the essential involvement of the brain in human (mental) capacities and the like are a much more important driver behind the success of neuroscience than are neuroscience’s achievements for bringing about such philosophical beliefs. If one holds on to the above victorious story of neuroscience, one will not easily see that the accomplishments of neuroscience have played a smaller role in bringing forth our current view of the role of the brain in what makes us human, than the idea *that the brain has a role in this* had in making neuroscience a success.¹¹

In an attempt to provide a more satisfactory account of the rise of neuroscience, this dissertation engages with the question of how neuroscience has accomplished the status it has today. It does so not by taking up neuroscience as a whole,¹² nor by studying sociologically the societal phenomenon of fluctuations in scientific hypes and hopes.¹³ The first would be too grand a project, the second would provide too narrow a view on the specificities of the science at hand, its contents and, particularly, its style—but more about this later. Rather, then, I will scrutinize neuroscience’s rise to prominence through a philosophical analysis of the exemplary and circumscribed case of the neuroeconomics of trust. And although full generalizability might be too much to ask for, the contention that neuroeconomics constitutes a sufficiently representative instance of the rise of neuroscience to get a handle on this event is easily made. First, neuroeconomics is exemplary of the way in which neuroscience moves into other fields of study that were previously or are currently elsewhere conceived of as the territory of wholly different tribes—in this case, tribes of social scientists or moral philosophers. Second, as far as its neuroscientific character is concerned, there is not much that sets neuroeconomics apart from other areas of study within the neurosciences that are crucially involved in the rise of the neurosciences.

**Trust**

This leaves me with the obligation to justify my choice to focus on trust as an object of research. In this regard I can be fairly brief. To begin, it comes from my determina-
tion to engage with a concrete, real-life object and the philosophical problems and puzzles that poses—as opposed to an abstract argumentative enterprise only comprehensible in the context of a contemporary philosophical debate. While trust, in itself, may not be all that concrete, it plays a central role as an object of neuroeconomic research. As such, it is in fact a very concrete object of research. All in all, this dissertation is neither an abstract exercise in the philosophy of neuroscience nor a systematic philosophical analysis of the nature of trust. Instead, it is a philosophical analysis taking off from neuroeconomic research into trust—research which, as I will show, "restyles" trust in the process.

By focussing on trust I follow in the footsteps of the neuroeconomists who have turned trust into one of their pet-topics. Their interest raises two questions: Why is trust such a favorable object of research in neuroeconomics and, more interestingly, how did neuroeconomists turn trust into such preferred object?

Neuroeconomists would answer the why-question by saying that trust is a vital feature of all economic transactions and perhaps even of social life in general—it is the proverbial glue which holds society together, the lubricant which keeps our economic system and, more generally, our societies going. If it were not for trust—in spouses, in banks, in sellers, in buyers—society would not be able to function as it does.

The how question addresses the intricacies of the research technologies, skills, assumptions, theories, models and so on, which, together form the material basis of neuroeconomic studies of trust. This concerns, in other words, the wide-ranging set of linkages that keep the project together. A partial illustration was already given in the opening section, but much of this dissertation consists of analyses that shed light on these linkages and bring out their epistemological and ontological significance. It is through such analyses that I will clarify how neuroeconomists stylize trust, the brain, and the social world.

Furthermore, the neuroeconomics of trust is of great interest here because where neuroeconomics investigates trust it arguably trespasses on the territory of the more traditional social sciences. Hence, the neuroeconomics of trust helps understand what, for instance, social interaction is supposed to be, how this ought to be studied and what implications the novel sciences of the brain might have for (more traditional) sciences of sociality—issues that are closely related to the conception of man that is enacted in neuroeconomics. In order to concretize the issue of the potential interference between conceptions of sociality from neuroeconomics and nonscientific ones, consider the contrast between the way in which trust is investigated in neuroeconomics and a more intuitive, more everyday notion of trust.
To this end, I will take my cue from the articulation of trust in a more traditional social science which, compared to neuroeconomics, is more dedicated to investigating the order of everyday reality in everyday reality: ethnomethodology. Ethnomethodology is a branch of social science founded by Harold Garfinkel and further developed by him and his followers from the 1960s onwards. Although not a mainstream perspective in social science, the number of self-professed ethnomethodologists has grown rather substantially since the publication of Garfinkel’s pioneering *Studies in Ethnomethodology* (1967) and it has proven to be a stable presence in the field ever since.

Distinctive of ethnomethodology is its claim to investigate the methods that the subjects investigated by ethnomethodologists—the *ethnos*—use to make sense of their actions and to give meaning to their life and to the world around them. Thus, ethnomethodology is not concerned with (nomological) explanations of social phenomena but instead seeks an understanding of meaning-making practices from the perspective of the practices investigated. The “method” in the term “ethnomethodology,” then, does not refer to a specific method which ethnomethodologists claim to have developed, but rather to the fact that this approach investigates the methodic ways in which members of communities give meaning to and provide structure for their practices, and organize and account for their practical affairs (cf. Garfinkel 1967, p.vii). Ethnomethodologists even explicitly claim not to employ any circumscribed method in their work and claim to simply “follow around” those whom they investigate. If anything, then, that is ethnomethodologists’ method: simply following people around.

Simply following people, however, is anything but simple. It requires that one refrains from putting to work any conceptual or theoretical schemes familiar from social theory—or from any other scientific discipline for that matter. It entails that ethnomethodologists ask their subjects very “straightforward” questions, where straightforward is between scare quotes because no one would ordinarily consider such questions as ethnomethodologists ask straightforward—they are too basic for that. They are what we might call ontological questions, as they simply inquire into what is going on in people’s daily lives. Asking such questions helps make visible what is so obvious that it would otherwise remain invisible. For this reason, asking such questions has been compared to performing Husserlian bracketing exercises (Watson 2009, p.479), in which also, but in the context of phenomenological investigations, all reliance on assumptions we commonly make about the world surrounding us is deferred. Instead of subsequently turning to social theory in order to develop explanations for the sense-making practices or ordering schemes laid bare, ethnomethodologists are content when they succeed in simply describing them—any deviation from the footsteps of the ethnos is to be avoided.
Another specific form ethnomethodological research has taken, and for which ethnomethodology has become rather (in)famous, is the so-called breaching experiments. These are experiments done in the ordinary flow of people’s lives, forcing upon them as it were a break with that flow. This can, for instance, be done by pretending to be a boarder in one’s own home: conducting oneself politely and circumspectly to one’s parents, siblings and children, and speaking only when spoken to (Garfinkel 1967, p.47). Or it can be done simply by denying the commonsensical nature of a conversation partner’s statements, as in the example below:

The subject [S] was telling the experimenter [E], a member of the subject’s car pool, about having had a flat tire while going to work the previous day.
(S) “I had a flat tire.”
(E) “What do you mean, you had a flat tire?”
She appeared momentarily stunned. Then she answered in a hostile way:
“What do you mean? What do you mean? A flat tire is a flat tire. That is what I meant. Nothing special. What a crazy question!” (Garfinkel 1963, p.221)

Garfinkel used such breaching experiments “to raise into visibility matters that are typically taken for granted” (Watson 2009, p.479), to investigate what it is that disappears when order breaks down—perhaps analogous to the way in which physiologists make inferences concerning normal bodily functioning from pathological cases. The ultimate aim of this technique is to reveal what subjects do in order to make sense of their surroundings in ways that are socially acceptable to the group they belong to. The rules of social order are constantly in the process of being made, maintained and repaired. Depicting this process is entirely distinct from investigating how “objective” norms act on subjects in order to inform their behavior or inquiring into the origins of such norms or into the functional role they fulfill in the grander scheme of social life.

Once something has been “raised into visibility” through the specific research technologies employed by ethnomethodology—asking ontological questions and performing breaching experiments—it can be turned into an object of research in itself. Such objects as are made visible in these ways, however, are of a well-delineated sort. They are what have been called “constitutive orders,” that is, all the presuppositions that together make up the natural attitude shared by members of some ethnos, and the mutual commitment to the relevant “rules of engagement” in the practice thereby constituted (Watson 2009, p.476).

Constitutive orders, furthermore, are to be understood in contradistinction with regulative or summary rule orders (cf. Rawls 1955, Searle 1995, Watson 2009) and are closely associated to the basic rules that are definitive of social practices. In regard to the breaching experiments discussed in Garfinkel’s article (1963), it turns out that Garfinkel’s answer
to the question of what it is that has disappeared on such occasions is trust. In Garfinkel’s view, trust is lacking in situations in which people can no longer make sense of what practice they are involved in, or when they have lost the feeling of being mutually committed to a common practice with their conversational partner.

“Trust,” then, is here understood as “compliance to the constitutive order” and, as such, as “a condition for grasping the events of daily life” (Garfinkel 1963, p.190). This conception of trust has been endorsed by other prominent ethnomethodologists, including Ann Rawls (2008) and Rod Watson (2009). Trust as conceived of by ethnomethodology, then, is not something akin to a psychological feature or a particular form of behavior. It is not a characteristic that may or may not be a part of some meaningful social exchange or other—put differently, it does not belong in the realm of preferred play. On the contrary, to ethnomethodologists trust “a condition of stable concerted actions” (Garfinkel 1963, p.196), a feature of the constitutive expectancies which order social life.

Watson captures some of the ideas at work here as follows:

The idea of constitutive features – such as trust – as actually prior to and constitutive of action and objects (within a practice) rather than as emerging from – as outcomes of – action stands as the essential difference between what Rawls (1955) called summary rule orders and constitutive orders. Garfinkel takes up trust as a necessary background condition in a constitutive order […] and this distinguishes his approach from most of sociology, which takes a formal analytic […] approach. (Watson 2009, p.479)

Ethnomethodologists thus give trust an essential and constitutive role in the routine, ordinary maintenance of social practices—practices that would not even be there if it were not for trust, because failing to trust terminates one’s participation in the practice at issue altogether (Garfinkel 1963, p.187). Trust, stylized in ethnomethodology, comes out as akin to a transcendental condition for the possibility of sociality, which explains why, according to Garfinkel, “there is no reason to look under the skull since nothing of interest is to be found there but brains” (Garfinkel 1967, p.190). We should not expect to find trust beneath the skull in this transcendental understanding, since only if trust were akin to a thing such as a mental state, something of which it makes sense to say that it has what Wittgenstein has called “genuine duration,” could we possibly find out that it has a physiological substrate that could be localizable anywhere at all. Moreover:

Trust in [an ethnomethodological] sense does not mean to trust the whole person in all of their aspects, but, rather, to trust only that they are committed to [the practice at issue], competent to perform it, and that they trust this of you. Not to trust in this way is to fail to participate […] altogether. (Watson 2009, p.478)
Rather than identifying trust with a particular shape that individual actions can take (as we saw neuroeconomists did in section 1.3), ethnomethodologists conceive of trust as “a necessary background condition of any mutually intelligible interaction” (Watson 2009, p.476), as a condition of “constitutive practices” such that “all parties to the interaction must understand that they are engaged in the same practice, must be competent to perform the practice, must actually perform competently and assume this also of the others” (Watson 2009, p.475). If these conditions are not met, there is simply no interaction taking place that we would consider to be part of an orderly social practice. And whenever an orderly practice can be observed, the ethnomethodologists’ version of trust must be present and doing its job.

One pertinent issue in the present investigation concerns the relationship between trust as it is stylized by neuroeconomists and the everyday concept of trust, as articulated by ethnomethodologists. Do they coincide, overlap or relate in another way? Is it the phenomenon ethnomethodologists individuate that neuroeconomists subsequently succeed in unraveling? What implications, then, could the neuroeconomic investigation of trust have?

Reality

In my earlier statement that neuroeconomists stylize trust, the brain and the social, several core assumptions of this dissertation can be discerned. First of all, by mentioning neuroeconomists rather than neuroeconomics it is indicated that science is an inherently human and social affair. There are two dimensions to this. First, science is a social affair in the sense that its results are achieved through collaborative practices and communicative interactions. Second, although neuroeconomists might do most of their work within the confines of their laboratories, offices and conference venues—secluded spaces relatively detached from the rest of society—their work can only artificially be disconnected from the societal context in which it takes place. This dissertation works towards the substantiation of the claim that each aspect of what goes on in these different spaces affects all other aspects, and that therefore such practices and means of communication that are at the heart of science cannot be excluded from epistemological and ontological analyses.

The second core assumption is that when neuroeconomists stylize their objects of research, they bring novelty to the world. When scientists are finished with the objects they investigate, those objects are different from how they were before. Science is an inquisitive enterprise and, of course, one might be tempted to neatly separate problems of epistemology from problems of ontology, arguing that all that scientists change in the process of research is our knowledge of the world. In this dissertation, however, I defend the (pragmatist) view that epistemological and ontological matters cannot be so easily severed. To do so, I will take my cue from the philosophy of Polish philosopher-scientist Ludwik Fleck
(1896–1961), which I will elaborate in chapter 2. Central to Fleck’s philosophical views is the notion of style, which helps us understand the simultaneously social and cognitive organization of science.

This, then, is the philosophical backbone of this dissertation, namely a particular conception of the interlinked relationship between science and reality which goes back to the groundbreaking work of Fleck. Pioneering the practice-oriented study of science, in his Entstehung und Entwicklung einer wissenschaftlichen Tatsache Fleck illustrated the fruitfulness of amalgamating sociological, historical and philosophical analyses of science. Although the book was originally published in 1935 and first translated into English as Genesis and Development of a Scientific Fact in 1979, his work remains relatively little known to this day, at least among philosophers. This is despite the fact that he initiated a practice that has proven to be productive from the 1970s onwards, as evidenced by such works as Bloor (1976), Callon (1986), Collins (1985), Kusch (1999), Latour (1987), MacKenzie (2006), Pickering (1993), Rose (1990) and Shapin and Schaffer (1985). Judged by this heritage, Ludwik Fleck ought to be ranked among the founding fathers of a particular view of science and of how to study it philosophically and scientifically.

A distinctive feature of Fleck’s approach is that it does not equate science with the facts it (allegedly) reveals and the theories it construes and refines. Instead, in the Fleckian view science refers to all those facts and theories—be it that they are conceived of differently—plus all the implicit assumptions, research-enabling technologies and skilled experts, as well as the complex ways in which all of these interact. In his view of science there is no place for straightforward dichotomies such as those between knowledge and belief, who knows and what is known, scientific knowledge and social context or theory and experiment. In this dissertation I will employ the Fleckian view, but only after I have explained it more thoroughly in chapter 2.

1.3 How neuroeconomics can be studied philosophically

The primary concern of this dissertation is what happens to trust when it is investigated by neuroeconomists. However, this topic does not stand alone; it is connected to a wide range of related philosophical problems and issues, spanning from questions regarding aspects of the production of neuroeconomic facts to the portrayal of man as (social) being and the conception of society enacted. As I mentioned previously, to examine such issues I will employ the philosophy of science developed by Ludwik Fleck.

To further clarify the nature of this endeavor, however, I appeal to a philosopher of science who would, on almost all counts, be quite contrary to the project at hand: Karl Popper. Insofar as its rationalism is considered, Popper’s view of science differs very much from
the one I adhere to and defend here. This becomes evident when one looks at Popper’s appraisal of experimentation relative to theory. For Popper theory is primary and experiments only serve to test theories (cf. e.g. Popper 2002, p.90). In my analysis, however, the relationship between theories and experiments is not assumed to have a fixed form, and a hierarchy is not imposed on the pair at the outset.

Also when it comes to what philosophers are and what they are or are not supposed to engage in explanations of the development of scientific knowledge, the Fleckian account differs from the Popperian one. On Popper’s view, the bulk of what matters is understood when one succeeds in applying the logical argument known as modus tollens to the body of scientific hypotheses under scrutiny. This procedure allegedly allows for an evaluation of whether or not the theory at issue is falsified—or can be falsified at all. Only what is internal to science matters, in Popper’s view, with the most importance given to logical relations between theories and hypotheses and the predictions that can be deduced from them. The Fleck-inspired approach to such matters, on the other hand, does not principally rule out any type of influence. By minimizing the importance of the distinction between so-called internal and external influences on the development of scientific knowledge, other types of developments—social, conceptual, philosophical, technological, economic, cultural—can, in principal, be presented in such accounts. Indeed, a premise of my dissertation is that what is considered to be internal or external to science is more the outcome of scientific developments than it is a source of them.\(^7\)

So what is it, then, in Popper’s view of science that can help clarify the nature of the current undertaking? It is something intimately related to Popper’s persistent fallibilism, namely, his belief that we can never be sure whether or not our scientific explanations get things right. In one of his more poetic formulations of this idea, he connects it to an antifoundationalist theme:

> Science does not rest upon solid bedrock. The bold structure of its theories rises, as it were, above a swamp. It is like a building erected on piles. The piles are driven down from above into the swamp, but not down to any natural or ‘given’ base; and if we stop driving the piles deeper, it is not because we have reached firm ground. We simply stop when we are satisfied that the piles are firm enough to carry the structure, at least for the time being. (Popper 2002, p.94)

In this formulation we immediately recognize Popper’s rationalism: science consists, first and foremost, of a structure of theories, which he then compares to a building constructed on top of piles driven into a swamp. Obviously, Popper’s imagery concerns the relationship between theories and empirical data, or the so-called empirical basis of science. In Popper’s view the empirical data do not constitute the kind of rock-solid foundation empiricists
of various stripes assume—objective science’s empirical basis “has [...] nothing ‘absolute’ about it” (Popper 2002, pp.93-94). Instead, Popper argues that, where theories come into contact with empirical data, all we have are the man-made decisions that this and that data justifies the acceptance of some basic statement. The piles in the above image, then, are the empirical conclusions drawn from the theories at issue, and what is at their base, what allows them to carry the weight of the structure above (the theories), are man-made decisions regarding their solidity.

Without adopting Popper’s view on the relationship between theories and empirical data, I wish to pursue the antifoundationalist idea Popper holds, that science does not rest on a definite and firm ground and that it takes (a large number of) “piles” to keep it in place. But unlike Popper I do not identify the better part of science as the bold structure rising above the swamp; instead, I see the piles, which prevent that structure from sinking into the swamp, as crucial part of science. In my rendering of the metaphor, those piles do not simply consist of the empirical conclusions drawn from the all-important theories. As I appropriate the metaphor, Popper’s piles are Fleck’s stylized active linkages at work in science. The building that rests on top of them is the stable set of passive linkages—reality as it is experienced as if it were independent from those linkages we actively engage in our attempts to uncover reality. To a large extent, then, this dissertation consists of laying bare and analyzing the piles on which the neuroeconomics of trust rests. For this, I have to step into the swamp of scientific practice.

Fleck’s philosophy is, of course, not the only viable framework for an endeavor such as this. A large variety of more standard philosophies could be found and employed in this analysis of neuroscience. Without wanting to spend too much time and space on what will not actually figure in this dissertation, I will briefly discuss some of the more established alternatives. To frame these views consider that, over the last century, philosophy of science has grown ever more fragmented. During the twentieth and twenty-first centuries, the trend toward specialization into specific disciplines of the sciences, which began at least as far back as the seventeenth century, continued as science saw enormous growth in the number of practicing students and scientists, the number of disciplines, and the amount of money involved. At the same time, philosophy of biology, philosophy of social science, philosophy of physics and the like became more relevant areas of study than general philosophy of science. With the advent of neuroscience in the 1960s and its subsequent growth in the 1970s, 1980s and onwards, philosophy of neuroscience also entered the family.

From neuroscience to philosophy

Probably the best-known philosophical project concerning neuroscience goes by the name of neurophilosophy. This is a naturalistic venture which puts neuroscience at the center of
many an investigative field that philosophers think of as theirs. The mind-body problem, free will, morality, religion, the nature of mental states, knowledge, belief and all such issues, according to the neurophilosopher, are to be dealt with, first and foremost, on the basis of neuroscientific knowledge (see e.g. Churchland 1986, Churchland 2007).

When discussing neurophilosophy, only one name can figure centrally, even if it denotes two distinct philosophers: Churchland. The Canadian philosophers Patricia and Paul Churchland have become famous for establishing or at least accelerating this field of enquiry. For while the neurophilosophical approach may have already been discernible in the mind-brain identity theories of Ullin Place (1956) and John Smart (1959), and in the eliminative materialism formulated by Paul Churchland (1981), the endeavor really only gained traction after the term for it was coined in 1986, with Patricia Churchland’s publication of a book by the very name of Neurophilosophy.

As a species of naturalism, neurophilosophy sees philosophy and the natural sciences as continuous projects. The two share objects of investigations, and when it comes to the methods considered appropriate, naturalists of this creed generally see no principled distinctions between what is philosophical and what is scientific. If anything, it appears to be the type of focus or the degree of zooming in on different kinds of problems that distinguishes neurophilosophy from neuroscience.

Recently, Patricia Churchland published a book with the telling title Braintrust (2011), which concerns the foundations of morality and puts trust center stage. In her insightful discussion Churchland, to a considerable degree, deals with the same scientific material that is also central to this dissertation. Thus, her book provides the perfect material for a comparison between neurophilosophy and the Fleckian approach to the philosophy of science espoused here.

Churchland has a profound and fairly classical philosophical objective in mind—to further our understanding of human morality. To this end, she examines “the foundations of mammalian sociability in general, and human sociability in particular” (Churchland 2011, p.10). Her hypothesis is that morality is a four-part scheme of (i) caring, (2) recognition of others’ psychological states, (3) social problem-solving and (4) learning social practices. Vital to her hypothesis are neuroeconomic studies since, in Churchland’s view, these reveal a plethora of facts concerning, mostly, (2), (3) and (4). This is especially the case with experiments investigating the neuropeptide oxytocin. As Churchland puts it:

The extension of caring to dependent infants, and then to mates, kin and affiliates, marks the crucial shift that makes us social. At the center of the intricate web of neural connections is oxytocin (OXT), a powerful peptide that in mammals has been recruited in organizing the brain to extend self-care to infants, and thence to a wider circle of caring relationships. (Churchland
Moreover, key to the extension of care from infants to all others is trust, as oxytocin works its spell through trust. Put differently, insofar as morality is understood as the thing that grounds social behavior (Churchland 2011, p.12), it is crucially dependent on the trust-inducing workings of oxytocin. Looking at the intertwined issues of morality and sociality in this fashion, Churchland, in her own view, has progressed significantly from what she considers conventional moral-philosophical perspectives, namely perspectives which bear no clear and explicit connection to understandings of evolution and the brain, and which for that reason are “in peril of floating on a sea of mere […] opinion” (Churchland 2011, p.2).

More so than in other neurophilosophical writings by the Churchlands, in this book, Churchland points out the limitations of the type of neurobiological and evolutionary explanations of the foundations of morality. Much cultural variation can be found in moral practices, if only because the issues requiring moral decisions differ from one culture to another. Regulation of stem cell research is typically something that has only become a concern in our current Western society, for instance. Even with regard to such “basic” moral issues as human sacrifice, war or murder, the historical record shows decidedly divergent practices. Churchland acknowledges that the cultural or societal variation of such matters cannot be explained solely in terms of neurobiology (Churchland 2011, p.106). In other words, understanding how people determine the best action to take with regard to many issues of moral concern, requires knowledge of more than just the neurobiology of moral decision making. Accordingly, “nurture” or cultural learning takes up a large part of her view of human morality (see especially Churchland 2011, chapter 6). However, hers remains a philosophical project, inspired and informed by current neuroscientific wisdom, as her principal questions lucidly express:

Where do values come from? How did brains come to care about others? (Churchland 2011, p.12)

Since the answers to these questions crucially involve trust and neuroeconomical experiments investigating trust, Churchland deals with much the same material as I do in this dissertation. But she does so with a completely different agenda. My goal is not to shine a philosophical light on the findings of neuroeconomists, where these findings, as such, are more or less accepted for what they are. Rather, I strive to understand how the practices of neuroeconomists and all the active linkages enrolled therein together lead to these findings, what the weight and function of each linkage is, and what the significance is of neuroeconomic findings on trust vis-à-vis the everyday conception of trust. Churchland’s neurophilosophical project regarding trust, hence, does not offer an analysis of the neuroeconomics of trust that competes with mine. In fact, it more closely resembles the type of
material I analyzed than it does my analysis of that material.

**From philosophy to neuroscience**

No less (in)famous but almost diametrically opposed to the neurophilosophical project epitomized by the Churchlands, is the work of another duo: Max Bennett and Peter Hacker. This partnership comprises a neuroscientist (Bennett) and a philosopher of Wittgensteinian creed (Hacker), who have co-authored or -edited three books (2003, 2007, 2008) and published several articles in the neuroscience journal *Progress in Neurobiology* (2001, 2002, 2005, 2006). Their prime achievement, though, is indubitably their *Philosophical Foundations of Neuroscience* (2003). In this book they first trace historically the conceptual roots of neuroscience. Then they thoroughly analyze neuroscientific investigations concerning such diverse human faculties as sensation, perception, knowledge, belief, thinking, emotion and volition. Similar analyses follow concerning the role of neuroscience in regard to consciousness and self-consciousness. And finally, they discuss the methods and goals pertaining to neuroscience, on the one hand, and philosophy (of neuroscience) on the other.

Bennett and Hacker argue at length that neuroscientists continually run up against the limits of language, trespassing its logical grammar to the effect of uttering meaningless statements. The prime source of neuroscientists’ leaps into the senseless is what Bennett and Hacker call the *mereological fallacy*. Qualities, functions, capacities and abilities are often ascribed to a part of a human being, while in fact it only makes sense to ascribe these to the human being (or other kind of animal) as a whole. Whereas Churchland writes phrases such as “our brains often make good decisions in figuring out some solution” (Churchland 2011, p.7), Bennett and Hacker are of the opinion that “there is no such thing as the brain’s thinking or reasoning, feeling pain or perceiving something, imagining or wanting things” (Bennett & Hacker 2003, p.379)—let alone brains figuring out solutions. According to Bennett and Hacker, it is the task of philosophy to point out such conceptual oversights and insensibilities to neuroscientists. Doing so will help neuroscientists to carry on their empirical work correctly—that is, without indulging in gibberish.

Thus, where Churchland turns to neuroscience in order to lay bare the foundations of human morality, Bennett and Hacker practice ordinary language philosophy and claim that in this way they can provide neuroscience with its philosophical foundations. Determining and explicating the logical grammar of our ordinary language is commonly the first step taken in this field of play. Second is ensuring that in some region of application, language is in fact used in accordance with this logical grammar by pointing out where it is not and showing how mistakes can be amended.

Critically scrutinizing neuroscience is at the core of their endeavor and as such this might
appear to be much closer to the approach I take than is neurophilosophy. However, there are vital differences between the ordinary language philosophy of Bennett and Hacker and the stylistic analysis I provide in this dissertation. As was the case with neurophilosophy, these differences concern the object investigated, the method used and the goals pursued.

The object of the ordinary language philosophy of neuroscience is formed by all the linguistic expressions found in the final products of neuroscience—in statements of fact and theory as they are formulated in academic articles. Bennett and Hacker are not concerned with how such expressions arrive in academic writings, but that is a topic I will explicitly address. And although language is the object of research, this does not mean that Bennett and Hacker pay attention to such devices as rhetoric or metaphors. Although they do discuss one overarching image of man—the dualistic picture promoted by Plato and Descartes—they do so only insofar as they consider this image to obstruct our understanding. Beyond their purview is the productivity of such an image or of metaphors (e.g. how such constructs or devices might help drive a research program). Moreover, unlike in the present dissertation, in Bennett and Hacker’s book (2003) no attention is given to the non-linguistic issues that are implicated in neuroscience—think of the social and material practices neuroscience embodies and the scientific methods and research-enabling technologies involved in it.

Bennett and Hacker analyze neuroscientific expressions by comparing them to the standards provided by their own account of the logical grammar of ordinary language. The leading question for them is whether or not what neuroscientists say makes sense. Often, however, this appears to lead to highly contentious judgments, because what makes sense to one person does not necessarily make sense to another; thus, just what that logical grammar of ordinary language is, may not be as straightforward as Bennett and Hacker assume (see also Dennett 2007). Also in this respect, then, my analysis diverges from that of Bennett and Hacker. Conceptual analysis has its role to play here too, but the question of whether or not the expressions uttered by neuroscientists—or, more precisely, neuroeconomists—make sense is not as crucial. The reasons for this is that, unlike Bennett and Hacker’s book, my project does not begin with the assumption of a settled logical grammar. In fact, one of the key assumptions underlying this dissertation is that the meanings of concepts and their relations to other concepts change almost continuously, as what appear to be the same concepts figure in different (scientific) practices that are informed by different interests, questions and technologies (cf. Fleck 1979, p.42). Indeed, the travels of the concept of trust through different social and intellectual circles and the impacts of these travels on the meaning of the concept are issues I will analyze.

Both their object of investigation and their method are directly related to the goal Bennett and Hacker work toward, namely clearing up the language in which neuroscience is
conducted. To this end they show that, in their view regrettably, much of neuroscience is informed by the aforementioned dualistic conception of man, which goes back to Plato, but which has found its most well-known and in the context of neuroscience, most influential and well-known proponent in René Descartes. In Bennett and Hacker’s view, the contemporary form which dualism takes is that of a radical and unbridgeable split between brain and body, rather than a split between two kinds of substances, one of which is characterized as thinking and the other as being physically extended. They attempt to show that this conceptual framework is incoherent and leads to meaningless gibberish, such as explaining “human perceptual and cognitive capacities and their exercise by reference to the brain’s exercise of its cognitive and perceptual capacities” (Bennett & Hacker 2003, p.72). However, this is completely meaningless, because “[o]nly of a human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees, is blind; hears, is deaf; is conscious or unconscious” (Wittgenstein (1953, §281); quoted in Bennett & Hacker 2003, p.71). In order to avoid such senselessness, they offer a conceptual scaffolding that harkens back to Aristotle, “supplemented by Wittgensteinian insights that complement Aristotle’s”, as this “is necessary to do justice to the structure of our conceptual scheme and to provide coherent descriptions of the great discoveries of […] cognitive neuroscience” (Bennett & Hacker 2007, p.131).

In this respect, also, the present undertaking contrasts with that of Bennett and Hacker. For even if the stylistic analysis that is to follow is critical at points—especially in regard to scrutinizing the strength of the relations between active and passive linkages (i.e., questioning whether facts indeed follow once a style-induced aspect of neuroeconomics is in place)—it is so without there being a clear and allegedly incontrovertible benchmark in place. To wit, if Bennett and Hacker were to analyze the neuroeconomics of trust, their analysis would have to start with a stringent idea of how trust relates conceptually to anything else. Here I take a more “agnostic” position. Instead of taking a stand on what trust is or what it does or does not relate to, I analyze how others define it and in the process, relate it to all kinds of things—emotions, physiological states, behavioral models or the like.

Ethical, legal and social issues in neuroscience

Another branch of philosophy of the neurosciences concerns itself with the variety of ethical, legal and social issues brought up by these sciences—ELSI research, as it is conventionally called. As an example, consider neuroethical work concerning sex in the brain. According to the prominent neuroscientist and expert on autism, Simon Baron-Cohen—not to be confused with the comedian Sacha Baron-Cohen—male brains are “hard-wired for understanding and building systems,” whereas the female brain is “predominantly hardwired.
for empathy” (Baron-Cohen 2004, p.1). This, so his neuroscientific explanation goes, is due to the differential average levels of fetal testosterone that exists in males and females from the eighth to the twenty-fourth week of pregnancy. Cordelia Fine and other neuroethicists have critically assessed this type of research with respect to a variety of aspects. Specifically, questions were raised concerning the reliability and origins of behavioural sex differences, the proposed links between foetal testosterone levels and later gender-typed behaviour, the existence of sexual dimorphisms in the brain and their relation to behaviour or cognitive style, and the simplistic conception of development to which [...] accounts [such as Baron-Cohen’s] implicitly subscribe. (Fine 2012, p.286)

Moreover, Fine suggests that Baron-Cohen’s conclusions are a perfect candidate for the “looping-effect,” as described by philosopher of science Ian Hacking (Hacking 1995, Fine 2012), and, for that reason, it deserves ethical attention. According to Hacking, the scientific understanding of particular groups of people may well impact on the character, behaviors and (self-)image of the members of such groups, which in turn can affect those understandings. In short, (at least) where humans are concerned, classifications affect what it is that they classify and the effects of the classifications can potentially feed back into newer classifications. A concern in this context is that, for scientific understandings to have such an effect, they do not have to be correct: false theories may have exactly the same impact as true ones (cf. Schwartz 1997). So, even if the criticism by Fine and others holds, the question is still open as to whether such criticized neuroscientific investigations of sex differences “merely” explain the status quo, or if they might also reproduce and strengthen it. Neuroscientists, Fine contends, have an ethical responsibility to consider such questions.

Of course, this is only one illustration of what ELSI research may look like. More than the other two examples of philosophy engaging with neuroscience, ELSI research is a very broad category, comprising work from various perspectives and using various methods—from conceptual analysis to experiments in ethics to empirical sociological or ethnographic work. The goals of all such work, however, are relatively confined. These are, as the label makes clear, to critically assess the ethical, legal and social implications of developments in neuroscience and its applications—ELSI is first and foremost a critical endeavor.

Compared to this type of work, the present project is somewhat more encompassing. Although questions regarding the implications of neuroeconomics for our conception of society, for our conception of human beings as social animals and for more traditional sciences of the social are important to this project, they are discussed as part of what, at its core, is an epistemological and ontological analysis. The question of how trust is stylized by
neuroeconomists is a question concerning a knowledge practice with evident ontological and social impact. However, I will consider this impact only insofar as our analysis of the epistemological and ontological issues permits this. Returning to Popper’s metaphor, more than anything else what I will analyze here is the shape and size of piles that are used to support neuroeconomic facts about trust. What claims, what technologies, what assumptions and what other facts are required in order to establish neuroeconomic facts about trust—that is what I will deal with. The piles that I find will be inspected thoroughly, and the implications of the facts they link up to will not be ignored. Nevertheless, this will be an analytical enterprise more than a critical one.

1.4 Road map

In the previous sections, I have briefly introduced the main issues I will discuss in this dissertation: First, there is the question concerning the conditions of possibility for neuroscience to say anything about trust. Second is the question of how, once neuroeconomics has emerged and trust is taken up as an object of research in that field, trust is stylized. Third are the implications for academic research and the division of academic labor and, more pertinently perhaps, the implications for our conceptions of social bonds and of what type of being man is. Last is the subordinate question concerning the appropriateness of my Fleckian approach to science.

Before it is put to work in chapters 4 through 6, I will first elaborate the Fleckian perspective in chapter 2 and illustrate its precepts in chapter 3. That is, chapter 2 will systematically introduce the Fleckian framework, while chapter 3 will build on the ever-growing historical, philosophical, sociological and anthropological literature on the neurosciences and their history in order to illustrate some of the ideas central to the Fleckian outlook on science. Simultaneously, chapter 3 will provide background concerning some of the philosophical and technological niceties of today’s neuroscience and, by extension, neuroeconomics. Here, I will begin to identify several of the active linkages implicated in neuroeconomic studies, specifically those that are grounded in the idea of localizationism—the idea that psychological functions can be localized in the brain.

Following in chapter 4, I will discuss the emergence of neuroeconomics first, by analyzing the “official” disciplinary history of neuroeconomics as described by Paul Glimcher and colleagues (2009b) and, second, by using a case that examines the promotion of neuroeconomics as an independent field of study. In doing so, I will offer a broader view of the style at work in neuroeconomics and will provide various other indications concerning the links actively engaged in the neuroeconomic line of work. Here we get, in other words, a better glimpse on the piles in the swamp.
Once I have presented a basic idea of what neuroeconomics is, where it came from and how it has developed, I will delve into the neuroeconomics of trust in chapter 5. In this chapter, I will discuss various lines of research concerning the topic of trust in order to create a clear picture of what neuroeconomists make trust out to be. Thus, I will focus on what facts about trust have been established—what passive linkages neuroeconomists have connected to or, to stick to the metaphor, what is erected above the swamp.

Subsequently, in chapter 6 I will move deep into the swamp. This chapter will deal extensively with the active linkages at work in the neuroeconomics of trust, (i.e., those linkages through which the previously discussed facts were reached). In this chapter, then, the style enacted in neuroeconomics will be discussed in detail, as will be its consequences for our conception of man as social animal.

In chapter 7, finally, I will draw conclusions regarding neuroeconomics, trust and reality and return to the broader motivations behind the endeavor at hand.