CONTRIBUTIONS TO ECONOMICS

Serena Sandri

Reflexivity in Economics

An Experimental Examination on the Self-Referentiality of Economic Theories

Physica-Verlag
A Springer Company

Reflexivity in Economics

Serena Sandri

Reflexivity in Economics

An Experimental Examination on the Self-Referentiality of Economic Theories

Physica-Verlag
A Springer Company

Dr. Serena Sandri
Humboldt-Universität zu Berlin
Wirtschaftswissenschaftliche Fakultät
Institut für Entrepreneurship
Spandauer Str. 1
10178 Berlin
Germany
serena.sandri@wiwi.hu-berlin.de

ISBN: 978-3-7908-2091-1 e-ISBN: 978-3-7908-2092-8
DOI: 10.1007/978-3-7908-2092-8

Contributions to Economics ISSN: 1431-1933

Library of Congress Control Number: 2008935108

Cover design: WMXDesign GmbH, Heidelberg

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

To Ali, my father-in-law

Acknowledgments

I would like to thank my supervisor, Prof. Marco Lehmann-Waffenschmidt, for his support and helpful comments.

Further, I especially thank Prof. Werner Güth for his constructive suggestions and for his supervision of the experimental part of this work.

For funding the experiments reported in this dissertation I gratefully acknowledge the Max Planck Institute for Economics in Jena and the University of Bari.

Contents

Introduction

Since the individuals are not just stimulus-response machines but more complex beings that think and are simultaneously conscious of their thought, reflexivity is potentially involved in all human acts of cognition and in all conceptualizations. On this basis, each human discourse can be characterized as a way of thought formulation and therefore, reveals a self-referring nature. On this level of reflexivity, the individual thought shapes beliefs and mental representations which give life to mental models and strive to predict future events and developments to support the individuals in their decision-making. Such mental models are reflected by the individuals themselves and on the situation they are confronted with. According to the result of this recursive application, the individuals will then decide which model they want to refer to, or in other words, which model they want to absorb.

Similarly, the individuals can make use of social theories and predictions which can therefore yield recursive effects and interfere with the phenomena they aim to depict. Revealed theories, if accepted, may influence the behaviour or the agents they focus on, either in the sense of validation of the theoretical content or in that of its rejection.

This dissertation tries to discuss the implications of the recursive or self-reflexive effects of economic theories on bounded rational economic behaviour and interaction. The mechanisms through which bounded rational actors perceive the self-referential nature of economic theories and might absorb their prescriptions will be focussed and deepened both from a theoretical and an experimental point of view, according to the evidence of two experimental studies.

First of all, the polymorphism of reflexivity and its involvement in any form of human understanding, activity and conceptualizing will be underlined, and the concept of "self-reference" will be defined on basis of different kinds of reference relations. Some common varieties and possible taxonomies for self-reference will then be discussed and some implications for formal and natural language presented. In order to test self-reference at logical consistency and to extrapolate some guidelines for its legitimacy, its relation with paradoxes will be deepened. Some considerations on the role of self-reference for human understanding as well as for social and individual decision-making conclude the first chapter.

S. Sandri, *Reflexivity in Economics*: *An Experimental Examination on the Self-Referentiality of Economic Theories*,

The investigation of the notion of "self-reference" will, in particular, indicate that neither social reality, nor its observation or description can be disentangled from their self-referential character, which means that reflexivity and its implications should be of central concern for the social sciences and research. In particular, it can be distinguished from two orders of reflexivity affecting social reality: a first order of reflexivity which involves social reality per se and consists of social phenomena which may recursively have an effect on themselves, and a second order which concerns the "discourses" on social reality, such as social sciences and theorizing.

The analysis of the first order of reflexivity will be explored in Chap. 2, which offers an overview of common reflexive social phenomena ranging from anthropology, linguistics, law, politics, sociology and psychology. The chapter concludes with examples of reflexivity which involves economic reality.

In order to discuss the implications of reflexivity for the social sciences and the predictability of social and economic reality, Chap. 3 adopts a constructivist perspective which considers reflexivity as an inescapable basis of all that can be thought and conceptualized and as a key concept for deepening human cognition and behaviour. On this basis the implications of reflexivity of the social sciences for the predictability of social and, in particular, economic reality will be widened because they emerge from the comparison between the processes of explaining and predicting social reality. The self-altering, reflexive effect of social predictions which emerges in this insight will be discussed as well.

Since the recursive effects of economic theories and predictions on the dynamics of an economic system essentially depend on their understanding and acceptance by the economic actors, the discrepancy between the neoclassical rationality standard and the observable cognitive limitations to the subjective rationality will be explicitly analyzed. Therefore, Chap. 4 focuses on the rationality debate in economics and discusses the unsolved dualism between rationality assumption and psychology of choice. The role of heuristics in orienting bounded rational decision-making will be investigated in Chap. 5 and presented together with some elements of the debiasing research which can serve as a design of a suitable framework for the experimental analysis of the self-referentiality of economic theories and their absorption.

Chapter 6 specifically addresses the concept of "theory absorption" which constitutes a tool for analysing the possible causal role of theories on bounded rational economic behaviour and for testing theoretical frameworks at their descriptive and normative validity. This concept will be discussed both in the spirit of the neoclassical theory and of the bounded rational analysis.

This analysis will, in particular, indicate that the recursivity of economic theories and their absorption differ from case to case, which means that the inquiry of this topic needs to be supported by empirical and experimental findings. A possible approach to the experimental analysis of self-referentiality and absorption of economic theories will thus be sketched in Chap. 7, which mainly focuses on the results of the two experimental studies of Morone, Sandri, and Uske (2008) and Fiore, Morone, and Sandri (2007).

The first experiment examines the absorbability of equilibrium predictions on guessing games, whereas the second discusses the absorbability of the informational

cascades' theory. Both settings have several attractive characteristics for testing theory absorption among bounded rational decision-makers. They namely permit the effects of rationality to be disentangled from social preferences and have at the same time a very simple economic interpretation. Further, both the guessing game and herding behaviour represent interactive settings in which the individuals, in order to achieve a satisficing result, have to anticipate the others' behaviour and in which individually optimal behaving does not per se ensure success.

In particular, the first experiment discusses aspects of full and partial theory absorption in repeated one-shot p-guessing games with changing parameterizations, while the second experiment argues whether providing individuals with theoretical information on informational cascades affects the overall probability of herding phenomena to occur. The second experiment also argues whether an incorrect cascade can be reversed because of bounded rational adapting to the theory's prescriptive.

Chapter 1
Reflexivity and Self-Reference

The phenomenon of referring is pervasive and regards all fields of human thought and activity, so much that it appears to be an inescapable basis of all that can be thought, conceptualized and expressed.[1] The human capability of referring creates the basis for ordering the subjective perception of the world, for interpreting events, for interacting with others, etc., thus creating the basis for all activities which regard human cognition and which are essential for individual survival. Being able to establish self-references is even a necessary prerequisite for self-change and behavioural adjustment. Furthermore, the reflexive capacity underlies basic problem-solving abilities and makes mental adaptiveness possible.[2]

Consciousness (in the form of self-consciousness) can be identified as the main source of reflexivity for human thought and action. Individuals think and are simultaneously conscious of their thought, so that all discourses are both directed to outward reality (the external world) and to the inner reality of the individual who formulates them, since she is conscious of expressing them. Therefore it can be said that each human discourse, being a human way of thought formulation, has a self-referring nature.

This chapter is dedicated to the analysis of the polyvalent concept of "self-reference." After its definition which will be accompanied by an overview of the different kinds of reference relations some common varieties and possible taxonomies for self-reference will be presented. The polymorphism of self-reference will be illustrated by its implications for formal and natural language. Logical consistency of self-reference in its different forms and contexts of appearance will then be discussed, in that the relation between self-reference and paradoxes will be deepened and some guidelines for testing the legitimacy of self-references will be extrapolated. The chapter concludes discussing the role of self-reference for human understanding as well as for social and individual decision making.

[1] Cf. Bartlett (1987, p. 5).

[2] Cf. Bartlett (1987, p. 6).

S. Sandri, *Reflexivity in Economics*: *An Experimental Examination on the Self-Referentiality of Economic Theories*,

1.1 Definition

Reflexivity is "*the quality or state of being reflexive*,"[3] whereas "reflexive" stems from the past participle of Latin verb *reflectere*, i.e. *reflexus*, which means reflective, turned back.[4] The concept of reflexivity thus applies to something capable of bending back a beam of light as well as to something that is directed or turned back upon itself. In this second meaning, reflexivity as reflection upon itself encompasses mental acts of thought, something capable of reflecting, or something relating an entity to itself.[5]

The concept of reflexivity applies to the broad class of self-reflections, though to all "self-x," whereas "x" can be any possible entity. "'*Reflexivity' is the generic name for all kinds and species of circularity. It includes self-reference of signs, the self-application of principles and predicates, the self-justification and self-refutation of propositions and inferences, the self-fulfilment and self-falsification of predictions, the self-creation and self-destruction of logical and legal entities, the augmentation and self-limitation of powers, circular reasoning, circular causation, cyclic and spiral powers, feedback systems, mutuality, reciprocity, and organic form. It includes the fallacious, the vicious, the trivial, and the question begging, but also the sound, the benign, the useful, and the inescapable*."[6] So, dealing with reflexivity means to deal with a heterogeneous class of elements, the only common feature among them is that they are based on a circular self-referential relation.

The related concept of "self-reference" "*is involved in a description which refers to something that affects, controls or has the power to modify the form or the validity of that description. […] In this general sense, self-reference establishes a circularity that may involve not only referential but also causal, interpersonal or instrumental relations and thereby constitute a unity of its own*."[7] Therefore, self-reference is strongly related to the concept of feedback and based on its mechanism, whereas "*a positive [resp. negative] feedback loop [is] a chain of cause-and-effect relationships [that] closes on itself, so that increasing any one element in the loop will start a sequence of changes that will result in the originally changed element being increased [resp. decreased] even more*"[8]. An example of a positive feedback loop is the increase of money in a savings account due to the interest rate.

If the synonymous use of the concepts "self-reference" and "reflexivity" is legitimate or not is still a quite debated question. Some[9] try to interpret self-reference in a restrictive sense as a subspecies of reflexivity. As their opponents do, they also adhere to the view that self-reference occurs whenever an object refers to itself. However they merely grasp such a definition as denotation or denomination, thus

---

[3] From the Merriam-Webster's Unabridged Dictionary (2000).

[4] Idem.

[5] Idem.

[6] Suber (1987b, p. 259).

[7] Cf. Krippendorff (1986), at http://pespmc1.vub.ac.be/Asc/SELF-REFERE.html.

[8] Cf. Bartlett (1987, pp. 21, 22).

[9] As pointed out for instance in Scheutz (1995, p. 19).

as something which aims at depicting or characterising itself. Therefore, they essentially restrict the application of the concept of self-reference to linguistic species of reflexivity, both regarding natural or formal languages. According to the supporters of this distinction, the concept of "reflexivity" labels the whole categories of reflexive entities, forms and structures, and should therefore resume all kinds of self-referential phenomena, self-references included. This distinction sounds somehow artificial and rather arbitrary, since splitting reflexive phenomena according to the entity that self-refers seems to be quite a trivial distinction. In order to be reflexive, a phenomenon needs, to be based on a reflexive relation, which can be of different kinds. It therefore makes more sense to distinguish among reflexive phenomena just when they rely on different sorts of reflexive relations and classify them correspondingly. Keeping with Bartlett and Suber's (1987) tradition, since both the concept of "recursivity" and that of "self-reference" are similarly based on a feedback loop mechanism, they will be henceforth used as synonyms.

The first studies on reference concerned mostly linguistic questions, whereas a self-referential sentence is "*a statement that refers to itself or contains its own referent*."[10] Lots of elementary linguistic forms present a self-referential character: definite descriptions or proper names, for instance.[11] There are different degrees of semantical self-reference, depending on if the sentence refers exclusively to itself, or also to itself as a member of the whole class of reference. An example of a totally self-referring sentence is "This is a short statement," while the sentence "All the sentences on this page are meaningful" is just partially self-referring. Besides, a statement can also be incidentally self-referring if it can be interpreted as a self-reference, though only if the statement itself belongs to the sub-class to which it refers; for example, "Some sentences on this page are meaningful."[12] A statement may also be self-referential if it refers to itself via another statement. An example of such an indirect self-reference is if on one side of a blackboard is written "The statement on the other side of the blackboard is true," and on the other side is written "The statement on the other side of the blackboard is false."

Reflexivity has often been interpreted[13] as a possible menace to logical reasoning and as potentially leading to paradoxes. Since there are different forms of self-references, it is not possible to draw general conclusions on the logical legitimacy of reflexive mechanisms. An attempt in this direction will be done at Sect. 1.7 of Chap. 1. The famous Liar's paradox, "p is false," leads to a contradiction, whether p is false or true, and therefore it constitutes an example of a misleading, malignant, self-reference. On the other side, the sentence "q is true" cannot be anything but true and valid, since the opposite (the Liar's paradox) is contradictory per se. Thus, this is the case of a harmless, benign, self-reference. Generalizing those examples, self-references that present a self-reinforcing character - thus relying on a positive

[10] Cf. Krippendorff (1986), at http://pespmc1.vub.ac.be/Asc/SELF-REFERE.html.

[11] Cf. Bartlett (1987, p. 5). Going more into details, there is a vivid debate on the legitimacy of self-referential sentences, as well as on the relations between linguistic self-referential sentences and paradoxes. See for more information Bartlett (1987) and Whewell (1987).

[12] Cf. Whewell (1987, pp. 32, 33).

[13] This relates particularly to the fields of philosophy, logic and scientific methodology.

feedback loop - have a stabilizing effect on the phenomena they concern, while reflexivities acting in a self-refuting way - i.e. based on a negative feedback loop - have a destabilizing nature.[14]

1.2 Reference Relations

A reference relation R connects a class of referring objects with a class of objects being referred to, and can be expressed on set-theoretical terms as:

$(a,b) \in R$ iff b is referred to by a

The set of a's for which there is a b-(a,b) belonging to R - constitutes the domain of R, which can be denoted as $dom(R)$. While the set of b's for which there is an a-(a,b) belonging to R - is the range of R, denoted by $ran(R)$.[15] As shown at Fig. 1.1, the reference relation R can be illustrated by $dom(R) \cup ran(R)$.

Obviously, if domain and range of R are strictly separated from each other, i.e. if $dom(R) \cap ran(R) = \emptyset$, no self-reference can occur. A self-referential relation requires $dom(R) \cap ran(R) \neq \emptyset$, otherwise there is no possibility for an element of the domain to refer to itself as an element of the range. Only if this prerequisite is fulfilled, a self-reference may establish.

Self-reference always articulates on circular basis, whereas it can be distinguished between direct and indirect self-reference respectively if there is a loop at a or if a is contained in a cycle.[16]

Such a difference can be illustrated, considering sentence T, "This sentence is true," as an example of direct self-reference. Sentence T relies on three references: first, it refers to the sentence itself by means of the term "this"; second, it refers to the "is"-relation; and third, to the concept of truth. The sentence and its reference relations can be visualized as in Fig. 1.2.

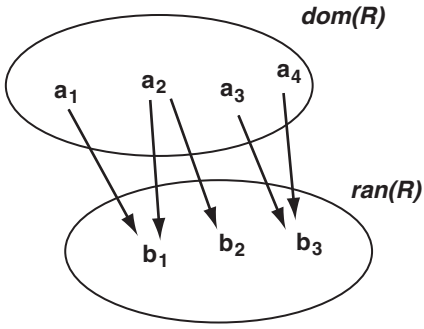

Fig. 1.1 Reference relation R (author's representation)

[14] Cf. Davis and Klaes (2003, p. 333).

[15] Cf. for definitions and notation Bolander (2002, p. 9 ff).

[16] Cf. Bolander (2002, pp. 12–14).

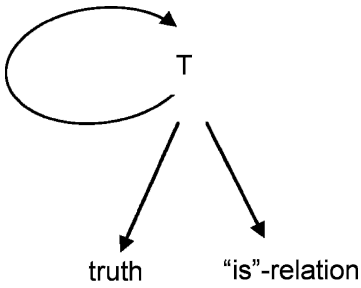

Fig. 1.2 Reference relations for "this sentence is truth" (Bolander, 2002)

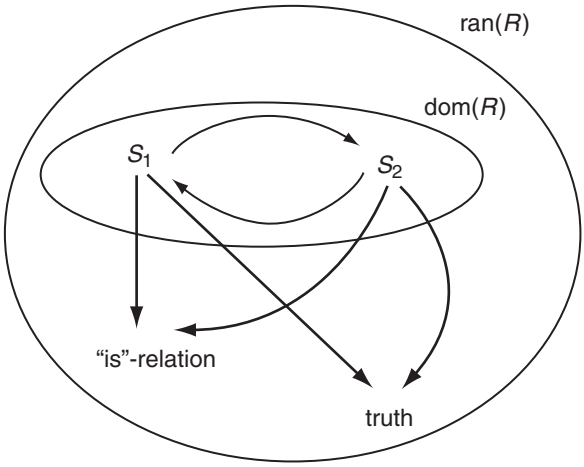

Fig. 1.3 Reference relations for S_1 and S_2 (Bolander, 2002)

The loop at T hints at a direct self-reference, because it means that T itself refers to the sentence T, so that there is a self-referential relation $(T,T) \in R$.

On the other hand, an indirect self-reference establishes when two separated entities refer to each other. Consider, for instance, the following sentences:

S_1: The sentence S_2 is true S_2: The sentence S_1 is true

Neither of these sentences is directly self-referential. Nonetheless, since $dom(R) \cap ran(R) \neq \emptyset$, the prerequisite for a self-reference to occur is given considering domains and ranges of both sentences. In addition, elements of the domain (a) are contained in a cycle in the reference relation where, as in Fig. 1.3, S_1 refers to S_2 and S_2 refers again to S_1, so that an indirect self-reference establishes.

Self-reference is particularly likely to take place in situations involving reflection, universality and ungroundedness.[17] Each of those properties, either occurring together or singularly, might account for a self-reference to establish as well as for its specific features.

[17] Cf. Bolander (2002, p. 10 ff).

1.2.1 Reflection

The notion of "reflection" has already been used in defining the concept of "self-reference," reflection meaning "bending back" and self-reference consisting of bending back to itself. Whenever something can be viewed back as in a mirror, it is commonly spoken of as reflection. In the eyes of the reproduced object, its mirroring represents an outward perspective on itself.

Thus, a reference relation R can be said to have a "reflection," if and only if for all elements involved in R there is a referring object (formally $dom(R) \subseteq ran(R)$).[18] In other words, for every element r of the domain $dom(R)$ the relation R associates an element q belonging to the range $ran(R)$.

Self-reference is just a special case of reflection, which happens whenever reflection links equivalent elements of range and domain with each other (i.e. $q = r$). As will be illustrated in the following paragraph, this always occurs when reflection combines with universality.

1.2.2 Universality

When statements regard the totality of entities they are discussing, they can be said to be universal. In other words, universal statements encompass a totality of elements, and formally, an object a belonging to the domain $dom(R)$ is called universal if $(a,b) \in R$ for every $b \in ran(R)$.[19]

A universal object a implies a self-reference whenever it belongs to the domain of a relation R that has a reflection (in the above specified sense).[20]

This result can be derived as follows: assume a to belong to $dom(R)$ being universal, as for instance the utterance "all sentences," and R to have a reflection, so that $dom(R) \subseteq ran(R)$. Because of that, and since $(a,b) \in R$ for every $b \in ran(R)$, $(a,a) \in R$, which is a self-reference, holds.

When reflection combines with universality creating a self-reference, paradoxes are most likely to appear. An example of this is Cantor's paradox, which proves an inconsistency in set theory, stating that there is no greatest cardinal number. Cardinal numbers as such must always admit for an ordering, while Cantor's paradox involves the universal concept of infinite sizes, which should be infinite per se, so that no greatest cardinal number can be determined.[21]

According to Whitehead and Russell (1910) the responsibility for similar logical inconsistency should be ascribed to universality rather than to self-reference. This

[18] Cf. Bolander (2002, p. 18).

[19] Cf. Bolander (2002, p. 11).

[20] Idem.

[21] Assuming C to be the largest cardinal number. Its power set, i.e. the set of all subsets of the set C, is 2^C, which has per definition a strictly larger cardinality than C. But this contradicts the premise, C being the largest cardinal number, from what follows, the largest cardinal number cannot exist.

is because, as stated by the theory of types, "*it is impossible or meaningless to state propositions which have an unrestricted possible range of values, or which, in any sense, are arguments to themselves*."[22]

1.2.3 Ungroundedness

Finally, self-reference is likely to appear in situations that show an ungrounded nature, i.e. in situations that rely on an infinite regress. An object belonging to the domain of a reference relation ($a \in dom(R)$) can be said to be ungrounded, if there is an infinite path in the graph of the reference relation R beginning from a. Otherwise, such path being finite, a is grounded.[23] Obviously, for ungroundedness to occur there must be some common elements between the range and the domain of a reference relation, i.e. the intersection between $dom(R)$ and $ran(R)$ should not be empty. Again, this is the same condition self-reference requires.

However, even if self-reference always leads to ungroundedness, ungroundedness is not necessarily accompanied by self-reference. It can be thought as a consequence of self-reference, since any self-referential object establishes a cycle, which inevitably results in an infinite regress, thus an infinite path passing through the entities involved in the self-referential relation.

Ungroundedness, either associated with self-reference or not, often links with paradoxes. Consider for example Yablo's paradox that posits, for an infinite sequence of sentences S_i, "All sentences S_j with $j > i$ are false, $i, j = 1, \ldots \infty$."[24]

Thus, when it comes to identifying possible sources of paradoxes, ungroundedness should be also considered.

1.3 Varieties of Self-Reference

It is possible to enumerate many types of self-references and several classifications of this concept have been proposed. In order to classify reflexivity, two different approaches can be essentially followed: it can be focused either on the entity involved in the self-referential relation, or on the kind of self-referring relation. Because of the pervasiveness of reflexivity, the first method may not offer manageable taxonomies, while the second may suffer from a lack of comprehensiveness.

In the spirit of the first approach this paragraph concentrates on a variety of self-referring entities. In offering an overview of some well-known forms of reflexivity without the claim of being comprehensive, pervasiveness and polymorphism of

[22] Cf. Weiss (1992, p. 37).

[23] Cf. Bolander (2002, p. 12).

[24] Cf. Bolander (2002, p. 13).

self-reference will be illustrated. Particular attention will be paid to the role reflexivity may assume in different contexts and for the various disciplines.[25]

The second approach for classifying self-reference differentiates the various forms of reflexivity according to the sort of recursive relation they imply and will be discussed in the next paragraph (cf. Sect. 1.4 of Chap. 1).

As reflexivity is involved in all discourses that seek to know the presupposition involved in knowing it applies to an incredible range of phenomena, spreading from the symbolic attribution of meaning to language codification and its acceptance leading to discourses referring to other discourses, for instance science of science, i.e. scientific methodology. According to Lorenzen[26] the acceptance of elementary sentences (thus of natural language) constitutes the condition of any discourse and has a recursive character per se. The symbolic process, which originates from both natural and formal language is a self-referring system, because it puts the basis for its own regulation. Symbols testify moreover self-awareness and self-consciousness of individuals: they were created to express human thoughts, so that individuals are at the same time users - i.e. subjects-and content - i.e. object - of symbolism.[27]

The first studies on reflexivity concerned linguistics and focused on both semantic reflexivity and self-referential elements of the natural language.

Natural language always grants for possibilities of referring to semantic concepts through linking language to a class to which language itself can refer, typically to its own content. Semantical self-references in particular may establish a combination of concepts truth and falsity since they establish a semantical link per se, and because they articulate on a self-referring basis.[28] Such a link can then become self-referential, if it relates a statement or its semantical content to itself, e.g. in the sentence "This is a true statement."

That natural language allows for semantical reflexivities, however does not add much to the legitimacy or congruence of semantical referring. A vivid debate has flourished about this topic,[29] which has become even more complicated, considering that there are different degrees of semantical self-reference as well as different ways a statement may semantically refer to itself. While aspects of semantical reflexivity, as well as some conditions for its logical consistency, will be more closely discussed in Sect. 1.5 of Chap. 1, it should be mentioned here that semantical self-reference may either be based on self-validating (respective self-refuting) dynamics, or rely on meta-logical linkages, for instance in meta-statements. Those are statements whose content applies to the content itself, or in other words, statements speaking about themselves, e.g. "This is not a statement about Socrates."[30] The first are statements involving judgement of themselves as expressed by sentences involving truth, falsity, plausibility etc.

---

[25] The main reference for this paragraph is Bartlett (1987).

[26] As mentioned by Bartlett (1987).

[27] Cf. Winrich (1984, p. 988).

[28] Cf. Winrich (1984, p. 988).

[29] For a contribution, see e.g. Whewell (1987).

[30] The example stems from Whewell (1987, p. 35).

When a sentence is applied for its own legitimization or defence, the threat of logical rudeness is always present. This occurs in ad hominem argumentations which cannot be falsified and are therefore irremediably sterile. Although the menace of logical rudeness can be mainly related to self-validating semantical reflexivity, it should be contemplated even for dealing when with meta-statements.

Self-referential elements are present in all natural languages, and they establish a circular recursive semantical relation and therefore they rely on similar characteristics. Several attempts have been made to reduce them to a common denominator. Peirce[31] refers to these self-referential elements as "indexical signs," Russell[32] as "egocentric particulars" and Reichenbach[33] as "token-reflexive words." All these elements refer to the argument of a sentence, whereas the subject, object, or conditions stated can be hereby meant. Self-referential elements are for example terms as "I," "here," "now" etc. These terms all share the property of referring, in a way that is relative to the speaker who uses them. While Peirce's analysis follows an enumerative approach, Russelll and Reichenbach strive toward reducing all such expressions to the utterance "this." Furthermore, it can be noted that recursive relations based on self-referential elements of discourse typically do not create the conditions for paradoxes to take place.

A further variety of reflexivity can be generated by dictionary reference. This occurs most in an indirect way, whenever a circular relation between two or more references is established. A dictionary provides a definition for each word, which cannot be expressed but in terms of other words. Thus, if a concept has been explained recurring to a second one, the latter might be of a certain likelihood used for defining the previous concept. In this case a direct self-reference would occur. To a greater probability, however, self-reference would establish indirectly, i.e. after a long chain of words linked together by mean of reference relations: a first concept is defined through a second concept, which is again explained through a third, a forth, a fifth concept and so on, up to when the first concept will be employed again. In this way an indirect self-reference would take place creating a cycle that links the concepts altogether in an ungrounded regress. Dictionary reference is not merely a speculation or a possibility, but something inevitable since languages do not possess but a finite number of words, so that any dictionary must contain words which are indirectly defined in terms of themselves.

As pointed out by Wittgenstein (1999), dictionary reference also articulates at the semantic level, mining with ungroundedness the entire semantic construction of language. A dictionary relates words to each other, so that without knowing any words it is not possible to learn anything from it. Thus, a dictionary constitutes a closed self-validating system and suffers from a logical point of view from ungroundedness. Ostensive definitions,[34] which convey the meaning of a term by pointing out the object itself or examples for the word to be explained would represent

[31] Cf. Peirce (1931–58).

[32] Cf. Russell (1940).

[33] Cf. Reichenbach (1947).

[34] Cf. Wittgenstein (1999).

a possibility to escape the self-referentiality of defining and of dictionary attribution of meanings. Their applicability however remains confined to pure speculative purposes and does not make much sense for practical usages.

Reflexivity plays a particularly significant role both for philosophical argumentation and for phenomenological analysis. "Petitio principii," "reductio ad absurdum," and "ad hominem arguments" are just some examples of circular figures that are often used in philosophy. Philosophy relies much on reflexive approaches in particular when it comes to arguments with transcending as well as transcendent orientations. These are respective arguments trying to delimit the realm of possibilities and arguments establishing a linkage between entities belonging to different ontological or logic levels.

The solution for such argumentations lays in many cases on self-application of principles or categories, self-validation (respective-refutation) of theories or their propositions, as well as on the self-supporting nature of inductive speculations. Reflexivity supporting philosophical argumentation operates at a procedural level and regards the form of the argumentation while reflexivity that may be involved in phenomenological elaborations concerns the content of the argumentation itself. E.g. in Husserl's phenomenology,[35] phenomenology can be reduced to a science of science, thus to a theory of theories. The radical constructivist interpretation of any phenomena as a construct of an observer's mind posits the recursive character of ontology and phenomenology, because it ascribes to an autopoietic origin of the mind.

Self-referential techniques have been exemplarily applied in proof theoretical approaches to mathematics and their analysis leads to un-reassuring conclusions since it indicates the paradox nature of mathematics. This is because, self-references are typically involved in mathematical reasoning when it comes to extend the analysis to the borders with infinite (thus universal) entities. Concepts such as "infinity," "non-closed system," "incompleteness," "non-excludability," etc. are most likely to create the basis for paradoxes arising on a self-referring basis since they involve both infinity and negation. The scale of the problem can be appreciated thinking that as soon as self-reference is ruled out, mathematics (that is in the end a formal language and as such a closed self-referring system) becomes unworkable. Corollaries of that nature are provided e.g. by Gödel's numbering,[36] and by Russelll's paradox developments of the axiomatic set theory.[37]

Other self-referring applications of mathematics can be found in the computability theory and are e.g. illustrated by the reflexive functions theory, as well as by research on artificial intelligence. Studies of artificial intelligence deepen the notions of self-correcting, -regulating, -organizing and -reproducing systems, of self-initiated learning, etc.[38]

A way to codify the perception of space and time when it comes to understanding infinite phenomena is to use recursive signs and representations.[39] The closed loop,

[35] Cf. Elveton (2000).

[36] Cf. e.g. Smullyan (1991).

[37] Cf. e.g. Winrich (1984).

[38] For an overview on artificial intelligence, see e.g. Luger (1995).

[39] More on that is presented e.g. in Priest (1987).

the Moebius strip, the Klein bottle and the Rimannian model of a closed universe are example of recursive structures which are used for spatial analysis. Further, closed temporal loops can constitute self-referring tools for capturing temporal cyclic dynamics and can be applied to the most various fields of analysis, that range from particle physics to the analysis of economic cycles.

Physics may be affected by reflexivity as well as illustrated by the puzzling results in quantum mechanics and general relativity. Despite traditional view of a natural scientist as perfectly disentangled from the reality she observes, experiments in quantum mechanics reveal that in some situations measuring system, observer and quantum phenomena to be measured "*constitute a system which itself reflexively defines properties of the phenomena which may be measured*."[40] Functionally interdependent descriptions represent a reflexive structure which is often used in physics for example in general relativity, density and gravitation are defined as functions of the space curvature, in models of closed universe, which is at the same time unbounded but finite.

Up to this point, all reflexivities mentioned have in common that they are to be more or less the product of human thought and activity. This is concurrent with the thesis interpreting self-consciousness as main source of reflexivity. It should be however noted that reflexivity may as well exist independent of human activity, as attested by numerous reflexive natural phenomena. For instance, recursivity is implied in viral reproduction, which can be interpreted as a self-replicating process. A further example is offered by the biological notion of the self-organizing system.[41]

1.4 Taxonomies of Self-References

Basically, it can be distinguished between self-references which establish circular (i.e. closed) or processual (i.e. open) relations. While incompleteness might affect the legitimacy of closed self-references, processual open self-references are mostly immune to such a threat.

Self-referential relations can establish tautological, set-theoretical, pragmatical or metalogical relations.[42] A tautological self-reference can be characterised as a static relation, or in other words as a self-reference that does not add anything but redundant information to its predicate. A set-theoretical reflexive relation generates mostly paradoxes and it appears when set-memberships are used in a reflexive way. Groucho Marx's gag: "I don't want to belong to any club that will accept me as a member," is an example of such reflexivity. When the content of a sentence and the sentence itself refer to each other, as in the sentence "there are no truths," it can be spoken of as a pragmatic or performative self-reference. Finally, the relation "*between a truth-functional referring proposition and the set of conditions which*

[40] Bartlett (1987, p. 13).

[41] This notion can be applied to many phenomena like e.g. the spontaneous folding of proteins, formation of lipid membranes and homeostasis. For more, see e.g Solé and Bascompte (2006).

[42] This classification of self-references is freely based on Bartlett (1987).

are necessary in order for the proposition to be capable of referring at all"[43] can be defined as meta-logical. Human understanding in general and scientific reasoning often involves this sort of referring.[44]

Another interesting taxonomy of self-references has been proposed by Davis and Klaes (2003). They distinguish between immanent, epistemic and transcendent reflexivity depending on which levels the reflexive relation involves. Immanent reflexivity means a reflection from an entity to itself, while epistemic reflexivity results from a conscious act of a subject referring to itself. Eventually the transcendent reflexivity almost coincides with the meta-logical self-reference presented above.

Scheutz (1995) restricts self-reference to linguistic forms of it and interprets reflexivity as the general class to which it belongs.[45] On that basis he distinguishes between reflexivity of symbols, syntactic and semantic reflexivity, set theoretical, pragmatical or performative, and metalogical or transcendental reflexivity.[46]

Even if it seems impossible to reduce the different way of classifying self-reference to a unique resuming taxonomy, an attempt can be made in this direction, since it can be noted that there is in fact a common denominator in analysing reflexivity, and that there are some categories that similarly appear in the different classifications. In particular, there is a certain consensus in recognizing the specificity of the mechanisms involved by semantical, pragmatical, metalogical and computational - in the sense of set theoretical - reflexivity.

Resuming, what emerges from the analysis of the concepts of "self-reference" and "reflexivity" is their occurrence in uncountable forms and their application to a broad spectrum of phenomena. After this introductory overview on varieties and general taxonomies the concept of self-reference will be related to those aspects that are particularly relevant for human conceptualizing, namely reflexivity in natural and formalized language. The first will be explained in the next paragraph (Sect. 1.5 of Chap. 1), where some of the possible ad hoc classifications will be presented and linguistic terms and general mechanisms through which self-reference may establish will be discussed. The latter form of reflexivity is discussed in Sect. 1.6 of Chap. 1, and deals with how self-reference can be expressed in formalised languages and with the consequences for the consistency of the system to which the formal language refers.

1.5 Self-Reference in Natural Language

In examining self-reference in grammar, Van Fraassen[47] distinguishes between two kinds of self-referring terms, namely context-dependent and -independent. The

[43] Cf. Bartlett (1987, p. 10).

[44] More on Sect.1.8 of Chap. 1.

[45] This has already been discussed in Sect. 1.1 of Chap. 1.

[46] For more, see Scheutz (1995, p. 14 ff).

[47] It is here mostly referred to Van Fraassen (1970).

difference among them lies in having a variable (i.e. "context dependent"), or an invariable (i.e. "fixed") addressee. Formulated otherwise, it can be spoken of a context-dependent self-referring term, when "*The extension of the term (or rather, the extension of an occurrence of that term) depends on the context in which it occurs.*"[48]

Each type of self-referring term implies that a different mechanism of self-reference will occur, in order to be distinguished between accidental or functional self-reference: "*Whenever self-reference consists in reference to a sentence by means of a term occurring in that sentence [...], and this term is such that its occurrences have the same extension, we shall speak of accidental self-reference. [...] When self-reference is produced through the use of terms with context-dependent reference, we shall speak of functional self-reference.*"[49] This taxonomy however has some shortcomings in that it exclusively sticks to the dimension of context-dependency without capturing other variables that can be made responsible of different ways for self-referring. For example, Van Fraassen's taxonomy does not enhance to distinguish between total and partial, between direct and indirect, or between benign and malign referring.

According to this insight, the classification of self-reference in natural language proposed by Martin[50] goes beyond Van Fraseen's, in that it contemplates two levels for evaluating self-reference: at the first level, the difference between benign and malign self-references is considered, while at the second self-references are classified in direct, indirect, empirical and general.

Martin belongs to the philosophical tradition interpreting self-reference as a constitutive part of human language and discourses so that its benignity or malignity may affect logical reasoning. A self-reference can be clearly stated as being malign, when contradiction follows, whereas it can be denoted as benign if it "*can be judged, in logically unproblematic way, to be straightforwardly true (or false)*".[51] The state according to which all self-referring statements cannot be but meaningless and necessary leading to paradoxes, was firstly advanced by Russell and supported by his theory of types.[52]

The second distinction which Martin introduces is centred on the difference between "*self-referential sentence type*" and "*self-referential token*" in the natural language:[53] "*a sentence type is self-referential if and only if every token of that type is self-referential, and [...] a sentence token is self-referential if and only if what one mentions with the subject expression of the sentence token is the sentence itself (type or token) which is being used*".[54]

---

[48] Cf. Van Fraassen (1970, p. 696).

[49] Cf. Van Fraassen (1970, p. 696).

[50] See Martin (1967, 1992).

[51] Cf. Martin (1992, p. 119).

[52] Cf. Russell (1903).

[53] Such distinction exclusively applies to natural language and cannot be extended to formalized language.

[54] Cf. Martin (1992, p. 133).

Whenever a sentence type is self-referring, i.e. whenever each of its token is in the sense of the previous definition self-referential, it can be spoken of direct self-reference. This is defined in Martin's words the "*crux version*" of self-reference. An example of a reference of this kind is the sentence: "This very sentence is so and so," since "*the expression 'this very sentence' refers to the whole expression of which it is a part*."[55]

Direct self-reference differs from the other categories, namely indirect, empirical, and general reference. These differ in that the first requires that its entire tokens token are self-referential, while the others include sentences only tokens of which are self-referential.

The category "indirect self-reference" always applies to a plurality of sentences, i.e. up to a pairs, as "The next sentence is so and so. The previous sentence is this and that," or as Jones saying "Most of Nixon's assertions about Watergate are false," Nixon replying: "Everything Jones said about Watergate is true."[56] Indirect self-reference, neither sentence being per se self-referential, is a sub-case of reference to itself by mean of other sentences. Indirect reference does not necessarily imply indirect self-reference, except when $dom(R) \cap ran(R) \neq \emptyset$.[57] "The next sentence is funny. The previous sentence is English"[58] constitutes an example of indirect referring sentences that do not imply a self-reference.

The sentence "Sentence (1) is so and so" furnishes an example of empirical self-reference, which is in the end a sort of hidden direct self-reference.[59] Recursivity occurs as the context, in which a sentence token is inserted, makes the token refer to the sentence, so that the sentence assumes the character of a self-referring statement. This is the case of a context dependent (in this sense "hidden") direct self-reference, since reflexivity is given by the context, which makes out of (at least) one token a reflexive term, which directly refers to the sentence.

General referring is applicable when a sentence occurs to be self-referential, because of referring to a whole class and being in the meanwhile member of it.

Because context dependency matters in Martin's classification, Van Fraassen's and Martin's taxonomies can be integrated, the latter classification represents an extension of the first. Van Fraassen's category of accidental self-reference applies to Martin's direct self-reference, whereas under the label "functional self-reference" all remaining varieties of self-reference, i.e. indirect, empirical and general self-reference can be subsumed. The distinction of which Martin's taxonomy is based – i.e. the difference between self-references caused by sentence types or tokens - seems to provide a good perspective for capturing the specificity of the different mechanisms self-reference in natural language involves. Unfortunately, Martin's classification cannot be applied to self-reference in formalized language because of the impossibility of formally finding an equivalent to the type-token

[55] Cf. Martin (1992, p. 133).

[56] This example refers to Kripke (1975).

[57] Cf. Sect. 1.2 of Chap. 1.

[58] Freely translated from Scheutz's examples. Scheutz (1995, p. 49).

[59] Cf. Scheutz (1995, p. 52).

distinction. The same applies to Van Fraassen's classification, as well. Again, the appealing possibility of classifying different varieties of self-reference adopting a unique pattern of classification is precluded because of the polymorphism of reflexivity.

In his article "Self-Reference and Meaning in a Natural Language"[60] Whawell distinguishes between benign and malign self-reference and is concerned with finding criteria for distinguishing what he calls "legitimate" from "illegitimate" self-references. He lists three possible ways for a sentence to be self-referential that are to be totally, partially or incidentally self-referring. In a more general form this distinction has already been mentioned at p. 3. It will be here more specific related to the context of natural language and will be then compared and integrated with the other taxonomies that have been proposed in this paragraph.

Once again Whewell's classification is based on the different degrees according to which semantical self-reference can occur: a sentence can either refer exclusively to itself, being thus totally self-referring, or also to itself as a member of the whole class of reference. For this to occur a partially self-referring character it should be assumed. In particular, "*A statement is totally self-referring if it refers explicitly to itself by means of a singular referring expression*,"[61] i.e. by mean of a term referring exclusively to itself. This allows for other terms of the totally self-referring sentence to be other-referring, as for instance in the sentence: "*this statement and the previous one are false*."[62] This is still the case of a totally self-referring sentence as the other-referring term does not modify the totally referring character of the sentence.

Whewell's classification focuses exclusively on semantic degree of reference, and grants explicitly the possibility of plural referring, but does not explicitly distinguish between direct and indirect referring. The category of totally self-referring sentences is intended to encompass even those cases that in Martin's taxonomy would have been interpreted as examples of indirect self-reference. For example if on one side of a blackboard is written "The statement on the other side of the blackboard is true," and on the other side "The statement on the other side of the blackboard is false."[63]

The second category in which Whewell divides self-reference is that of partial self-reference. In his words, "*A sentence is partially self-referring if it is about a whole class of statements of which it is itself a member. For instance, 'all the statements on this page are true' and the statement 'every meaningful but non-tautological statement must be in principle empirically verifiable*.' "[64] For this sort of reflexivity to occur the context -in a broad sense – plays an important role, because no referring tokens, terms or expressions are strictly necessary for such a partial reference. The only requirement being that the sentence establishes a reference to itself via its meaning. Whewell refers to that by stating that it can be sometimes

[60] Whewell (1987).

[61] Cf. Whewell (1987, p. 32).

[62] Idem.

[63] The examples have been literally quoted from Whewell (1987, p. 32).

[64] Cf. Whewell (1987, p. 33).

necessary "*to go outside the statement in order to know whether it is self-referring or not*"[65]. Equivalence with Martin's general referring can be easily inferred.

The third category of self-reference is that of incidentally self-reference. This could also be interpreted as a sub-species of partial self-reference considering that a statement can be incidentally self-referring even when its self-referring nature emerges. This is not only because the statement is a member of its class of reference, but also because it belongs to the sub-class of members to which it refers, as well. This can be illustrated by the sentence: "Some sentences on this page are meaningful."[66] The other classifications previously presented did not provide a separate category for this sort of reflexivity and consider it equivalent to partial self-reference. It seems convenient for analytical purposes to rely on such a distinction, being the self-referring character of incidental self-referring out of doubt, but at the same time is not very pronounced, as based on second-order set-membership.

Eventually, Mackie (1973) classifies self-reference on two different levels, by distinguishing between total and partial, and between explicit and implicit self-references. A total self-reference is the strictest way of referring, because it requires a referring term or phrase, e.g. "this sentence." As posited by Whewell, Mackie considers partial self-reference to origin from set-membership, but does not further distinguish between a partial and an incidental form of referring. Whenever self-reference is established because of the presence of a referring singular terminus it can be spoken of explicit self-reference. On the contrary, if a quantifier or a general terminus is responsible for a self-referential occurrence, it can be spoken of an implicit self-reference. These two orders of self-referential relations integrate, in that: "*In general, an explicit self-reference will be total, but it need not be: 'This statement and others…' or 'This statement and the previous one…' and so on. And in general an implicit self-reference will be only partial; but it is only contingently that an implicit self-reference is a self-reference at all […] and what might have been only a partial self-reference could contingently be total, e.g. 'Something written on this page…' where nothing else happens to be written on the page in question*."[67]

1.6 Self-Reference in Formal Language

A formal language[68] (L) is defined by mathematical or machine processable formulas. It can be characterised as the set F of finite sequences of elements drawn from the finite set A of symbols, so that $L = \{A,\ F\}$. A further differentiation that won't be discussed here can be drawn between a formalized logistic system (calculus) and

[65] Idem.

[66] Cf. Whewell (1987, p. 32, 33).

[67] Cf. Mackie (1973, p. 286).

[68] Cf. Wikipedia at the voice "formal language," http://en.wikipedia.org/wiki/Formal_language, 23th January 2007.

formalized language system (interpreted language).[69] This could be of a certain interest for example for assessing and rethinking the linkages between reflexivity and human thought.

Self-reference in formalized language is a particular delicate topic, because it is often linked with logical inconsistency. It will be briefly illustrated in the following how formal languages may allow self-referential devices, how such possibilities may be responsible for inconsistency and under which conditions consistency might be regained.

Many formal languages are equipped with the possibility of establishing self-referential statements. Smullyan's approach to this concept it to define a chameleonic term σ, which refers to the same formula in which it occurs.[70] In other words, the term σ is provided of a context dependent semantics, so that it can be viewed as a formalized translation of the term "this (sentence)."

Barwise and Etchemendy[71] adopt a similar mechanism for establishing self-references in formalized language, that consists in the definition of the formal term "this," denoting the proposition in which it occurs.

Similar ways of establishing direct self-referring relations require however quite a sophisticated formalized language, and cannot be applied in elementary languages, as e.g. in first-order predicate logic. Given the tools of a simple logic, self-references cannot be established but in an indirect way, and has to involve diagonalization or equivalent procedures resembling indirect self-reflection in natural language.

The diagonalization method involves a diagonalization function $d : N \rightarrow N$ defined by $d(\phi(x)) = \phi(\phi(x))$. If the function d is recursive it then follows that there exists a formula $D(x_1, x_2)$ representing d in Q. Given the formula $D(x_1, x_2)$ the diagonalization lemma can be stated as follows: Let $\gamma(x)$ be a formula of predicate logic with only x free. Define the formulas $\alpha(x_1)$ by $\alpha(x_1) = \forall x_2(D(x_1, x_2) \rightarrow \gamma(x_2))$ and β by $\beta = \forall x_2(D(\alpha, x_2)) = \alpha(\alpha)$.[72]

The diagonalization lemma relates to self-reference in that it expresses the same semantic principle involved in the sentence α: "Sentence α has the property expressed by β." As it characterizes itself as being on posses of property β, sentence α makes a statement about itself.

1.7 Logical Consistency of Self-Reference

The history of reflexivity is tied with that of paradoxes and logics in general, because the issue of reflexivity has often emerged in combination with the occurrence of paradoxes.

---

[69] For more on that, see Martin (1992, p. 75), who refers to the contributions of Carnap (1942) and Church (1951).

[70] Cf. Smullyan (1984).

[71] Cf. Barwise and Etchemendy (1987).

[72] Cf. Bolander (2003, p. 70 ff), who refers to Mendelson (1997).

Paradoxes have always attracted philosophers and logicians, because their analysis is often pushed to the boundaries of human thought and of the validity of logical constructions. In some cases exploring the insurgence of paradoxes furnished "*the occasion for major reconstruction at the foundation of thought*."[73] Reflecting on ambiguities which underlie paradoxes is not just a sterile speculation as it might sound at first sight, but a feasible method to achieve scientific advances.

1.7.1 Self-Reference and Paradoxes

The concept of a "paradox" can be applied to statements that appear plausible and true, but in fact imply something contradictory. A paradox is defined as "*a statement that is actually self-contradictory and hence false even though its true character is not immediately apparent*"[74] or "*an argument that apparently derives self-contradictory conclusions by valid deduction from acceptable premises which defies either intuition or logics*."[75] The word paradox stems from the Greek and combines the preposition "para," which means "against,", "beyond" and the substantive "doxon," which stands for "thought," "belief," "opinion," so that etymological paradoxes refer to something that is "contrary to expectation." The use of this concept can be traced back to Plato's Parmenides where Zeno during a conversation with Socrates and Parmenides criticises some accepted understandings by means of commenting on their own basis to paradox conclusions. The most famous of these paradoxes is the one which says that as long as a tortoise continues to move, it won't be overtaken by the fleet-footed Achilles. Though it is now easy to solve Achilles' paradox through the notion of convergent series, it should not be forgotten that this notion was introduced in mathematics only in the 19th century. So, what has long been regarded as an unsolvable incongruence, resulted in the end to be based on a falsifiable premise.

Paradoxes, whose contradictory conclusion simply stems from a fallacy in their demonstration, can be called, according to Quine's classification (1962), "falsidical." A falsidical paradox can be in other words solved, and represents no real danger for logical reasoning. Besides Zeno's paradox, another well known falsidical paradox is that of the disproof "$1 = 2$," which is based on a wrong division.[76]

Another class of paradoxes is that of "veridical paradoxes," which appear to be absurd, but can be demonstrated to be valid and true. Quine (1962) refers in this insight to an anecdote about the protagonist of "The Pirates of Penzance," who after five birthdays is 21 years old, being born - here the solution is of apparent absurdity - on February 29th.

---

[73] Cf. Quine (1962, p. 21).

[74] Cf. Marriam's Webster Unabridged Dictionary (2000) at the voice "paradox."

[75] Idem.

[76] For $x = 1$, $x = x_2$, so that $x_2 - 1 = x - 1$. Now, dividing both sides by $x - 1$ the supposed result is $x + 1 = 1$, from which it follows $x = 2$, so that $1 = 2$ should come out. It is easy to find out the error in the division of $x - 1$ by itself, whose result is 0 and not 1. Cf. Quine (1962, p. 22).

There is a possibility that paradoxes may also belong to a third class, which is in fact the one which "*brings on the crises in thought*."[77] This is the class of "antinomies."

An antinomy is a paradox of a particular kind, which "*produces a self-contradiction by accepted ways of reasoning*."[78] That is, it is the acceptance of the principles stated in the antinomy itself that leads to paradox conclusions. Antinomies are typically universal statements that, once applied recursively, lead to an internal contradiction, because they are being the undermined principle in the rule coincident with the one sustained by the antinomy itself. In this sense, the paradox nature of antinomies can be ascribed to its application to itself, i.e. in its self-reference. Antinomies might represent a threat for logical reasoning since they require a rethinking about the principle stated in the antinomy, which says that it has either to be abandoned or somehow restricted.

Some examples of well known antinomies will now be presented. From the exploration of the self-referential mechanisms on which they are based, some tentative conclusions on how self-reference can affect logical thought will be discussed.

Grelling's paradox[79] deals with the notions of the "autological" and "heterological." The concept "autological" applies to self-describing adjectives, such as the adjective "short," which is in fact a short adjective, "polysyllabic," which is in fact polysyllabic and so on, so that every autological adjective can be said to be "*true of itself*."[80] On the contrary, the category of "heterological" applies to all adjectives that are not suitable for their own description. The adjective "long," for example, does not result to be a long adjective at all, or "tall," "red," "monosyllabic" etc. Such adjectives cannot be applied to their own description, as they are not true of themselves. Now, what can be said for the adjective "heterological?" It cannot be autological since this would imply that "heterological" is self-depicting, true of itself, which would then require that "heterological" be heterological, but how can it be true of itself (i.e. autological) if it is not heterological? This chain of reasoning cannot be reduced but ad absurdum, leading to what is known as Grelling's paradox.

Another famous antinomy is known as the "Liars paradox" and is given by the Cretan Epimenides saying all Cretans to be liars. To the same category of antinomies belong the paradox sentences "I am lying" or "This sentence is false."

Similarly, the exception rule states that there is an exception to every rule. But, if every rule admits for an exception, what about the exception rule itself? Is there an exception to the rule saying that there is an exception to every rule? The application of the rule to itself leads in this case to paradox conclusions, so that the validity of the rule seems questionable.

Russell's paradox concerns set-theoretical membership and seriously threatens the axiomatic fundaments of set theory. It became popularised by the cover story of

[77] Cf. Quine (1962, p. 23).

[78] Cf. Quine (1962, p. 23).

[79] The first formulation of this paradox can be ascribed to Kurt Grelling and Leonard Nelson who formulated it in 1908. Cf. Quine (1962).

[80] Idem.

a village's barber, who is an adult male who shaves all male villagers who do not shave themselves or anybody else. The question now is: who shaves the barber? In set-theoretical terms, Russell's paradox deals with the question, whether the set of all sets which do not contain themselves contain itself?

1.7.2 Harmless and Harmful Self-Reference

It has been shown, that contradictions in antinomies are connected with self-reference. Self-reference is not the sufficient condition to constitute the paradox nature of an antinomy. Though the paradox nature of antinomies emerges in their application to themselves, other factors or characteristics must apply to get to paradox outcomes. In this sense, antinomies can be extremely useful for spreading some light on the presumed viciousness of self-reference for logical thought.

A common feature of antinomies is to deal with universal statements, and this is what makes self-referential devices unavoidable. However, universal statements can be done without incurring in an antinomy because neither universality nor self-reference can be made responsible for the paradox arising.

The paradox essence of antinomies can be rather ascribed to the combination between universality and a negative device of some kind. Grelling's paradox arises because the concept "heterological" is only possible as a negative definition, such as "not-suitable to self-description," "not-truth of itself," that is, "not-autological." Similarly, the Liars paradox is also produced by the concept of falsity, the negation of truth. No contradiction would be produced. The Cretan Epimenides said that all Cretans speak the truth: even if a similar statement has a universal character, it can be self-reflected without generating any logical contradiction. Moreover, both Grelling's and Epimenides' paradox involve the truth concept, which has been seen to have a self-referential character in itself, whereas, again, only its negation can be made responsible for the antinomy. In the same way, the paradox emerging from the exception rule relies on "exception" being essentially a negative concept, which means "*exclusion or restriction […] by taking out something that would otherwise be included*."[81] Russell's paradox also inquires negative set-membership.

Such result can be generalized to mean that self-reference can be considered to be a vicious cycle for logical thought whenever it involves a negative device, thus whenever it implies a negative backlash. On the contrary, self-reference in conjunction with positive concepts can be said to be innocuous for logical reasoning. It has been seen that an essential feature of self-reference is that it is accompanied by infinite regressions as in a feedback-loop. A self-reference will then assume a harmless or a harmful character, depending on the direction in which the regression works. In particular, a self-reference which presents a self-reinforcing character, based on a positive feedback loop, will have a stabilizing dynamic on the entities involved.

[81] Cf. Merriam-Webster's Unabridged Dictionary (2000) at the voice "exception."

On the contrary, reflexivity which acts in a self-refuting way, based on a negative feedback loop, always works in a destabilizing sense.[82]

1.8 Reflexivity in Human Understanding

Human understanding is inevitably confined within the bounds of its own conceptual structure. Such limitations cannot be transcended and inevitably arise in conjunction with understanding, conceptualizing and information procession of any kind, therefore every activity involving intellectual faculties.

To put it another way, this implies that we inevitably reflect the limits of our understanding on all our intellectual elaborations so that the products of our intellectual operations suffer from the same limitations we do.

In particular, there are two sorts of limitations affecting human understanding per se, namely pragmatic and metalogical limitations.[83] Both can be linked to a self-referential regress because they are generated by the circular structure which is established among cognitive faculties that constitute the premises of any understanding, as well as among their constructs or elaborations.

Recalling the definitions presented in Sect. 1.4 of Chap. 1, pragmatic or performative self-reference regards the commitments involved in making a certain assumption. It often comes to light in revealing the premises underlying a statement. A critical use of this kind of self-reference can be done in taking it as a measure of internal consistency, as there should not be incompatibility between the content of a theory and the commitments involved in its formulation. Therefore, proving pragmatic self-reference at its own consistency could be a first test at formal validity of the statement in which it occurs. Whenever pragmatic self-reference leads to inconsistency, a constructive use of the proven inconsistency could be thought to refute this thesis. Because it is self-referential, logically weak fundaments could be a first argument in supporting the validity of its negation.

Similarly, human understanding can be critically or constructively formulated by its meta-logical fundaments, which cannot but develop in a self-referential way. Meta-logical fundaments constitute the very fundaments of logic, which, as they stay beyond logics, give it shape and consistency. As said on Sect. 1.4 of Chap. 1, meta-logical reference occurs in evaluating the set of conditions which make truth-functional propositions feasible. Meta-logical self-reference can represent a limitation, whenever it implies self-falsifying or self-refusing dynamics, while it can be interpreted constructively, when it involves self-validation.[84]

---

[82] Cf. Davis and Klaes (2003, p. 333).

[83] Cf. Bartlett (1992, p. 3 ff).

[84] For more on the constructive versus critical use of self-reference involving human understanding see Bartlett (1992, p. 3 ff).

1.9 Self-Reference in Social and Individual Decision-Making

The following paragraph analyses the role of self-reference in social and individual decision-making and is based on a constructivist approach to human cognition and decision making[85] combined with elements of cognitive psychology.[86]

The mechanism through which human cognition tries to make the environment meaningful and predictable is one of the storing representations and establishing relations. This mechanism relies on several cognitive heuristics, one of the most relevant being causality. Mental representations connected within each other constitute maps of cognition. According to the constructivist approach, cognition can be depicted as an input-output process with a feedback mechanism of proof and update. Because of this feedback, which aims to validate the output of the cognitive process, cognition is involved in a circular confrontation with the external environment and therefore is articulated in a self-referring way. This acts in the sense as a reinforcement cycle of the kind of "*learning by success*,"[87] according to which "*[i]f a pattern of cause and effect is applied successfully, the perceived accomplishment will fortify the relations*"[88]. Success is meant here in a constructivist interpretation, a provisional validity or viability. In particular, this does not require an internal (mental) representation to reproduce reality exactly, since such a metaphysical ideal would have no operational content. Instead, a rule simply has to assure that a specific model is provisionally "valid" or "viable" for the action of an individual who acts in given conditions.

It is to this extent that the constructivist approach introduces the notion of "cognitive equilibrium" as a sort of measure to which the viability of a mental representation can be related. An individual can be said to be in "cognitive equilibrium" if the actions generated by her internal environment are consistent with her objectives, given the environmental responses.[89]

To make the concept of provisional viability more operative, the notion of cognitive equilibrium could be related to that of satisficing, since it allows for heterogeneous mind constructions to coexist and to meet subjective aspirations. Viability is based on individual experience. Experience should not be interpreted here as observation, but as action, since it is only through action that the individual is able to test her own mental models. The test result operates on the construction of knowledge as feedback, which leads either to the validation of actual mental models or to their modification. In this way, knowledge is a representation of its experience rather than a representation of the world.[90] This feedback mechanism enables the representation to characterise the cognitive process as self-referential because every construct,

[85] The perspective of constructivism will be deepened later on Sect. 1.1 of Chap. 3.

[86] It will be in particular referred to the work of Neisser (1976).

[87] Handlbauer (1997, p. 763).

[88] Idem.

[89] Idem.

[90] Cf. Maturana and Varela (1987).

once confronted with the subjective experience of the external environment, shall be reflected to the mind which originated it, i.e. it shall be self-reflected.

Thus, human cognition is steadily recursively engaged in the elaboration of mental models out of external stimuli and in their evaluation according to the individual experience of the external world, either to consolidate a (subjective) viable construct or to modify a non-viable one.

Both cognitive maps and cognition work in a self-referring way. The self-referring structure of decision making, which can be characterised as a continuous path-dependent self-referential process (as in Fig. 1.4), adds to the self-referring structure of cognition.[91] Path dependency is both the result of experience, to which cognitive maps are related for the sake of their validation, and of prior perceptions and constructs, by means of which provisional validity is weighted. In this sense, cognitive maps accumulate experiences and prior perceptions in a meaningful structure.[92] Therefore decision-making can be seen as the application of cognitive maps which have in most cases a random structure.

The meaning of cognitive maps stems from the circularity of learning-by-success on which they are based, that is in the continuous proof at their adaptability to external environment. Updating and revising of cognitive maps, as well as decisions, initially focuses on viability, then sense-making. In this regard, it should be noted

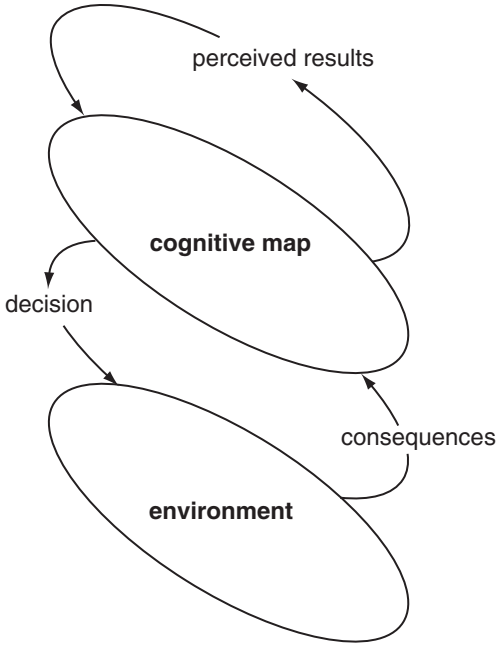

Fig. 1.4 The self-referring structure of decision-making (Handlbauer, 1997)

[91] Cf. Handlbauer (1997, p. 764).

[92] From Handlbauer (1997, p. 764).

that plausibility matters much more than accuracy.[93] Among other things, this can be related to basic findings on cognitive dissonance,[94] for example, individuals are prone to seek information that will reduce dissonance and avoid information that will increase dissonance.[95]

Heterogeneity of behaviour can easily be observed in real life situations and reflects heterogeneity of mental maps and constructs. Such heterogeneity stems from partiality of human cognition,[96] since a purpose oriented individual,[97] who relies on given, limited cognitive heuristics and capabilities, cannot but simply strives for what she perceives to be necessary to manage her usual environment. So, contingency as well as past experiences determine path-dependency and possibility of cognitive structures[98] and consequently of behavioural patterns.

In order to find a viable course, individuals involved in interactive situations have to make assumptions on the others and on their behaviour. This implies, in terms of cognitive maps, that individuals have to add assumptions and beliefs to their cognitive patterns about the others, about their cognitive maps and their behaviour. Such assumptions rely both on the interpretation of behaviours observed and on introspection, since individuals are not able to think the way others do. Hereby, considering the social position and the role of the interacting individuals may matter, the structure of sense-making is also based on social context.[99]

Direct communication is in principle never possible because separated cognitive entities need a means of establishing communication among themselves. Mediate communication is a construct which suffers from the limits of subjective understanding. This is because the communicated message has to be filtered by subjective cognition and integrated in one's own cognitive maps. Therefore, the observation of the others' behaviour and introspection are the key mechanisms for inserting beliefs about the others in an individual's own map of thoughts.

Even for collective decision-making, continuous proof of viability of the cognitive maps regarding both an individual's own and others' cognitive structures. When such constructs are perceived to be viable, a cycle of mutual reinforcement may arise.[100] The question if such empowerment leads collective maps and shared cognition to develop remains unanswered. Even if the role of intensive exchange of information in promoting general acceptance of cognitive constructs can be empirically confirmed, cognitive structures remain fundamentally affected by individually specific autopoietic cognitive processing.

In observation of social interaction it has emerged that there are collective accepted behavioural patterns which rely on coherency between individual mental

[93] Cf. Weick (1995).

[94] See e.g. Festinger (1957).

[95] Cf. e.g. Frey (1986).

[96] Cf. e.g. Tamborini (1997).

[97] Constructivism interprets individuals as purpose oriented. For more see Sect. 3.1.2 of Chap. 3.

[98] The possibility of the cognitive structure indicates the possible coexistence of heterogeneous cognitive maps.

[99] Cf. Luhmann (1984, p. 580).

[100] Cf. Handlbauer (1997, p. 766).

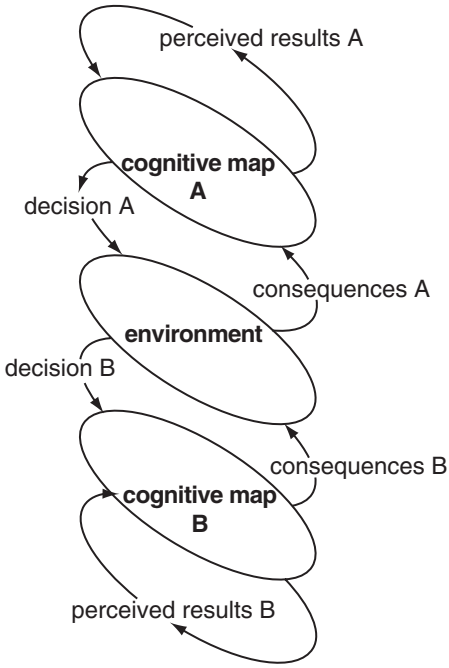

Fig. 1.5 Simple structure of individual decision-making in a social situation (Handlbauer, 1997)

maps regarding the individual herself and those regarding the others (Fig. 1.5). Common acceptance of mental patterns may also work in a self-enforcing way, as sometimes acceptance can make constructs viable independently of their intrinsic value.

Individuals belonging to a social group or community may benefit in several ways from the shared acceptance of mental patterns. The benefits include an increase in the predictability of a social situation, a simplification of individual decision-making in the specific social context and a reduction in the intensity of cognitive effort necessary, which in turn makes partial knowledge suffice. Furthermore, social affiliation promotes the convergence of behavioural patterns by means of elimination of disrupting behaviours, and works in this sense in favour of the survival of the social group's identity.

Those considerations help in understanding the primary role of institutions in ruling social interactions. According to Scott, institutions consist "*of cognitive, normative and regulative structures and activities that provide stability and meaning to social behaviour*".[101] Institutions are mainly concerned with cognitive stabilisation among social groups. They create a loop in the individual decision-making which evaluates the viability of the cognitive maps in the institutionalised environment (cf. Fig. 1.6). Institutions shape the social environment and get perceived as a fixed part of it. They do not directly belong to individual cognitive maps but influence their validation because they work towards cognitive stabilisation by consolidation of socially viable constructs.

[101] Cf. Scott (1995, p. 33).

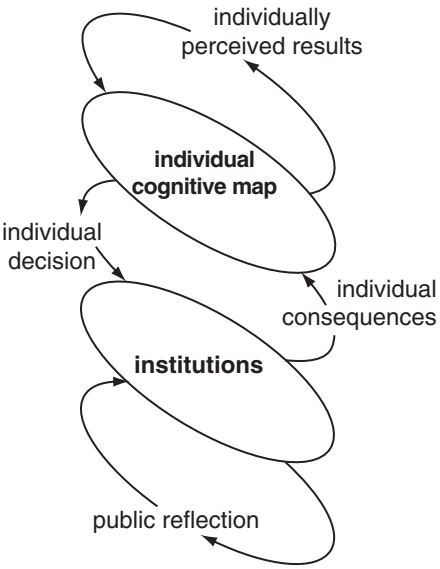

Fig. 1.6 Decision-making in institutionalised environment (Handlbauer, 1997)

Institutions reduce uncertainty and enable individuals to disregard several alternatives that could be potentially possible. Thus, individuals can concentrate on institutionally accepted cognitive patterns that become routinized through this process.

Institutions are at the same time a product of the social process and a constitutive part of the environment. They are typically perceived as fixed and durable. The objectivization of institutions acts in self-reinforcing way, inducing the routinization of cognitive processes and further encouraging behavioural standardization. Self-empowerment may however disrupt whenever viability of institutional constructs and responses get seriously questioned: this induces institutional change, which "*is supposed to occur if the individual cycle of perception and the public cycle of reflection can no longer be meaningfully connected*".[102]

[102] Cf. Handlbauer (1997, p. 769).

Chapter 2
Reflexivity of Social Reality

The previous chapter has underlined the polymorphism of self-reference and its involvement in any form of human understanding, activity and conceptualizing. As neither social reality, nor its observation or description can be abstracted from their self-referential character, reflexivity and its implications are of central concern for social research in general. Accordingly, this chapter focuses on the reflexivity of social reality and phenomena and discusses some of its implications.

In particular, two different orders of reflexivity can affect social reality: a first order of reflexivity involves social reality per se and consists of the social phenomena that are self-referential in that they may affect themselves, as they can for example imply, control, or modify their own dynamic or development. A second order of reflexivity concerns the "discourses" on social reality, such as social sciences and theorizing. The present chapter offers an overview on common reflexive social phenomena, while the Chap. 3 focuses on the second order of reflexivity concerning social sciences and theories.

It can be essentially premised, that the first order of reflexivity, which invests social reality, depends on the autopoietic character of social phenomena that create themselves on the basis of their inescapable systemic character. The second order of reflexivity can be fully appreciated if related to the constructivist observer-observation scheme, as it will be discussed in more detail in the Chap. 3.

One of the first difficulties that the analysis of the reflexivity of social reality posits is represented by the huge range of phenomena which can be subsumed under the label "social reality." Therefore, the analysis will begin with specifying the notion of "social reality," which will be conceptualized in opposition to "natural reality." After that, some notes on the reflexivity of social reality will introduce an overview on some common and widely analysed reflexive social phenomena. This overview does not have the claim of being comprehensive and ranges from anthropology, linguistics, law, politics, sociology, and psychology, and then concludes with examples of reflexivity involving the economic reality, as illustrated by the dynamic of financial markets and of the business cycle.

This overview is conceived as a natural first step for the successive enlargement (in the Chap. 3) to the reflexivity affecting social research and theorizing.

S. Sandri, *Reflexivity in Economics*: *An Experimental Examination on the Self-Referentiality of Economic Theories*,

The difference between reflexive social phenomena and social discourses that lead to reflexive social phenomena, which has been condensed in the notions of first- and second-order reflexivity, is not so clear and sharp as one could think at a first glance. Social reality is in the end by human activity, which is coined by human thought and consciousness, so that the separation between action and cognition, between practical and intellectual activity, is something fluent, a continuum rather than a dichotomy. Reflexivity of social reality is due to the fact that the intentional action of the individuals defines the course of the social system the individuals belong to.

2.1 What is Social Reality?

Social reality encompasses the human aspects of the world and it is constituted by tenets, beliefs, principles and opinions which may inspire the behaviour of a community. Social reality can be better defined negatively, i.e. by distinguishing it from everything that does not belong to its realm, though from the natural reality.

In the tradition of realism, objective facts can be divided in two categories: natural and social facts. The first exists and follows its course, independent of human perception, thus independent of what individuals may think about it, while the latter due to its relation to human thought, essentially depends on its perception and conceptualization.[1] Natural facts, which Searle (1995) expressively calls "brute facts," remain the same whether individuals realise them or not, such as whether mankind exists or not. They have a physical presence that, at least for the supporters of realism, cannot be doubted. The discussion of the eventual ontological priors of natural facts and if the notion of "objective facts" is tenable at all has been and still is vividly debated. However, such a discussion goes beyond the purposes of the present analysis, so that it will simply be omitted. A naïve interpretation of the concept "natural facts" will be adopted, as illustrated by the following example: A piece of paper has a physical presence, so that it can be depicted as a "brute," natural fact. However, independent of the meaning it might assume for a certain community, it may become a social fact as well. For example, it can be a money bill, if it represents a monetary value, but it can also be a property certificate, thus representing a property right. This points to the relation between social and natural facts: social and physical reality are not completely separated and independent from each other, but they overlap each other.[2] It can be argued that social facts can exist without natural facts. Social reality is embedded in physical reality although the borders among them are not always univocal. The questions of whether there can be social without natural facts and where the demarcation appears between the "brute" essence of reality and the perception by means of which it can be connoted, are raised. Critical examples in this sense are verbal agreements, ideas, copyright, etc.

[1] Cf. Searle (1995).

[2] Cf. Searle (1995, p. 35).

This short discussion can be summarized when social facts are considered a human product that are essentially in existence because of human construction. A social fact gets created when a community of individuals determines a function for a physical object or for certain circumstances, a community defined as that group of people who accept the assigned function and conform to it.[3]

2.2 Recursivity of the Social Reality

Recursive relations play an important role for social phenomena and are almost pervasive both of social reality and of the social sciences.

Reflexivity of social reality is due to the fact that individuals involved in a social system act intentionally, for example, they try to reach a certain end state from a given initial state, and, in the process, define the course of it. So, the individual mental representations and expectations play a decisive role in shaping the social reality. Examples of recursive phenomena can affect the different aspects of social reality and have always attracted the attention of social analysts.

A central notion for appreciating the essential reflexive character of social reality and recursive phenomena is that of "autopoiesis," the process whereby a system or organism produces itself, its own components and by means of which can distinguish itself from its environment. Autoiesis indicates in other words self-creation, i.e. a self-creative act or process and can be applied to the description of a social system. Social systems are essentially autopoietic, as they define their own identity and steadily reproduce it in their interaction with the external environment, in order to maintain the system's survival. The notion of "autopoiesis" was first introduced by Maturana and Varela (1973) as an attribute to describe the nature of living systems, e.g. the biological cell, that are able to produce all the components needed for the maintenenance of the living system out of an external flow of ressources.[4] The application of the concept of "autopoiesis" to social reality can be then traced back to Luhmann's work on systemic theory,[5] according to which social systems are autopoietically closed, because they rely on a flow of resources from their environment in order to continue maintaining their specific identity and to differentiate from the environment. The self-referring nature of any social system can be related to its autopoiesis, as the survival and maintenance of social facts relies on the infinitum regressum of filtering and processing information from the environment.

---

[3] Cf. Searle (1995).

[4] An autopoietic machine is defined as "*a machine organized (defined as a unity) as a network of processes of production (transformation and destruction) of components which: (i) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and (ii) constitute it (the machine) as a concrete unity in space in which they (the components) exist by specifying the topological domain of its realization as such a network.*" Cf. Maturana and Varela (1980, p. 78).

[5] Cf. e.g. Luhmann (1984).

Some elements which corroborate this thesis will be provided by the overview of reflexive social phenomena which follows.

2.2.1 Reflexivity in Anthropology

The thesis which sustains that language coins and anchors thought is known as Sapir-Whorf hypothesis, from the name of the researchers who provided it with robust scientific fundaments.[6] In its core argument the Sapir-Whorf hypothesis refuses the view that language merely mirrors culture and habits and argues that the relationship between language and thought is one of mutual influence. In other words, the characteristics and grammatical structures of a certain language influence and shape the understanding and behaving of its speaking community. This implies in particular that the mother tongue of an individual has a decisive impact on her way of thinking and processing information.

Speculations on the reflexive relation between language and thought can be traced back to vivid debates among Indian linguists up to the sixth century AD. In Europe, one of the first research studies that contributes to the reciprocity of relation between language and thought can be ascribed to Humboldt.[7] Boas, commonly recognised as the founder of anthropology in the United States, studied some of the languages of Native Americans and found that they often belong to different linguistic families. Sapir carried on Boas' researches and observed how different languages could give life to different habits and behaviour. Stating that human beings "*are very much at the mercy of the particular language which has become the medium of expression for their society*"[8] Sapir stressed the importance of social constructs, in general, and of linguistic structures, in particular, in shaping the reality in which humans act and interact. Language conventions and habits filter and orient the interpretation of external reality and therefore influence individual perception and experience of it. This goes so far, that "*the 'real world' is to a large extent unconsciously built upon the language habits of a group. No two languages are ever sufficiently similar to be considered as representing the same social reality. The worlds in which different societies live are distinct worlds, not merely the same world with different labels attached*."[9]

Whorf structured those ideas further, arguing that: "*We dissect nature along lines laid down by our native languages. The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face; on the contrary, the world is presented in a kaleidoscopic flux of impressions which has to be organized by our minds - and this means largely by the linguistic systems in our minds. We cut nature up, organize it into concepts, and*

[6] Cf. Sapir and Mandelbaum (1986) and Whorf and Carroll (1964).

[7] It is here referred to von Humbolt's essay "Über das vergleichende Sprachstudium in Beziehung auf die verschiedenen Epochen der Sprachentwicklung." (Von Humboldt, 1945).

[8] Cf. Sapir (1929, p. 69).

[9] Cf. Sapir (1929, p. 69).

ascribe significances as we do, largely because we are parties to an agreement to organize it in this way - an agreement that holds throughout our speech community and is codified in the patterns of our language … all observers are not led by the same physical evidence to the same picture of the universe, unless their linguistic backgrounds are similar, or can in some way be calibrated."[10]

Whorf's principle of linguistic relativity criticizes the interpretation of thought as unilaterally influencing language, that reduces language to a means of expression of what is already coherently formulated as thought. In spite of that, Whorf's analysis points at the repercussions that the different grammatical structures of different languages may have on the mental constructs of a speaking community.

For example, one of Whorf's most famous research[11] focused on the differences between Standard Average European's and Hopi's linguistic structures. It emerged that while the first linguistic structures tend to analyse space and time in a spatial static sense, the latter have a more dynamic conception than that, which is based on processes rather than on points. Whorf argues that this could be responsible for different understanding of mathematics, of spatial metaphors etc.

Among recent studies dealing with linguistic differences among populations, which provide evidence for the linguistic relativity hypothesis, Gordon's 2004s research can be mentioned. This research was conduced on a Brazilian tribe, whose language only contemplates three counting words, namely one, two and many. The complete inability of the tribe's members to learn how to count has been ascribed to this peculiarity of the tribe's language and interpreted as a sign of the shaping role the social construct language may have on the community which originated it.

The Sapir-Whorf hypothesis implies in particular that although language as a social fact is a construction of the human thinking activity, it may have an influence on human thought. This indicates that a circular relation between language and thought establishes and hints at the recursive effects between the two entities. Language and thought can be therefore interpreted as staying in a self-supporting stabilizing reference relation, linguistic determinism being its product.

Reflexivity among language and thought has often tickled human fantasy, as shown by the conspicuous fictional presence of linguistic determinism. In Orwell's "1984," for example, language is interpreted as a means for pursuing political totalitarian aims. In Orwell's novel, the idea underlying the formulation of "Newspeak" is that by abolishing words such as "freedom," people will no longer strive for it, or by not knowing the meaning of "revolution" they will never rebel.

As the existence of a social fact depends on the assignment of a certain function and meaning to certain physical objects or situations (a natural fact) as well as on its acceptance among a community social group, relations of mutual maintenance and influence can often occur in the realm of social reality, as the proceed of the analysis illustrates.

[10] Cf. Whorf and Carroll (1964, pp. 212–214).

[11] This research was mostly conduced during the 1930s, cf. Whorf and Carroll (1964).

2.2.2 Reflexivity in Linguistics

Reflexivity of the natural language constitutes an interesting case of study and is worth careful analysis, since it can concern two different levels of the semantic allowance of self-references and of the recursive dynamics of linguistic changes and maintenance. The first level has been already addressed in 1.5, which focused on the terms and the constructions through which natural languages admit self-referential devices and either tolerate the semantic they imply or not. The second level concerns autopoietic dynamics of language and thus addresses self-enforcing and self-defeating developments that determine the evolution of language. Languages are constructions of the human thought, but at the same time they can anchor or affect their own "creator," though they can recursively act on the human thought. The evolution of a language can be characterised as a circular bi-directional process, because language evolves as a result of modifications in its speakers' community but at the same time may induce some of the modifications that determine its evolution. The conception underlying the considerations on linguistic reflexivity that follows is that the evolution of a language comprehends both reflexive stabilization and subversion of norms and can be therefore characterised as the result of those two opposite circular processes.[12]

The establishment of a language as the means of communication of a certain community relies on self-supporting dynamics. This is because the acceptance of the semantics and of the norms of a language is the necessary prerequisite for a rough combination of sounds to become a word. Since the living norms of a linguistic community are defined by the convergence of social practices, language norms represent at the same time cause and effect of this convergence process, which can then be said to be self-referential.

Conventional usage of language is based on the self-enforcing compliance of the members of the speaking community. This is because the acceptance of a norm by the majority or by a conspicuous part of individuals belonging to the linguistic group of reference represents per se a reason not to deviate but to conform to that norm. Deviations tend to be sanctioned through misunderstanding and/or criticism, while compliance is rewarded by efficacy of communication, understanding, social acceptance and feeling of belonging to a certain speaking community. These mechanisms put the basis for the self-stabilization of language, which is a process that ensures the efficacy of language as a major communication means among individuals. Furthermore, this self-stabilization process stresses how once a linguistic norm is established, its survival is caused by the existence of the norm, whereas the existence of the norm decisively depends on its survival. In this spirit, the maintenance of linguistic norms can be at the same time alleged to the conventional usage of language and is motivated by such a conventional usage. Linguistic norms can be thus interpreted as being the creator and at the same time as being created by linguistic conventions and can be furthermore characterised as "*historical, or immanent, or constitutive a*

[12] Cf. Suber (1989).

posteriori."[13] In other words, the validity of certain linguistic norms refer to certain historical, immanent conditions and are neither eternally valid nor immune to changes and they constitute almost paradoxically a posteriori, because they "*structure not only what we approve but what we understand and how we act*."[14] Their "*normative priority to experience arises ('congeals') from the flux of history and passes away again*."[15]

The other dynamics which are involved in the evolution of a language involve the reflexive substitution of language norms and grants in this way for a reflexive dynamics of change.

There are many reasons for linguistic changes to take place and correspondingly many mechanisms through which linguistic changes may occur. Lending words from other languages (both foreign and technical), introducing new words to cope with new needs and circumstances, "*playfulness, imitation, laziness, ignorance [...]*"[16] are just few examples of the numerous mechanisms through which linguistic changes may happen.

In general, the processes by which linguistic norms get substituted, revised or commuted can be said to be based on a logics of "*amendment through violation*".[17]

Furthermore, the mechanisms through which new language norms are established, or through which old norms get changed, can be divided into phonetic and non-phonetic ones. While linguistic changes happening on a phonetic base can be lead back to "mispronunciation," the class of non-phonetic mechanisms of linguistic change is more articulated and illustrates clearly the role of errors and violation in the process of linguistic change.

In all languages it is possible to find cases in which systematical mispronunciation of words induces their spelling modification, so that a word substitution occurs. A mispronunciation becomes only then systematic if it does not get sanctioned, whereas sanctioning linguistic deviations has been discussed as an essential base for a language to gain the status of means of communication. It should be hereby premised that the notion of "mispronunciation" implicitly requires a standard of correctness. For a language such a standard consists in a family of acceptable pronunciations. In addition, the authoritative source for specifying the notion of acceptability among a certain speaking community is represented by its members, whereas no hierarchy among them can be in principle justified, if the conventional basis of language is accepted. This allows for the possibility of a norms violation, whenever the boundaries of mutual understanding do not get overwhelmed. Mutual understanding is a relative notion that can only be specified with reference to a particular speaking community, whereas speaking communities "*overlap each other, admit of innumerable borderline members, and may take very different shapes depending on*

[13] Cf. Suber (1989).

[14] Idem.

[15] Idem.

[16] Idem.

[17] Idem.

what case of usage we are investigating."[18] Official languages coexist with slang, dialects, technical languages, familiar lexicons etc.

Examples of language changes due to mispronunciation are the English words with Old French origin "naperon," which changed into "apron," or the consonant inversion between "s" and "p," which occurred in "waeps," which was the original word for the insect named "wasp." In several other cases letters have been added in order to simplify the pronunciation, for example "thunder" did not originally contain the "d," showing more clearly its relationship with its German equivalent "Donner."[19]

Non-phonetic changes can stem from popular but wrong etymologies, misleading backformation, according to which the prior between a verbal and a substantive form get exchanged and consequently some syllables get improperly added or omitted, or per analogy with other similar words.[20] The latter mechanism based on heuristic reasoning and analogical thinking and is nicely illustrated by the children's way of speaking. Children typically infer regularities in the language and they often apply them improperly, not being able to mind for exceptions and irregularities. Whenever a conspicuous number of members of a speaking community conform to the improper application of a rule such a violation may become more widely accepted, till it might amend the norm. For example, the past form of the English verb "to snow," which was originally irregularly built as "snew," got changed into the 'regular' "snowed," while the original regular past form of the "to strive" ("strived") progressively changed into "strove," which was most likely derived per analogy with other similar verbs.[21]

Finally, even the reflexive subversion process of a linguistic norm requires a reflexive stabilization process in order to establish ex proprio vigore, so that linguistic changes can be interpreted as emerging from the combination of loops working in opposite directions and respectively relying on a self-disrupting and self-enforcing dynamic.

It can be therefore concluded, that "*the mutability of language and its norms is a result of the balance of two reflexive processes. One is the self-reinforcing stability we called self-stabilization, and the other is the reciprocal causation and reflexive hierarchy we see in any norm of vulnerable to change from the posterior usage it structures*."[22] Those opposite processes balance, in the sense of leading to a stable development, essentially because of their different time horizons: while norms' constraint is a day to day process, norms' change typically takes years to occur, or it takes a period over which changes in the speaking community may become relevant.

---

[18] Cf. Suber (1989).

[19] The examples refer to Suber (1989), to which it can be forwarded for further details.

[20] Cf. Suber (1989).

[21] The examples refer to Suber (1989).

[22] Cf. Suber (1989).

2.2.3 Reflexivity in Law

Self-reference casts its shadows also in law and may work both stabilizing and disrupting factors into the development of a legal system. Paradoxes or infinite regresses might arise, when laws contain self-referential statements, which, as it will be discussed in this paragraph, cannot be totally avoided. Paradoxes and recursive dynamics would always have tremendous destabilizing effects if juridical praxis had not found in many cases pragmatic solutions and routine procedures for escaping the contradictions they in principle always imply. Still some self-referring instances or sentences can actually give life to logical inconsistency that puzzle not only logicians and philosophers but create shortcuts in the juridical system. Some other self-references might however even represent a useful way to avoid infinite regresses.

For example, the principle of validation of each law through its legitimation by prior or higher law should ensure the legitimacy of a legal system but implies an infinite regress.[23] To escape the shortcuts such an infinite regress would imply, juridical systems have to admit some laws to be valid ex proprio vigore. A process of validation by its own strength, by its own validation clearly implies a recursive dynamics. This common legal device illustrates a case in which a self-referring clause acts to correct another self-referring device.

Similarly, according to the so-called "bootstrap doctrine," courts must have some forms of self-applicable jurisdiction at their disposal, in order to be legitimate to rule some questions concerning their own jurisdiction without forwarding them to another court, which if put in the same situation would have to forward them to another one and so on.[24]

Self-legitimating laws and courts ruling on their own activity represent cases in which recursive circular structures are employed to interrupt different potentially destabilizing loops and to grant the smooth functioning of a juridical system.

As the rule, however, legal circles are more likely to create problems of ungroundedness, instead of solving them. A "renvoi" occurs if a court has to consider and to deliberate on the law of another state.[25] Of course, even if in principle a renvoi or every other circular remand cause a loop that cannot be escaped or solved, juridical systems are usually able to find a pragmatic solution to avoid jurisdictional stagnation. Perfect legal norms and perfect contracts cannot exist, since it is not possible or at least not economically sustainable to foresee all possible future states and developments. Therefore, legal systems are always provided with residual clauses that enable them to solve conflicts in a finite time and with a reasonable disposal of resources.

---

[23] Cf. Suber (1990).

[24] Cf. Suber (1990).

[25] This could happen for example when a contract that has been stipulated in a certain state gets violated by one of the parts involved in another state. There are cases in which, despite the contract explicitly states that such controversies are competence of the state in which the violation takes place, the laws of that state want such controversies to be solved according to the laws of the countries in which the contract has been stipulated. For further examples see Suber (1990).

Another formally controversial matter in law which is based on a reflexive mechanism is the constitutional amendment. The concept of "amendment" is used to depict the alteration of a law, whereas legal systems with "rigid" or "entrenched" constitutions rule the amentment of their constitutional law through a specific procedure, which differs from the one adopted for the other laws.

An amendment needs an amending clause to be formally admitted by a system. Because of its universality, any amending clause is feasible due to its recursive application. Therefore, self-amendment is in principle always possible. This resembles the so-called "paradox of omnipotence:"[26] a deity to be as such has to be omnipotent but can she create a so heavy stone she cannot lift? It can be similarly argued to what extent an amending clause can amend itself.

Self-amendment of constitutional norms stems from the need to balance between the two opposite instances of the stability of a legal system as a product of its formal consistency and the possibility of its modification., on the one hand to overwhelm shortages that may follow from the incompleteness of law and on the other hand to cope with new necessities.

2.2.4 Reflexivity in Politics

The logic underlying amending clauses in laws reflects the necessity of limiting power in modern political systems. Political power derives from sovereignty and is exercised either directly or through delegating representatives by the sovereignty holders. Every modern political system has to rule both the self-limitation and the self-augmentation of the political power. This involves the definition of limitations the political power has to underline, both concerning the time (e.g. time for the representatives to govern) and the content of its exertion (e.g. inviolable rights).

The political course may be affected both by stabilizing and disrupting recursive dynamics. For example, a shared fundamental ideological background underlies each political system and creates a self-validating system of beliefs that works self-reinforcing and self-isolating.[27] This shared ideological background - exemplified by national symbols, flag and patriotic creed - increases cohesion and sense of belonging among the members of the political system and promotes the acceptance and maintenance of its institutions. Such an ideology is an expression of a certain political system and affects at the same time the system's dynamics in a stabilizing way. Similar considerations obviously apply to every kind of ideology, so that ideologies can be interpreted as self-enforcing systems of beliefs.[28] Ideologies can also work in a destabilizing way on a political system, leading to structural changes that might even assume a revolutionary connotation. This indicates that revolutions can

[26] Cf. Suber (1990).

[27] Cf. Bartlett (1989, p. 15).

[28] Similarly, also religious belief-systems can be said to be self-referential. Cf. Bartlett (1989, p. 15).

also be interpreted as reflexive phenomena, in concomitance with whom the internal dynamics of a political system becomes self-destructive.[29]

The legitimacy of a political system and of its institutions is indissolubly tied to its perception among the members of that political system. The system's stability as well as the stability of its institutions may also depend on the coherence between reality and its perception, so that a significant gap between reality and perception might erode legitimacy of institutional arrangements. This implies that in certain situations, even without objective relevant social changes, the perception of certain institutional arrangements as inadequate might, in itself, spread among the members of the political system and call for their modification.

From the discussion of some of the reflexive phenomena affecting politics it emerges again how self-referential phenomena involving positive feedback loops work in a stabilizing self-reinforcing way, allows for continuity and does not introduce contradictory elements, whilst self-referential phenomena which rely on negative feedback loops have a destabilizing character and challenge logical thinking.

The balancing between positive and negative self-referring dynamics represents a particularly critical issue in the perspective of a political system. Similarly to the processes of linguistic change, subversive political dynamics owes its survival to a successive stabilization process, whose regulative efficacy depends on the one hand on the degree to which it self-enforces and on the other hand to the disruptiveness of the initial negative loop. Even if gradual changes can also lead to deep modifications in a political system, more radical changes can be needed in some cases, especially when it comes to subverting the balance of power or the established pecking order, as illustrated by the famous quote "*If you strike at a king, you better kill him*."[30]

2.2.5 Reflexivity in Sociology

Within the social sciences, sociology is maybe the only one that directly attempts to confront the problems of reflexive social phenomena. This is because sociologists are conscious that social phenomena might occur simply because of its common acceptance among a certain community. The existence of social facts is tied with their perception among the social actors and socially relevant developments which might decisively depend on the degree with which they are thought to be true and accepted.

The difference between the two orders of reflexivity that can affect social reality becomes particularly slight if applied to the realm of sociological analysis. Sociology deals with all social phenomena, with discourses about social reality included. Sociology focuses both on social facts and on the opinions and beliefs the social

[29] Cf. Bartlett (1989, p. 15).

[30] Cf. Holmes (1980), who ascribes this sentence to the philosopher, essayst and poet Ralph Waldo Emerson.

actors hold about them. The constitutive role of opinions and beliefs in shaping social reality represents a core topic in sociology. For example, the research program of the sociology of scientific knowledge aims at the analysis of social influences on science and addresses explicitly reflexivity in sociology. To simplify things, one main thesis underlying sociology of scientific knowledge is that social factors play an active role in shaping the development of science.

Plenty of examples of self-fulfilling as well as of self-destroying dynamics can be mentioned. A published prediction may affect the predicted event and either work toward the self-fulfilment or the self-destruction of the prediction. The disclosure of a public opinion survey or the publication of scientific results can purposefully strive for certain reactions among the public.

While reflexive predictions will be more specifically approached in the next chapter (par. 3.2.3), the reflexive effects of beliefs and opinions will be here illustrated by the so-called "Pygmalion effect."

The Pygmalion or Rosenthal effect refers to situations in which pupils, who are expected to perform better than others, will indeed perform better. This effect was first examined in a study by Rosenthal and Jacobson (1968, 1992), in which some teachers were misleadingly told that some children had a higher-than-average IQ. It was shown that the expectations of the teachers led the children to an actual enhancement of their performance. The name refers to Ovid's tale of the sculptor Pygmalion, who created a statue of perfect beauty and fell in love with his own creature.

As said, the thematic of reflexive predictions will be more specifically approached in the next chapter (par. 3.2.3).

2.2.6 Reflexivity in Psychology

Psychological states of individuals can have a decisive impact on reality, as different beliefs or attitudes might ceteris paribus lead to the realisation of opposite developments and produce alternative outcomes.

For example, the cognitive bias in which a researcher may occur, if she expects certain results can be mentioned: the researcher could either unconsciously manipulate the research method and the experiment or misinterpret the data and therefore actually obtain the expected results. This bias is known as the "observer-expectancy effect."[31]

Rosenthal (1998) provided evidence that an experimenter can subtly and unintentionally communicate the experimental subjects' own expectations and biases through a conspicuous number of cues or signals. These cues can significantly affect the outcome of the experiment and artificially create the conditions for expected developments to occur.[32]

[31] This effect is also known in the literature as "experimenter-expectancy effect," "experimenter effect," or "observer effect."

[32] Cf. Rosenthal (1998).

This effect is a clear example of how even science, despite all methodological premises that should grant for the objectivity of enquiry and results, cannot transcend its own dimension of human enterprise and is therefore "held in check" by the consciousness of the human mind. In other words, the reflexivity of human understanding implies the reflexivity of all its elaborations. Scientific knowledge makes no exception.

A possibility of debiasing the observer-expectancy effect is to rely on a double-blind methodology. This consists in concealing both experimenters and subjects from which subjects are assigned to which group (control and test group) until the end of the study. This procedure can best be applied in computerised experiments but also presents the disadvantage of being quite costly.

Also the participants to an experiment or to a research study can manipulate the results of the analysis. In similar cases it can be spoken of "subject-expectancy effect." Plenty of evidence of this effect can be found in psychotherapy and medicine. The effects of the subjects' expectancy on the efficacy of a certain cure can even imply healing processes to be accelerated or to fail. That the patient's symptoms can be alleviated, just because of the belief in the efficacy of an otherwise ineffective treatment, is the well-known placebo effect. The opposite effect, the so called "nocebo effect," can occur as well, if a patient disbelieves an effective treatment.

The usual procedure to prevent biased behaviours of the experimental subjects consists in running single-blind trials, i.e. not to reveal to the subjects whether they belong to a test or to a control group.

When the experimental results confirm the experimental hypothesis, but not for the expected reasons it can be spoken of Hawthorne effect. The assumption underlying this effect is that the results simply occur because of the subjects' awareness of participating in an experiment, with no other plausible explanation. The Hawthorne effect is thus a reaction of the experimental subjects to the fact that they are objects of analysis. It therefore represents a case in which the act of observing induces a modification in the observed entity that would not have taken place without the observing act and provides evidence for a recursive causal relation between the act and the object of an observation to establish.

The Hawthorne effect first emerged from investigations on the influence of different work environment characteristics on productivity, which were conducted between 1924 and 1932 at the Hawthorne work of the Western Electric Company in Chicago.[33] It could be observed, that the modification of different factors of the work environment, e.g. pay, light levels, rest breaks etc., all induced an increase in productivity, independently on the direction of such changes (increase or decrease). It was also found that even the return to original conditions had a positive effect on the productivity and therefore any plausible explanations could be formulated except that the awareness of being an object of a study per se motivated the experimental subjects to modify their behaviour.

[33] These studies were conducted by Fritz J. Roethlisberger and William J. Dickson.

2.2.7 Reflexivity of Economic Reality

Reflexive dynamics can also affect the course of the economy, as the actors interacting in an economic system act purposefully and can on this basis have a part in determining economic events or developments. It can be thought of several examples of self-fulfilling as well as of self-destroying dynamics concerning economic systems and interactions.

For example, the disclosure of a market research can influence both investment and consumer behaviour. It can assume a different predictive value independent of its dissemination status and of how trustworthy it is perceived to be among the economic actors involved.

Similarly, the German Federal High Court recently had to decide on the responsibility of the Deutsche Bank for the bankruptcy of Kirch's corporation; Kirch accused the bank of having caused its bankruptcy by publicly doubting its creditworthiness.

It can be said, that the expectations of the market's participants can have a decisive influence on the market's performance. Evidence for similar dynamics can be observed in many contexts. Examples can be found that inflationary and deflationary developments often show a self-fueling character, and that the compounding of reinvestment can influence economic growth. Reflexive dynamics can also be exploited as a way of achieving economic goals, as illustrated e.g. by self-investment as possible management strategic device. They have also been contemplated to different extents from a theoretical point of view, as attested e.g. by game theoretical prescriptions that can reveal both a self-undermining and self-guaranteeing character.

2.2.7.1 Reflexivity of Finance

It is unquestionable, that the development of the financial markets is influenced and mirrors the historical course of the events: political stability or shortcuts, profound social changes as migrations, conflicts etc. inevitably reflect on the dynamics of financial markets and decisively coin their functioning. However, because the markets are a product of the interaction of purpose oriented actors, they can sometimes anticipate real developments and therefore enforce or accelerate their occurrence. Financial interactions are social facts that are embedded in the social reality, which means that mutual influences between them and political, social as well as cultural events and developments can never be excluded. This implies that causal relations between financial interaction and other social facts have to be interpreted as a circular recursive process.

Soros' (1994) conception of the financial markets and their functioning focuses exactly on this point, namely on the relation of mutual influence in which financial markets stand with the social system they are embedded in. This has been quite frequently discussed in financial literature. It is commonly known and accepted both among managers and operators and among researchers and analysts[34] that manag-

[34] Cf. Copeland, Koller, and Murin (2000, p. 52).

ing expectations of markets' participants can have a decisive influence on market's performance.

This is also illustrated by the three forces that are typically considered to govern the market's course. These forces are the fundamentals (revenue, cost, capital etc.), the technicals (price and volume data) and the market behaviour, whose dynamic is not easy to predict.

Operating in a financial market requires thinking about the system and the events one is participating in. The thinking contributes to shape the object of thought, which is social reality "*not independently given, but contingent upon our decisions*."[35] As thought constitutes an element, a part of the social reality, it should be considered in examining social dynamics. Since "*our thinking forms part of the reality we think about, the separation between thinking and reality is breached. Instead of a one-way correspondence between statements and facts there is a two-way connection*."[36] In this spirit Soros criticizes the view that the market should always be right. In Soros' eyes the market can only be wrong, as it is always, inevitably biased. The market's distorted view of future events works in two directions. First, market participants rely on such a distorted view. Second, this view can influence its future developments. The impression that markets predict the truth can actually be inferred. But only because of this bidirectional influence does the impression appear to be sustained by evidence. This relation of mutual influence implies for the one side, that the view of the market's participants contributes to shape the course of the market, while from the other side, concurrently the course of the market enforces or corrects the participants' view.

Soros (1994) applies the concept of "reflexivity" to the description of this mutual relation of causality between the course of the market and its participants' expectations. In Soros' view the application of the methodology of the natural to the social sciences is condemned and its failure has to be ascribed to this reflexivity. As economics is the social discipline with the highest ambition of resembling formalisms, axiomatic construction and method of the natural sciences, the effects of reflexivity might have devastating consequences on its scientific validity. This is particularly evident if one considers the notions of rationality and perfect understanding, which are assumed to inspire the behaviour of the economic actors. These assumptions are clearly inadequate to stylize the real individual cognitive capabilities that can actually be observed. Imperfect understanding, limited processing and computational capabilities seem to inform the behaviour of individuals who deal with economic problems and interactions.

Financial markets are always biased, which is confirmed by the continuous fluctuations that steadily perturb their long term dynamics. The force acting beyond such perturbations is, in Soros (1994) view, mainly constituted by the participants' perception or preconception about the dynamics of the market.

The reflexive relation, which links the thought of the actors involved to the situation they are part of, can be expressed by means of a system of recursive functions

[35] Cf. Soros (1994).

[36] Cf. Soros (1994).

consisting in a "cognitive" or "passive" function, $y = f(x)$, and a "participative" or "active" function, $x = \phi(y)$.

The cognitive function expresses the participants' perception and opinion on the market, for example, y, as a function of the independent variable, x, which summarizes the main features of the market, i.e. the trends which underlie the market. In this way, the function y points out that the markets' participants form their beliefs and opinions on the dynamic of the market on the basis of the observation of the market. The beliefs as a result of the thinking activity of the individuals will then reflect on the dynamic of the market. Therefore, participants' beliefs constitute the argument of the participative function, $x = \phi(y)$, which thus catches their influence on the course of the market.

Such a recursive system of functions ensures that changes in the cognitive function are captured by the participative function, which describes the course the events will take. Let $y = f(x)$ be a "cognitive" or "passive" function, which expresses individual beliefs about the situation that they are confronted with and $x = \phi(y)$ a "participative" or "active" function, which catches the repercussions of such beliefs on real developments.

The reflexivity of the system constituted by $y = f(x)$ and $x = \phi(y)$ can be appreciated by substituting the arguments of the two functions with each other, so that both $y = f[\phi(y)]$ and $x = \phi[f(x)]$ become a function of themselves. This is illustrated by Fig. 2.1, which points out that the market's underlying trends, x, influence the participants' beliefs, y, as expressed by the cognitive function $y = f(x)$, while at same time the participants' beliefs, y, affect the market's trends, x, as captured by the participative function $x = \phi(y)$.

This reflexive functional system can either develop toward the equilibrium or generate changes and invert trends. While pre-programmed routinized cognitive patterns and procedures can be expected to inform the decisions of the market's participants in usual settings, historical events are most likely to induce significant changes in the participants' beliefs. Therefore, the participants will act in a stabilizing way toward the market equilibrium when they perceive the course of the market to be normal. They will react and contribute to the deviation from the equilibrium,

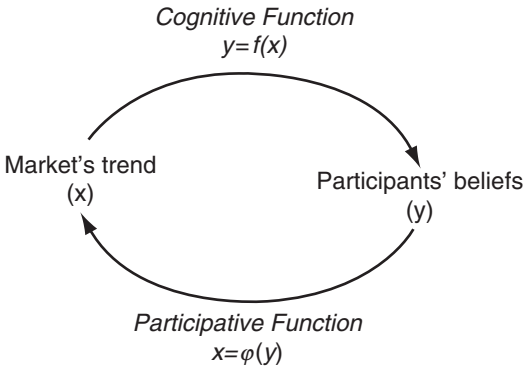

Fig. 2.1 Reflexive interaction between cognitive and participative function (author's representation)

whenever they perceive the market to be challenged by unusual developments. This yields for the coexistence of stabilizing and destabilizing reflexive perturbations, the combination of which inform the course of the social reality.

The stock market represents an ideal setting for analysing reflexivity and appreciating its biasing influence. This is because the stock market resembles the basic feature of a market with perfect competition: it is a centralized marketplace, with a high number of buyers who do not make the price and do not have a direct immediate control on it. In addition, in the stock market entrance barriers are practically absent. Homogeneous goods are traded with very low transaction costs, almost in absence of transport costs. Furthermore, brokers' reports provide a plausible basis for inferring something about the beliefs that are held by the market's participants. These reports can convey a picture of largely shared opinions about the market in a certain moment. However, it is, in principle, impossible to understand exactly which beliefs and opinions are mostly shared by the market's participants. It is even more difficult to infer such beliefs ex post on the basis of the observation of how the stock market has developed. It can be inferred that positive optimistic beliefs must have been shared by the majority of the markets' participants, if ceteris paribus stock prices grew. On the contrary, the fact that ceteris paribus stock prices failed can be ascribed to negative pessimistic beliefs and expectations. However, even the specification of the ceteris paribus condition is only possible on the basis of contrafactive speculations and therefore, different ways to infer the participants' beliefs have to be explored.

The course of the stock market is illustrated by the development of the stock prices. This is again shaped by an "objective" and a "subjective" component. While the first component is represented by the development of fundamental variables of the market, the latter reflects the prevailing beliefs of the market's participants. In this perspective, the dynamics of the stock market can be simplified to the product of interaction between the trends in the fundamentals and the prevailing opinions and beliefs that are held by the market's participants. The development of the stock prices - from boom to collapse – can be related to the interaction between the fundamentals trend and the prevailing market participants' opinion. In this way, the reflexivity that influences the market will be closer examined.[37]

Reflexivity can be extrapolated from the comparison between the development of the market's and of the fundamentals' trends. Reflexivity can be interpreted to work in a stabilizing way whenever those trends reveal similar dynamics. Reflexivity can be alleged to have biased the course of the market, whenever there is a discrepancy between the market's underlying trend and the trends in the fundamental.

Under the label "fundamentals" are summarized all those variables which may affect the market. Among them are e.g. dividends, patrimonial values, cash flow and earnings. For simplicity, in the following model of the stock market, the earnings per share have been chosen as representative for the fundamentals because they can be interpreted as a variable which results from the market's development.

[37] Cf. Soros (1994).

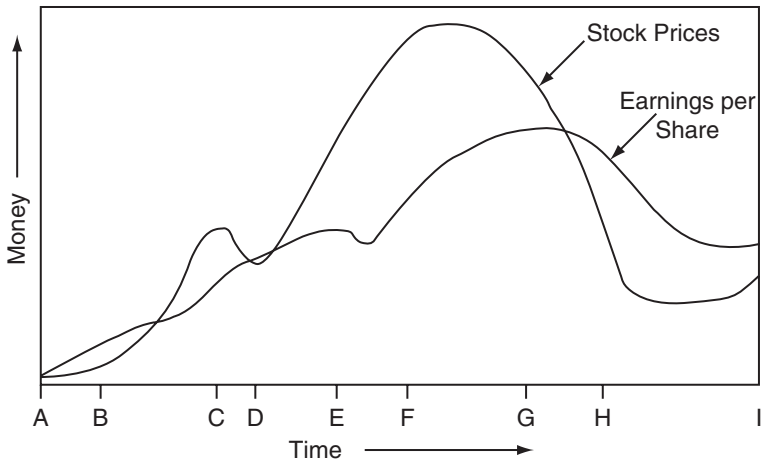

Fig. 2.2 Reflexivity in the Stock Market (Soros, 1994)

Therefore, the stock market's development will be modelled (see Fig. 2.2) by observing the development of stock prices (as result of the market's course) and the development of earnings per share, which have been chosen as representatives for the fundamentals.

Typically, market cycle is articulated over the following steps. At the beginning (AB in Fig. 2.2) a certain trend positively affirms and affects the per share earnings. Most likely, market's participants hold spread beliefs about the market's future developments, so that no prevailing opinion can be outlined. As time passes, the market's participants gain awareness on the developing trend and enforce it through conform expectations (segment BC in Fig. 2.2). The trend further develops with a certain stability and is not decisively disturbed by slight transitory changes that can occur in the participants' expectations, as illustrated e.g. in CD. The stability and persistence of the market's trend finally breaks participants' last resistances and enforces trend conform expectations. Expectations remain optimistic and sustain the stock prices in spite of a deceleration in the earnings' trend (DE). In EF participants' optimistic expectations are the only force sustaining the stock prices and are no longer confirmed by reality, as it is illustrated by the inversion on the fundamentals' trend. In this insight, "*The essence of a speculative bubble is a sort of feedback, from price increases, to increased investor enthusiasm, to increased demand, and hence further price increases. The high demand for the asset is generated by the public memory of high past returns, and the optimism those high returns generate for the future. The feedback can amplify positive forces affecting the market, making the market reach higher levels than it would if it were responding only directly to these positive forces*."[38]

Eventually, the market's participants recognise their optimistic expectations to be erroneous. This then induces the prevailing opinion to change (FG). The stocks

[38] Cf. Schiller (2001, p. 3.)

loose their last support and their price fails, whereas both the expectations and the fundamentals' trend work in the same negative direction and amplify each other's effect (GH). It can be observed, that "*Just as the euphoria of a boom exacerbates investors' preference for making abnormally large returns, a bust exacerbated the despondency of suffering deep portfolio losses*."[39] Finally, as the pessimism of the markets' participants becomes too big, the market slowly stabilizes (HI).

The cycle that has been sketched is just one of the infinite possible configurations the dynamic of the stock market can assume. On its basis, the interaction between beliefs and expectations of the market's participants and the market's "objective" development can be observed. In this frame, reflexivity emerges as the residual force that affects the market's dynamics and can find expression in the difference between the fundamentals' trend and the market's course.

In "The Alchemy of Finance," Soros systematically applies this model of reflexivity to the analysis of several cases, e.g. to the currency market, to the Real Estate Investment Trusts, to the venture capital boom, and to the conglomerate boom.

The boom that many conglomerate companies experienced in the late 60s was based on a very simple mechanism, the rising of per share earnings which should have reflected the good management of a company. But in the end, realty was produced by the acquisition of companies with typically high dividends but low price-earnings ratio.

Because of the positive expectations of investors, the shifting of the conglomerate from high tech to consumer goods does not get reflected in a lower price-earnings ratio. The positive expectations motivate the investors to buy more stocks and this further artificially sustains their price.

In more detail, when a high tech company with high price-earnings ratio starts diversifying, it typically acquires consumer goods companies with high dividends but low price-earnings ratios. This makes the earnings of the conglomerate company rise and improves its market value per share. This further increases the borrowing possibilities of the conglomerate company for further expanding, so that further consumer goods companies can be acquired. The continuous growth of the per-share earnings of the expanded company will induce the investors to buy, attracted by the high price-earnings ratio. The investors' behaviour will further enforce such a blown up upward trend, which is only supported by the overall earnings ratio and does not reflect the fraction of the conglomerate company that operates in low price-earnings ratio businesses. When eventually the investors realize this, the stock price will fall to match the real characteristics of the company.

Finally, Soros' analysis is centred on the self-referential character of the financial markets. Their dynamics can therefore be described by a reflexivity theory, according to which the market process is determined by a two-way feedback loop between fundamental values, on the one hand, and subjective beliefs and estimations of the market's participants, on the other hand. Soros focuses on the interaction between cognition and action of actors involved, which yields for reflexivity.

[39] Cf. Calandro (2004, p. 53).

2.2.7.2 Reflexivity of Business Cycle in Politicised Markets

The dynamic of the business cycle is shaped by the interaction between the objective features of the economic system, i.e. its fundamentals, and their subjective perception by the participants in the economic system. This interaction informs market behaviour. In this perspective, the business cycle can be seen as steered by the adjustments between the fundamentals and the way they are subjectively perceived and processed by the market's participants.

Par. 2.2.7.1 sketched the dynamic of a typical boom-bust cycle on the stock market, which has been approximated as being substantially free from "external" interferences. Its course has been interpreted as mainly determined by its own fundamentals and the decisions of the people engaged in the exchange. However, the inference of politics on economics cannot be understated, as it creates the institutional frame in which economics develops.

Political intervention on economics pursues precise aims and interests and can be motivated both by regulative purposes and opportunistic instances, e.g. re-election. The reduction of the interest rate below its market-determined level is a frequent example of a political device which serves opportunistic concerns. This induces the market's participants to spend more which then leads in the short-term to a boom. The boom is blown up by the increased selling amount, which continues until the market's participants finally realize the discrepancy between the biased fundamentals and those that would reveal the real condition of the market.

The boom-bust model presented in par. 2.2.7.1 will be sketched again and discussed together with some of the typical political interventions that may occur in each of its phases. Hereby, the reflexive implications of economic policy making will be focussed, as they emerge when political intervention creates mechanisms that affect the political course either in a self-enforcing or in a self-defeating way.[40]

The market's participants initially need some time to recognize the trends in the fundamental. Since the electors typically consider a positive conjuncture to be favoured by the political action, a good trend in the fundamentals (as illustrated by segment AB in Fig. 2.2), this typically yields the politicians an increasing popularity. The politicians are as a rule reluctant to waste their potential of interventions in the very beginning of a boom, where positive further developments of the business cycle can be expected. In addition, it is in this initial phase that the central bank typically intervenes by diminishing the interest rate. However, as soon as the end of the legislature approaches, the politicians' incentives for adopting business cycle encouraging manoeuvres are high and therefore similar interventions (e.g. tax reductions) are most likely to be considered even in the initial phase of the business cycle.

Segment BC in Fig. 2.2 illustrates price appreciation. It drives market's participants' expectations to grow and the political leadership typically is enjoyed over a spreading consensus. Despite some perturbations (cf. CD in Fig. 2.2) the positive trend holds. The upward dynamics of the business cycle continues even when the

[40] Cf. the analysis of Calandro (2004).

fundamentals' trend inverts downward as due to the only support of the market's participants' unrevised positive expectations.

In this phase, politicians and institutions can be interested in artificially sustaining such positive expectations and could for this reason choose to affect some fundamental substitutes in order to influence the investment climate.

To do that it can sometimes be enough to encourage certain accounting practices to spread, as for example focussing on spurious gain indicators in order to induce an overestimation of the real profits.[41] Similar devices achieve the artificial sustainment of the cycle's upward dynamics, in that they induce the economic actors to "*replace fundamental analysis with fundamental substitute analysis that will support the boom*."[42]

In similar settings expansive manoeuvres of economic policy are likely to be undertaken, e.g. credit expansion. The boom could then be ascribed to wealth effects created on the basis of the new economic manoeuvres or even to new economic conditions,[43] as for example the debate on the new economy pointed out.

Finally, the market's participants recognise the prevailing negative course of the market. Their expectations reconcile with the fundamentals' trend and as a consequence the price fails uncurbed (see segment FG). Typically and in particular in those cases where the boom had been artificially amplified, politics should interfere in order to avoid a market crash.

A contraction in the central bank's monetary stance, e.g. an increase in the interest rate, then sets a bust in motion. As such a bust develops expectations, it becomes more negative. The market behaviour gets more and more biased toward what can be called "*irrational despondency*"[44] and this leads to massive disinvestment and mass selling (GH). Finally, the market stabilizes progressively, while prices and fundamentals adjust toward their pre-bubble levels (HI).

[41] Cf. Hayek (1970, p. 95).

[42] Cf. Calandro (2004, p. 52).

[43] Cf. Hazlitt (1996, p. 158).

[44] Cf. Calandro (2004, p. 62).

Chapter 3
Reflexivity and Predictability of the Social Sciences

Reflexivity is inevitably involved in all social facts and discourses about social reality because social discourses are inevitably made by an observer who is part of the system of observation. Discourses on social reality are elaborations that transcend meta-logically the reality they focus on, as they are formulated by a subject that cannot be disentangled from the observed object.

This chapter addresses the unavoidable reflexive nature which characterizes all kinds of discourses about social facts and reality. This can be ascribed to the involvement of the observer in the system she observes and constitutes an inescapable condition in which all human conceptualizations concerning social facts are caught. In set-theoretical terms, human discourse's self-referring character is given because its domain (the subject of the analysis) is embedded in the range to which it refers (the object of the analysis).[1]

Reflexivity of this kind, which is implied by the so-called "observer-observation problematic," constitutes a peculiarity of the social sciences and finds no real equivalent in the natural sciences. Its most similar problem is represented by the Heisenberg's uncertainty principle.[2] This can be traced back to the observer effect and ascribed to the interference of the act of analysing with the analysed object. Starting from these premises, constructivism extends the effects of reflexivity to any form of human analysis because the act of human observation inevitably concurs to shape the object to which it refers. Constructivism states that all that can be experienced or thought is "constructed," so that reflexivity constitutes an inescapable basis of all that can be thought and conceptualized. Reflexivity belongs therefore to one of the central concerns of the constructivist analysis and represents a key concept for deepening human cognition and behaviour.

Therefore, the first part of this chapter focuses on the perspective of radical constructivism. Its essential fundaments will be illustrated discussing the observer-observation scheme and the construct of the self. The specified constructivist perspective will then be adopted for discussing and modelling the cognitive processes.

[1] Cf. Sect. 3.2 of Chap. 1.

[2] This principle will be briefly addressed at p. 68.

S. Sandri, *Reflexivity in Economics*: *An Experimental Examination on the Self-Referentiality of Economic Theories*,

Some implications of reflexivity for social research and economics, as they emerge from the approach of constructivism, will then be addressed.

The second part of the chapter addresses the implications of reflexivity of the social sciences for the predictability of social and, in particular, economic reality. After an introduction on different approaches to social predicting, the processes of explaining and predicting social reality will be compared. The self-altering, reflexive effect of social predictions which will emerge in this insight will then be developed, and reflexive predictions will be discussed.

3.1 Constructs and Reality

On a toujours cherché des explications quand c'était des représentations qu'on pouvait seulement essayé d'inventer (Valéry)[3]

The concept of "constructivism" bears a plurality of different approaches which extend to several disciplines and fields of analysis. One of the first forms in which constructivism has been formulated is represented[4] by the sociological theory of social constructionism. This theory has been developed based on Hegelian ideas by Durkheim and then became prominent due to the work of Berger and Luckmann (1967).

As it is suggested by the word "constructivim," its different approaches all share the thesis that all that can be experienced or thought is "constructed" and does not necessarily reflect any external transcendent real facts.

The introduction to constructivism which follows refers to the radical constructivist tradition, whereas elements of operative[5] and of social constructionism[6] will be combined. In opposition to a realist stance, constructivist analysis shifts the attention from reality to its subjective perception and reveals therefore a pronounced epistemologic rathen than ontologic orientation. The main thesis on which constructivism is based is the unsondability of reality by human knowledge. Reality cannot be permeated by human knowledge, since it is based on mental representations, which are constructed out of subjective experience and therefore do not necessarily refer to ontologic priors. That individuals elaborate mental representations out of their subjective experience is corroborated by neurological evidence on the activity of the human brain.[7]

Radical constructivism fundamentally relies on the following three theorems, namely the "observer," the "construction," and the "validity theorem."[8]

The "observer theorem" states that knowledge is essentially a human construction, so that it does not make sense if disentangled from the human constructive

[3] We have always sought explanations when it was only representations that we could seek to invent (author's translation).

[4] For an introduction on costructivism, see e.g. Fischer (1995).

[5] Cf. e.g. Luhmann (1984) and Maturana and Varela (1987).

[6] Cf. e.g von Foerster and von Glasersfeld (1999) and von Glasersfeld (1995).

[7] Cf. e.g. Hoyenga and Hoyenga (1988) and Haberlandt (1998).

[8] Cf. Rusch (1999, p. 8 ff).

activity which generated it. As "*there is no knowledge without a knower*" and "*the knower personally participates in all acts of understanding*"[9] knowledge is conceived by the subject who elaborates it. It therefore constitutes an observation that as such inevitably reflects the subjectivity of its observer. As an observation is influenced by the subjective way of perceiving and processing the external stimuli, by past and contingent experiences of the observer, as well as by the boundaries of the subjective cognition, it cannot be appreciated without considering its observer.

The "construction theorem" posits that knowledge is only individually possible. This is because knowledge gets constructed by individual cognitive instruments and repertoire which are specific and highly subjective, so that it will reflect the subjectivity of the knower who carries its limitations and mirrors its experience.

The "validity theorem" underlines the relativity of knowledge and its subjectivity. Because of the highly subjective character of knowledge, as it is posited by the first two theorems, the realm of its validity is also limited to the the individuality of the knower, of the observer. Validation or invalidation of constructs can be interpreted only in a subject-related way, so that a construct can only be assessed as valid or not if related to the individual and to the concrete circumstances from which it originated.

These theorems enable the profile to have the specific epistemological stance of radical constructivim. It can be inferred that knowledge is considered to be articulated parallel to reality. It is only able to penetrate reality in an "operative" way and not in an ontologically consistent way. Constructivism thus redefines the role and meaning of knowledge, which lowers the epistemological capabilities of the individuals. In a constructivist perspective such capabilities no longer convey a consistent picture of reality. At most they enable the formulation of "operational" knowledge, i.e. of knowledge that aims at supporting human interaction with external reality. In other words, knowledge primarily reveals an operative nature because it enables the individuals to deal with the environment they are confronted with by means of defining cause-effects relationships that build the mental representations and models.

In a constructivist approach the essence of the facts and reality is relative to the act of observation on which their definition relies. In other words they do not exist, if not in the way they are processed by the mind of their observer. While the most radical constructivist positions refuse to believe in the existence of an objective reality, other currents have a more agnostic stance and merely question the permeability of reality by the human act of knowing.[10] Despite these differences, the centrality of the act of observation in defining reality brings together the different interpretations of constructivism. Observing reality is a creative act which implies a "facere," an activity which is constitutive of reality and creatively defines factuality.[11] Consequently, the concept of truth or at least its accessibility cannot be thought but as contingent to the subjective construction process.

---

[9] Cf. Kincheloe (1991, p. 26).

[10] For a further discussion see e.g. von Foerster and von Glasersfeld (1999) and Rusch (1999).

[11] Cf. Rusch (1999, p. 9).

Altogether the separation of knowledge and reality and the substitution of realistic stances of knowledge verification with more context dependent operative validity introduce elements of cognitive relativity. As truth cannot be defined in an absolute sense, there are no objective criteria for the validation of constructs. However, in a constructivist perspective the possibility of evaluating the resemblance between constructs and reality is ruled out. This means that there are no objective criteria to appreciate the degree to which constructs mirror reality and therefore to appoint a certain construct as "true" or "false." Instead, constructivism posits that constructs can be evaluated according to their capability of granting human survival, i.e. according to their "viability."[12]

The category of "viability" can be applied to constructs concerning both sensorial perception and intellectual elaborations. In particular, a construct concerning sensorial perception is said to be viable if it supports the individual in her (intentional) interaction with the external environment, granting therefore the individual's survival given the state of the environment the individual is confronted with. Constructs that regard intellectual elaborations, strive for gaining knowledge and involve speculation and abstraction, will be assessed as viable if what they conceive does not contradict other viable constructs.[13]

Viability can be therefore profiled as a dynamic criterion which neither relies on objective nor absolute standards. An operative definition of viability cannot be formulated because viability is a highly contingent notion. It is based on the contingency created by the concrete possibilities of actions of a certain individual, which are given by the actual states of the environment and the sensorial and cognitive equilibrium of that individual. For this reason, viability accounts for the provisory and individual validation of constructs. Constructs mirror their own genesis. They reflect the feature of the cognitive process from which they origin. They are the sensorial or intellectual product of a certain individual facing a specific situation in a certain time and as such they also reflect the individual experience and social history.[14]

The substitution of the objective truth-based validation criterion for constructs with the dynamic criterion of viability allows for the possible heterogeneity of constructs, whose implications will be further discussed at Sect. 3.1.2 of Chap. 3.

However, underlining the subjectivity of constructs and admitting their possible heterogeneity, does not lead to solipsismus. Human knowledge, although it is not permitted to penetrate reality, is vital to individuals because it supports their behaviour and enables their survival. Despite the subjectivity of constructs, their interpersonal validation of constructs can occur if they are accepted as viable among a certain community. The interpersonal validation of a construct is produced and enforced by the interaction between individualsand leads to the social construction

[12] The concept of "viability" assumes a central importance for the constructivist approach. Cf. e.g. von Glasersfeld (1995), Maturana and Varela (1987) and Rusch (1999).

[13] Cf. von Glasersfeld (1995, p. 122).

[14] More on viability will be discussed at Sect. 3.1.2 of Chap. 3.

of reality. Such a shared system of viable constructs creates a sort of "objectivity," i.e. allows for the individuals belonging to that community to be in cognitive equilibrium.[15]

3.1.1 Observer, Observation and the Construct of the Self

Individuals do not have a direct immediate access to reality. Their approach to it can only be filtered by their own perception, which informs their way of cognitively processing it. Therefore, cognition and knowledge do not represent acts toward the discovery of reality, but toward the construction of cognitive reality.

Individuals cannot know reality, but only the observation they make out of it. The differences between phenomena, objects or events, as they are perceived by an observer, only play in their own minds. Human constructs inevitably carry and reflect elements of the subjectivity of the individual who formulates them. Constructs are therefore relative and related to their subject and mirror the subject's biological determination, experience and social history.

It follows that observations cannot be abstracted from their observer, since they only exist and assume their sense of being within the observer's cognitive reality, in which the observer is embedded and to which she belongs as well.

Radical constructivism differentiates between the concepts of "internal" and "external observer:" while the first applies to the individual whenever involved in her own observation, the latter depicts the individual observing something external to her own cognitive reality.

It is interesting that even as an internal observer, the individual does not gain any privileged ontological stance. She can just process a construct of the "self," but cannot have any immediate knowledge of herself, i.e. not filtered through her own cognition.

The construct of the "self" assumes a central role for the coordination and evaluation of the individual's behaviour. This is because the self-reflection of an individual is the basis for the development of consciousness, which enables the perception of the individual's own behaviour as such and permits the individual to define cause-effect patterns between her own actions and their consequences.

The human brain is interpreted as a subsystem which is steadily involved in the reception and interpretation of external stimuli. Cognitive neuroscience describes perception in a way which corroborates this view. It has been proved that external stimuli hit the sensorial receptors and get communicated to the brain in the form of electrochemical signals. All of these signals are qualitatively equivalent, independent from which sensorial receptor they come from.[16] They reveal the same amplitude and only differ in their frequency, which expresses their intensity.[17] This

[15] Cognitive equilibrium is not an eternal (durable) state, but is contingent on the particular state of the environment which is faced.

[16] Cf. Lehmann-Waffenschmidt (2002, p. 23).

[17] Cf. von Foerster (1992, p. 56 ff).

indicates, that even in a physiological perspective, perception can be characterised as a process by which external stimuli are codified,[18] whereas the brain does not have any direct access to the entity from which such stimuli were generated. It can be therefore said that the brain works in a self-referential and self-explanative way in the sense that "*alle Bewertungs-und Deutungskriterien muss das Gehirn aus sich selbst entwickeln*."[19] Memory assumes a central role for the perception process, as it stores previous sensomotorical experiences and constitutes the key for their cognitive evaluation.[20]

Other results of cognitive neuroscience support the constructivist interpretation of human perception and cognition. They provide evidence that three different sets of brain regions can be related to three different neuronal processes. A first set of brain regions, namely the higher order sensory cortices, processes the perceptual representation of stimuli. A second set, formed by the amygdaly, striatum, and orbitofrontal cortex, is responsible for the association of perceptual representation with emotional response, cognitive processing and behavioural motivation. The third set which consists of the higher cortical regions completes the construction of an internal model of the social environment.[21]

The human brain has an undeniable physical presence, and can therefore be analysed as any other external entity,[22] that means individuals can observe it as external observers. There is an essential difference between the "real" and "cognitive" world, i.e. between outward and inner reality. Individuals simply have access to the cognitive world that they themselves create. Everything they perceive, all information they process and every interaction they take part in, take place within the realm of cognitive reality. Cognition acts as a filter and at the same time as the only link between inner and outward reality, as it coordinates the processing of the stimuli from the external reality.

Cognitive reality is closed in itself and it is only in its realm that the individual experiences the difference between inner and external reality. The perception of the individual's own physical presence, of her own body, brain and brain activity can, without exception, only be experienced and processed by the individual within the bounds of her own cognition.

The cognitive reality an individual lives is the expression of a complex learning process which builds from the interaction of the individual with the external environment. This learning process can be configured as a continuous experience evaluation, memorization and selection of successful behavioural patterns. The cognitive process and specifically, the link between behaviour and cognition will be further explored in the following paragraph.

---

[18] Cf. Roth (1995, p. 59 ff).

[19] "The brain has to develop all evaluative and interpretative criteria out of itself." Author's translation, cf. Schmidt (1987, p. 15).

[20] Cf. Roth (1985a, p. 239).

[21] Cf. Adolphs (2003, p. 166).

[22] The brain as perceivable entity, that can be as such object of analysis, has been often used as critical argument against constructivism. Roth's argumentation (which follows) responds to such criticism. Cf. Roth (1985a) and (1985b).

3.1.2 A Constructivist Approach to the Cognitive Processes

The framework of constructivism can provide a philosophical basis for deepening the general mechanism through which individuals form their beliefs as well as an explanation of their heterogeneity. As this approach merely allows for partial subjective knowledge of the world, it takes into account both "*the intrinsic limits of the human mind in terms of computation and prediction capabilities… [and] the heterogeneity of agents in their beliefs and information endowments*;"[23] this is therefore coherent with the bounded rationality approach. Relying on the constructivist approach, it seems possible to outline a framework in which the heterogeneity of knowledge and information and the subjective rationality of the individuals can be modelled in a mutually consistent way.

"Knowledge" should be distinguished from "information"; the former being a "*map from action to consequences… [which] is activated whenever the system changes its state*"[24] the latter being the identification of a given state.[25]

In a constructivist way,[26] "*(i) knowledge is the output of active elaboration of the subject ranging from the selection of external inputs to the constructions of "models of the world"; (ii) the subject is continually engaged in the empirical control of such models, which thus act as a feedback mechanism in the construction process*."[27] This implies, in particular, that the individual "models of the world" cannot be isomorphic with the external world. Also according to the bounded rationality approach, cognitive activity produces stylized subjective mental models to support the individual in her decision-making. The individual will then evaluate the feedback such models receive from the external environment and, if satisfied, she will rely further on those models; if not, she will modify them.

The cognitive process can be depicted as an input-output process with a feedback mechanism (cf. Fig. 3.1):

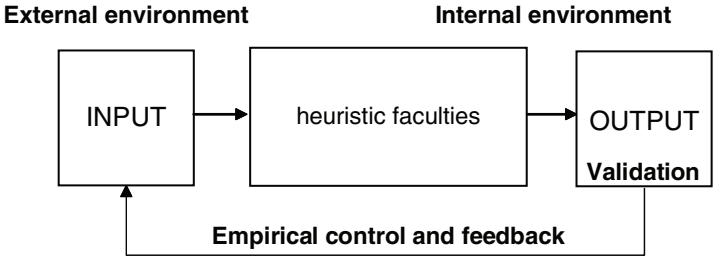

Fig. 3.1 The cognitive process in the constructivist approach (Tamborini, 1997)

[23] Cf. Tamborini (1997, p. 255).

[24] Cf. Tamborini (1997, p. 257).

[25] Idem.

[26] Here is meant the approach of Watzlawick (1981).

[27] Cf. Tamborini (1997, p. 257).

Stimuli from the external environment, interpreted here in the form of physical signals without any prior ontological assumption, are the inputs of the cognitive process and are elaborated in the internal environment. The internal environment, i.e. the "mind," includes all thinking faculties.

To represent its functioning, the computational approach[28] will be followed. Such an approach focuses on the mapping from external to internal states, whilst the latter can be interpreted as logical ordering of syntactic elements, "*like steps in a computer programme*."[29]

An alternative approach could have been the "neural approach," which focuses on the physical disposition of the internal environment, thus on the particular configuration of the neuronal networks.[30] While the neuronal approach seems to have a better explanatory power for the unconscious mental processes, the computational approach suits the representation of conscious thinking and decision-making better.[31]

According to the computational approach, the mapping from external to internal states is based on a series of heuristic faculties, among which abstraction and causation are essential for rational decision-making. The internal environment produces as output a representation or a mental model of the world, i.e. "*a set of causally ordered relationships [. . .] among selected objects or events, aimed at explanation and prediction*."[32] Causal ordering, i.e. how the human mind creates an efficient order for action,[33] plays a central role in human explaining and predicting. Rational actions are based on causal models that forecast in the individual's mind the consequences of her actions.

Consistently, the bounded rationality approach also relies on the notion of mental models. To elaborate her mental models, the individual selects the external signals and combines them according to pre-existing patterns of configuration. The combinations of signals have then to match with such patterns in order to be recognized.[34] This process of selective perception can be called "abstraction" and regards both physiological (e.g. of an object) and conceptual perception (e.g. of an immaterial object or of a social situation).

The last stage of the cognitive process is represented by the validation of the mental models. In the spirit of constructivism, this validation does not require an internal (mental) representation to be an exact reproduction of the external reality because this would just be a metaphysical ideal, deprived of any operational content. Rather, the individual simply needs a rule that establishes that a certain model is provisionally "valid" or "viable" for action.

---

[28] Cf. Newell and Simon (1972), Simon (1977, 1981).

[29] Cf. Tamborini (1997, p. 258).

[30] Cf. McClelland and Rumelhart (1986).

[31] Cf. Tamborini (1997, p. 258).

[32] Tamborini (1997, p. 259).

[33] Cf. Lorenz (1973).

[34] Idem.

The constructivist approach introduces the notion of "cognitive equilibrium" as a sort of measure to which the viability of a mental representation can be related. An individual can be said to be in "cognitive equilibrium" if the actions generated by her internal environment are consistent with her objectives, given the responses from the external environment.[35]

To make the concept of provisional viability more operative, the notion of cognitive equilibrium could be related to that of satisficing, since it allows that different mind constructions can coexist and meet the subjective aspiration levels. Viability is based on what the individual experience, whereas experience here should not be interpreted as observation of events, but as action, because it is through action that the individual tests her mental models. The result of such testing operates as feedback on the construction of knowledge, which leads either to the validation of actual mental models or to their modification. In this way, knowledge is not directly a representation of the world, but a representation of the experience of the world.[36]

This feedback mechanism lets one characterise the cognitive process as self-referential, because every construct, once confronted with the subjective experience of the external environment, shall be reflected to the mind which originated it, i.e. it shall be self-reflected. Human cognition is recursively engaged in the elaboration of mental models out of external stimuli and in their evaluation according to the individual experience of the external world, either to consolidate a (subjective) viable construct or to modify a non-viable one.

Knowledge in the constructivist view can be characterised through the attributes of partiality and possibility, as opposed to the objectivist ideal features of completeness and necessity. This allows for heterogeneity and coexistence of mental models.[37] Partiality stems from the conception of the individual as purpose oriented and as such is guided by "interest," i.e. any purpose (in a broad sense) that can motivate ("cause") the individual's action. Thus interests "*(i) elicit agents' action and (ii) direct agents' heuristic procedures in construing an [intentionally] valid model of the external environment. In cognitive terms, interests provide the focus for 'conscious devices' aimed at reducing complexity through pattern creation and signal-pattern matching*."[38] Interest does not only provide a motivation for action, but can also direct the cognitive process on which rational action relies. An immediate consequence is that no one needs more knowledge than what she needs to manage the situations she is usually confronted with; another one is that, since no isomorphic representation of the external environment is possible, the selection of an absolutely valid rule can be excluded. So, partial knowledge is the intentional result of the cognitive activity of an interested individual and not just an exogenous constraint. As such, it explains the persistence of heterogeneous mental models and the consequent behavioural patterns.

---

[35] Idem.

[36] Cf. Maturana and Varela (1987).

[37] Cf. Tamborini (1997, p. 261 ff).

[38] Cf. Tamborini (1997, p. 261).

Since constructivist knowledge is insolubly connected to a particular experience, it gives life to a constellation of different mental models. Their convergence to collective shared mental models may take place or may not. If individuals interact for a sufficient time, their possible (heterogeneous) mental models may eventually converge to a common one (with common knowledge of this). However, there is also evidence[39] that, if the exchange of information is imperfect, the convergence of beliefs toward a common mental model may be excluded.

Though admitting the possible heterogeneity of beliefs and mental models, constructivism assumes isomorphism of the human mind; this means individuals do not differ in the way they know. Also, isomorphism can be integrated with the bounded rational approach: individuals assume isomorphism or symmetry between themselves and others and they use the tool of introspection to form beliefs about others.

3.1.3 Scientific Research and Reflexivity

The constructivist stance to knowledge contrasts the objectivist and rationalist conception of a world which exists independently of human involvement in it and which can be known in its constitutive objective features apart from the human knowing activity. Constructivism challenges such a view in that it underlines how research constructs what it claims to analyse: what is described by research studies does not exist independently of such studies and their involvement with their own objects of analysis.

Aside from an ontological perspective, it is here sustained that all that is known can only be claimed to be known as a construct, i.e. as a model or a modelling process that can be neither abstracted from the constructing process from which it originates nor related to an external independent reality. The natural system and its processes cannot be investigated without dealing (though without interfering) with them and can only be known through the categories and distinctions made by an observers' community. This relates to the mechanism by which every distinction insolubly links its components, as no outside can be defined without its boundary.

The research process should therefore be seen as socially constructing a world, researchers being an integral part of it, as illustrated by the following statement: "*As inquirers and researchers, we create worlds through the questions that we ask coupled with what we and others regard as reasonable responses to our questions*."[40]

In this perspective, reflexivity is an inescapable feature of research. It and cannot be avoided and constitutes a source of distortion that affects different steps of the research process. The awareness of it and of its implications calibrates the validity of scientific explanation, in the sense that science cannot transcend the limits of human

[39] Cf. Geanakoplos (1989).

[40] Cf. Steier (1991a, p. 1).

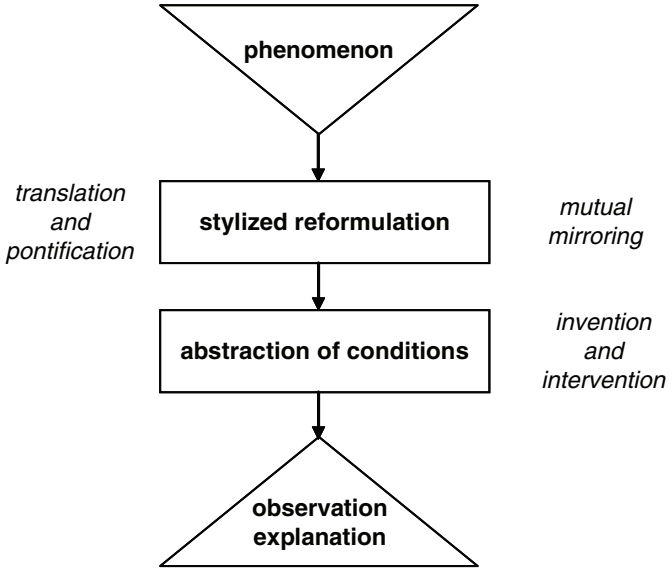

Fig. 3.2 The Reflexivity of Research (author's representation)

constructing. Its validation can therefore be more realistically interpreted relying on the relative and contingent category of viability.

Research aims at explaining phenomena and observation constitutes its first step. The process of scientific explanation is typically articulated by explaining social reality which means at first to observe a phenomenon and then to codify it in a stylized way (see Fig. 3.2). This "stylized reformulation"[41] is the expression of the current mainstream codification of the situation in question. Such reformulation gives evidence of some "abstract conditions" that can properly depict the situation in a standardized and comparable manner, so that a suitable "theoretical framework" – a theory or a model - may be found or worked out and an "explanation" of the phenomenon can be achieved.

In formulating a phenomenon in a stylized way, researchers inevitably shape what they are going to investigate, by means of assessing which features have to be underlined as relevant, which ones have to be intensified by the analysis and which ones can be neglected. This kind of construct of an observed phenomenon reflects its creator (as all constructs do). Because of the mutual reflection process established between observer and observation, research can therefore be characterised as mutually mirroring.[42] The self-referring character of mirroring can be appreciated as a dialectical device: if the individual is deprived of the possibility of penetrating reality, she can merely perceive it in antithesis to herself, which means that every

[41] "*This stylized reformulation is the actual common sense understanding of the scientific community of the real situation which gives the subject of the analysis*." Cf. Lehmann-Waffenschmidt (1996, p. 46).

[42] Cf. Steier (1991b, p. 173).

discourse about reality cannot transcend the limits of her understanding. The individual is therefore mirrored by her own construction, which represents an outward view of the individual and bends back, reflects the individual's features.

When research focuses on social facts, mutual mirroring "*works both ways when dealing with human/social systems whose members are capable of creating meaning and context for those situations in which they are embedded*."[43] The members of a social system construct the situations they are confronted with and at the same time intervene and shape the course of the events they are participating in by means of their constructs. The researchers' constructs overlap with the members' ones and may influence each other mutually. It can also come to parallel processes[44] between researchers' and social group members' constructions, in the sense that researchers are not immune to the social dynamics they investigate.

A scientific observation is based on specific abstract conditions which rely on the standardized notation which is commonly shared by a certain scientific community. This standardized notation includes categories, definitions and measures that "*have been built up and make sense based on the kinds of distinctions felt necessary for activities of a particular research community*."[45] In this sense, research resembles a work of translation as well as a reformulation of perceived fact in the standardised language of the research community. This holds both for the inquiry of natural facts, whereby natural phenomena and patterns have to be linked with standardised categories, and for the analysis of social facts. These are social constructs or at least they reflect them. Translating is means for a constructionist researcher to be "*faced with the dilemma of wanting to understand 'how others construct meaning or make happenings' (in their 'language'), while recognising that he or she is a member of community of particular researchers who have a particular language with which they demand to be addressed, and that these two languages are not the same*."[46]

Social research translates the constructs of a certain group into constructs of another group, thus creating bridges, i.e. "pontificating,"[47] between these two groups. As every reference relation, the one established by social research develops between the elements of a range and those of a domain. Each individual belongs to several different social groups embedded in each other (the set of all those sets existing as well), social scientists do not constitute an exception. Because of the non emptiness of the intersection between range and domain, the reference relation which social research establishes fulfils the requirements for being self-referential in a strict sense, as it has been discussed at Sect. 1.2 of Chap. 1.

Research defines and classifies phenomena. Creating an order is a way to construct new features of reality, which whenever perceived, can affect the course of

[43] Cf. Steier (1991b, p. 173).

[44] Cf. Smith and Crandall (1984) and Smith, Simmon, and Thames (1989).

[45] Cf. Steier (1991b, p. 175).

[46] Cf. Steier (1991b, p. 175).

[47] Cf. Steier (1991b, p. 175).

the system observed. Analysis, as a way of interpreting phenomena, constructs its object and contributes to shape them.

Because of its reflexive character, scientific research can be interpreted in an ecological sense, whereas "ecology" here stands for "*a context constituted by a fitting together of ideas, ideas that here include a researcher (co-) constructing (with reciprocators) a world*."[48] Mirroring indicates that a researcher should be located in such an eco-system.

3.1.4 Science as Language Game

When focussing on science as a human construct, specifically on the involvement of a researcher in the system which constitutes the object of her research as well as on the similarities between research and translation, science can be characterized as a language game.

It is just through the individual's presence that words and sentences which build a language assume meaning. Language is a human construction. As every construct, it depends on the subject who formulates it and reflects some elements of the subjectivity of its creator.

Consensus regarding the meaning of a certain sentence requires an interpersonal coincidence of interpretation of the words it is composed of, as well as the linguistic structures involved. Such coincidence develops among members of the same linguistic community, as they take part in a continuous interaction process. This is a process, which articulates over both linguistic and non-linguistic communication forms and puts the basis of a complex learning process, which aims at granting individual survival in social interactions. As a result, it is then possible to define certain "language games"[49] which are specific to a community and depict simple forms of language "*consisting of language and the actions into which it is woven*."[50]

Individuals speaking with each other are not exchanging meanings or information but merely words and texts. Each of the individuals involved in the conversation will then associate words with mental constructs and representations, though the meaning of such words are assumed in their minds.[51] This is essentially a subjective operation, which cannot transcend its subjective character even among individuals belonging to the same linguistic community. This is due to the fact that their experience and behavioural repertoire might differ throughout.[52]

Language as a means of communication assumes a central role for science, as science relies on it to be expressed and transmitted. Typically, scientific texts (both

[48] Cf. Steier (1991b, p. 180), referring to the interpretation of Bateson (1972, 1991).

[49] Cf. Wittgenstein (1953).

[50] Cf. Wittgenstein (1953).

[51] Cf. Schwegler (1999, p. 18).

[52] "*Was unsere Partner verstehen, […] kann sich nur in den Bedeutungen verwirklichen, die sie augrund ihrer Erfahrung mit den Klangbildern der Wörter verknüpfen, die wir gebrauchen – und ihre Erfahrung ist nie identisch mit der unsrigen*." Cf. Von Glasersfeld (1996, p. 92).

written and oral) take the form of descriptions or observations of phenomena in the way they are perceived.[53]

Science is based on a specific methodology, which codifies how observations should be made and which parameters and procedures should be respected. Such operative criteria direct the interpretation of scientific statements and texts, because it restricts the possibility of associating words and/or symbols with meanings. However, the definition of such operative criteria does not completely eliminate interpersonal degrees of freedom, as the assignation of meaning involves subjective mental processes, according to which the scientific description is perceived as a phenomenon and is processed. The scientific description will be faced with individual mental models and with already existing viable constructs, so that individual experience and social history are relevant for the process of understanding, interpreting and eventually accepting scientific statements and texts.

3.1.5 Constructivism and Economics

Constructivism represents a useful framework for analysing economics. On the one hand this is due to the fact that economic models and theories are essentially constructs and can therefore be analysed according to the observer-observation scheme. On the other hand constructivism applies to the analysis of economics because of the pronounced self-referring character economic theories and models reveal which indicates the autopoietic character of the economic systems and the possible causal role of theories on the behaviour of the economic actors.

Economic models and theories are constructs that order phenomena and events and typically try to establish causal relations among them. A creative operation is to set a causal relation, because it is based on some assumptions of contingency, which, if accepted by the actors belonging to the economic system, can influence the dynamic of the system described. Individuals who face economic problems act intentionally and inform their behaviour to their own mental representations and models. The construction of causality in economics therefore represents a particularly sensitive field because accepted theories are integrated with the individuals' mental models and affect their behaviour.

Causality between events is articulated over the continuum between the two extremes of perfect causality and of full casualty. Perfect causality between events, $E_1, \ldots E_n$, implies that the events are linked and follow each other in a deterministic way, so that from a certain event E_i the whole sequence of succeeding events $E_i \rightarrow E_j,\ j > i$, can be determined. On the contrary, full casualty can be spoken of between events if the events are independent from each other and succeed each other in a random way.[54]

[53] Cf. Schwegler (1999, p. 20).

[54] Cf. Lehmann-Waffenschmidt (2006a, p. 6).

The "post-hoc-ergo-propter-hoc" fallacy[55] occurs when a coincidental correlation is falsely interpreted as causality. Causality between events can be inferred in a fallacious way if it is assumed that the preceding event must be the cause of its successor, i.e. that given the sequence $E_i \rightarrow E_j,\ j > i$, E_i causes E_j. If a fallacious relation of causality is believed by the individuals, they will inform their behaviour to it and might make it become true. "*If men define things as real, they are real in their consequences*."[56]

Economic thought further implies constructivist elements whenever it comes to relate actual economic behaviour and decisions with subjective variables as intentions, purposes and expectations of the decision makers. This is for instance the case of management and organization theory, as well as marketing, finance and game theory.[57]

Elements of constructivism can be also applied for the analysis of strategic thinking. For example, the strategic repertoire of individuals belongs the so-called "strategic thinking of higher order." An individual tries to anticipate the behaviour of her counterparts, putting herself in the others' shoes and reflecting on what she thinks the others think she thinks etc. The ability of forming hypotheses and construct mental models about the beliefs and the intentions of the others, considering that they can differ from their own, is something potentially innate in the individuals.[58] This ability is depicted in cognitive psychology as "theory of mind" (tom).[59]

Tom implies that individuals who think about others' beliefs and mental constructs reflect the constructs that they assume to be held by their counterparts and insert them in their own constructs and mental representations. The supposed constructs of the others will therefore contribute in shaping the individual's perception of reality.

3.2 Recursivity of Social Theorizing and Predictability of Social Reality

Social theorizing can be affected by recursivity in two different ways: first, a social scientist inevitably is part of the system she analyses, and second, theories can affect the state and the evolution of the social system they aim to describe.

Though the first sort of recursivity regards all forms of human theorizing (as the scientist can never be completely disentangled from the reality she examines),[60] the

[55] Cf. Fulda, Lehmann-Waffenschmidt and Schwerin (1998, p. 366).

[56] Cf. Thomas and Thomas (1928).

[57] Cf. Lehmann-Waffenschmidt (2006a, p. 3).

[58] Several studies provided evidence that this ability develops in 3 or 4 years old children. Cf. e.g. Wimmer and Perner (1983).

[59] For a discussion of "tom" in philosophy, see e.g. Davidson (2001) and Dennett (1987) and for its application to cognitive psychology see e.g. Carey (1985) and Wellman (1990).

[60] Consider for example the indeterminism problem for quantum physics.

fact that theories may interfere with the dynamic of the system they aim to describe represents a peculiarity of the social sciences.

Reflexivity of this kind is implied by the so-called "observer-observation problematic," and stems from the coincidence between subject and object of the analysis. It therefore does not find a real equivalent in the natural sciences, as it can be associated with the human analysis of certain natural facts rather than with the analysed facts per se.

The most similar problem that can be discussed in this insight is represented by's uncertainty principle.[61] This principle, whose most popular version is the one of Schrödinger's cat,[62] emerges as a consequence of wave-particle duality and states a trade-off between the accuracy of paired measurements of canonical conjugate quantities. The accuracy of the first of the paired measurements is taken at the expense of the accuracy of the second measurement within a mathematically predictable quantitative range. This yields for the reflexivity of the act of measurement because it introduces an irreducible uncertainty, i.e. it affects the system it refers to. There is however an essential difference which remains between the reflexivity implied by the Heisenberg's principle and the one which affects the social sciences. This consists of the fact that Heisenberg's uncertainty principle only arises in concomitance with and because of the human act of measuring and can, in the end, be traced back to the observer effect. The uncertainty relation does not affect the course of the entities per se, but it only affects the possibility of their measurements. While in Heisenberg's relation reflexivity is due to the presence of an external observer, in the social sciences reflexivity can be ascribed to the individuals both in quality of observers and of object of observation.

3.2.1 Social Predictions

Humans have always strived for a reliable way of divining the future, because being able to anticipate future events is an important capability which can increase the chances of human survival and improve the adaptation to the environment.

In the antiquity predicting the future was mostly related to religion. The key to anticipating the future was represented by the connection with the divine. The method consisted of the interpretation of natural signs. That was, for example, the case of augurs investigating animal entrails or "reading" the bird's flight. Some equivalent manifestations however still survive today, e.g. gipsies reading the future on a crystal ball, or future-tellers reading the tarot.

There are essentially two ways of predicting the future on a scientific basis, namely either an empirical or a theoretic approach can be adopted. Morgenstern (1972)[63] illustrates the two approaches using weather forecasting as an example.

---

[61] Cf. Heisenberg (1927, 2000).

[62] Cf. Schrödinger (1935).

[63] Cf. Morgenstern (1972, p. 705), who quotes Granger and Morgenstern (1970, p. 5).

The empirical approach consists of registering the weather map of a certain day (e.g. today), indicating several parameters for example air pressure, temperature, wind direction and speed, then searching among previous weather maps for a similar one. The weather map of the day after will be then used as prediction for the coming day. This method is "*strictly empirical in the sense that it requires no theory, i.e., no understanding of the mechanism which produces weather changes. It merely makes the assumption that there exists some correlation of the weather at time t and at time t + 1.*"[64] The limits of the empirical approach can be found in the huge amount of data it requires and in the difficulty of finding an almost perfect duplicate of the actual situation in the past.

The second approach to prediction overcomes this shortcoming and is more operative. It is the one which is typically applied in scientific predicting. It consists of inserting the data, e.g. today's weather map, in a mathematical model which expresses meteorological regularities in the form of equations. The solution of the mathematical model represents the prediction of tomorrow's weather. Approaching prediction in a theoretical way the accuracy of the forecast depends on the theoretical knowledge in possession and the possibility of collecting precise data.[65]

3.2.2 Explaining and Predicting the Social Reality

Explaining and predicting natural phenomena can be seen as symmetric processes, as the natural course is not affected by human perception and knowledge ("*Nature does not care – so we assume – if we penetrate her secrets*"[66]). The possible reflexive implications of knowledge constitute, as already discussed, a peculiarity of the social sciences.

Explaining is the codification of a particular real situation or event by means of abstracting some conditions and theoretical laws that apply. Predicting aims to proceed in the opposite logical direction, i.e. starting from the observation of the occurrence of certain conditions in the particular situation examined. By means of application of the theoretical laws, which suit the conditions, the development of the observed situation will be elaborated.

The common methodological ground between explaining and predicting can be found in the explicating process, the process which leads to a theoretical substitute for a pre-theoretical concept.[67]

As represented in the flow chart in Fig. 3.3, even if the social and natural sciences are both based on the scientific methodology of explaining and predicting, reflexivity disturbs the symmetry of those processes for social theorizing.

---

[64] Idem.

[65] Cf. Morgenstern (1972, p. 705).

[66] Cf. Morgenstern (1972, p. 707).

[67] Cf. Güth and Kliemt (2001, p. 1). For a definition of "explication", cf. Carnap (1956).

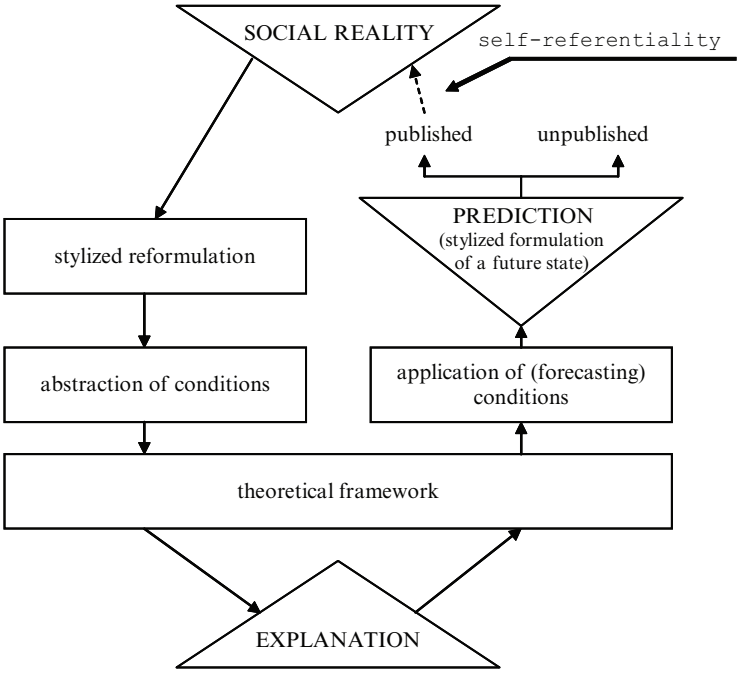

Fig. 3.3 Explaining and predicting the social reality (author's representation)

On the left side of the flow chart, the process of scientific explaining of social reality is depicted, while the right side represents that of social predicting.

Explaining social reality means at first to observe a social phenomenon and then to codify it in a stylized way. Anew, this "stylized reformulation" (*"This stylized reformulation is the actual common sense understanding of the scientific community of the real situation which gives the subject of the analysis*") [68] is the expression of the current mainstream codification of the situation in question. Such reformulation gives evidence of some "abstract conditions" that can properly depict the situation in a standardized and comparable manner, so that a suitable "theoretical framework" – a theory or a model - may be found or worked out and an "explanation" of the social phenomenon can be achieved.

Predicting social reality can be characterized as a successive step of the scientific procedure, and be interpreted as a first test of the validity of scientific explaining. For this reason, the process of social forecasting can be read as articulating in the opposite logical direction of that of explaining. Consequently, the forecasting process goes top-down on the flow chart and starts from the "explanation."

The first step in order to forecast a social phenomenon is to insert it (again) into the "theoretical framework" which suited the "explanation". The "abstract conditions" that the phenomenon can be reduced to should be compared with those that led to the choice of the theoretical framework. If they are coherent to each other,

[68] Cf. Lehmann-Waffenschmidt (1996, p. 46).

a "stylized formulation of a future state" (i.e. a "prediction") can be elaborated. As already stated, it makes a great difference if a social prediction is published or not. A published social prediction may lead social actors to modify their behaviour, either in order to fulfil or to destroy the prediction's content. In this way, social predicting can influence its own object, i.e. the forecasted reality.

Every step of explaining and predicting the social reality can be affected by some distortions, which mainly stem from the cognitive and computational limitations of the individual. Going top-down, the theoretical framework can be affected by the problem of the validity of the theories or of the models chosen; they may be misspecified or even ideologically distorted. Both the abstraction of conditions in the explaining process and their application in the predicting process can, of course, be affected by the application problem. Further, as every stylization is the reduction of the redundant aspects of a complex reality, such an operation is in itself arbitrary. Scientific stylizations, even if they are based on strong codified and standardized methods, make no exception. Thus, both the stylized reformulation of a complex real phenomenon (on which the explanation relies) and the stylized formulation of a future state (the prediction) can only be selective and may be distorted and undermined by information deficiencies.

A last sort of distortion is related only to the predicting process and it deals with the publication of social predictions. Once again, if a prediction gets communicated to social agents, it leads them to modify their behaviour. Whether the prediction gets fulfilled or destroyed, the reaction of the social agents invalidates it on its essence. Even a further prediction capturing the individuals' reaction is condemned to be similarly biased.

In general, social sciences face three fundamental problems concerning predictions. [69] First, there is an essential trade-off between accuracy of a prediction and the likelihood of it happening. A second problem concerns the necessity of assuming the continuity of the regularities in which the prediction is based. The "continuity assumption" requires that the situation, defined through several concrete conditions for which a future development can be forecasted, will actually stay unchanged till the predicted event occurs (or even not). Whereas the falsification of a prediction does not prejudice the predicting activity per se, admitting the necessity of the continuity assumption points at one of the fundamental epistemological limitations of predicting.

While these first two problems apply in an undifferentiated way to the predictability of natural and social facts because they are related to intrinsic limitations of human predicting, a third problem exclusively refers to social forecasting. This problem, which is known as the "Oedipus effect," is represented by the reflexivity of social predictions, which is responsible for the possible self-altering effect a social prediction may have.

[69] Tietzel refers to them as "Geltungs-, Anwendungs-und Ödipusproblem" that can be translated as validity, applicability and Oedipus problem. Cf. Tietzel (1989, p. 546 ff).

3.2.3 Reflexive Predictions

A social prediction can act in a reflexive, self-altering way if its knowledge among a certain community alters its predictive accuracy ceteris paribus. A reflexive prediction is a prediction that fulfils the following conditions. First, the prediction has to be revealed to the decision makers. It speaks in this sense of the "dissemination status" of a prediction and can be distinguished between "made public" or "kept private."[70] Second, in order for the self-altering effect not to be trivial, the prediction's informational content has to differ from the mental representations about the future developments which are otherwise held by the individuals.[71] Only in this case can an eventual modification of behaviour be alleged to the prediction. A third condition is that the prediction has to be believed by the decision makers.[72]

According to Buck (1963):

"*A prediction is reflexive if and only if*:

1. *its truth-value would have been different had its dissemination status been different*;
2. *the dissemination status it actually had was causally necessary for the social actors involved to hold relevant and causally efficacious beliefs*;
3. *the prediction was, or if disseminated, would have been believed and acted upon; and finally*,
4. *something about the dissemination status or its causal consequences was abnormal, or at the very least, unexpected by the predictor or by whoever calls it reflexive, or by those to whose attention its reflexive character is called*."[73]

A further crucial point regarding the self-altering effect of revealed social predictions is whether they are believed or not. The "compliance" with a prediction is tied to the perception the agents have of its validity. For example, theoretically erroneous statements could become self-fulfilling solely because they are believed by individuals, or supposed to be believed by the majority of them. Similarly, the concept of "sunspots equilibrium"[74] refers in economics to situations in which market outcome or allocation of resources depends on variables that only matter because individuals believe they do.

The study of reflexive predictions can be traced back to the nineteenth century. Venn (1966) discussed, for example, the concept of "suicidal prediction." A vivid sociological debate flourished on the reflexivity of Marx's work, as illustrated by the contributions of Merton (1936), Roshwald (1955) and Grünbaum (1956).[75]

---

[70] Cf. Buck (1963).

[71] Cf. Tietzel (1989, p. 554).

[72] Idem.

[73] Cf. Buck (1963, p. 361–362).

[74] Cf. Cass and Shell (1983).

[75] For a further interesting contribution of reflexive predictions and the possibility of economic forecasting see Bombach (1962).

The fundamental sociological principle, known as "Thomas Theorem," ("*if men define things as real, they are real in their consequences*"[76]) explicitly refers to the possible creative role conceptions and beliefs may have in shaping reality. This implies that a situation can occur solely because of being conceived and illustrates the mechanism on which reflexive predictions rely.

Merton (1936) interprets social public predictions as something which fundamentally biases the axiomatic construction they are built upon. For example, the ceteris paribus condition, which is implicitly assumed by every forecasting, gets inevitably violated because of the following prediction: "*Public predictions of future social developments are frequently not sustained because the prediction has become a new element in the concrete situation, thus tending to change the initial course of developments*."[77] This gives account for the frequency and pervasiveness of self-altering predictions. They typically perpetuate the "*reign of error*"[78] they are responsible for, because the occurrence of the predicted events apparently confirms the validity of the prediction.

Reflexive predictions can either act self-fulfilling or self-destroying. A self-fulfilling prophecy is "*in the beginning, a false definition of the situation evoking a new behaviour which makes the original false conception come true*."[79] Self-destroying prophecies are "suicidal" prophecies which do not survive their own dissemination and acceptance. If the individuals are informed of the prediction and believe in it, they will act which invalidates its predictive content.

A critical issue in estimating the degree to which a prediction is reflexive is represented by the counterfactual comparison between what actually occurred after publishing the prediction and what would have occurred otherwise. The difficulties that are associated with such an analysis are evident, because speculative tasks of counterfactual analysis are involved.[80]

The possible causal effect of predictions on the predicted events has been related by Popper to the legend of Oedipus. "*Oedipus, in the legend, killed his father whom he had never seen before; and this was the direct result of the prophecy which had caused his father to abandon him*." Popper therefore suggested "*the name 'Oedipus effect' for the influence of the prediction on the predicted event (or, more generally, for the influence of an item of information upon the situation to which the information refers), whether this influence tends to bring about the predicted event, or whether it tends to prevent it*."[81] In Popper's view, the reflexivity of social predicting is so pervasive that potentially every social prediction can be invalidated by the reactions it provokes among the social actors.

---

[76] Cf. Thomas and Thomas (1928, p. 571–572).

[77] Cf. Merton (1936, p. 903–904).

[78] Cf. Merton (1936, p. 477).

[79] Cf. Merton (1936, p. 477).

[80] Counterfactual analysis is enjoying in the last times certain favour in the historical literature. See e.g. the contributions of Demandt (2001), Ferguson (1999) and Salewski (1999).

[81] Cf. Popper (1957, p. 13).

In economics, the theme of reflexive predictions was first emphasized in 1928 by Morgenstern's seminal essay on economic forecasting, which originated a vivid debate in the economic literature.[82]

Morgenstern approached the problem of economic forecasting in an extremely direct way. While economic theory aims at enabling accurate predictions, the predictive power of the theory represents the ultimate test for the theory's validity. The point is that even false theories may often reveal a great predictive power.[83] According to Morgenstern, predictions are only possible on basis of general theorems (i.e. "*in the large*"[84]). Therefore, they cannot deliver really useful information because they merely consist of observations of the essence of a general theorem as it applies to concrete settings and confirms the regularities it states. More precise and informative historically concrete predictions ("*in the small*"[85]) are not possible in the rule, because "*the data with which the economic forecaster must deal are of such a nature as to make it certain that the prerequisites for adequate induction must always be lacking*."[86] In addition, "*economic processes [. . .] are not characterised by a degree of regularity sufficient to make their future course amenable to forecast*"[87] and furthermore "*forecasting in economics differs from forecasting in all other sciences in the characteristic that, in economics, the very fact of forecast leads to 'anticipations' which are bound to make the original forecast false*."[88]

By strict logical reasoning, Morgenstern[89] came to the conclusion that social phenomena cannot be foreseen, since revealed social predictions influence the analysed system in a way that can in principle never be correctly evaluated. Morgenstern's argumentation is that every social prediction is followed by a behavioural adjustment of the social actors. Even a reformulation of this prediction, which takes into account this feedback, will be followed by another adjustment etc. This infinite readjustment process is known as the "Morgenstern Paradox" and can be represented as in Fig. 3.4.

Given a prediction (P_1) about a social event, it is reasonable to assume that its object will react to it. Such a reaction invalidates the original prediction. Assuming that the reaction (R_1) is known to the forecasters, a new prediction, P_2, which takes into account R_1, should now be formulated. P_2 will also generate a reaction to itself, R_2, so that another prediction, P_3, will be necessary. From a purely logical point of view, this infinite recursive process between prediction and reaction makes it impossible to correctly deal with social predictions and also with the theme of the self-referentiality of social theory. About the question of whether the Morgenstern Paradox also has an empirical and not merely a logical validity, a vivid debate

[82] See e.g the contributions of Grunberg and Modigliani (1954) and Bosse (1957).

[83] Even if fundamentally false, the Ptolemean theory could calculate eclipses accurately and sometimes even astrology guess. Cf. Morgenstern (1972, p. 705).

[84] Cf. Morgenstern (1972).

[85] Idem.

[86] Cf. Margert (1929, p. 313–314).

[87] Cf. Margert (1929, p. 313–314).

[88] Idem.

[89] Cf. Morgenstern (1928, 1935).

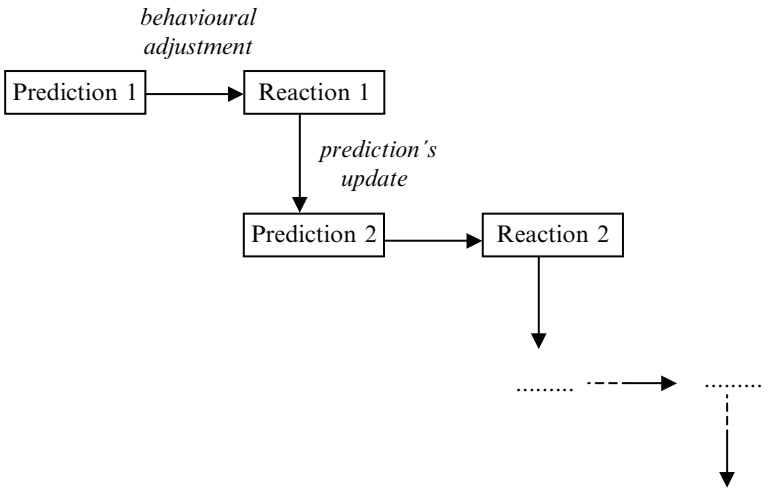

Fig. 3.4 The Morgenstern process (Lehmann-Waffenschmidt, 1990)

flourished.[90] It can be demonstrated however (both in a mathematical and a pragmatic way) that the conclusions to which the Morgenstern Paradox leads, are not sustained from the evidence.

As Grunberg and Modigliani (1954) demonstrate, the Morgenstern Process may also converge to a limit point, which represents the correct prediction of the phenomenon, so that, as in Lehmann-Waffenschmidt (1996), the Morgenstern Process only refers to one of the two conceivable cases which does not admit a finite solution.

A pragmatic solution of the Morgenstern Paradox can also be formulated, considering that in reality nobody can perform infinite reflection processes. Two sorts of limitations occur: the bounded rationality of the subjects and natural restrictions.

Bounded rational subjects are not able to perform infinite steps of recursive reasoning and are aware that their counterparts won't be able to do that, as well. This is proved, for instance, by the experimental evidence from beauty contest interactions: in those situations where individuals have to guess what the others are going to choose, subjects usually perform only a few (two or three) reflection steps.

Natural restrictions refer to time restrictions (the real choice-making process cannot take an infinite time horizon) as well as to cost restrictions (time has an opportunity cost). This explains why even fully rational subjects could not perform infinite reflection processes.

Keynes' analysis of reflexivity essentially concerns the issues of beauty-contest-like economic interactions, conventional expectations and animal spirits.[91] Although Keynes does not explicitly address the theme of reflexive predictions, his

[90] Among others see Bosse (1957) and Grunberg and Modigliani (1954). For a review cf. Lehmann-Waffenschmidt (1990).

[91] More on Keynes's approach to reflexivity in economics, cf. Mackinnon (2003).

analysis corroborates the idea of an infinite regress which is established between the economic theories, the individuals' own and their counterparts' beliefs. In order to take a decision the agents consider what they assume the others are going to do and are aware that the others are engaged in similar reflections. The individuals are therefore engaged in a potentially infinite regress between expectations, individual possible choices and others' reactions. In particular, the notion of "conventional expectations" underlines the creative and manipulating power of the "prevailing view" on individual behaviour, which can on this be basis assimilated to a self-fulfilling prophecy, as it reveals a self-enforcing dynamic.

Plenty of examples of self-fulfilling as well as of self-destroying dynamics can be mentioned. The disclosure of a public opinion survey can manipulate its results by acting in a self-fulfilling way. Similarly, the German Federal High Court recently had to decide on the responsibility of the Deutsche Bank for the bankruptcy of Kirch's corporation; Kirch accused the bank of having caused its bankruptcy by publicly doubting its creditworthiness. Strikes warnings are typically associated with the prediction of scarcity in the commodity supplied by the striking sector. They sometimes provoke panic-stricken reactions among a population that make the feared scarcity crisis become true and increase the prices. Similar hysteric reactions occurred, for example, in Italy after the September 11, 2001 attacks in New York and the US-attack on Afghanistan. As a consequence of media reports warning against scarcity in the oil supply, which were not supported by real evidence, oil prices rose.[92]

[92] Cf. Lehmann-Waffenschmidt (2006a).

Chapter 4
On the Rationality of the Economic Actors

Modern economic theory follows the neoclassical approach and largely relies on the paradigm of rational choice. The paradigm of rational choice interprets human decision making as fully rational and based on Bayesian optimization of subjective utility. This rational way of modelling human decision making stems from the axiomatic characterization of utility and subjective probability rather than from the direct empirical observation of the human economic behaviour.[1] There is plenty of evidence which questions the rational choice approach, both for its interpretation of decision-making and for the assumptions on which it is based. Decisive critiques also come from interdisciplinary studies which integrate economics with findings from psychology, neurology, research on artificial intelligence and cognitive disciplines in general.[2]

From these criticisms alternative characterizations of the rationality of the economic actors have been formulated, so that different "*visions of rationality*"[3] coexist in economic literature and assume different standards of rationality (see Fig. 4.1).

While models of full or perfect rationality assume that "*the human mind has essentially demonic or supernatural reasoning power*,"[4] models of bounded rationality underline the boundaries of human reasoning. Perfect rationality of the economic actors is interpreted as their capability of performing unbounded rational reasoning and is modelled by probability theory and concretized in the optimization under constraints.

Models of bounded rationality stem from the empirical observation of economic decision-making and are essentially informed by the systematic violation of the rational choice paradigm. Sub-optimal outcomes are not just odds that get eliminated because of the principle of survival-of-the-fittest. They configure in many cases stable solutions. Admitting the boundaries of the subjective rationality means

[1] Cf. Selten (1999).

[2] For a critical approach to full rationality see e.g. Kahneman (2002), Gigerenzer and Selten (2001), Güth and Kliemt (2004b), March (1994), Simon (1990, 1957). A more philosophical approach is discussed e.g. in Kliemt (2001).

[3] Cf. Gigerenzer and Todd (1999, p. 7).

[4] Cf. Gigerenzer and Todd (1999, p. 7).

S. Sandri, *Reflexivity in Economics*: *An Experimental Examination on the Self-Referentiality of Economic Theories*,

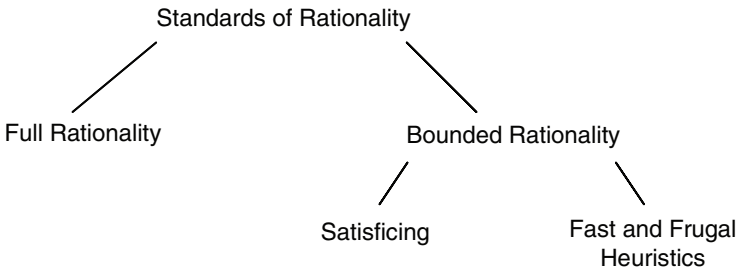

Fig. 4.1 Approaches to the rationality of the economic actors (author's representation)

to interpret the deliberate decision-making process as "*the ability to construct new representations of problems*"[5] and stresses the "*distinction between two types of cognitive processes – the effortful process of deliberate reasoning on the one hand, and the automatic process of unconscious intuition on the other*."[6] As represented in Fig. 4.1, there are essentially "*two main forms of bounded rationality: satisficing heuristics for searching through a sequence of available alternatives, and fast and frugal heuristics that use very little information and computation to make a variety of kinds of decisions*."[7]

This chapter offers an overview on the rationality debate in economics. After discussing the unsolved dualism between rationality assumption and psychology of choice, some of the solutions adduced in defence of the neoclassical paradigm will be presented. The bounded rationality approach will then be introduced and it will be focussed in particularly on the pioneering contributions of Simon and of Kahneman and Tversky. The illustration of some principles of human problem-solving, framing-effects and prospect theory will then conclude the chapter.

For convenience, the analysis of heuristics (both as sources of behavioural biases and as ecological aspects of bounded rationality) will be examined in the next chapter. This will serve as an introductory discussion on the debiasing research. The debiasing research will be considered as a possible inspiration for developing a suitable framework for the experimental analysis of the self-referentiality of economic theories and theory absorption.

4.1 Questioning the Descriptive Validity of Rational Choice Theory

The first criticisms of the theory of rational choice developed parallel to the discovery of some of its seminal findings, as shown e.g. by the 1944's Von Neumann-Morgestern's utility function and Allais' experiments which question its validity. The experiments were run shortly after.

[5] Cf. Egidi (2005, p. 18).

[6] Idem.

[7] Cf. Gigerenzer and Todd (1999, p. 7).

In the Fifties, the development of linear and dynamic programming enhanced the interest towards rational choice theory. At the same time, however, the increasing complexity of the models enforced the doubts on the plausibility of assuming real agents to be able to perform in such a sophisticated way and to correctly solve complicated analytical tasks.

A fundamental dilemma facing economics is constituted by the question whether constrained maximization conveys a proper description of the economic decision-making or whether its validity has to be rather confined to a normative level. Attempts towards the solution of this dilemma have been elaborated both by supporters and by sceptics of the rational choice theory. While the former essentially defends the neoclassical paradigm by stating the negligibility of the psychology of choice when dealing with economic decision-making, the latter formulates on the basis of the internal limitations of the subjective rationality alternative (more ecological and less standardised) approaches to decision making.

4.1.1 The Neoclassical Defence

Initially, "*utility, be it total or marginal, was considered a psychic reality, a sensation that became evident from introspection, independent of any external observation [...] with directly measurable proportions*."[8] In this sense, utility was not considered to be an intrinsic quality of the alternatives faced by the actors, but something dependent on their subjective evaluation. Its measurability and ordinal rating represent therefore critical issues and even create problems which involve the psychological mechanisms underlying the individual decision-making. Substituting the notion of "utility" by that of "preference" constituted a device for separating rational choice from its underlying psychology and represented an answer of the neoclassical approach to the critique of its axiomatic fundaments.

The substitution of "utility" with the much more manageable concept of "preferences"[9] enhanced the construction of the axiomatic model of choice by assuming the preferences' completeness, transitivity, continuity and independence.[10] In addition, it was attempted to anchor the theory of rational choice to empirical data observing the behaviour of large aggregates of individuals. For example, the S-shaped expected utility curve[11] was elaborated on this basis, whereas it did not resist the test with data form representative samples.

The attempts of defending the neoclassical paradigm explicitly avoid the question whether the individuals are in possession of the proper formal tools to adequately optimise or not and strive toward the definition of new analytical instruments for increasing the range of situations to which rational choice can be applied. All in all,

[8] Cf. Schumpeter (1954), interpreting Menger and Böhm-Bawerk's stance on utility.

[9] This can be ascribed to Pareto.

[10] Cf. Egidi (2005, p. 3).

[11] Cf. Egidi (2005, p. 4), referring to Friedman and Savage (1952).

they fail to address the problem of the descriptive validity of the standard model of rational behaviour at its core.

An exception in this sense is represented by Friedman's defensive line of the "as if" approach.[12] Friedman's main argument is that even if individuals are not in possession of the proper analytical skills to correctly solve tasks of constrained optimization, they behave "as if" they were. The conception underlying the "as if" approach is that individuals might not be aware of the mental processes actually involved in their decision making, as tacit knowledge plays a considerable role in human choice making. Similarly to bikers keeping themselves in equilibrium without being necessarily aware of the physical laws of motion, individuals confronted with economic problems solve them in a rational way, without being aware of how their decision emerged.

The "as if" approach further postulates that the individual preferences are not observable because the individuals are not able to express them consciously and explicitly. In this perspective the only variable that can be observed is the decision itself, while the process of decision-making remains a black-box. It cannot be investigated because of the impossibility of empirically checking both the way decisions emerge and the individual preferences. An additional problem is the unconsciousness of the individuals upon these variables. Their analysis is even not considered to be of interest for economics, as selection ensures rational behaviour to be established. Sub-optimal non-rational decisions will be gradually eliminated; agents who do not perform constrained optimization correctly are punished and progressively excluded. Because of the principle of survival-of-the-fittest optimal outcomes eventually represent the only sustainable result of the economic decision-making.

Until today the "as if" approach has enjoyed a broad consensus among economists. It represents an elegant but fallacious attempt of justifying the axioms underlying rational choice theory by stating the non reliability of the individual's expression of her own preferences as well as of non-consciousness of her own decision emergence. Its way of simply avoiding the psychology of choice has been seriously challenged by Allais' and Simon's criticisms, as it will be explained as follows.

4.1.2 Allais' Experiments

Running an experiment in which subjects were faced with pairs of binary choices, Allais showed systematic violations of some of the axioms underlining the expected utility theory.

In Allais' experiment the participants were first asked to choose between the following two alternatives:

(A) Certainty of receiving 100 million (francs)
(B) 10% chance of receiving 500 million, 89% of receiving 100 million and 1% of earning zero

[12] Cf. Friedman (1953).

Second, subjects had to choose between alternatives C and D:
(C) 11% chance of receiving 100 million and 89% of receiving zero
(D) 10% chance of receiving 500 million and 90% of earning nothing[13]

According to the expected utility theory, the individuals who prefer A to B should prefer C to D as well, which is what was systematically violated in Allais' experiment.

Allais' results did not have the destabilizing effect that could have been expected because they were qualified as referring to an extreme case which concerned a hypothetical choice. Despite of the mainstream scepticism about Allais' results, later experiments, in which participants were paid, proved that the inconsistency in the individuals' preferences occurred systematically.[14] The initial reaction to Allais' research was that even more sophisticated versions of the expected utility theory under uncertainty were formulated, trying to generalize its results and relax some of its underlying axioms.

Allais' results began to enjoy a broader acceptance by the mid-1970s, when they became supported by many other relevant studies which seriously questioned the descriptive validity of expected utility theory. Among them "*the ubiquity of EU [expected utility] violations in choices, process data, and elicitation procedures; the elegance of Kahnemann and Tversky's batch of new paradoxes; and Machina's (1982) assimilation of some of the empirical evidence against EU and introduction of sophisticated tools for doing economic theory without the independence axiom*"[15] can be mentioned.

During the 1980s, further revisions of the expected utility theory coexisted with the growing interest for alternative paradigms that insisted on the psychological aspects of choice. Among the firsts, the weighted utility theory,[16] the regret theory[17] and the disappointment theory[18] can be mentioned, although it should be noted that none of them received a statistical confirmation over their full domain of applicability.[19]

4.1.3 Simon's Bounded Rationality Approach

A decisive attack to the paradigm of rational choice came from the theory of bounded rationality, which was developed by the research group lead by Simon at the beginning of the 1950s. Simon and associates were enquiring administrative and managerial behaviour. Their critique of rational choice builds on the unsuitability

[13] Cf. Allais (1953, p. 527), from Roth (1995, p. 8).
[14] Cf. Camerer (1995, p. 626).
[15] Cf. Camerer (1995, p. 626). Cf. Camerer (1995) for the references, as well.
[16] Cf. Chew and McCrimmon (1979).
[17] Cf. Loomes and Sudgen (1982).
[18] Cf. Gul (1991).
[19] Cf. Egidi (2005, p. 7), referring to Hey (1991).

of this theory firstly, for describing decisional processes in organizations and secondly, for expanding to more general aspects of psychology of choice and decision emergence.

The extreme complexity of constrained optimization immediately emerged as an insurmountable limit for modelling the results of the field studies on economic organizations conduced by Simon's research group. In particular the analytical and computational tools required for optimizing properly turned out to be unrealistically demanding for individuals dealing with concrete decisional tasks in organizations.

Simon inferred from the observation that neither the individuals' knowledge nor their computational abilities allows them to optimize correctly, that both internal (i.e. mental) and external (i.e. environmental) constraints limit human rationality. He therefore characterized human rationality as essentially "bounded."

Internal and external limits of the subjective rationality are linked to each other, so that "*human behaviour […] is shaped by a scissors whose two blades are the structure of the task environments and the computational capabilities of the actor*."[20] The individuals have to take their decisions in a complex environment in which it is very difficult and costly to explore and evaluate all available alternatives. Additionally, individuals are also subject to time limits. Exploring the environment constitutes therefore the first step of the decision process "*through which the relevant information is gathered and the appropriate knowledge is structured*."[21] The decision inspires to the results of the exploration of the environment and further depends on the cognitive and computational tools the individuals are in possession of.

4.2 The Bounded Rational Revolution

The focus of Simon's analysis lies on the notion of "subjective rationality," in contraposition to the objective, absolute standard of rationality postulated by the neoclassical theory. In Simon's words, "*in a broad sense rationality denotes a style of behaviour that is appropriate to the achievement of given goals, within the limits imposed by certain conditions and constraints. […] The conditions and constraints […] may be perceived characteristics, or they may be characteristics of the organism itself that it takes as fixed and not subject to its own control. The line between the first case and the other two is sometimes drawn by distinguishing objective rationality, on the one hand, from subjective or bounded rationality, on the other*."[22]

The special emphasis given to the notion of subjective rationality finds among its most influential precursors Popper (1967) and points at the discrepancy between the Olympic standard of full rationality and the behavioural patterns as they can be observed in the reality. On this basis, the subjective logic and the reasoning processes

[20] Cf. Simon (1990, p. 7).

[21] Cf. Egidi (2005, p. 6).

[22] Cf. Simon (1982, p. 8).

which underlie individual decision-making are sought to be understood rather than qualified as irrational or based on erroneous premises.

One of the most innovative contributions of the bounded rationality approach is the attention it pays to the explorative phase of the decision-making process. In this way, both the constructive activity of human reasoning and the influence subjective mental representations have on the outcome of the decision process are in the foreground. The bounded rationality approach does not restrict its critique to the rational choice theory to the too demanding assumptions that are posited to the individual computational abilities. Instead, it extends its criticism to the neoclassical hypothesis viewing the individual problem agenda as exogenously given: "*Neoclassical economic theory assumes that problem agenda, the way in which problems are represented, the values to be achieved (utility function), and the alternatives available for choice have all been given in advance. It has no systematic way of explaining how problems get on the agenda, […] what it is that people value and how values change, or how action alternatives are created […].*"[23] Trying to discover the logic behind decision making reveals that only a "*theory that deals with problem formation and with the design of solution alternatives can provide the basis for a theory of economic change and development*."[24]

For investigating the mechanisms underlying human cognition, Simon developed a new method of analysis, based on the proof principle according to which "*if a problem can be clearly described with appropriate language, then it can be transferred into a form computable for a machine*."[25] The artificial reproduction of human thought can convey a picture of how individuals cognitively select and process information in order to build the subjective frameworks on whose basis they will then take their decisions.

This has been carried on by Simon e.g. through the observation of individuals playing chess. The theoretical analysis of the game of chess and the observation of the players' behaviour was chosen as an ideal setting for investigating the individual computational capabilities and their limits. In addition, the players' verbalizations on their own reasoning during the game were protocolled and helped the understanding of how decisions emerge.

Simon's analysis reveals that individuals approach decision-making totally differently than rational choice theory assumes they do. As illustrated by chess playing, "*chess strategies are inter-temporal decisions which require players to elaborate and re-elaborate their analyses; their decisions are based on a process of learning and mental model building repeatedly at odds with perfect rationality*."[26] Decision-making can therefore be characterized as an extremely complex process which involves the interaction between the mental activities of induction, reasoning and problem-solving. Problem-solving can, in particular, be depicted as central feature of the individual mental processes.

---

[23] Cf. Simon (1992, p. 5).

[24] Idem.

[25] Cf. Egidi (2005, p. 6).

[26] Cf. Egidi (2005, p. 7).

A further feature of the individual approach to decision-making, which starts to be profiled in Simon's research, is that human decision-making is both strived for by deliberate and non-deliberate mental processes. Obviously, using artificial intelligence to simulate and reproduce human reasoning can solely capture the deliberate mental processes, i.e. those processes individuals are aware of and are able to express by means of introspection. The decomposition of rationality in its constitutive components of intuition and reasoning represents the core of Kahnemann and Tversky's research and will be discussed later in Sect. 4.3 of Chap. 4.

4.2.1 Adaptive and Satisficing Behaviour

In Simon's words, "*human reasoning, the product of bounded rationality, can be characterised as selective search through large spaces of possibilities. The selectivity of the search, hence its feasibility, is obtained by applying rules on thumb, or heuristics, to determine what paths should be traced and what ones should be ignored The search halts when a satisfactory solution has been found, almost always long before all alternatives have been examined*."[27]

Bounded rational decision-making can be characterised as a creative process, whose results depend on the paths followed by the individual in her exploration of the environment and of the possible alternatives of choice. Because of subjective (i.e. related to the bounds of subjective cognition and computation) and objective (time and costs related) constraints, individuals are not likely to explore the complete range of all feasible alternative choices. They would rather carry on their search until they find a satisfactory alternative and then interrupt it. Therefore, "satisficing" behaviour is based on a completely different mechanism than optimising behaviour.

Optimising behaviour does not represent a feasible decisional procedure for bounded rational individuals because it would require them to be much more sophisticated and computationally skilled than what in fact they are. Constrained optimization in particular necessitates the abilities to know the whole domain of all possible outcomes, to asses the respective conditional probabilities and to rank them according to the subjective preferences. In addition, each alternative has to be related with its corresponding expected utility. This operation, together with the requirements of full knowledge of outcomes domain, of complete ordinal preference structure and of appropriate probability estimating is without any doubt too cognitively demanding to provide a valid description of how human decisions emerge.

The bounded rationality approach decisively simplifies the abilities that are involved in its description of the decisional process. It assumes the costly exploration of the outcomes domain is interrupted as soon as the individuals evaluate an outcome as "satisfactory," i.e. as good enough to meet their subjective aspirations. The satisficing approach to decision-making is therefore based on extremely simple

[27] Cf. Simon (1992, p. 4).

dynamics, which does not involve complex analytical and computational capabilities. Individuals behaving according to the satisficing principle have to solely rely on a two-valued function because they simply have to evaluate the outcomes in a binary way, i.e. as satisfactory or not.[28]

An individual behaves adaptively, if she adjusts her behaviour to extant conditions in order to fit the current conditions of the environment she is confronted with. In this sense, "to satisfice" means to adapt to the environmental conditions, i.e. to reach an outcome that suffices the subjective aspirations. This underlines how the process of decision-making is shaped by the individual perceptive and cognitive capabilities on the one hand and by the structure of the environment on the other.

The notion of "environment" here does not apply to the description of the physical characteristics in the context in which the decision occurs, but rather encompasses all the aspects of the context which can be relevant for the individual and her mental representation of the faced situation. The specification of the environment will therefore "*depend upon the 'needs,' 'drives,' or 'goals' of the organism, and upon its perceptual apparatus*."[29]

For convenience, the implications of the satisficing rule can be effectively illustrated in a very simple setting: Consider a simple organism that has a single need or goal, namely getting food, and has the choice between three possible activities, which are resting, exploration and food getting.[30] The environment which is faced by this organism can be represented as a bare surface on which the organism can move and over which food is randomly spread in one-meal piles. The organism's internal limitations are that the organism is not able to view the totality of the environment at a glance, but just to a circular portion of it and that it has fixed maximum rates of motion, of food storage and metabolizing.

This setting is particularly simple, because of the following features: the organism has the single goal of getting food, which does not compete with alternative needs, its aspiration level remains unaffected by successes or failures and, while food is randomly being distributed over the surface of motion, no searching pattern is in principle better than the other, so that the organism does have to plan its search. Finally, motion does not serve any other aim than reaching food.[31]

Because of the extreme simplicity of the setting and of the organism's needs, choice will be simple as well. In its search for food, the organism (a) explores the environment at random and (b) whenever it sees a food pile it proceeds towards it and eats it. Since it is able to rest, it is plausible to assume that "*(c) if the total consumption of energy during the average time required, per meal, for exploration and food getting is less than the energy of the food consumed in the meal, it can spend the remainder of its time in resting*."[32]

---

[28] Cf. Simon (1955).

[29] Cf. Simon (1956a, p. 40).

[30] Cf. Simon (1956a, p. 41).

[31] Idem.

[32] Cf. Simon (1956a, pp. 41, 42).

According to the satisficing approach, even in more complex settings decision-making articulates over the three processes of aspiration formation, satisficing and aspiration adjustments.[33] Individuals first set their aspirations, which can be interpreted as discrete levels of goal achievement (process of aspiration formation). Second, they try to fulfil such aspirations by using the means they possess which they consider appropriate for the achievement of their goals (process of satisficing). Third, they evaluate the feedback their choice receives and either confirm or adapt their aspiration levels (process of aspiration adaptation or adjustment). Aspiration levels can represent both "*aspects of preferences and restrictions of decision alternatives, illustrating that in bounded rationality theory the separation of means and ends, as known from the rational choice approach, may not apply*."[34]

Compared with the bounded rational approach, rational choice theory has the advantage of being an axiomatic elegant construction which can be easily applied in a task transcending way. Once specified the individual structure of preferences,[35] the expected utility theory can be universally applied to predict the outcome of decision-making in different settings. Despite the questionability of the predictive content of rational theory, the advantage of its task of transcending applicability is what perhaps motivated the resistance of mainstream economics to abandon such paradigm.

The satisficing approach and the bounded rationality theory in general do not offer operative frameworks which can be applied in an undifferentiated way to a variety of decision settings. Rather, "*the satisficing approach so far offers only an intuitive and natural terminology – already quite an achievement – but hardly any specific guidance when trying to predict (economic) decision behaviour or to give advice like in teaching and consulting*."[36] Some recent experimental studies are now discussing questioning the extent to which the satisficing approach is task transcending.[37]

Similarly, the models of adaptive behaviour are extremely simple but highly task specific. The plausibility of their way of modelling economic behaviour can be alleged to their simplicity, while their specificity hampers their undifferentiated applicability to a broad range of tasks.

Adaptive behaviour can be modelled on the basis of the psychological "Law of Effect."[38] Given the individual initial propensity of choosing certain behaviours, future choices will be influenced by already experienced responses. Behaviours both in the form of simple actions and rules of thumb which could be associated with positive outcomes will reinforce and be more likely to be repeated.[39]

---

[33] Cf. Güth (2006, p. 1).

[34] Cf. Güth (2006, p. 2).

[35] Whereas considerations of the non-plausibility of explicability and completeness assumptions still apply.

[36] Cf. Güth (2006, p. 2).

[37] Cf. e.g. Güth (2006) and Fellner, Güth, and Martin (2006a).

[38] Cf. e.g. Bush and Mosteller (1955) and Roth and Erev (1995).

[39] Cf. Stahl (1995, p. 304).

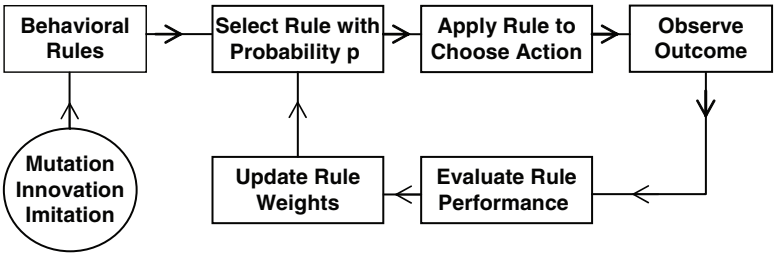

Fig. 4.2 Simple model of adaptive behaviour (Stahl, 1995)

Figure 4.2 illustrates a simple model of adaptive behaviour. This is centred on the evaluation of the feedback a certain behaviour yields for, to which the individual rules are updated.

A consistent part of the research on bounded rationality can be interpreted as an attempt of making simple adaptive behaviour models (such as the one presented at Fig. 4.2) operative. By doing that, three fundamental questions are addressed. "*First, what is the empirically relevant class of rules? Second, what are the initial propensities: Do all individuals have the same initial propensities, or is there significant diversity among individuals? Third, what are the specific dynamics: how is reinforcement quantified, how much weight is given to old information and new information, how much experimentation/search is employed, how are new rules introduced and how significant are computational errors?*"[40]

In this light, the "heuristics and biases" program of Kahnemann and Tversky[41] together with the "ABC" research program on "fast and frugal heuristics" lead by Gigerenzer and Todd[42] can be interpreted as striving toward the deepening of the first question. The analysis of individual propensity, the measure of their determinism as well as the factors influencing their formation is object of conspicuous research on cognitive psychology.[43] Research on aspiration formation and satisficing emerged to answer the first part of the third question[44] because they focus on the mechanisms of feedback evaluation, while intertemporal mechanisms for weighting information developed from the research programs on learning.[45]

4.2.2 Principles of Problem-Solving

One of the peculiarities of the bounded rational approach is to underline the centrality of problem-solving activities in individual behaving and decision-making.

[40] Cf. Stahl (1995, pp. 304, 305).

[41] Cf. e.g. Kahneman and Tversky (1974) and Kahneman, Slovic, and Tversky (1982).

[42] Cf. e.g. Gigerenzer and Goldstein (1996) and Gigerenzer and Todd (1999).

[43] For a review see Rabin (1998).

[44] Cf. e.g. Güth (2000), Simon (1956a, 1956b, 1990), Todd and Miller (1999).

[45] Cf. e.g. Camerer (2003a), Nagel (1995), Stahl (1995).

Human problem-solving is an essentially creative operation whose results emerge as a discovery and cannot be therefore a priori predicted. They cannot even really be simulated by reproducing the exact initial conditions under which the problem-solving activity takes place. This adds to the difficulty of positively modelling bounded rational behaviour and bounded rational decision emergence.

Some general features and principles of human problem-solving can be stylized from the empirical and experimental investigation of the individual behaviour in puzzles and games as well as in complex tasks like e.g. playing chess, mathematical problems, understanding instructions, detecting and inferring patterns between numbered and numerable elements.[46]

Problem-solving involves the selective search through large spaces of possibilities. The complete exploration of all the possibilities, which would be necessary for an optimal outcome to be reached, requires too much time and is therefore not feasible. In comparison with complete trial-and-error search, selective search has the advantage that it enables the individual to find and settle on an outcome within a reasonable time horizon.

Selective search is based on rules of thumb or heuristics, whereas heuristics are simple rules that provide easy replicable methods for solving complex problems.[47] Of course, "*the more difficult the problem and the less efficient the heuristics, the more search will be involved*."[48] Certain heuristics can only be applied to specific tasks, while others are quite general in their domain of application. Whenever existent and accessible to the individual knowledge, domain-specific heuristics will be preferred to general ones, so that the individuals rely on them whenever they can. Otherwise, they apply more general heuristics. These are also called "weak methods," as they do not make specific use of the information about the specific domain of the task to be solved. Individuals facing new problems, which typically demand exploratory and creative skills, often have to rely on general heuristics. In all likelihood, because of the novelty of the tasks, the individual cognitive repertoire does not contain suitable specific heuristics and patterns of solution; as a result individuals typically inform their problem-solving to weak methods.

Some bounded rational heuristics can be qualified as domain-specific or general, depending on the amount of domain-specific knowledge which is available to the individual. For example, the mean-ends analysis heuristic, which is one of the most widely applied, always relies on the same mechanism of comparing between the given initial state of the environment and the individual goals. For the individual using memory cues is a possible means to reduce the gap between the given initial state and the final individually striven configuration. The mean-ends analysis can either be characterized as a weak or strong method, respective to either the individual's possession of little domain-specific knowledge or a conspicuous amount of such knowledge.[49]

---

[46] Cf. Simon (1988, p. 108).

[47] Heuristics will be analysed in more details in Sects. 1 and 2 of Chap. 5.

[48] Cf. Simon (1988, p. 108).

[49] Cf. Simon (1988, p. 108 ff).

In Simon's terminology, domain-specific knowledge is stored in the memory in the form of "productions," or "*actions (A) paired to conditions (C)*."[50] If a problem-solver recognises certain conditions to suit the situation she is confronted with, she searches in her memory for similar conditions and then calls back the actions that are associated with them. In this sense each "*execution of a production is an act of recognition*."[51]

Experts' skills derive also from the large number of productions stored in the individual memory because domain-specific knowledge is obviously larger among those individuals that are acquainted with a certain problem. In this sense, individual expertise can promote the rapid achievement of satisficing solutions. It should be noted, that "*in this picture of expertise, there is no sharp boundary between 'insight' or 'intuition' and analysis*"[52] both of which are "*acts of recognition based on the stored knowledge of the domain*."[53]

The experimental examination of the self-referentiality of economic theories, as will be presented in Chap. 7, can be viewed as an attempt at testing whether adding to the theoretical knowledge of the experimental subjects makes the number of productions available increase and their recognition abilities improve.

Individuals make use of the three capabilities, insight, intuition and recognition in different proportions, depending on the sort of the task to be solved: "*for many simple, everyday problems, recognition alone may be enough, and little analysis may be necessary. In more complex situations, recognition of salient clues allows analysis to take larger and more appropriate steps than if the heuristic search has to depend on real methods alone*."[54]

4.2.3 Problem-Solving in Games and Puzzles

Some other interesting features of human problem-solving can be inferred by observing how individuals behave in games and try to solve puzzles.

Games are typically situations in which information is scattered, incomplete and difficult to process.[55] The individuals therefore face the challenge of collecting information, distinguishing relevant from superfluous information and trying to reduce the computational complexity associated with information processing.

The individual behaviour in games is strategic, although neither perfectly forward-looking nor fully rational. Despite that, game theory assumes that given the initial configuration of the game and its rules the agents can achieve perfect knowledge of the game tree. This means that they are supposed to be able to infer

[50] Cf. Simon (1988, p. 109).
[51] Idem.
[52] Cf. Simon (1988, p. 109).
[53] Idem.
[54] Cf. Simon (1988, p. 109).
[55] Cf. Egidi (1992, p. 148).

all possible future configurations and developments of the game, to associate each possible outcome with the corresponding pay-off and to assess the probability of each outcome to be chosen. From the exploration of the game tree, a rational player can therefore fix the whole sequence of moves she is going to make from the beginning. In doing that, she will take into account the countermoves she expects the others to do.

Because of the boundaries that constrain subjective rationality, during the game theoretical assumptions cannot be fulfilled except in particularly simple settings. As soon as complexity increases, the complete exploration of the game tree and perfect foresight become too demanding assumptions for a bounded rational individual. It seems more realistic to consider the individual exploration of the game as guided by search procedures[56] which are heuristic procedures that use cues to orient the search. They reduce the computational complexity by introducing different parameters to which the subjective evaluators of future configurations can be informed. A lower amount of analysis and simpler analytical skills are involved in such a selective exploration of the game. The players simplify the analysis by exploring only certain segments of the game tree and by excluding others which do not seem to fulfil their own subjective evaluators as they cannot be associated with the cues orienting the search.

Search procedures assume the form of algorithms,[57] which consist in a finite set of instructions for accomplishing a certain task. Search algorithms prescribe ways for choosing one out of a subset of outcomes. They are finite and work recursively. General search procedures ("*blind procedures*"[58]), which enable the complete exploration of the game tree, require unlimited computational capabilities, while selective search procedures ("*satisficing procedures*"[59]), which only insist on a part of the game tree, heuristically orient the search according to subjective criteria and parameters.

Puzzles, e.g. the Rubick's cube or Sudoku, are complex problems which given an initial configuration can generate a huge number of final configurations. The task they posit consists in reaching a certain final state by sequentially applying the typically very simple puzzle rules.

Similar to solving games, when solving puzzles, the individuals cannot explore all the configurations that are possible starting from the initial state. In solving puzzles, the individuals inform their search to the use of different heuristics. One of the most commonly applied is decomposition. As its name suggests, it aims to decompose a complex problem into simpler sub-problems. Decomposition can be recursively applied to all sub-problems, until each of them is decomposed into elementary and very simple problems.

By decomposition local optimal solutions can at best be identified. But the discovery of all local optimal solutions does not lead to finding the global optimum

---

[56] Cf. Egidi (1992, pp. 150 ff).

[57] For a discussion on definition and characterization of "algorithm" see e.g. Sipser (2006).

[58] This terminology refers to Egidi (1992, p. 154).

[59] Idem.

for the original problem.[60] This explains the lock-in of human problem-solving that motivates the persistence and stability of sub-optimal solutions.

A further reason for the emergency and persistence of sub-optimal solutions as well as erroneous beliefs can be found in the path-dependency of learning. Learning is path-dependent because the order in which alternative solutions to a problem are considered affects the decision and influences the future problem-solving. This is because, discovering "*a solution to a problem by searching in a particular way, it will be more likely to search in that way in future problems of the same type*."[61]

What emerges from the analysis of the individual approach to games and puzzles is that human behaviour depends on the two "extreme typologies" of application of routines and creative thought.[62] The individuals facing a certain situation can learn to implement different sequences of actions mechanically, as if they were executing a program.[63] As long as this way of behaving assures satisficing results, it is not likely to be changed. Otherwise, the individuals will make use of their creative, explorative skills and look for new strategies, routines or methods to solve the problem in an alternative way.

4.3 Decomposing Rationality

Kahnemann and Tversky's seminal contribution to the analysis of the subjective rationality[64] focuses on the factors that determine the persistence and systemacy with which violations of the rational choice paradigm occur.

Among the central findings are that thoughts differ in their accessibility ("*some come to mind much easier than others*")[65] and that choice can be either based on intuitive or on deliberate mental processes.

4.3.1 Intuition and Reasoning

Intuition and reasoning are two different modes of thinking and deciding. They can inform judgements and decisions in that those which are taken on the basis of the intuitive mode occur automatically and need a very short time, while those based on reasoning, i.e. the "controlled" mode, are deliberate and slower.

---

[60] For a demonstration of that, see e.g. Egidi (2003).

[61] Cf. Cyert and March (1963, p. 174).

[62] Cyert, Simon and Trow (1956) introduced in this insight the distinction between "programmed" and "non-programmed" decisions, while in Nelson and Winter (1982) speak of "routines" and "innovative decisions."

[63] For a discussion of elements of path-dependency see e.g Egidi and Narduzzo (1997).

[64] Cf. e.g. Kahneman and Tversky (1974, 1979) and Kahneman (2002).

[65] Cf. Kahneman (2002, p. 449).

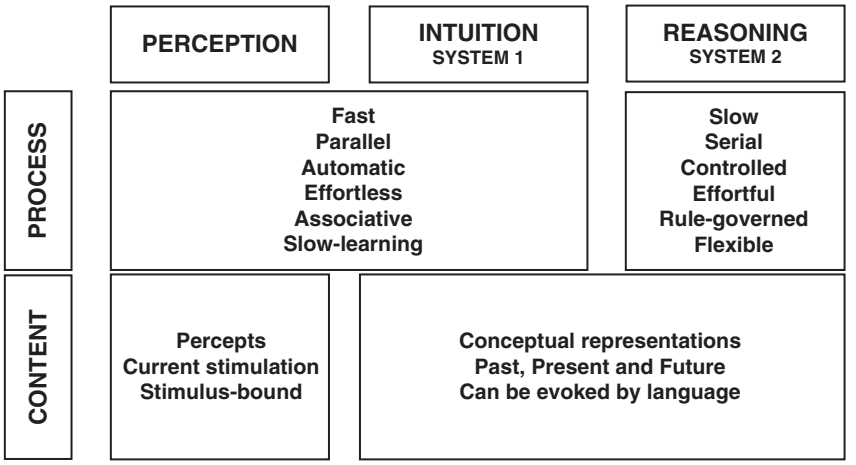

Fig. 4.3 Intuition and reasoning (Kahneman, (2002)

The distinction between "automatic" and "controlled" processes was initially developed by Schneider and Shiffrin (1977). It underlines that human thought both involves deliberate operations and automatisms and that, in the terminology introduced by Stanovich and West (2002), cognition can therefore be seen as consisting of two systems. As shown in Fig. 4.3, the first system is responsible for intuition (System 1), the latter for deliberate reasoning (System 2).

The conception underlying this representation of cognition is that "*intuitive judgements occupy a position [. . .] between the automatic operations of perception and the deliberate operations of reasoning*."[66]

The operating characteristics of intuitive judgements are similar to those of the perceptual processes. Both do not involve much cognitive effort and are fast and automatic, which means that they can also be processed simultaneous to other cognitive activities. Further, perception and intuition are both highly subjective, in the sense that they cannot be easily communicated or transmitted among individuals, and are difficult to control or modify. Intuition and perception differ in their content. While perception is the immediate codification of information as it is originated by a current sensorial stimulation, intuition is a cognitive mode which aims at the formulation of conceptual representations that can be stored in the memory and evoked by language.

Also reasoning aims at the elaboration of conceptual representations. Contrary to intuition, the cognitive mode of reasoning develops on the basis of slow, serial and controlled processes. Reasoning requires cognitive effort and therefore takes more time. It is rule-governed and can be transmitted by learning more easily than intuition. Further differences between perception and intuition, on the one hand, and reasoning, on the other hand, rely in the degree to which they are voluntary and explicit: "*the perceptual system and the intuitive operations of System 1 generate*

[66] Cf. Kahneman (2002, p. 450).

impressions of the attributes of objects of perception and thought. These impressions are not voluntary and need not to be verbally explicit. In contrast, judgements are always explicit and intentional, whether or not they are overtly expressed. Thus, System 2 is involved in all judgements, whether they originate in impressions or in deliberate reasoning."[67]

Further, System 2 supervises the mental operations of both systems and is responsible for the cognitive self-monitoring to which further application or modification of certain behavioural patterns is informed.[68] The cognitive self-monitoring of System 2 is however quite lax. Far from being perfect, it might often allow erroneous intuitive judgements to enforce. The reason for that is that individuals often tend to trust plausible judgements (as intuitive judgements inevitably are) and do not engage in further strenuous reasoning if the reached outcome fulfils their aspirations. For example,[69] students of prestigious US-universities were asked to solve the following puzzle: "A bat and a ball cost $1.10 in total. The bat costs $1 more than the ball. How much does the ball cost?" The large majority of the students followed the initial tendency to answer 10 cents without further checking the accuracy of such an intuitive response.

4.3.2 Accessibility

Intuitive judgements and thoughts come spontaneously to the mind because they are particularly "accessible" to human cognition. The key to understanding intuition therefore lies in capturing the features that make some thoughts more accessible than others. For this reason, the concept of "accessibility"[70] which "*subsumes the notions of stimulus salience, selective attention, and response activation or priming*"[71] was introduced.

For example, out of the three blocks in Fig. 4.4 three towers of equivalent height could be erected. The three blocks differ however in their accessibility. While the first block immediately conveys the idea of its height, which can be caught almost intuitively, the perceptual impression suggested by the second block is related to its area and requires more effort to distinguish its representation as a tower. Finally, the third block's tower has a different accessibility, which changes in dependency of the task.

This shows that accessibility also influences the assessment of relational properties among items, e.g. making comparisons. In Fig. 4.4 the similarities between the first and the third item seem to be more than between the first and the second.

---

[67] Cf. Kahneman (2002, p. 451).

[68] Cf. Gilbert (2002) and Stanovich and West (2002).

[69] Cf. the experimental evidence in Frederick (2003).

[70] Cf. Higgins (1996).

[71] Cf. Kahneman (2002, p. 453).

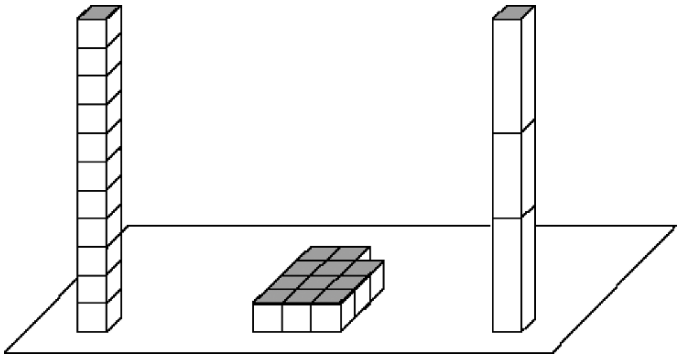

Fig. 4.4 Shape and accessibility (Kahneman, (2002)

Accessibility is not a dichotomous concept, but develops on a continuum between "*rapid, automatic and effortless operations*"[72] from the one extreme (as the activities that can be ascribed to System 1 are) and "*slow, serial and effortful operations*"[73] from the other.

Accessibility depends on the particular properties and attributes of the issue over which a judgement has to be taken, which means that its determinants are complex to identify and enumerate.[74] Among them, physical salience, e.g. size or the position of an item, plays a central role, since the bigger an item or the more prominent the place it occupies, the higher its chances to attract attention are, in comparison with a small item placed aside. The physical salience of an element positively influences the accessibility with which the element is perceived.

Further determinants of accessibility, which can even overcome physical salience, are motivational or emotional clues. Some items can be, independent of their physical qualities, more accessible than others, either because they evoke emotions and can therefore be spontaneously associated with emotional clues, or because the individuals are motivated, for instance through instructions, habits and so on, to pay them special attention.

Accessibility cannot be seen as an invariant property of an item and might also change over time, since it "*also reflects temporary states of priming and associative activation*."[75] Motivational and emotional clues, for example, influence accessibility much more when they are related to immediate emotions or current needs.[76] At the same time, however, accessibility is also related to "*enduring operating characteristics of the perceptual and cognitive system*."[77]

---

[72] Cf. Kahneman (2002, p. 453).

[73] Idem.

[74] The determinants of accessibility are also discussed referring to Kahneman (2002, p. 453 ff).

[75] Cf. Kaheman (2002, p. 454).

[76] Cf. Loewenstein (1996).

[77] Cf. Kahneman (2002, p. 454).

Some attributes, which have been defined "natural assessments,"[78] are "*routinely and automatically registered by the perceptual system or by System 1, without intention or effort*."[79] Among others, they include "*in addition to physical properties such as size, distance and loudness, […] more abstract properties such as similarity (e.g., Tversky & Kahnemann, 1982), causal propensity (Kahneman & Varey, 1990; Heider, 1944; Michotte, 1963), surprisingness (Kahneman & Miller, 1986), affective valence (e.g., Bagh, 1997; Cacioppo, Priester & Bernston, 1993; Kahneman, Ritov & Schkade, 1999; Slovic, Finucane, Peters & MacGregor, 2002; Zajonc, 1980), and mood (Schwarz & Clore, 1983). Accessibility itself is a natural assessment – the routine evaluation of cognitive fluency in perception and memory (e.g., Jacob & Dallas, 1981; Johnson, Dark & Jacoby, 1985; Schwarz & Vaughn, 2002; Tversky & Kahneman, 1973)*."[80]

The same item can make different thoughts accessible depending on the context it belongs to. Obviously, the possibility of ascribing different meanings or ambiguously interpreting such item in the rule is not perceived by the individuals. For example, if only one of the two rows in Fig. 4.5 would have been shown, the ambiguity by which the second sign of each row can be interpreted, which depends on the context (letters or numbers), would not have been in all likelihood noticed.

Framing effects and prospect theory will be shortly discussed. Both can be related to accessibility as well. Framing effects occur when the same problem suggests different thoughts and inspires different reactions when it is formulated in different ways. Prospect theory relies on the observation of the different accessibility of stimulations when they are expressed as absolute values rather than as changes.

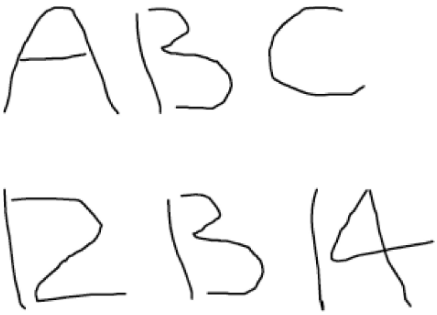

Fig. 4.5 Effects of the context on accessibility (Kahneman, 2002)

[78] Cf. Kahneman and Tversky (1983).

[79] Idem.

[80] Cf. Kahnemann (2002, p. 454), who refers to the listing in Kahnemann and Frederick (2002). Refer to them for the exact references.

4.3.3 Framing Effects

The phenomenon of framing regards decisions under risk and occurs when the modification (ceteris paribus) of the context and/or the description of a problem affects the decision. Alternative descriptions of the same problem can namely suggest different impressions and therefore make different thoughts more accessible than others. As it affects the subjective interpretation of the problem, it could have some repercussions on the perception of outcomes and contingencies which are associated with the different decision options. In this way, the same individuals can make different choices in similar problems just because of their respective frame.

The well-known "Asian disease problem" illustrates a case in which framing effects typically take place:[81]

"*Imagine that the United States is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows*:

- *If Program A is adopted, 200 people will be saved*
- *If Program B is adopted, there is a one-third probability than 600 people will be saved and a two-thirds probability that no people will be saved*

Which of the two programs would you favour?"[82]

Most of the respondents confronted with this problem are prone to choose program A, from which the clear preponderance of risk-averse individuals can be inferred. This is however contradicted by the behaviour of respondents which received the same cover story, but a different description of the two options:

- "*If Program A is adopted, 400 people will die*
- *If Program B is adopted, there is one-third probability that nobody will die and two-thirds probability that 600 people will die*"[83]

In this formulation, most of the people favoured the risk-seeking alternative, i.e. program B. This clearly violates the axiom of invariance of preferences (also known as the extensionality axiom)[84] which assumes that irrelevant features of options or outcomes do not affect the individual preferences.

In the Asian disease problem alternative formulations of equivalent options evoke different impressions, just because they insist either on the positive or on the negative attributes of the possible outcomes (respective survival and death). Although the probabilities associated with the different options in the two versions of the problem are actually the same, they are intuitively perceived in a different way. Program A is

[81] Cf. Kahnemann and Tversky (1983).

[82] Cf. Kahnemann (2002, p. 457).

[83] Idem.

[84] E.g. in Arrow (1982).

emotionally more appealing in the first version, since "*the certainty of saving people is disproportionately attractively and the certainty of deaths is disproportionately aversive*."[85]

Relating framing effects with accessibility and recalling the considerations on the imperfection of control by System 2 over System 1's operations, it is not surprising that the frame of a problem can even bias the choice of expert decision-makers. McNeill, Pauker, Sox, and Tversky (1982) show for example that if asked to choose between surgery or radiation therapy, both professionals' and patients' decisions could be framed by describing outcomes in terms of survival or mortality rates.

Framing effects can bias problem-solving activities, as well. For example, it has been observed[86] that the representation of a problem affects skill transferring. The amount of skill transfer among problems depends more on the perceived rather than on the objective similarities between them, so that isomorphic problems may be perceived as different and unrelated to each other. There is evidence proving that individuals are susceptible to the frame in which a problem is presented as they tend to adapt passively to its proposed formulation and do not reduce the problem to its stylized, canonical representation which would reveal eventual analogies with other (already known or solved) tasks.

The framing effects represent a systematic violation of the rational choice theory embodied in its invariance assumption, which means they seriously challenge the descriptive validity of the neoclassical approach and give a special emphasis to the role of accessibility in human problem-solving and decision-making.

4.3.4 Prospect Theory

A further central finding which questions the validity of rational choice theory is the reference-dependency of perception, on which basis the framework of the prospect theory relies.

Perception is reference-dependent, because "*the perceived attributes of a focal stimulus reflect the contrast between that stimulus and a context of prior and concurrent stimuli*."[87]

This property of perception can be illustrated considering the two enclosed squares of Fig. 4.6 whose brightness is erroneously perceived as differing, because of the brightness of the surrounding area. This influences perception in that it fixes a sort of benchmark to which perception refers, which is called "reference value" or "adaptation level." Not only current stimuli reaching the brain determine perception, but also prior stimulations as well as the history of adaptation.

Being guided by perception, the evaluation of decision-outcomes is reference dependent, too. This contradicts what is posited by the expected utility theory.[88]

---

[85] Cf. Kahnemann (2003, p. 457).

[86] For an overview, see e.g. Mark and Eysenck (2005).

[87] Cf. Kahneman (2003, p. 459).

[88] Expected utility theory has been first formulated by Bernoulli (1738).

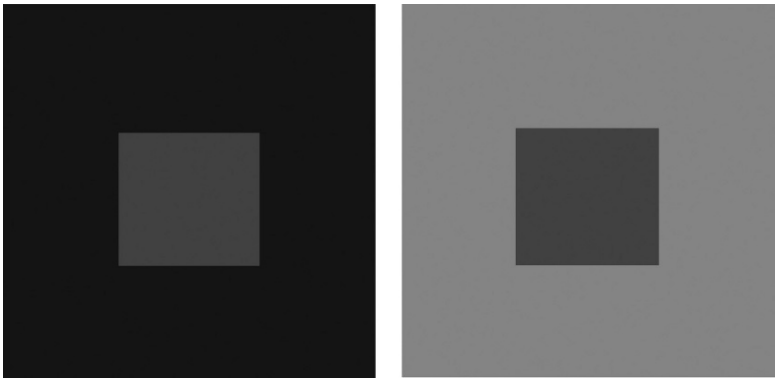

Fig. 4.6 Reference-dependency of perception (Kahneman, 2002)

This approach, which still remains the dominant approach to risky choice, assumes a logarithmic form of the utility function of wealth, postulating therefore that an increase in utility due to an increase in wealth is indirectly proportional to the initial wealth state.

In this framework, the maximization of the utility of wealth implies risk aversion and yields an explanation for the differences in the risk attitudes of individuals with differing initial states of wealth (high or low, i.e. rich or poor). Although Bernoulli's expected utility theory considers that different states of wealth have an effect on the individual attitudes toward risk, it does not relate risk attitudes to the reference provided by the initial state of wealth and does not account for the different way individuals are prone to experience losses instead of gains.

The centrality of the reference value in determining the individual risk attitude is proved by empirical and experimental evidence,[89] which reveals how "*the carriers of utility are likely to be gains and losses rather than states of wealth*."[90] Such evidence can be related to the principle of reference-dependency of perception, whereas "*the effective stimulus is not the new level of stimulation, but the difference between the existing adaptation level*."[91] Modelling risky choice in a psychological realistic way means to allow for the reference-dependency of perception.

In this spirit, Kahneman and Tversky (1979) developed the framework of the "prospect theory." Prospect theory relies on the idea "*that carriers of utility are changes of wealth rather than asset positions*"[92] which implies among other things "*that choices are always made by considering gains and losses rather than final states*."[93] The different accessibility of gains and losses frames the perception of changes in wealth, as it is intuitively illustrated by the two problems discussed in Fig. 4.7.

---

[89] See e.g. Kahneman and Tversky (2002).

[90] Cf. Kahneman and Tversky (2002, p. 461).

[91] Cf. Kahneman and Tversky (2002, pp. 460, 461).

[92] Cf. Kahneman and Tversky (2002, p. 462).

[93] Idem.

Problem A	**Problem B**
Would you accept this gamble?	Which would you choose?
50 % chance to win 150 $ 50 % chance to loose 100 $	Lose 100 $ with certainty OR 50 % chance to win 50 $ 50 % chance to loose 200 $
Would your choice change if your overall wealth were lower by 100 $?	Would your choice change if your overall wealth were higher by 100 $?

Fig. 4.7 The wealth frame (Kahneman & Tversky, 2000)

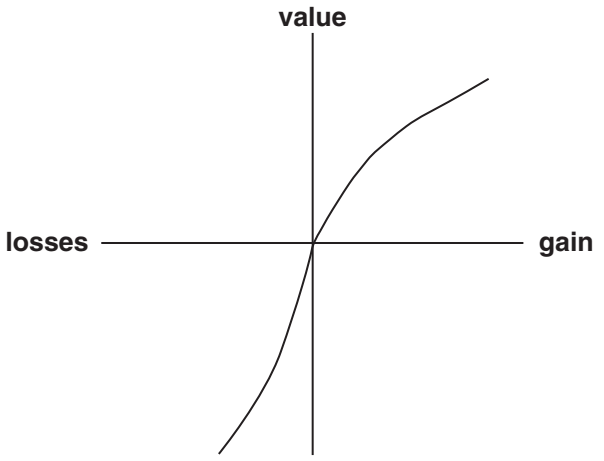

Fig. 4.8 The value function (Kahneman & Tversky, 2000)

It can be intuitively inferred that most of the individuals who face problem A will in all likelihood not accept the gamble and therefore reveal risk aversion.

The lottery presented by Problem B looks appealing. Evidence confirms that the majority of the individuals confronted with a similar choice typically accept the gamble, thus behaving in risk-seeking way. The evidence further shows that even a $100 increase in wealth does not induce them to revise their choice.

The expected utility theory considers problem A and B to be essentially equivalent and cannot therefore account for the differing risk attitudes revealed by the individuals in these two settings. Instead of that, Kahneman and Tversky's approach, modelling "*an alternative theory of risk, in which the carriers of utility are gains*

and losses – changes of wealth rather than states of wealth,"[94] provides a suitable framework for explaining how the perception of gains and losses frame the individual risk propensity.

The prospect theory predicts that the individuals favour risk averse choices when gains are involved, while it expects them to prefer risk-seeking behaviours when losses are involved. The value function represented in Fig. 4.8 visualises these results.

The value function is defined on gains and losses. While it is concave in the domain of gains, expressing therewith risk aversion, it is convex in the domain of losses. The value function models the different attitudes toward risk in a dichotomous way, in that it is sharply kinked at the origin, which constitutes the reference point for choice. The slope of the curve, which is steeper for the losses than for the gains, further indicates a higher sensitivity of risk propensity in the domain of losses rather than in that of gains.

[94] Cf. Kahneman and Tversky (2002, p. 462).

Chapter 5
Heuristics, Biases and Methods for Debiasing

Prediction is involved in many recurrent tasks individuals are confronted with and can be seen as the result of the interaction between "*judgement, intuition, and educated guesswork*."[1] Even when forecasts rely on mathematical methods the central role of intuition cannot be denied as it supervises e.g. the choice of variables that belong to the model, their initial value and their functional specification.

This chapter deepens a central aspect for human cognition and problem-solving, namely that individuals make use of bounded rational heuristics for taking decisions under uncertainty. Heuristics are simplified procedures for assessing probabilities. They are based on rules of thumb. They rely on mental clues which selectively orient the search process and enable the individual to reach her goals when time, informational and computational capabilities are constrained.

Although in some cases bounded rational heuristics can be made responsible for the sub-optimality of outcomes and for behavioural biases. In some other cases it represents an essential support for carrying on inference when complexity overloads the individual cognitive and computational capabilities. It enables the individual to reach better solutions than otherwise.

There are mainly two different approaches to subjective judgement and bounded rational heuristics, namely the "heuristics and biases" approach, pioneered by Kahneman and Tversky,[2] and the "ecological rationality" approach, with Gigerenzer[3] as one of its most influential proponents.

While the "heuristics and biases" approach underlines "*that inference is systematically biased and error-prone, powered by quick and dirty cognitive heuristics*,"[4] the "ecological rationality" approach rather insists on ecological aspects of fast and

[1] Cf. Kahneman and Tversky (1983, p. 414).

[2] See e.g Kahneman (2002), Kahneman and Tversky (1974, 1979), (2000) and Kahneman, Slovic, and Tversky (1982).

[3] See e.g. Gigerenzer, Todd, and the ABC Research Group (1999).

[4] Cf. Chase, Hertwig, and Gigerenzer (1998, p. 206), referring to Kahneman, Slovic, and Tversky (1982).

S. Sandri, *Reflexivity in Economics*: *An Experimental Examination on the Self-Referentiality of Economic Theories*,

frugal heuristics, which exploit the regularities of the physical and social environment and enable individuals to make inferences which rely on a manageable amount of clues.[5]

The essential difference between the two approaches consists in the interpretation of biases as systematically affecting the individual judgement or rather related to exceptions and therefore not seriously compromising the human capability of assessing inferences.

Heuristics are a constitutive feature of human thought as they are deeply internalised and constitute particularly resistant feature of the human cognition. They cannot be easily modified, eliminated or substituted, even when they lead to erroneous judgements which inform sub-optimal choices. This aspect emerges significantly from the attempts at debiasing the subjective judgement and behaviour, which often reveals a moderate to high resistance of several perceptional and judgemental biases.

As follows, the main features of bounded rational heuristics will be underlined. The heuristics of representativity, availability, anchoring and adjusting will be deepened. The analysis will focus on some of the aspects underlined by the heuristics and biases approach but also consider some of the critiques moved by the ecological rationality approach. An overview on the research on debiasing, whose methods and main results will in particular be related to the recursivity of economic theories and the possibility of testing their absorption, will conclude the chapter.

5.1 Bounded Rational Heuristics

Many decisions have to be taken under conditions of uncertainty, i.e. where different events are likely to occur. In similar cases decision-making will be informed to the estimation of the likelihood with which different outcomes can be predicted.

The bounded rational revolution deeply questions the conception of human inference as adherent to the laws of logic and probability theory, as it states the centrality of bounded rational heuristic rules in guiding the individual assessment of probabilities.

The Greek word "heuristic" refers to something which serves to find out or discover.[6] Its appearance into English can be traced back to the early 1800s where it was applied to depict useful processes alternative to the postulates of logics and probability theory for solving puzzling problems. The concept of heuristic received a negative connotation in Einstein's seminal paper of 1905 in which the term "heuristic" was used for indicating incomplete and erroneous ideas. It was nevertheless interpreted as useful and supportive for directing human thought. Simon and Newell kept on with the Dunckerian interpretation of heuristics as (mostly useful)

[5] Cf. Chase, Hertwig, and Gigerenzer (1998).

[6] This introduction on the notion of "heuristic" refers mainly to Gigerenzer and Todd (1999, p. 25–29).

tools for orienting the search for information[7] and modelled heuristics as computational methods.[8] In this way they underlined the supportive role heuristics may assume for human cognition and decision-making as well as the character of rules of thumb, simplified procedures and useful mental devices. In the 1970s, however, a more negative connotation of this concept found its re-emergence. It can be ascribed to its usage in studies analysing the psychology of decision-making, which adopted the term heuristic to depict mostly dispensable cognitive processes people erroneously choose to apply instead of relying on logic and probability theory.[9] However, the almost synonymous association between heuristics and biases was not in the original spirit of the homonymic research program of Kahneman and Tversky (1974). Their research was rather conformed to the consideration that even if heuristics may systematically bias the human cognitive and decisional processes, they nevertheless represent essential tools for supporting decision-making in situations in which the boundaries posited to the subjective rationality do not allow a decision to be taken on a different basis.

Profoundly different spirits respectively animate the "heuristic and biases" and the "fast and frugal heuristics" approaches: whilst the latter "*sees heuristics as the way the human mind can take advantage of the structure of information in the environment to arrive at reasonable decisions*,"[10] the first insists more on those cases in which heuristics can be viewed rather "*as unreliable aids that the limited human mind too commonly relies upon despite their inferior decision-making performance*."[11] Despite that, both programs stress the crucial role heuristics plays in guiding human thought and cognition.

5.1.1 Building Blocks of Bounded Rational Heuristics

Heuristics supports bounded rational decision-making in all its different phases. It specifies particular different principles for guiding and stopping search, whose results will then inform the decision. The principles posited by bounded rational heuristics tend to orient the selective search for information on the basis of cues rather than relying on extensive computational procedures. Searching for cues can be either random or ordered; whereas for an ordered search some simplified heuristic criteria are applied.

The search has to be selective oriented and stopped because of internal (cognitive and computational) and external (referred to time and costs of search) constraints. Heuristic principles establish some criteria for terminating the search at

[7] Gigerenzer and Todd (1999) refer in this insight to Duncker (1935).

[8] See e.g. Simon (1955).

[9] Gigerenzer and Todd (1999) remands for the early usage of the word "heuristic" to Groener (1983), Polya (1954) and for its 1970s usage to Kahneman and Tversky (1974).

[10] Cf. Gigerenzer and Todd (1999, p. 28).

[11] Idem.

an appropriate point. These principles can either be cue-based[12] or referred to an individually specified aspiration level.[13] They provide the decisive advantage of dispensing the individuals of evaluating any cost-benefit trade-off, to which optimization under constraints would be on the contrary based.

After information has been (selectively) collected and the search for it stopped, a final set of heuristic principles orients the evaluation of the results of the search process and informs the inference assessment and the individual choice.

These different sets of principles are separable and can be combined with each other. They can be considered as quite autonomous blocks, building the toolbox of the subjective rationality. This feature represents a further simplification which is introduced by bounded rational heuristic reasoning, which results in the possibility of reducing to an extremely manageable amount the numbers of sets of heuristic principles that has to be stored in the individual memory and applied for making decisions under risk. The possibility of a combination among different sets of principles additionally enhances the adaptability of heuristics to a broader range of situations and problems.

Heuristics can in other words be combined, nested and integrated. This allows for the reduction of the capabilities involved in bounded rational decision-making and problem-solving.

5.1.2 Main Features of Bounded Rational Heuristics

Heuristics is a product of human cognition. Shedding some light on the mechanism through which it emerges and by which it is stored in the cognitive repertoire would involve deepening of the processes which rules human cognition and its extremely complex interaction.

Heuristics can however be simplified to be stored in the cognitive repertoire by means of choosing a certain option (it won't be discussed whether this happens randomly or following pre-existing mental clues here) and then observing its outcome.

Heuristics can thus be assumed to be learned by experience[14] and specifically by a trial-and-error process. The results of which originally refer to the specific problem faced and are then inductively generalised to be applied to a range of other situations that are perceived to be similar. The inductive nature of the process by means of which heuristics is inferred from experience and extended to a certain set of situations implies that heuristics has the properties of context-specificity, task-generality and robustness.

Bounded rational heuristics are specific in that they are highly context dependent. Because the individual understanding of a problem is ruled by its accessibility, its objective features are not the only factors that could affect the way it is perceived.

---

[12] Cf. e.g. Gigerenzer and Goldstein (1996).

[13] Cf. e.g. Simon (1956a, 1956b, 1990) and Todd and Miller (1999).

[14] Cf. Campbell (1969).

Exterior elements might play a role, too, because the way a problem is displayed and presented can frame its interpretation among individuals.[15]

The property of context-dependency does not rule out the generality of heuristics over tasks; in absence of that there should be as many rules as concrete situations. A certain heuristic can therefore be said to be specific but general at the same time, because given the set of the problem's isomorphs[16] the heuristic will be applied to the subset of isomorphs that are actually perceived as such. To all the other isomorphs different heuristics will be applied, thus yielding different solutions and outcomes.

The property of generality is articulated over different degrees of dependency of the number of tasks that are solved by means of applying the same rule. Generality can be rated on a continuum where the one extreme is represented by heuristics with a very restrictive domain of application, up to heuristics defining rules on how to generate rules.

The heuristics of representativeness, availability, anchoring and adjusting can be situated at this second extreme, as they "*direct the way in which specific rules can be formed to solve problems*."[17] They are therefore also labelled as "meta-heuristics."

Meta-heuristics allows for the high level of generality which is implied by the task of rule generation and at the same time "*for the important effects of context, wording, response mode, and so on*."[18]

The particular feature of bounded rational heuristics which makes it suitable for leading the individual decision-making in an ecological way, i.e. in a way that fits the structure of the environment, is constituted by its solution of the trade-off between speed and accuracy of judgement in which any deductive reasoning is inevitably trapped. On this basis heuristics can often be at the same time fast and accurate to an adequate degree. This is because heuristics is articulated on a different dimension: the trade-off between specificity and generality.

As a consequence, heuristics reveals different degrees of robustness or strength, which diminishes as its fitting to the structure of the environment increases. In particular, the strength of certain heuristic influences its reinforcement. This is ruled by the feedback the heuristic received as a result of the individual experience, which can vary according to how and how often such a feedback occurs, which size it has, with which magnitude it is perceived and which methods are implemented by the individuals for checking on the outcomes achieved by means of rule application.

A critical point in this insight is that even inadequate and incorrect heuristics can be reinforced. This can occur when the individuals are unaware of the task structure and cannot thereforeperceive alternative or more appropriate, better ways

[15] Experimental evidence on that is conspicuous and wide accepted, see e.g. the contributions of Grether and Plott (1979), Lichtenstein and Slovic (1971), Simon and Hayes (1976), Tversky and Kahneman (1980).

[16] The expression "problem isomorphs" was introduced by Simon and Hayes (1976).

[17] Cf. Einhorn (1982, p. 271).

[18] Cf. Einhorn (1982, p. 271).

of solving the task. In a similar setting the concrete outcomes cannot be related to any superior benchmark of a solution, but can be only adjusted toward the individual aspiration level.

After plotting the experienced action-outcome combinations on an imaginary diagram, the outcomes will be compared to the individual aspiration level and consequently assessed either to the category of success or failure. Such an assessment can however be biased for several reasons: because of the absence of a complete map of experienced action-outcome combinations, because such an incomplete map becomes in the rule only sequentially available and is sometimes widely scattered over time, as well as because the systematic search for evidence disconfirming the heuristic applied is typically missing or is at least very rare.

Consider for example a waiter serving in a crowded restaurant.[19] Given the impossibility of offering good service to all customers, he might decide to dedicate more attention to those customers he assesses to be more likely to leave a higher tip. Treatment effects however (i.e. good service influences the tip other things being constant) elevate the probability of the waiter to confirm himself in his heuristic following. It does not seem plausible to assume from the waiter that he disentangle the treatment effects and give good service to some of those customers he does not expect to pay a generous tip. This would require awareness of the task structure on the one hand and is on the other hand not likely to be practised because of lacking motivation: why should the waiter risk a loss of income for doubting a rule that seems to work?

From the properties of specificity and generality of bounded rational heuristics as well as from the modalities by means of which outcomes are subjectively assessed, it can be inferred that the validity of heuristics should be evaluated according to correspondence rather than to coherence criteria. This means that its evaluation should insist on "*measures that relate decision-making strategies to the external world rather than to internal consistency, such as accuracy, frugality, and speed [...].*"[20] Heuristics has the function of enabling reasonable adaptive inferences given the concrete structure of the environment, within time and informational constraints, so that as long as it fulfils this function there is no reason to require internal coherency.

The purpose of this chapter is to show evidence of some parallels between the analysis of theory absorption and that of debiasing. Therefore, the discussion of bounded rational heuristics which follows mostly relies on the heuristics and biases approach and presents only a selection of cases of heuristics which lead to biased behaviour.

For a discussion on the differences and similarities between the heuristics and biases approach and the ecological rationality approach, attention should be directed to the interesting debate between Kahneman, Tversky and Gigerenzer.[21]

---

[19] Cf. Einhorn (1982, p. 282).

[20] Cf. Gigerenzer and Todd (1999, p. 22).

[21] For a further critical overview, see Gigerenzer and Todd (1999, p. 28).

5.2 Heuristics and Biases

According to the classification proposed by Kahneman, Slovic, and Tversky (1982), representativeness, availability and anchoring heuristics will be presented as the basilar principles supporting bounded rational probability assessment and value predicting. The most frequent biases that can be associated with the application of heuristics will be mentioned as well.[22]

5.2.1 Representativeness Heuristic

Common tasks involving likelihood estimations can be of the following type: "*What is the probability that A belongs to B?*"; "*What is the probability that A originates from B?*"; "*What is the probability for B to generate A?*"[23]

In order to solve similar problems, the representativeness heuristic is typically applied. This is a heuristic which evaluates probabilities by the degree to which A is representative of B. On this basis, whenever A's perceived affinity or similarity with B is high (respective low), the probability for B to generate A is judged to be high (respective low) as well.

Choosing the affinity between events as a criterion for assessing probabilities can lead to systematic mistakes. This occurs in particular when providing the individuals with the description of a certain phenomenon where no specific evidence is given and asking them to assess the probability of a list of events concerning the phenomenon described. They will tend to base their inference on the representativeness of the different events with the description provided. This occurs for example in the many cases in which people make judgements on the basis of cliché, rather than on real informative elements: take the example of the higher propensity to associate a researcher or a scientist with the cliché of an introverted and reflexive person, rather than with a sociable funny and easy-going guy.

Representativeness heuristics has been explained as the application of "intensional" instead of "extensional thinking"[24] This happens when problems are represented by means of referring them to individual mental models, to prototypic and similar representations, instead of being conceptualized according to formal notions and parameters.

When representativeness is relied on, other factors which are significant for inferring probabilities are neglected. In this way, it can be shown that individuals are systematically prone to the insensitivity to prior probabilityof outcomes (base-rate

[22] This discussion of heuristics and biases mostly relies on the analysis of Tversky and Kahnemann (1982, p. 3 ff).

[23] Cf. Tversky and Kahnemann (1982, p. 4).

[24] Cf. Kahneman and Tversky (1983).

bias), to sample size (law of small numbers) and to predictability. In addition individuals tend to mis-conceptualize chance and regression and underline the illusion of validity.[25]

5.2.2 Availability Heuristic

In estimating probabilities, a typical clue which individuals rely on is availability. Availability implies that the likelihood of certain events is estimated by recalling the frequency of their occurrence among their own acquaintances.

Even if, on the one hand, availability represents a useful criterion for inferring probabilities (since "*instances of large classes are usually reached better and faster than instances of less frequent classes*"),[26] on the other hand, it can be affected by factors that do not depend on probability and frequency. The use of availability heuristics can therefore systematically lead to biased judgements which can be e.g. alleged to the retrievability of instances, to the effectiveness of a search set, to imaginability as well as to illusory correlation.[27]

5.2.3 Adjustment and Anchoring

Anchoring refers to the tendency of overly relying on a specific piece of information and consequently to adjust the intuitive judgement to the value provided by that piece of information which then serves as a basis for inferring other elements of the problem.

The adjustment and anchoring heuristic implies that the individuals anchor their judgemnt to an implicitly suggested reference point for then adjusting toward that point. The tendency to anchor judgement persists even when no reference point is given because individuals are prone to estimate a reference value. They might even infer this value on the basis of incomplete computations.

Among the biases that can be related to the adjustment and anchoring heuristic, insufficient adjustment can be mentioned, as well as biased evaluation of conjunctive and disjunctive events, and erronous assessment of the distribution of subjective probability.[28]

---

[25] All of those biases are threaten in more details by Tversky and Kahnemann (1982). The same applies for the biases mentioned in Sects. 2.2 and 2.3 of Chap. 5.

[26] Cf. Tversky and Kahnemann (1982, p. 11).

[27] See for more details Tversky and Kahnemann (1982).

[28] See note 356.

5.3 Debiasing

Included within debiasing methods are all the strategies for testing the robustness of an observable bias by attempting to eliminate it under controlled conditions.

The robustness of a bias can be tested by means of controlled modification of the design in which such bias is observable. The sense of such a controlled modification of the design is to push it towards and beyond the limits of its validity, i.e. up to its specimen's failure. Borrowing a typical engineering expression, debiasing could be interpreted as a sort of destructive test which reveals under which conditions the design fails or at what point unbiased judgement and decision-making emerge.[29]

The search for the conditions under which the biased behaviour disappears (or is reduced in occurrence) allows for the mechanisms to be isolated which are responsible for the bias.

It is in this sense that research on debiasing constitutes one of the possible ways to investigate the mechanisms ruling human perception, cognition and judgement. Debiasing methods can further represent useful devices for testing the hypothesis concerning such cognitive mechanisms and mental structures. In this insight, both successes and failures of debiasing may yield relevant findings, in that "*(a) failure helps clarify the virulence of a problem and the need for corrective or protective measures, and (b) the overall pattern of studies is the key to discovering the psychological dimensions that are important in characterizing real-life situations and anticipating the extend of biased performance in them*."[30]

A first affinity between the research on debiasing and the investigation of the self-referential effects of a theory can be stated in this insight: a method for debiasing can only be effective if it is able to capture the psychology underlying the individual choice; a theory won't have self-referential effects and cannot be absorbed among certain individuals if it is not able to penetrate the way such individuals interpret the economic problem they face. This will be discussed in more details in chapter 6.

The different debiasing methods can be classified "*according to their implicit allegation of culpability*,"[31] as biases can be induced by the formulation or the accessibility of a task. They can also be due to erroneous judgements. However biases can also originate from the mismatch between task structure and judgement. Corresponding to the factors the bias is supposed to be the result of, debiasing can be strived for by means of applying different strategies as summarized in Table 5.1.

The specific structure of a task could be responsible for biased behaviours if it is e.g. unfairly formulated, based on confusing instructions, unfamiliar stimuli, lacking incentives to perform well, disbelief in experimental assertions on the task nature, mistrust of the declared payoff-structure and so on. Whenever similar conditions occur it is the case of an unfair task. Checking on the conditions which are perturbing the behaviour of the individuals confronted with an unfair task would serve

[29] Cf. Fischoff (1982b, p. 422).

[30] Cf. Fischoff (1982b, p. 423).

[31] Cf. Fischoff (1982b, pp. 422, 423).

Table 5.1 Debiasing methods according to underlying assumption (Fischoff, 1982b)

Assumption	Strategies
Unfair tasks	Raise stakes Clarify instructions/stimuli Discourage second-guessing Use better response modes Ask fewer questions
Misunderstood tasks	Demonstrate alternative goal Demonstrate semantic disagreement Demonstrate impossibility of task Demonstrate overlooked distinction
Faulty judges	
Perfectible individuals	Warn of problem Describe problem Provide personalized feedback Train extensively
Incorrigible individuals	Replace them Recalibrate their responses Plan on error
Mismatch between judges and task	
Restructuring	Make knowledge explicit Search for discrepant information Decompose problem Consider alternative situations Offer alternative formulations
Education	Rely on substantive experts Educate from childhood

to improve the scientific hygiene of the study, rather than attempting a debiasing operation in a narrow sense.

It can however be difficult to draw a clear dividing line between a misleading description which artificially biases the individual judgement and a description which frames the individual interpretation of the situation faced. In this sense, biases can also be due to the specific description or formulation of the task, which is susceptible to misunderstanding by the individuals. Task misunderstanding can be for example checked and eventually avoided by using alternative descriptions, underlying different key words, pointing explicitly at alternative goals, or by advising the individuals to concentrate on overlooked distinctions.

"*Once the task has been polished and the bias remains*"[32] a second approach to debiasing can be followed. The responsibility is then definitively alleged to the individuals performing the task.

The individual judgement can be either perfectible or incorrigible. Whenever the latter case applies, the subjective judgement is not sensitive to any debiasing training and device. The shortcomings introduced by the bias can only be overwhelmed by flanking the decision-makers with artificial support.

[32] Cf. Fischoff (1982b, p. 426).

Obviously, in case of incorrigible judgement, testing for debiasing strategies would not make much sense for deepening the analysis of the individual bounded rational judgement. Debiasing attempts make much more sense in the case of perfectible judgements. In these cases, debiasing can be achieved by providing the individuals with artificial experience, whereas searching for the most suitable methods (i.e. that can be understood, internalised and applied or in other words "absorbed" among the individuals) enables some light to be shed on the mechanisms ruling perception and cognition. Experience can be simulated and artificially provided to the individuals for example through different training programs. These training programs consist of warnings about the possible bias occurrence and offer personalized feedback or problem descriptions.

As it will be discussed later in more details (see chapter 7), the absorption of a theory can be empirically tested by means of providing the individuals with theoretical information concerning the concrete situation they face. The analysis of theory absorption can therefore be methodologically related to the debiasing of perfectible judgements, in that it similarly implies the training of the individuals in order to support their subjective rationality.

A third source of biases is constituted by the mismatch between judgement and task. The two possible approaches to debiasing that can basically be followed when a bias is implied by the discrepancy between judgement and task are "restructuring" and "education."

The procedure of restructuring can be endeavoured if the compatibility between individual skills and structure of the task is potentially given, but fails for some reasons to get exploited. If there is a subset of individuals who are able to exploit the compatibility between task and skills and another group who are not capable, debiasing can be achieved by educating the unsuccessful respondents.

Restructuring can in other words be applied if the respondents are assumed to be in possession of the skills required to solve the task, but they are observed so that not make use of them in solving the task. Inviting the respondents to express their knowledge and/or reasoning explicit, decomposing the problem, considering alternative situations or formulations of the task, making discrepant information more accessible etc. are among the different restructuring strategies that can be mentioned.

Educating the respondents is a debiasing procedure that suits the cases in which a bias is assumed to occur even if the task does not involve cognitive and computational capabilities that are available to the respondents. Such an assumption can be e.g. tested by comparing the biased performance of the respondents with the benchmark of individuals that are already acquainted with the task and have therefore developed a certain expertise. If this assumption seems to apply, debiasing could be achieved by means of "educating" the respondents. "Education" is meant here to differ from training measures concerning the faulty judgement category, in that it focuses "*on developing general capabilities rather than specific skills*."[33]

A similar educating procedure can be specified for the analysis of the self-referential effects of a theory and its absorbability, as well. As it will be specified in

[33] Cf. Fischoff (1982b, p. 427).

the coming chapters (Chaps. 6 and 7) testing for theory absorption will be based on the procedure of providing the respondents with theoretical information concerning the concrete task to be solved.

Although some general patterns or approaches to debiasing can be inferred, concrete debiasing procedures essentially differ from one bias to the other because of specificity and task dependency of heuristics. For this reason, as well as for the different attention which has been paid in the literature to the diverse cognitive biases, it is not easy to organize a systematic overview on debiasing. In addition, a certain degree of arbitrariness is always present when it comes to evaluating the relevancy of a certain debiasing attempts. Whether or not the efficacy of a debiasing experience or its methodological fundaments should be privileged, can be argued.

Three considerations essentially oriented the selection of the studies that are going to be mentioned, namely (1) to systematically privilege, even when it is at the expense of comprehensiveness; (2) to focus on the contributes providing some elements that can be interpreted as preliminary for the analysis of theory absorption; (3) in respect of the criteria posited by Fischoff (1982a, 1982b), to restrict the attention to studies published after peer review which offer some empirical evidence so that both theoretical suggestions and anecdotic evidence are ruled out.

A selection of debiasing studies which meet these requirements and focus respectively on the representativeness, availability and adjustment meta-heuristics will follow.

5.3.1 Debiasing the Representativeness Heuristics

When debiasing the representativeness heuristic, two alternative approaches are mainly followed: the procedures either aim at affecting the representation of information or at the training of the respondents.

5.3.1.1 Debiasing the Insensitivity to Prior Probability of Outcomes

Focussing on the insensitivity to the prior probability of outcomes, Rakow, Harvey, and Finer (2003) demonstrated for example that it is possible to reduce the occurrence of the base-rate bias if the respondents are allowed to discuss the situation in small groups. The study proved group discussion to draw the respondents' attention to the base-rate information provided because it stimulates the individuals to think more carefully about the information they received. Moreover, a second experiment revealed the beneficial effects of providing the respondents with the range of probability responses for the task. Of the range of probability responses, graphical instead of numerical representation yielded for different results in improving the accuracy of judgement.[34]

[34] Cf. Rakow, Harvey, and Finer (2003).

According to Gigerenzer, Hell, and Blank (1988) the systematic neglection of base-rate information can be alleged to the internal problem of representation rather than the use of the representativeness heuristic. This implies that "*by manipulating presentation and content, one can elicit either base rate neglect or base rate use, as well as points in between*"[35] and reveals therefore that "*representativeness is neither an all-purpose mental strategy nor even a tendency, but rather a function of the content and the presentation of crucial information*."[36]

Roy and Lerch (1996) rely on three alternative strategies for correcting the insensitivity to the prior probability of outcomes: different presentations of the information, training of the respondents in strategic devices for processing information and replacement of human decision-makers with a model or system following normative rules of probability theory.

Moreover, while Cole (1989) could diminish the insensitivity to priors by educating the respondents through the provision of feedback information, Gebotys and Claxton-Oldfield (1989) lowered the impact of base-rate neglecting by adding to the theoretical knowledge of the respondents. By a follow-up test 5 weeks after the experiment they could additionally prove that the training group maintained a superior performance.

5.3.1.2 Debiasing the Conjunction Fallacy

The conjunction fallacy occurs when specific conditions are assumed to be more likely than general ones because they are intuitively judged to be more representative than the general ones.

The attempts at debiasing the conjunction fallacy are mostly centred on pointing at the concept of "specificity" as a subset of that of "generality,"[37] as well as on increasing the respondents' acquaintance with elementary notions of set theory.

For example, Agnoli and Krantz (1989) developed a training method which combines external pictorial representations (Venn diagrams) with learning by doing. As their results suggested, the efficacy of debiasing seems to be strived by inducing the individuals to shift from the mode of intensional to that of extensional thinking. In the end this can be related to a reduction in the intuitive components involved in the judgement.

In 2005 a pilot experiment was conducted on undergraduate students at the University of Dresden which aimed at debiasing the conjunction-effect fallacy by means of providing the respondents with theoretical information. The experiment merged typical elements of debiasing (in form of graphical aids) with abstract theoretical information and focussed on the analysis of theory absorption. As a result, the almost complete elimination of the conjunction-effect bias could be achieved. Further details on this pilot experiment will be discussed shortly in Sect. 4.1 of Chap. 7.

---

[35] Cf. Gigerenzer, Hell, and Blank (1988).

[36] Idem.

[37] Cf. e.g. Moutier and Houdé (2003).

5.3.2 Debiasing the Availability Heuristic

Stapel, Reicher, and Spears (1995) codified and tested different moderators on the impact of biases connected with the availability heuristic. In doing that they emphasized the extreme context dependency of heuristics and judgement and insisted on the perceived relevance of the information provided for the task to be solved as well as on its different memorability. What they showed was that simply introducing a different cover story for the same task can lead to different judgements, as contextual factors can steer the use of certain information. They further provided evidence for high context-dependency of the perceived relevance of information, which rules its memorization and therefore influences its availability.

The hindsight bias expresses the propensity to see past events as predictable and reasonable as well as the overconfidence in the individual's own predictive abilities. The hindsight bias can be ascribed to a plurality of psychological factors (e.g. focal thoughts, accessibility, selectivity of memory, overconfidence etc.) and is based on complex psychological mechanisms that are deeply rooted in the human mind. For this reason, the insight bias occurs highly systematically and is particularly resistant to debiasing.

Debiasing procedures developed on the content of the prediction whose perception is biased by hindsight, reveal only a moderate efficacy,[38] while procedures which act on the interaction between focal thoughts and accessibility[39] enhance more effecting to lower the bias' occurrence.

5.3.3 Debiasing Adjustment and Anchoring

Mumma and Wilson (1995) developed and tested three procedures for debiasing anchoring effects in clinical-like judgements, each of which correspond to an alternative explanation of anchoring.

The first procedure (the so called "bias inoculation approach") consists in training the respondents about the anchoring heuristic and providing them with feed-back information on the primacy bias.

The second procedure (the "encoding elaboration model") invokes the role played by opposite information in inducing the respondents to realize differences among initial and late information. This second procedure has also been applied (in form of the so-called "consider the opposite" procedure) to the debiasing of the overconfidence and hindsight bias.[40]

The third procedure tested by Mumma and Wilson (1995) finally focuses on the "attention decrement" as a possible reason for the anchoring effect and accordingly consists of asking the respondents to write some notes on the information received.

[38] See e.g. Fischoff (1982a) and Pohl and Hell (1996).

[39] Cf. e.g. the contribution of Sanna and Schwarz (2004).

[40] See e.g. Russo and Schoemacher (1992) for the overconfidence bias and Sanna, Stocker, and Schwarz (2002) for the hindsight bias.

When applied to single-cue anchoring, all these three procedures showed differentiated but altogether encouraging debiasing results if applied to single-cue anchoring. They revealed almost no efficacy if applied to the more complex case of sequence-anchoring.

5.4 Concluding Remarks on Debiasing and Some Implications for Theory Absorption

The studies on debiasing that have been sketched so far generally reveal a moderate robustness of the cognitive biases. Overall, the reviewed contributions provide further evidence for the systematicity with which biases occur. In addition they show how, in most of the cases, biases cannot be reduced to a mere artefact of a faulty task or interpreted as a result of the triviality of the task which causes a lack of motivation and interest among the respondents. Debiasing efforts could be typically proven to achieve significant modifications in the biases' occurrence whenever they were able to modify the subjective approach to the task by acting on the mechanisms ruling the individual perception of the task.[41]

Whereas the research on debiasing fixes some common patterns for elaborating corrective procedures, it still leaves several questions open. The overall result which emerges is that people "*are skilled enough to get through life, unskilled enough to make predictable and consequential mistakes; they are clever enough to devise broadly and easily applicable heuristics that often serve them in good stead, unsophisticated enough not to realize the limits of those heuristics*."[42]

Whereas given the impossibility of evaluating subjective judgement in a generalized way because of its irremediable specificity and context-dependency, the research on debiasing does not yield any results that can really be generalized and properly applied to a very wide spectrum of biases, some of its major findings can be summarized as follows:[43]

1. Biases mainly arise from the mismatch between the deterministic proceeding of the human mind and the probabilistic nature of the tasks.[44] When the individuals are asked to assess inferences, they typically rely on subjective procedures and rules of thumb (the bounded rational heuristics) instead of on the laws of logics and probability theory. The discrepancy between the probabilistic structure of the environment and the non-probabilistic solving procedures applied can mainly be made responsible for the systematic occurrence of mistaken judgements.
2. Under conditions of uncertainty the most delicate phase of decision-making is represented by the information interpretation which is necessary for structuring the task, rather than by information collecting. Biases are therefore most likely

[41] Cf. Fischoff (1982b, p. 440).

[42] Cf. Fischoff (1982b, p. 442).

[43] Cf. Fischoff (1982b, pp. 442, 443).

[44] Cf. Fischoff (1982b, p. 442).

to be concomitant to the (selective) interpretation of information. On this basis, the fact that in experimental settings the respondents can typically avoid the collecting phase, as information is typically provided, does not constitute a serious shortcoming of the experimental method and does not detract from its validity as a precious instrument of the debiasing research.

3. This can be related to the fact that biases can mainly be ascribed to the individual approach to the task. They are thus shaped by the interaction between the individual mental representation, accessibility of the task and its informational structure. The impact of biases can be reduced by acting on these variables, so that biases can be said to be non-substantive.[45]
4. Normative theoretical standards can be inferred as a benchmark for correct inferences. It follows that, when inference assessment is required and given the structure of the environment, an optimal response to which actual individual responses can be compared exists.
5. Debiasing only makes sense as long as it exclusively acts on individuals' intuitive judgements. Correcting procedures should therefore strive toward the improvement of the quality of subjective judgements by inducing the individuals to discover them as more favourable patterns of behaviour and to recognise them as such. Providing responders with external computational aids of various nature (e.g. calculators, computers or similar) does not belong to the instrumental repertoire for debiasing inference judgements.
6. Biases are meant to have a cognitive and not motivational nature. It means that a systematic mistake can be defined as a bias only if it can be alleged to shortcomings in the individual judgements rather then to incentives' or tasks' triviality.

From these findings it is possible to infer some considerations that can inspire and inform some of the procedures and methodological fundaments for the analysis of the self-referentiality of economic theories and in particular for what the testing of the absorbability of the so-called "mainstream theories"[46] may concern.

1. The first finding from debiasing points to one of the difficulties of the analysis of the absorbability of mainstream economic theories among bounded rational individuals. This is constituted by the discrepancy between the neoclassical rationality standard, which underlies such theories and the observable cognitive limitations which constrain the subjective rationality. As the mismatch between the individuals' mental tools and the nature of the task can often be made responsible for inadequate and biased behaviours, the mismatch concerning the rational standard respective of the mental tools which actually belong to the subjective cognitive repertoire and of the tools which would be required for properly conceptualizing and correctly estimating the theoretical support provided by the theory could interfere with the absorbability of that theory.
2. Neoclassical theories typically privilege the information integration rather than its discovery. Although neglecting the phase of the search for information

[45] Cf. Fischoff (1982b, p. 442).

[46] With "mainstream theories" are here simply meant theoretical paradigma inspired by the neoclassical approach and assuming the full rationality of the economic actors.

undoubtedly represents a limit of the neoclassical approach and seriously affects its descriptive power, insisting on the phase of interpreting and integrating information, which represents one of the utmost needs of the individuals in their decision-making, could raise the appeal of theories' prescriptive and strive toward their absorbability.

3. The third result from debiasing seems to provide empirical support to Morgenstern's intuition that the degree to which a theory will be absorbed decisively depends on its formulation and accessibility.
4. As the absorption of a certain theory points to the evaluation and acceptance both of its predictive and of its prescriptive content, it results in this way in testing the survival of a theory to its own acceptance (as can be expressed by the individuals' compliance with it). The problem of how bounded rational actors are able to process the content of theories of full rationality is worth being explicitly analyzed and considered even when looking for its possible application as an instrument for enhancing efficient economic advising. Normative prescriptions for economic advising could also be taken from theories of bounded rationality, if it could be proven that such theories enjoy a broader acceptance by the bounded rational actors and survive to their absorption.
5. While the strive for debiasing rules out the usage of computational aid for supporting the individuals in their decision-making, the investigation of theory absorption could make use of similar instruments. The automatic implementation of theoretical content could represent a useful way of testing the individual willingness of absorbing a certain theory because it would answer the question of whether the individuals would choose to comply with a theory if they only were able to.
6. The last result sounds quite reassuring because it states the disposability of the individuals to "work hard" on the task faced. This implies that they will most likely be willing to have a look at eventual theoretical information provided and try to assess it as feasible decisional support or not.

Some preliminary considerations on the absorbability of economic theories of full rationality could be based on the comparison between experts' and lay people's performance. This could enable researchers to specify if and under which conditions possessing superior knowledge really constitutes an advantage.

One of the advantages of experts and professionals is their experience on a set of relatively homogeneous tasks, which enables them to reduce the cognitive efforts needed and to rely on habitual patterns of solution and routinized procedures. Task-specific reinforcement might add and imply the definition of a manageable range of decisional clear-cut criteria that can be in many cases made responsible for on-target judgements. It should however be remembered that experts are in the rule prone to the same biases as lay people when they predict quite intricate problems on intuitive basis.[47]

[47] Cf. Tversky and Kahnemann (1982, p. 18).

In conclusion, debiasing and theory absorption are two research programs that present some similar features because they both represent an attempt to deepen the understanding of the human cognition and of the subjective rationality.

Effective debiasing requires the development of procedures which act in the "proper way." Similar procedures should be accessible to the human mind and affect judgment in the wished direction by acting on the real determinants of judgement. Otherwise, they would not permit the systematic blockades of the intuitive judgment to overcome. In this sense, both debiasing and theory absorption act where cognition experience fails. They also both insist on the outward boundaries of cognition.

Besides serving the pure speculative intent in exploring the psychology of choice, these two research programs could yield useful application e.g. in supporting the development of effective decision support systems, teaching methods, training procedures, giving policy advice etc. They could also enjoy a broad acceptance among bounded rational decision-makers.

Both the inquiry of debiasing and that of theory absorption emerge from balancing theoretical and empirical instances: "*Good practice will require better theory about how the mind works. Good theory will require better practice, clarifying and grappling with the conditions in which the mind actually works*."[48]

[48] Cf. Fischoff (1982b, p. 444).

Chapter 6
Self-Referentiality of Economic Theories and Theory Absorption

Bounded rational social actors are not just stimulus-response machines but complex beings, whose actions are led by their own beliefs and mental representations. Such representations can shape mental models, subjective theoretical frameworks that predict the course of the social system the actors are involved in and that establish cause-effect relations that the individual uses in her decision-making. Individuals can modify their mental models when they are not satisfied with the results of their application. This operation requires that the individuals are able to reflect the theoretical statements on themselves and on the situation they are confronted with. According to the result of this reflection process, they will then decide which theory they want to refer to, or in other words, which theory they want to absorb.

Economic theories aim at the description and prediction of economic behaviour and interactions, but at the same time interfere with the phenomena they aim to depict. Revealed theories, if accepted, may influence the behaviour of the agents they focus on, either in the sense of validation of the theoretical content, or in that of its rejection.

This analysis tries to discuss the implications of those recursive, or self-reflexive effects of economic theories on bounded rational economic behaving and interacting. In particular, a distinction will be made between the perception of the self-referentiality of a theory by bounded rational individuals (i.e. the perception of its applicability to a concrete setting) and its absorption (i.e. the compliance of the decision makers with the prescriptions of the theory).

The discrepancy between the neoclassical rationality standard and the observable cognitive limitations that constraint the subjective rationality further complicates the evaluation of the role of mainstream theories in influencing economic interactions and behaviour. The problem of how bounded rational actors process the content of theories of full rationality undoubtedly is worth being explicitly analyzed because it could particularly yield interesting results e.g. for enhancing efficient economic advising. However, normative prescriptions for economic advising could also be taken from theories of bounded rationality if it could be proven that such theories enjoy a broader acceptance by the bounded rational actors and survive to their absorption.

S. Sandri, *Reflexivity in Economics*: *An Experimental Examination on the Self-Referentiality of Economic Theories*,

The different approaches to the rationality of the individuals facing economic problems develop two almost opposite views of economics: while, on the one hand, the neoclassical teleological orientation codifies economics as a sophisticated theory of reasoning about action; on the other hand, the second view which insists on evolutionary (in the sense of adaptive) elements as the key for interpreting the human approach to economic problems, permits theorists to characterise economics as an empirical behavioural science.

Merging teleological with evolutionary aspects of analysis is a task which deeply challenges economics. Each of these set of aspects implies a profoundly different approach to the economic issues and therefore, yields substantial differences both in the formulation of economic theories and models and in the assumptions in regard to their absorbability. While according to the neoclassical approach, a strong form of theory absorption should be unquestionable; focussing on the boundaries of individual rationality allows interpreting theory absorption in a more adaptive way.

In the discussion of the absorbability of economic theories, the assumptions on which such theories rely have to be considered. Therefore, this chapter will be introduced by a short digression on established economic methodology. The concept of "theory absorption" will then be defined and discussed both in the spirit of the neoclassical theory and of the bounded rational analysis. The contributions of Morgenstern (1972), Morgenstern and Schwödiauer (1976) and Dacey (1976, 1981) will be considered in the case of neoclassical theory while the pioneering work of Güth and Kliemt (2004a) will illustrate the bounded rational analysis. After some tentative conclusions, economic policy advising will be discussed as one of the possible applications of the research on theory absorption.

6.1 Economic Methodology

This short digression discusses some fundamentals of economic methodology and presents (without claiming to be exhaustive) some of the most established conventions and methodological principles that are widely used and commonly accepted in economics. For a systematic critical overview on economic methodology, Hands (2001) can be suggested.

For simplification, the scientific method can be characterized either as based on an empiricist or rationalist conception and philosophical approach to the human knowledge.

While empiricism contemplates knowledge as something insolubly dependent on empirical evidence, rationalism interprets reason to be the main source ofknowledge.

In an empiricist view, any scientific conceptualization must be derived from individual direct experience and observation. On this basis scientific hypothesis can be developed by means of induction, which means that general conclusions are

inferred from particular concrete premises. Because of their higher generality, such conclusions are however not necessarily in possession of the same degree of certainty as the premises from which they have been generated and therefore, have to underlie empirical testing.

On the contrary, rationalism appells reason to be the main source of knowledge or justification.[1] Therefore, in a rationalist perspective reason represents the scientific criterion of truth while deduction is the key mechanism for deriving knowledge and making science. Deduction involves a sort of concatenated reasoning, i.e. a reasoning whose conclusions necessarily follow the stated premises which consists of previously known facts and issues.

These two different approaches to the philosophy of science are not mutually exclusive. The scientific praxis actually integrates them and has made use of both inductive and of deductive reasoning.

The epistemologic conception which underlies the two approaches is that of naïve realism, which does not question the existence of an ontological truth and believes in the possibility of asymptotic approach toward it but not of its complete capture.[2]

Other epistemologic conceptions question this stance and correspondingly posit different degrees at which science can penetrate reality, providing therefore alternative conceptions both of science and of reality. Realism, instrumentalism and constructivism are examples of the most influential traditions. From the relativist position of constructivim, which has been sketched in Sect. 1 of Chap. 3, realism opposes the view that scientific statements actually resemble objective reality, while instrumentalism interprets scientific statements as useful instruments for expressing the human experience of reality but does not necessarily reflect the world as it is. As pointed out by Dacey, "*A (methodological) instrumentalist views a theory T merely as a device of convience in organizing data, i.e., a theory is a mere instrument. A (scientific) realist takes a theory literally, adhering to the view that the theory (more or less accurately) describes reality*."[3]

Economics investigates what on the basis of human behaviour in economic situations. For doing that it adheres to the epistemological fundaments of the contemporary scientific method. Economic situations are represented by all those problems and circumstances involving scarcity of resources. In other words, economic problems are faced whenever decisions have to be taken because of and under given constraints.

Neoclassical economics typically relies on the fundament of methodological individualism which explains social phenomena as the result of the aggregation of individual decisions. From this stance neoclassical economics develops the axiomatic costruction of the "homo oeconomicus," interpreting the individual as a rational and self-interested utility maximizer.

---

[1] Cf. Lacey (1996).

[2] For an interesting discussion on the relation between theories and reality see Albert (1972).

[3] Cf. Dacey (1976, p. 254).

Following the homo oeconomicus approach, neoclassical economics conceptualizes individual decision-making and strategic interaction in the perpsective of rational choice. Therefore, it assumes an economy to be populated by rational actors who are in command of "*unlimited ressources to analyze models of their interaction situation and to draw inferences from those models*."[4]

Modeling choice as rational implies that: (1) rational actors choose how to act in a concrete situation on the basis of a model of their own action; (2) rational actors are perfectly able to ascribe causality of events either as due to their own action or to factors not belonging to the realm of the own activity; (3) they are capable of intentional and purposeful planning which relies on the anticipation of the consequences of the planned actions; (4) choices are made and evaluated according to individual maps of preferences which are represented by complete, continuous, convex, transitive and independent functions of the subjective utility; (5) the individuals assume the others to be endowed with the same cognitive tools and rationality standards they are.[5]

The model of rational choice is thus based on the counterfactual assumptions of common knowledge of rationality and rational expectations which solve the question of the external validity of the rational choice theory on their own axiomatic basis. Once full rationality of the individuals, common knowledge of that, and expectations of rationality conform behaviour are assumed, rational choice theory will be perfectly absorbable among rational individuals.

Rational theory assumes itself to be known (or better inferable on an educative basis) by all the individuals, who are all aware of that. None of them will therefore behave differently than postulated by the rational choice theory. This yields the predictive accuracy of the rational choice paradigm and accounts for the strong notion of absorption that can be assumed for it. In other words, rational choice theory justifies itself because of its own axiomatic fundaments and in this sense, resembles the mechanism of metalogic reflexivity.[6] As any deductively correct model can explain what is stated in its own premises and assumptions, observing the absorbability of the rational choice theory in real settings is a device for testing the descriptive validity of its assumptions and its predictive sharpness in relation to the individual (bounded rational) foresight.

In principle, i.e. without the empirical test of its absorption, an explicatory validity, rather than an explanative one, can be ascribed to the rational choice theory. Fully rational individuals are able to infer each of its precepts in an educated manner, which "*thereby 'explicates' what '(ideal) rationality' means but it does not 'explain' choice-making in any nomological sense*."[7]

In this sense, discussing theory absorption among fully rational individuals is equitable to the proof of the theory's internal validity, i.e. of its logical consistency,

---

[4] Cf. Berninghaus, Güth, and Kliemt (2003a, p. 486).

[5] Cf. Berninghaus, Güth, and Kliemt (2003a, p. 487).

[6] Metalogic reflexivity has been defined in Sect. 4 of Chap. 4.

[7] Cf. Berninghaus, Güth, and Kliemt (2003a, p. 489).

while relaxing any of the assumptions underlying full rationality (as for example that of common knowledge of rationality) yields interesting implications for the theory's descriptive validity.

6.2 Self-Referentiality of Economic Theory and Theory Absorption

To depict the fact that a social theory which is known to the actors who interact in a social system may affect the course of the social system itself, Morgenstern (1972) introduced the concept of "theory absorption." Although any economic theory can potentially be absorbed for the resolution of a concrete problem,[8] its absorption differs from case to case, depending on its formulation, its understanding, and its acceptance by the members of the economic system, as well as on its accessibility.[9] Furthermore, past experiences and learning may also matter when it comes to evaluating the absorption of a certain theory.

Among the elements that determine or influence theory absorption, self-reference plays a preliminary role in the sense that it can be characterized as a prerequisite for the absorption of a theory. Individuals self-refer a theoretical framework to support their decision-making and then choose whether to rely on it or not, in other words, whether to absorb that theory or not.

Thus, theory absorption is a consequence of the self-referentiality of the social theory. An individual self-refers a theoretical statement and, according to the results of such reflection, she will then choose either to rely on that theoretical framework, or on a different one.

In an ideal setting (i.e. populated by unbounded rational social actors) a theory of rational choice will be absorbed universally, such theory being at the same time descriptive and prescriptive of the full rational behaviour.

In a real setting, thus populated by bounded rational social actors, things are slightly different and more complicated. A requisite for a theory to be absorbed is that it can be understood by the individuals (i.e. it does not overstretch their bounded cognitive and computational capabilities) and that it can be integrated with subjective beliefs and mental representations. In particular, a theory can be absorbed if its content does not violate the normative components of beliefs as well as the beliefs about the others.

Because the recursivity and self-referentiality of a certain theoretical framework constitutes the logical premise of its absorbability, the notion of "self-referential theory" will be first of all specified. Then, the original formulation of the concept of "theory absorption" will be presented and discussed in its implications for a population of fully rational individuals. After relaxing the assumptions posited to the

[8] Cf. Dacey (1976, p. 248).

[9] Cf. Morgenstern (1972, p. 707).

individual rationality by the neoclassical approach, some issues concerning theory absorption among bounded rational individuals will be finally considered.

6.2.1 Self-Referential Theories

The prerequisite for the absorbability of a social theory has to be sought in the potentially inevitable self-referring nature of any conceptualization about social phenomena. The notion of "self-referential theory," which therefore assumes a central meaning in discussing theory absorption, is however feasible for different interpretations depending on the degree to which a theory's reflexive implications can be appreciated.

Some of the considerations that follow can be associated with what has been previously discussed about the reflexivity of human theorizing and knowledge (cf. Sect. 8 of Chap. 1). These considerations are further informed by some results and conclusions from the "sociology of scientific knowledge"[10] approach.

Science can be characterized as a social activity which is done in communities. Therefore, it cannot avoid reflecting some of the paradigms that are collectively held among a certain community,[11] which will at least inform the research agenda. In this sense, scientific theories do not differ from any other intellectual artefact in that they constitute of socially held beliefs, which are thus coined by the society and at the same time coin the society to which they refer and by which they have been created.

Relying on that, the approach of the sociology of scientific knowledge postulates that there is an "*explanatory symmetry between scientific and other social beliefs*."[12] This symmetry raises the issue of which theory ought to be applied for examining scientific beliefs because they specify themselves the fundamentals on which their own analysis has to be based.

This can be referred to theory absorption, as well, since the absorption of a certain theory sets in motion a recursive mechanism between the content and the object of that theory. The reflexivity implied by the back-coupling between a social theory and the social phenomena it aims to describe, therefore, adds to the reflexivity involved in the analysis of scientific beliefs and statements.

The back-coupling between the theory and its object is influenced by many different factors (e.g. understanding, acceptance etc.) but is essentially determined by the perception the social actors have of its self-referring character in the concrete situation they face.

There can be different notions of self-referential theories, at first because there is much hidden behind the concept of "theory" (as it has also been pointed out by the introductory considerations on the scientific and economic method) and then because of the polymorphism recursive relations may assume.

---

[10] For more on the sociology of scientific knowledge, cf. e.g. Davis and Klaes (2003) and Hands (2001).

[11] Cf. Hands (2001, p. 173).

[12] Cf. Hands (2001, p. 173) (italics omitted).

"Theory" can be meant as one or several statements that describe a property, a peculiarity or, more generally, a certain aspect of a certain object. An economic theory will be conceived here as a conditional generalization, i.e. as a statement of the sort: "*for every x, if P of x, then Q of x*."[13]

In the perspective of radical constructivism,[14] every sort of theorizing would be self-referential because of the interdependence between observer and observation. The approach of the sociology of scientific knowledge yields similar conclusions in that it questions the view of a scientist standing in a disentangled relation to the world which constitutes, in the end, the object of her analysis.[15]

However, in a strict sense, a theory should be said to be self-referential if it refers to itself, i.e. if it contains sentences or theorems related to the theory itself, as illustrated for example by meta-theoretical considerations. In this acceptation, a self-referential theory is a theory which applies to itself or which in other words constitutes one of the objects belonging to its own realm of application.

In a broader sense, a theory that deals with something that can modify or affect the validity of its content, also implies a recursive relation and thus can be said to be self-referential. All conceptualizations, where presuppositions involved in knowing are sought, belong to this category of recursivity. This implicates a meta-logical referring and occurs whenever observer and observation are part of the same system.

The social sciences are intrinsically exposed to this sort of recursivity since they are the product of the reflection of individuals upon selected "facets" of the individual in her social system. Social sciences aim to describe a system made by an observer, who inevitably is part of the system observed. Considerations on meta-logical reflexivity undoubtedly suit the social sciences and indicate their self-referential nature because the kind of social theory which is known to the social actors who are interacting in a social system can affect the social system itself. "*There is thus a "back-coupling" or "feed-back" between the theory and the object of the theory*"[16] depending on which social theorizing can have both a self-supporting or a self-refusing impact on the social actors and on the social system it aims to analyze. As social actors can react opportunistically or opposite to the theorizing about themselves, every theoretical statement can be either invalidated or reinforced by the actors' behaviour, as in a feedback loop.

6.2.2 Introducing the Notion of Theory Absorption

The notion of "theory absorption" was first introduced by Morgenstern (1972) and accompanied by the following considerations:

---

[13] Cf. Dacey (1976, p. 249).

[14] Radical constructivism has been discussed in Sect. 1 of Chap. 3. For an introduction see e.g. Rusch (1999) and Schmidt (1987).

[15] Cf. Woolgar (1992, p. 334).

[16] Morgenstern (1972, p. 707).

"*Nature does not care – so we assume – whether we penetrate her secrets and establish successful theories about her workings and apply these theories successfully in predictions. In the social sciences, the matter is more complicated and in the following fact lies one of the fundamental differences between these two types of theories: the kind of economic theory that is known to the participants in the economy has an effect on the economy itself, provided the participants can observe the economy, i.e. determine its present state. […]*

However, the distribution of the kind of theory available, and the degree of its acceptance, will differ from one case to the other. This in turn will affect the working of the economy. There is thus a 'back-coupling' or 'feedback' between the theory and the object of the theory, an interrelation which is definitively lacking in the natural sciences. […]

In this area are great methodological problems worthy of careful analysis. I believe that the study of the degree of 'theory absorption' by the members of the economy and the study of the above mentioned embedding relationship will make of all us more modest in judging how far we have penetrated into the economic problems."[17]

Its first appearance in Morgenstern's 1972 article "Descriptive, Predictive and Normative Theory" does not reflect the real history of Morgenstern's research program on theory absorption. With this concept Morgenstern addressed one of the topics that had been a central concern all his life. In this regard, "theory absorption" can be characterized as a subsuming label and as a "*refinement of ideas*"[18] condensed in a single notion, a conspicuous amount of speculations.

Concerns that the feedback which is established between the economic theory and its argument can result in the unpredictability of the economic system already find expression in Morgenstern's 1928 essay "Wirtschaftsprognose."[19] In the core thesis of this essay is the consideration that in economics "*the very fact of forecast leads to 'anticipations' which are bound to make the original forecast false*."[20] In this sense, economic predictions can only be self-falsifying since, as soon as they are captured by the agents belonging to the economy, they induce them so that they opportunistically adjust their behaviour to the forecast.

Speculations in regards to the degree at which agents are aware of a theory, play a central role in Morgenstern's discussion of the relations between the economic theory and economic policy, as he discussed in his 1934s contribution "Die Grenzen der Wirtschaftspolitik." His 1935 contribution further developed this thematic underlining of the incompatibility between "Perfect Foresight and Economic Equilibrium."

Morgenstern did not elaborate a formalized definition of theory absorption but rather focussed on the plurality of elements which might affect the way the economic actors interrelate with a certain theory and which should therefore be considered for the analysis of the recursive effects of a theory and its absorption.

---

[17] Cf. Morgenstern (1972, pp. 706, 707).

[18] Cf. Dacey (1981, p. 111).

[19] Some further aspects of Morgenstern's position on economic forecasting have been shortly addressed in Sect. 2.2 of Chap. 3.

[20] Cf. Margert (1929, pp. 313, 314).

6.2.3 Determinants of Theory Absorption

Discussing the absorption of a theory among the members of an economic system involves the consideration first of the methodological fundaments of economics, second of the standards met by the human rationality, and finally of the individuals' in- and foresight capabilities.

As theory absorption can be essentially characterized as a meta-theoretical operation, its inquiry should include considerations on the methodology on whose basis theories are formulated, i.e. on the assumptions and simplifications they rely on. A problem in this insight is constituted by the embedment of the economic reality in the social reality. While any statement on a certain economic phenomenon inevitably reflects its abstraction from other aspects of reality,[21] "*the economic 'reality' is never a whole, as in the physical universe: Economic phenomena are embedded in a political, legal, moral and ideological world from which influences come, also determining events*."[22]

Evaluating to which extend a theoretical substitute applies to its pre-theoretical concept[23] is a further difficulty worth being considered in a methodological point of view. It expresses the degree to which an explication stands for the explicated phenomenon and informs the resemblance of the theory with the object it describes, as well as its internal coherency. It is hereby sought to appreciate the extent to which a theoretic model can be truly descriptive of the phenomenon it models. It can be pragmatically conceived that, in principle, theories may be descriptive of reality and that from case to case a certain theory "*may get separated from reality either because its assumptions are at fault, have become obsolete, e.g. for reasons of evolution, or because the concepts used in the description can be replaced by better ones*."[24]

Theory absorption depends on the real, actual degree to which a certain theory is reflexive and thus depends on the individual hindsight, foresight and mental capabilities, so far as those elements may affect the "acceptance" of a certain theory.

In this sense, a theory is feasible to be reflexive for certain individuals if it handles a setting which can be related to the real situation the individuals face. Such a theory can then be said to be reflexive for those individuals if they actually perceive it as being related and applicable to the circumstances they are involved in.

Further, the acceptance of a theory by certain individuals is influenced by their degree of confidence with it. In order to absorb a certain theory, the individuals have to appreciate it as a normative benchmark on which to inform their decision-making and have to consider it convincing.

---

[21] This is in Morgenstern (1972, p. 707), explicitly related to predictive theory and expressed as follows: "*Thus the prediction of economic phenomena implies statements about their separability (or lack of it) from the other social world of which they are only a part*."

[22] Cf. Morgenstern (1972, p. 707).

[23] For a definition of "explication" see Carnap (1956).

[24] Cf. Morgenstern (1972, p. 703).

The way in which any descriptive theory can assume a normative character is ruled by its acceptance, as the "*normative property is based on the acceptance of the theory*."[25] According to Morgenstern, there are essentially two points which determine the acceptance of the normative order proposed by a theory: "*The first point relates to the use of theory: if the individual, for example expresses the desire to behave 'optimally' in a specific situation or environment, the theory will tell him how he ought to behave. The individual will then follow this advice if the theory is 'absolutely convincing'. This 'convincing' means the intellectual (and practical) acceptance of the theory for its predictive worth. The individual or agent […] has to understand the theory in order to develop this degree of confidence. Only then will he allow his behaviour to be guided, and possibly be changed by the theory*."[26] The second point refers to the fact that the acceptance of the normative of a theory "*involves the acceptance of institutions*."[27]

Foresight capabilities influence the acceptance of a theory and therefore, rule its absorption, since purposeful action depends on the subjective predicting capabilities. As Morgestern (1935) discusses, the assumption of perfect foresight might invalidate theoretical equilibrium predictions. This especially applies to those cases where the equilibrium predictions do not resist their absorption, i.e. when they are based on a self-destroying dynamic. In similar cases, only admitting individuals who are not perfectly forward-looking (i.e. not fully rational) could account for the equilibrium stability.

Finally, the acceptance of a theory can also be affected by individual hindsight, that is, by the shadow of the past. Morgenstern considers that the formulation of a theory, its understanding and accessibility by the members of an economy, "*[…] will be variously distributed as far as particular theorems are concerned as well as regards concrete circumstances of the economy in question*."[28]

6.2.4 Some Theoretical Implications of Theory Absorption

Neoclassical economics assumes theory absorption in its strongest form, i.e. as full theory absorption. This is allowed because of the neoclassical interpretation of choices as emergent from the individual rational pursuit of ends with common knowledge of that.[29]

From the perfect teleology of action, which is implied by the neoclassical paradigm, follows that the absorption of a theory can be conceived as a process which is

[25] Cf. Morgenstern (1972, p. 712).

[26] Cf. Morgenstern (1972, p. 711).

[27] Cf. Morgenstern (1972, p. 712).

[28] Cf. Morgenstern (1972, p. 707). The text continues as follows: "*For example, the population of a government of a country having just gone through a period of inflation – and therefore having accumulated experience – will react differently to a new inflation than a country having had no such recent experience. […]*."

[29] Cf. Güth, Berninghaus, & Kliemt (2003b, p. 2).

steered by induction.[30] Rational actors are defined axiomatically and can therefore only act rationally, so that the "*teleological theory explaining the actions of rational actors is commonly known and is in fact applied by them in choosing their own actions*."[31] Therefore, rational choice theories consider themselves to be inferable on an educative basis by every (rational) actor with common knowledge of that. Thus, in a neoclassical perspective, theory absorption can be characterised as an ultimate test of the validity of economic theories.

Since a theory of rational choice should constitute a rational standard of behaviour, it should survive its own acceptance and not reveal itself as being self-defeating. As Morgenstern and Schwödiauer (1976) point out: "*One of the main requirements of a satisfactory, or at least acceptable, concept of solution (i.e. rational behaviour) is that it ought not to be invalidated by the knowledge of the theory on part of the participants in the game – it should be immune against 'theory absorption'*."[32]

On this basis they criticize the interpretation of the core as an equilibrium solution for a market game, arguing its non-absorbability. The absorption of the core as the set of competitive equilibria would imply collusive re-contracting against which the core would not resist. Therefore, it would loose its equilibrium predicting validity because the "*knowledge of the core on part of the traders may result in a collusive stabilization of some dominated imputations*."[33]

The notion of the core of a market game is the set of non-dominated price-profiles and its introduction can be traced back to Hedgeworth's oligopoly solution of the "contract curve."[34] Competitive contracting should come to an end in the core because the profit maximizing traders won't find any trading arrangement which yields higher profits than those belonging to the core. They will thus agree on one of the core's price profiles.

As is the usual case of behavioural theories, the core does not survive its knowledge and acceptance as equilibrium prediction. It is, in other words, not immune to its absorption, in the sense "*the traders are rational – in the sense of always striving for higher profits – and if they knew that the core were the only stable outcome of the bargaining process, then the core would not be stable*."[35] Knowing that the core represents the equilibrium prediction would in all likelihood induce rational profit-seeking traders of the same type to collude, that is, to pre-contract and commit to otherwise dominated agreements. This implies, that "*competition can only prevail if the behaviour of the traders is characterized by a peculiar mixture of rationality, complete information about the opportunities and the market offers, and short-sightedness*."[36] As this clearly violates the rationality assumptions on which

[30] Cf. Morgenstern (1972) and Dacey (1976, 1981).

[31] Cf. Güth, Berninghaus, & Kliemt (2003b, p. 2).

[32] Cf. Morgenstern and Schwödiauer (1976, p. 218).

[33] Cf. Morgenstern and Schwödiauer (1976, p. 217).

[34] Cf. Hedgeworth (1981).

[35] Cf. Morgenstern and Schwödiauer (1976, p. 219).

[36] Idem.

the model of the market is based, the prediction of the core as feasible equilibrium is a self-defeating prophecy. It constitutes therefore "*an example of a theory of rational action the knowledge of which (on part of the actors) destroys its predictive validity*."[37]

The argument for the stability of the core as the set containing the equilibrium brings into question its suitability as rational standard of behaviour. The alternative solution concept proposed by Morgenstern and Schwödiauer is that of a stable-set solution, as conceived by von Neumann and Morgenstern (1944).

A stable-set solution can be defined as a closed set of price profiles which do not dominate each other ("internal stability" condition) but which can be dominated by some profiles not belonging to it ("external stability" condition).[38]

While the property of internal stability applies to both the stable-set and to the core solution (the different price vectors belonging to the core indifferent to a rational trader), the property of external stability is a prerogative of the stable-set solution because it is not always granted by the core. The strength of the stable-set solution and its sustainability as a rational standard of behaviour is that it "*provides a consistent and defensible rule of division not only for the case of competitive behaviour (the core is always part of the solution) but also for all conceivable collusive arrangements*."[39] The stable-set is immune to its absorption because its knowledge and acceptance as a solution concept would not violate its predictive validity.

Therefore theory absorption can be also considered to be a tool for testing theories of full rationality at their internal consistency and stability of equilibrium predictions. The full absorbability of a theory represents a necessary condition for a theory to represent a rational standard of behaviour in a neoclassical perspective. In this perspective, testing the theory's absorption is an operation which essentially involves the logic of a second-order conceptualization, as it questions how rational individuals would behave once they become aware of a certain theory.

Relaxing the full rationality assumption, i.e. allowing for boundaries in the individual rationality, testing the absorbability of a certain theory (both of bounded and of unbounded rationality) yields different results. It reveals that if bounded rational decision-makers perceive that certain theory to be reflexive (i.e. to be applicable to the situation faced), if that theory induces the individuals to modify their behaviour, and if it is considered to represent a normative standard, the individuals would like to conform to it.

However, the implications the enquiry of theory absorption may yield also depend on the assumptions that are chosen to model the logic through which individuals make inferences.

Dacey's contribution centres on this aspect and explores which consequences the choice of the logic of inference may have for the absorption of a theory from a philosophical point of view.

---

[37] Idem.

[38] Cf. Morgenstern and Schwödiauer (1976, p. 229).

[39] Cf. Morgenstern and Schwödiauer (1976, p. 219).

Dacey formalises the absorption of an economic theory as an inductive acceptance process resulting from the maximization of epistemic utility functions. In doing that, he leans on the philosophical approach which explains the process of scientific inference by means of decision-theoretic concepts[40] and interprets "*the acceptance and rejection of scientific hypothesis as a process of maximizing epistemic utilities*."[41] Epistemic utilities "*represent preferences over cognitive objectives of scientists, for example, truth, information (in the technical sense of 'amount of information'), explanatory power, and simplicity*."[42] Dacey applies this notion to the individual choice of absorbing a theory and assumes this choice to be based on individual cognitivist inductive logic. This is defined by the pair $I = \langle P, U \rangle$, whereas P is a measure of the inductive probability[43] associated with certain sentences of a language[44] (which consists of conditional generalizations and represents in the end a theory) and U indicates an (expected) epistemic utility function.[45] The (expected) epistemic utility function should provide a measure of the logical content of a generalization g relative to the evidence e, and can be therefore specified as

$$U(g|e) = P(g|e) - p(g) \tag{6.1}$$

A theory T is a sentence of the language over which P is defined if it introduces a new concept to that language.

Dacey's analysis demonstrates that the test of a theory at its absorption is in principle never possible. He corroborates his thesis in that he formally discusses "*three variants of absorption: absorption on the basis of factual evidence alone and both instrumentalist and realist absorption on the basis of factual evidence and new (theoretical) conceptual evidence*."[46] Whereas the first variant seems to point to theory absorption in its most neoclassical acceptation (rational actors interacting with other rational actors cannot act but rationally and cannot expect but rational behaviour of the others), the remaining two variants emerge as a consequence of assuming that different individuals rely on different logics of cognitive induction.

This difference can be formally stylized by introducing different functional specifications of the epistemic utilities, each of which expresses a different Weltanschauung toward the cognitive elaboration and processing of additional knowledge. Thus, he distinguishes between two forms of cognitivist logics, the instrumentalist and the realist ones. In particular, if a theory T is conjoined with the evidence e, an individual may incorporate it into her epistemic utility in two different ways, which are:

$$U_1(g|e\&T) = P(g|e\&T) - P(g) \text{ and } U_2(g|e\&T) = P(g|e\&T) - P(g|T) \tag{6.2}$$

[40] Dacey mentions e.g. the works of Hempel (1960, 1962), Levi (1967a, 1967b).

[41] Cf. Dacey (1976, p. 250).

[42] Cf. Dacey (1976, p. 250, 251), in reference to Niiluoto and Tuomela (1973).

[43] For more on the properties of inductive logic see e.g. Hintikka, Suppes (1966).

[44] Moreover, the language is assumed to be a monadic, first-order language. See for more Dacey (1976, p. 251; 1981, p. 116).

[45] Dacey relies on the notion of epistemic utilities as introduced by Hintikka and Pietarinen (1966). Cf. Dacey (1976, 1981).

[46] Cf. Dacey (1976, p. 254).

The adoption of U_1 rather than U_2 reflects the individual philosophical stance on the relation between theory and reality. In particular, "*U_1 seems fitted to the position of a (methodological) instrumentalist, whereas U_2 seems natural for a (scientific) realist*."[47]

Each of the two different cognitivist inductive logics informs an alternative logic for the absorption of a theory which can either be an instrumentalist or realist logic of absorption.

The practical relevance of Dacey's elegant formal construction can be argued to a great extent. The questioning starts with how far a process of maximization of epistemic utilities may contribute to the better understanding of the individual approach to theories for the solution of practical problems. Dacey's analysis brings about the attempt of explicating (what in a neoclassical spirit has possibly been too often interpreted as equivalent to formalizing) Morgenstern's notion of theory absorption. By analysing some of the implications theory absorption yields, economics and economic theorizing, he discusses the possibility of the testability theory absorption.

Dacey further investigates several of the factors which result, at a purely theoretical level, in the uncontrollability of the absorption of a theory, focussing, among other things, on the disruptive consequences of inexact (ambiguous) information. Overall, his analysis "*suggests that the concept of theory absorption leads to major problems in the methodological foundations of inference as well as in the foundation of the social sciences*"[48] and "*reinforces Morgenstern's recommendation for modesty on the part of the economists (and social scientists generally)*"[49]

6.3 Theory Absorption among Bounded Rational Decision-Makers

Several attempts have been made to reduce the bounded rationality approach to the homo oeconomicus approach, mostly relying on the consideration that the social actors, though not fully rational, act "as if" they were. Among others, the generally accepted view that "*The aim of a good theory is prediction and in prediction lies the ultimate test of validity*"[50] has added some plausibility to the "as if" argument. But, in spite of that, the fact that in simple settings bounded rational best replies may coincide with optimal responses, is far from endorsing the "as if" approach. Moreover, there is no evidence that satisficing behaviour can simply be read as a step towards optimization.[51]

Furthermore, it can be argued whether the proof of a behavioural theoretical framework, thus its validation, can solely rely on its accuracy of prediction, instead

[47] Cf. Dacey (1976, p. 254), italics omitted.

[48] Cf. Dacey (1981, p. 133).

[49] Cf. Dacey (1976, p. 248).

[50] Cf. Morgenstern (1972).

[51] For more cf. Güth and Kliemt (2004a, p. 522, 523).

of also trying to describe the way a decision emerged. In other words, the validity of a behavioural theory should be proved "from inside," and in this sense the proof of the absorption of a certain theory could also be interpreted as an ultimate test of the validity of such a theory in a concrete setting. As stated, the prerequisite of theory absorption is a self-application of a theoretical framework. Its conditions are its coherency with the subjective mental representations. The fact that individuals absorb a theory implies that they accept that they will rely on it - i.e. that they are first able to conceptualize it in order for them to share the rationality standard underlying such an approach - and then that they adhere to its logic. A theory that passes such a test can be said to assume a real descriptive validity of the behaviour of "human beings" dealing with economic decisions and not just of stylized economic subjects.

A theory is said to be absorbed by an individual if that individual internalizes it in her own mental models and chooses to act according to its logical content. In interactive contexts, theory absorption will also be strongly related to the supposed mental models of the others. It can be distinguished among unilaterally-, partially- and fully-absorbable theories, depending on the number of individuals - from one to all - who follow its prescriptions and are satisfied with the result.[52]

6.3.1 Individual Absorbability of Theories

A theory can be said to be unilaterally absorbable by a bounded rational decision-maker only if that individual considers herself to be in possession of the theory and chooses to comply with it, will be satisfied with the theory's predictive and prescriptive.

This means that both the extent to which future developments are contemplated and predicted by the theory and the prescriptions that are claimed, meet the individual aspiration level and can therefore inform choice and decision-making. In other words, a theory is individually absorbed if for that individual who absorbs the theory there is no need or reason for deviating from the theoretical prescriptions.

Because of the teleological perspective underlying any theoretical frameworks, predictive and prescriptive components of a theory are tightly related to each others. Both of these components matter for the absorbability of theories in that a purpose-oriented and forward-looking individual can be reasonably expected first to prove the theoretical predictions in principle meet the aspiration level and eventually to choose to behave according to the theory's prescriptive.

As the prescriptive content of a theory associates certain individual behaviours to certain consequences and future developments, the mechanism of individual theory absorption can be decomposed into the following steps: (1) only the individual who is in possession of the theory considers the problem she faces to belong to the theory's domain of application, which means, the individual perceives the theory as self-referring; (2) the theory's predictive, in principle, fulfils the individual aspira-

[52] Cf. Güth and Kliemt (2004a, p. 523).

tions; (3) the corresponding behavioural prescriptions are manageable to the extents of the individual's subjective rationality. Whenever this holds, (4) the individual will choose to rely on the theoretical framework and will thus comply with the theoretical prescriptions. If the individual is satisfied with the feedback she receives from the compliance with the theory, her behaviour will definitely stick to the theory's prescriptive and won't deviate from it.

A theory can be individually absorbed when its absorption exclusively depends on the individual who is in possession of it. In particular, when deciding about absorbing the theory, the individual does not have to build beliefs about other individuals who interact. Assuming the individual to be fully rational, not only theories of individual decision-making can be individually absorbed, but also those theories prescribing dominant strategies in strategic interaction settings which grant such possibilities. For example, in an ideal setting where a strict dominant strategy exists for a certain individual, a theory which advises that the individual choose such a strategy profile must be absorbed by that individual.

However, as soon as bounded rationality is admitted, things become slightly different, because the "*practical value of advice to a world of rational being is quite precarious in a world of less than fully rational individuals*."[53] Hereby, the individual absorbability requires in addition that the theory's prescriptive has to be perceived as "*good advice*."[54]

In trying to make the notion of "good advice" more operative, considerable difficulties are encountered. Good advice should sound "reasonable" to bounded rational individuals, should be workable, processible and understandable. This implies, in particular, that it does not have to overwhelm their limited cognitive and computational capabilities.

A theory is absorbed in a real setting if it is apt to guide the individual in her decision-making and inform her choice, meaning that the theory's advice can be considered as a standard of behaviour. This can be brought into action only if it does not overload the individual computational capabilities, which means it is unlikely that a too sophisticated theory will be absorbed. For example, simple heuristics such as "Take the Best"[55] or one-reason decision-making[56] in general could be favoured as stopping rules for making inferences to the rational choice prescription of extending the search until the marginal benefits of information acquisition are perceived to exceed the marginal costs.[57]

Theory absorption could hereby represent a criterion of theory selection, whereas if the inquiry encompasses both theories of bounded and of unbounded rationality, uniqueness of theory selection and absorption might not hold.[58]

---

[53] Cf. Güth and Kliemt (2001, p. 4).

[54] Cf. Güth and Kliemt (2001).

[55] Cf. Gigerenzer and Todd (1999).

[56] Idem.

[57] Cf. e.g. Stigler (1961).

[58] This motivates, among else, the choice of focussing on testing the absorbability of rational choice theory in the experimental part of this study (see Chap. 7).

Following Güth and Kliemt (2004a), some related aspects of individually absorbable theories can be illustrated considering the absorbability of routine behaviour in the secretary problem.

In the secretary problem[59] an individual has to select one of different items that are presented sequentially, whereas rejection is an irrevocable choice.

The standard formulation for this problem of optimal stopping is the choice of hiring a secretary: interviewing as many candidates as possible is useful for collecting information but is a costly process. Adding to that, waiting too much to decide whom to employ could imply that the best candidates be hired by someone else.

This illustrates a task, in which in each of the $N(\geq 2)$ rounds, the uppermost of a staple of $N(\geq 2)$ randomly shuffled cards, each of which yielding a different monetary value, is shown to a subject. In each round, the subject can choose between accepting the card for gaining the monetary reward specified on it or rejecting that card, so that a new card will be turned. In the final round, the monetary value of the last card will be paid to the subject.

According to the theory of rational choice, an individual facing a similar task should first specify a distribution for the values the cards may assume and then determine an optimal stopping rule. This represents a quite demanding procedure for lesser mortals who could more likely inform their decision-making to the formation of aspirations and to the search for a satisficing option. It is further plausible to assume that aspirations can be adapted, in particular, as more final rounds are approached. Proving the absorbability of this routine of bounded rational behaviour would corroborate its coherence with the individual mental models and with the bounded rational search stopping procedures.

6.3.2 Full Absorbability of Theories

In interactive settings a theory can be said to be fully absorbable if all the actors choose to follow the theory and are satisfied with the results of its application, so that anyone of the actors have a reason for modifying her behaviour or the theory to comply with.[60] This, in particular, requires that the fact that all actors comply with the theory does not per se provide an argument for deviating from its prescriptions.

A fully absorbable theory has therefore to admit an equilibrium which is behaviourally sustainable among a population of bounded rational individuals. Thus the theory has to state the "*minimum conditions for reaching an inter-personally sustainable 'reflective equilibrium' [...].*"[61]

In principle, all theories of rational choice should acknowledge their full absorption since by assuming rational expectations they grant the self-reinforcement of the

[59] Cf. Bolle (1979).

[60] Cf. Güth and Kliemt (2004a, p. 523).

[61] Cf. Güth and Kliemt (2004a, p. 528).

compliance with their prescriptions ("*rational actors expect the theory to be true and by behaving accordingly make it true*.").[62]

In addition to the requirements necessitated by individual absorbability of theories, full absorbability requires that the expectations concerning the behaviour of the others conform to theoretical predictions.

However, among a population of bounded rational actors it can be argued, on the one hand, whether considering a theory which prescribes full rational actions it is reasonable for an individual to expect all of her counterparts to obey the theory and behave in a rational way. On the other hand, a theory of bounded rationality can be fully absorbable only if it is apt "*to describe how actors are influenced by the theory itself and how they expect others to be so influenced etc*."[63] In this sense, full absorbable theories have to be reflexive and able to contemplate their own reflexive implications.

This is another way of looking at the dilemma challenging economics and its methodology: even if, on the one hand, the solution proposed by the neoclassical approach on basis of the assumptions of full rationality and its common knowledge are not sustained by the evidence, on the other hand, the economic theories, which assume individual bounded rationality, are obviously not provided on the axiomatic basis for correcting the anticipated behaviour of the others. A theory of bounded rationality which relies on "*the premise that boundedly rational actors assume other boundedly rational actors to behave according to the theory and themselves behave according to that theory*"[64] could offer a possible solution. It could be experimentally tested, for example, to confirm whether or not this assumption is sustained by the evidence.

In interactive situations the individuals think strategically. They form beliefs about the behaviours of the others and then (boundedly) best respond accordingly.

Among fully rational individuals, rational and mutual consistent beliefs lead to the same rational behavioural patterns which enable an (unique and stable) equilibrium to be identified.[65] The assumption of full rationality can therefore neither explain the possible heterogeneity of behaviour (as all individuals share the same rational beliefs), nor its persistency (as sub-optimal behaviours are eliminated by optimal rational ones).

Beliefs may be non homogeneous because they stem from lots of different, highly subjective factors.[66] They can be mistaken and also modified[67] if agents are not satisfied with the result of their application. The factors that shape the subjective beliefs cannot be classified exhaustively, essentially because the individual, as a

---

[62] Cf. Güth and Kliemt (2004a, p. 528).

[63] Idem.

[64] Idem.

[65] The general mechanism through which beliefs are updated is learning, which takes place, when according to Camerer (2003a) a change in behaviour due to experience can be observed.

[66] Among them can be mentioned past experience, knowledge, expertise, social norms and individual perception of them, risk propensity and its frame, etc.

[67] Cf. Camerer (2003a, p. 265).

social actor, is a non deterministic product of her social history.[68] The beliefs about the others, though also influenced by the individual social history, can be much more characterised as the product of introspection.

It seems reasonable to assume that the individuals perceive themselves not as perfect but bounded rational, and that they suppose the others possess the same cognitive and computational capabilities.[69] To form beliefs about the others, the individuals can use the tool of introspection. This is the only way to try to predict the others' behaviour since the bounded rational behavioural pattern cannot be logically inferred. Obviously, being the expression of bounded rational individuals, introspection can only be bounded, too. Through it an individual is only able to conceptualize a finite number of iterations of strategic thinking and will then respond strategically to what she presumes to be the last iteration of thinking performed by the others. In this sense, each individual perceives herself as being the most sophisticated.

Anew, a theory is fully absorbable if its predictions, as well as its expected absorption do not question its predictive content. Full absorbability holds for strategic equilibria[70] by definition, since it is optimal to behave according to the expected equilibrium. Guessing games exemplify a case in which the correct anticipation of the others' behaviour improves a player's chances of winning more than the knowledge of the equilibrium predictions alone. The absorbability of the guessing game theory by bounded rational individuals will be discussed in Sect. 5 of Chap. 7 in more details and related to the main findings of an experimental study.

In some other cases, however, theoretical equilibria predictions could be questioned on basis of their absorbability. This is, among others, illustrated by theories which focus on the discrepancies between individual and collective rationality (e.g. theories on the provision of public-goods and social traps in general),[71] as well as by behavioural theories (e.g. theories of oligopolistic competitive equilibrium).[72]

For behavioural theories absorbability may not be given: for example, in bilateral negotiations[73] a theory, which predicts an agreement at the means of the initial demands, can be confirmed with some intuition and empirical observations.[74] This would, in all likelihood, inspire outrageous initial demands, to which the theory would no longer apply.

Further, theories of herding behaviour explain rational basis phenomena according to which the rational choice of conforming might lead to collective irrational outcomes exactly because of the compliance with the theoretical advice, i.e. because of its absorption. This is, for example, the case of informational cascades, whose discussion in Sect. 6 of Chap. 7 will be informed by the results of an experimental examination.

---

[68] See e.g. Berg, Dickhaut, and McCabe (1995) and Mistri (1997).

[69] There is also evidence showing a clear tendency of overestimating the coincidence between the own and the others' motives.

[70] Cf. Cournot (1838) and Nash (1951).

[71] Cf. e.g. respective Coase (1974) and Rothstein (2005).

[72] As discussed in Sect. 2.4 of Chap. 6, according to Morgenstern and Schwödiauer (1976).

[73] Cf. e.g. Raiffa (1982).

[74] Cf. e.g. Pruitt (1981) and Sebenius (1992).

Overall it emerges that, in bounded rational settings, the common knowledge of a theory does not eliminate strategic uncertainty. Among bounded rational individuals, the common knowledge of a theory does not imply its absorption, which represents the necessary condition for strategic uncertainty to be ruled out, providing a reliable basis for predicting the behaviour of the others.

6.3.3 Partial Absorbability of Theories

A theory is said to be partially absorbable if among the population of $n(\geq 2)$ individuals, there is a subset $m(\geq 2)$ of subjects whose compliance with the theory would not provide any reason for modifying the own behaviour or for revising the theory to rely on.

While manifestly in a full rational setting partial theory absorption cannot hold, partial theory absorption is conceivable among bounded rational actors, as, for example, illustrated by a repeated public-good game with voluntary contribution.[75]

Assuming full anonymous voluntary contribution and that the subjects are solely informed of the total number of contributions after each round, the individual dominant strategy is that of non-contribution. It would lead, if adopted by all of the players, to a Pareto inferior result. A possible stylization for this problem is that of a repeated n-person prisoner's dilemma in which for the individual i, contribution yields a payoff of $f_i(k)$, while defection that of $g_i(k)$, with k being the number of contributing individuals in the round $(0 \leq k \leq n-1)$.

If $(k-1)$ individuals contribute, the payoff of the individual i will be $f_i(K-1)$ if she cooperates and $g_i(k-1)$ otherwise, whereas $f_i(k-1) < g_i(k-1)$ holds for each of the n individuals and for every $k > 0$. From the individual point of view free-riding is better than contributing for any number of contributors, while the condition $f_i(n-1) > g_i(0)$ states the result of universal non-contribution to be Pareto-inferior to universal contribution.[76]

While a fully rational individual would in all likelihood stick to the dominant strategy of non-contributing because she would expect the others to do the same, a bounded rational individual could condition her choice to cooperate or not in a certain round to the number of individuals who cooperated in the previous round.

Such a behaviour could be modelled by a (bounded rational) decision rule advising the individual to contribute if in the precedent round less than k' individuals $(1 < k' < n)$ contributed and to defect otherwise. Manifestly, a similar decision rule would not hold among fully rational actors: full rationality implies symmetry among individuals and homogeneity in the standards of their rationality.

If a subset of individuals conceptualizes a certain rule and behaves on its basis, while another subset conditions on an alternative principle, it is conceivable for both rules to be absorbed, although only partially, i.e. only by a subset of the totality of

[75] This example has been taken from Güth and Kliemt (2004a, p. 532 ff).

[76] Cf. Güth and Kliemt (2004a, p. 533).

the individuals. A prescription based on asymmetry among the subjects involved can only be partially absorbed. Partial absorption implies that for the subset of the absorbing individuals, there is no reason to change the theory or the behaviour because the outcomes achieved can be assessed as satisficing.

The decision of contributing instead of free-riding would be conditional on the predictions of the others' choice. This could, for example, be inferred assuming that the number of contributors from one round to the successive remains the same if there were at least k' contributing individuals. A number of contributions equal or higher than k' yields a collective satisficing outcome, so that the theory is able to resist isolated deviations. In particular, "*as long as people do not have a suspicion that others might bring cooperation levels down they may be willing to cooperate and may go on to do so voluntarily as long as the results are satisfactory*."[77] In this sense, a solution that could not hold among maximizing individuals could be even robustly absorbed by satisficing individuals.

However, the stability of theory absorption in its partial form remains a highly arguable issue. Partial theory absorption can only resist up to a certain threshold value. In the example, spreading uncertainty about the maintenance of the threshold can drop the rate of contribution.

Some mechanisms for diminishing strategic uncertainty (e.g. commitments, reputation formation, ideological background etc.) could have a stabilizing effect, as they could consolidate beliefs on the stability of the asymmetric equilibrium achieved by partial theory absorption. In other words, a partially absorbable theory "*must be absorbed by sufficiently many sufficiently influential individuals who must remain under its spell and individuals must be assured that sufficiently many will remain so*."[78]

As partial theory absorption can only be accounted for by relaxing the conventional common knowledge assumptions. It is definitely ruled out by theories of full rationality. Ad absurdum, a theory of rational behaviour, which would acknowledge its partial absorbability but would not fulfil the axioms on which it is based, should therefore be rejected because of logical inconsistency.

On the contrary, admitting the boundaries of human rationality means to recognize the subjectivity of the aspirations, cognitive abilities and mental models and therefore, allows the asymmetry some settings may imply to be captured.

Indeed, many real-life situations can more realistically be described allowing for the asymmetry in the characteristics of the individuals involved. For example, it frequently occurs that individuals with different skills interact: some of them can be, for instance, more experienced, smarter or simply have a better theoretic knowledge of the situation they face.

It could seem obvious and would also adopt the neoclassical paradigm of the homo oeconomicus that a superior endowment of capabilities (experience, expertise, cleverness…) constitutes an advantage. In spite of that, what really matters in interactive situations is to be able to accurately predict the choices of the others in

[77] Cf. Güth and Kliemt (2004a, p. 534).

[78] Idem.

order to be able to best respond to their actual (in the observed case of asymmetry probably inferior) behaviour. In such situations "superiorly endowed" individuals could misunderstand what motivates the "less endowed" counterparts and so behave in a too clever, but unsuccessful way. Among others, results from the ultimatum game in Güth et al. (1982) as well as findings from the "less-is-more" effect[79] corroborate this thesis.

For those reasons, individuals who possess superior knowledge, expertise or capabilities could paradoxically decide not to rely - or at least not exclusively - on them, but on their common sense, in order to predict what a "representative" bounded rational individual might do.

6.4 Applications of Theory Absorption to Economic Policy Advising

Although reflexive phenomena have been widely observed in economics, the analysis of the mechanisms that lead bounded rational individuals to accept and eventually comply (although in a bounded rational way) with the theoretical prescriptions is still at its beginning. Bounded rational individuals can only process the content of a theory in a bounded rational way. A better understanding of the mechanisms on which theory absorption relies could help specifying bounded rational expectations. Among other things this could enhance the specification of bounded rational best replies and economic forecasts and could be used, for example, for training economic professionals and for economic policy advising[80].

Economic policy advising aims at promoting the settlement for political rational decision. It does not merely refer to the consultation of politicians by economic professionals, but can be extended to the activity of advising all actors dealing with political decisions,[81] e.g. citizens (either individually or aggregated), institutions and organisations.

As discussed in this chapter, the boundaries of the subjective rationality influence the absorption of a theory at each of the steps involved in the absorption process. Similar to the economic actors, policy makers are at most bounded rational and therefore, are only able to process the advice they receive in a bounded rational way. Absorbable advice should acknowledge both the boundedness of its addressee and that of the behaviours on which it has to be ruled in order to set effective normative prescriptions and offer realistic descriptive predictions. When bounded and full rationality reactions to a certain policy are discrepant, advice that is based on theoretical models of perfect rationality may prove to be incorrect and may fail the political purpose it was asked to achieve. For this reason, "*Explaining to bounded rational policy makers on the basis of bounded rational behavioural assumptions*

[79] Cf. Gigerenzer and Goldstein (2002).

[80] For some introductory remarks on economic policy advising see e.g. Franz (2000).

[81] Cf. Pies (2000, p. 291).

why and how certain measures may (or may not) work will render policy advice more acceptable than conventional advice based on welfare maximization."[82]

The traditional economic approach to policy advising blames ideologies without considering that they are indeed at the centre of policy-making.[83] Therefore, the understanding of the role of ideologies and beliefs should be of crucial concern for providing truly informative and effective advice.

Approaching policy-making in a cognitive and evolutionary perspective aims at the analysis of the psychological aspects which rule the perception, interpretation and cognitive processing of the political actors, as well as the related facets of power, interests and subjective incentives which often motivate political decisions.[84]

The resemblance between the cognitive and evolutionary approach to politics and the bounded rational approach to economics is evident. Making allowance for the bounded rationality of the political actors means to recognize that "*beliefs and ideologies are at the centre of economic policy-making, and can, therefore, not be overcome by simple proposing 'efficient solutions' to policy problems, nor by fostering 'solid analytical knowledge' alone*."[85]

In particular, there are three main questions which challenge the cognitive-evolutionary view to economic advising: "*(i) What beliefs do citizens and policy-makers actually hold and how do they contrast with economists' beliefs? (ii) To what extent do economic beliefs guide economic behaviour and what happens when economic and political actors behave according to their 'deviating' beliefs? And perhaps more fundamentally: (iii) How do people acquire beliefs, mental models of theories, and how do they evolve?*"[86]

Results of theory absorption could enrich such a research agenda, hopefully providing some evidence of the kind of theorizing which is compatible with the actors' beliefs and mental models and which, therefore, the actors are more prone to accept and comply with.

The most widely accepted views on economic policy advising (among them the philosophic constitutional approach of Buchanan and the economic constitutional approach of Rawl are e.g. worth mentioning)[87] typically reveal a decisionistic stance. In such a "decisionistic approach"[88] the role of a professional adviser consists in suggesting efficient means for achieving certain goals and purposes that are given by the policy-makers, as illustrated by Fig. 6.1.

The decisionistic approach relies on assumptions which can be to a great extent assimilated to the hypothesis of full rationality of the political actors. Therefore, it often fails to convey a realistic description of the political process and its propelling forces.

---

[82] Cf. Güth and Kliemt (2004a, p. 538).

[83] For a survey see Cassel (2000). For a discussion of the effects of economics on policy-making see e.g. Frey (2000) and Wehner (1995).

[84] For more on a cognitive evolutionary approach to policy-making see e.g. Meier and Slembeck (1994), and Pelikan and Wagner (2003).

[85] Cf. Slembeck (2003, p. 128).

[86] Cf. Slembeck (2003, p. 137).

[87] Cf. e.g. Pies (2000).

[88] Cf. Slembeck (2003).

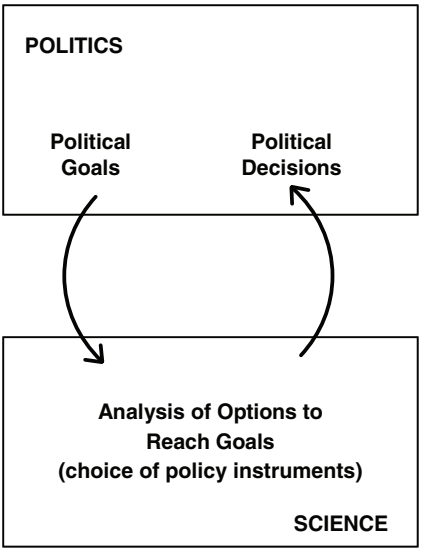

Fig. 6.1 A decisionistic approach to economic policy advicing (Slembeck, 2003)

The decisionistic approach in particular requires: (1) policy goals to be fixed and specified unambiguously; (2) the existence of a first- or second-best solution and the possibility of profiling it; (3) knowledge of all pros and cons and their correct assignment to the alternatives advised; (4) compatibility of the advice with the prevailing ideologies and beliefs,[89] allowing for its non-revolutionary essence.

As it is quite evident that reality rarely fulfils such ideal requirements, it can be argued, whether a similar interpretation of policy advising can lead to the formulation of a really effective and efficient advice. By interpreting the definition of political goals as an endogenous variable, that is paying more attention to the emergence of the political problems and to the setting of the political agenda, this could provide the basis for a more realistic view on policy making. Besides, considering political problems and aims as decisively shaped by individually, collectively held beliefs could enhance the formulation of more acceptable economic advice.

The cognitive-evolutionary view on economic advice[90] draws upon similar considerations, whereas one of its basic tenets is that "*problems are not fixed and given, nor do they exist independent of individual perception. Problems emerge through the perceptions and interpretations of individuals. With regard to policy-making this means that problems arise at the individual level in that the individual perception of problems initiates the political process*."[91]

This leads to conceive policy advising in a procedural, process-oriented way, as illustrated by Fig. 6.2.

---

[89] Cf. Slembeck (2003, p. 150).

[90] It is here referred to its conceptualization by Slembeck (2003).

[91] Cf. Slembeck (2003, p. 142).

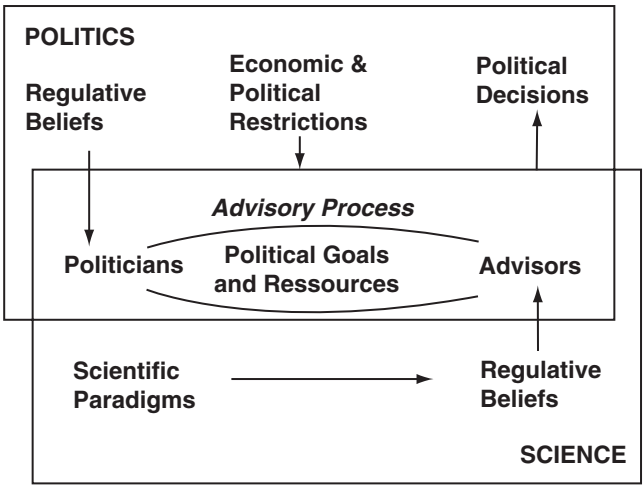

Fig. 6.2 A process-oriented approach to economic policy advising (Slembeck, 2003)

In the light of the absorption of certain advice, a procedural approach to policy advising calls for the need to define an interactive role for the economic advisors. This is compelled by the acknowledgement of the endogenous nature of goal setting, on the one hand, but also by the admission that not all of the proposed means have the same probability of being accepted (and thus implemented) by the at most bounded rational policy-makers on the other hand.

Finally, for an interesting application of how the self-referentiality of economic theories can be empirically tested Lehmann-Waffenschmidt (2006b) can be referred to. He infers conclusions for professional advisers on optimal advising in the case when self-referential effects may occur. Lehmann-Waffenschmidt (2006b) also presents an application of this idea to the case of underpinned advice which is confronted with delaying reaction behaviour of the addressees. Considering the propensity of the decision-maker to postpone the execution of given advice in dependency of the perceived urgency of its underpinning argument, Lehmann-Waffenschmidt (2006b) discusses the objectives of professional advisers as follows: "*Naturally, his first objective is to choose such an advice and such underpinning arguments that the advice really will be taken by the addressed agents (argument justification objective). This is closely related to the problem of the predictability of social events [...]. The adviser's second objective [is that] of being right with his underpinning arguments and his third objective (is) [...] his potential self-interest in the ultimate outcome*."[92]

[92] Cf. Lehmann-Waffenschmidt (2006b, p. 3).

Chapter 7
On the Absorbability of Economic Theories – An Experimental Analysis

Although reflexive phenomena have been widely observed in economics, the analysis of the mechanisms that lead bounded rational individuals to accept and eventually comply as far they could with theoretical prescriptions is still at its beginning.

The point is that even if, in principle, the reflexive implications of social theorizing on the economic actors can never be ruled out, their occurrence actually depends on the understanding, acceptance and coherence between a certain theory and the individual mental models. The recursivity of economic theories and their absorption will therefore differ from case to case, which means that the inquiry of this topic cannot only rely on theoretical speculations but needs to be supported by empirical findings. In particular, the analysis has to address the following issues: (1) how real economic actors perceive the recursive character of economic theorizing; (2) if and under which conditions economic theories affect the behaviour of the economic actors in a self-referential way; and (3) how the self-referentiality of economic theories can be empirically tested.

The experimental part of this study can be conceived as an attempt to approach the wide and complicated field of the recursivity of economic theories and to explore the possibility of testing the validity of such theories relying on their absorbability among bounded rational individuals.

In the first section of this chapter, after some introductory considerations on the experimental method in economics, a possible approach to the experimental analysis of the self-referentiality and absorbability of economic theories will be proposed. In this regard, some related research concerning absorbability and task transcending of satisficing will be discussed. The indicative results of some preparatory attempts at testing the self-referential implications of theories that have been conduced in the form of pilot and classroom experiments will be briefly presented. In consideration of their scarce statistical validity and questionable methodological hygiene, however, extreme caution is advised in interpreting the hints they provide.

The second section will then be dedicated to the analysis of two experimental studies. The first is focused on the absorbability of equilibrium predictions on guessing games (cf. Sect. 5 of Chap. 7), while the second discusses the absorbability of

S. Sandri, *Reflexivity in Economics*: *An Experimental Examination on the Self-Referentiality of Economic Theories*,

informational cascades' theory. Each of the two studies will be discussed by introducing previous related experimental evidence, by presenting the specific design of the experiment with its main results. Some tentative conclusions which can be inferred on the basis of the experimental evidence so far conclude the chapter.

7.1 The Experimental Method in Economics

While for the so-called hard sciences (like physics, chemistry and biology) the experimental methodology constitutes the prevailing technique for inferring and testing theories, the application of the experimental method to the enquiry of economics represents a clear rupture with the traditional non-experimental approach. With this approach, economics essentially conceived as a non-experimental subject and posits accordingly that economic data can only be collected in a passive way, by observing real developments and phenomena under exogenously given circumstances, i.e. not under controlled conditions.

Experimental economics, whose first attempts can e.g. be traced back to the work of Thurstone (1931), Preston and Baratta (1948), Mosteller and Nogee (1951), Allais (1953), Edwards (1953a, 1953b), May (1954), Davidson, Suppes and Siegel (1957), Davidson and Marschak (1959),[1] shows indeed that economic data can be generated in the laboratory under controlled conditions. This makes for better quality data, available at lower costs and in reasonable amounts of time.

The conception underlying experimental economics is "*to use the lab as a hothouse for cultivating theory: Existing lab data guides model construction. Once constructed, the model is subjected to a new round of tests. Depending on the results, the model is either refined or abandoned*."[2]

Experiments can serve different purposes, which have been colourfully labelled and are well-known as "*Speaking to Theorists,*" "*Searching for Facts*," "*Searching for Meanings*," and "*Whispering in the Ears of Princes*."[3] They strive respectively for testing formal theoretical predictions, for studying the effects of variables not contemplated and/or not fully seized by existent theories as well as for developing mechanisms of effective policy advising.

The setting of an experiment consists in a controlled economic environment, whereas in an "economic environment" one or more individual economic agents interact together through an institution.[4] Whilst achieving control over the institution is a straightforward operation in an experimental environment, extending the control to the agents involved requires the codification of some additional assumptions and procedures. An essential point in this regard is to strive toward the emergence of constitutive economic characteristics, such as preferences, technology, resource

[1] For a review, see e.g. Kagel and Roth (1995).

[2] Cf. Bolton (1998, p. 258).

[3] Cf. Roth (1995, pp. 22, 23).

[4] Cf. Friedman and Sunder (1994, p. 13).

environment and information. For achieving that, experimental economics steers on the reward medium which relies on the so-called "induced-value theory." [5]

Induced-value theory posits, as sufficient conditions for controlling agents' characteristics, the three requirements of monotonicity, salience and dominance of rewards. They state respectively: (1) that subjects' preferences should be strictly monotone which increases the reward medium (which is normally achieved by using domestic currency as the reward medium); (2) the reward received should be related to the action taken in a way that is intelligible to the subjects; (3) that, in the experiment, changes in subjects' utility are predominantly due to the reward medium while other influences are negligible. These conditions are, as a rule, made operative by choosing a performance-dependent reward scheme whose average rewards can be estimated to exceed the subjects' average opportunity costs. In addition, the subjects have to be provided with simple, clear instructions, drawn in neutral terms, whose understanding can be e.g. checked in preliminary dry runs and/or control questions. Privacy concerning the individual performance and payoffs must be granted, and the experimental aim must be kept secret. Experimenters cannot deceive or lie to the participants in an experiment regarding none of the experimental features.

The evidence which is gained from economic experiments can then be transferred into theory according to the so-called "parallelism precept," which states that "*[p]ropositions about the behaviour of individuals and the performance of institutions that have been tested in laboratory microeconomies apply also to nonlaboratory microeconomies where similar ceteris paribus conditions hold*."[6]

Experiments have been revealed to be an essential support for economic and, in particular, microeconomic theorizing. So far, a myriad of experiments have been conduced, deepening the economic behaviour in almost all microeconomic situations contemplated by the economic theory. For a survey on milestone-contributions in experimental economics and a systematic discussion of their main findings and achievements see, for example, Kagel and Roth (1995) and Hey (1994).

7.2 A Possible Experimental Approach for Testing the Self-Referentiality and Absorbability of Economic Theories

The experimental investigation of the self-referentiality of economic theories and their absorption should include a large spectrum of experiments. As said, this is because even if absorbability of economic theories is, in principle, always possible, its concrete modality can only be highly context specific and follows the complex mechanisms ruling bounded rational cognition and decision-making. Therefore, the experimental testing of theory absorption should, in an ideal case, be articulated over a broad range of experimental situations so as to observe how the specific faculties

[5] Cf. Smith (1976).

[6] Cf. Smith (1982, p. 936).

involved in different scenarios of economic problem-solving and decision-making affect the perception of a certain theory and eventually create the conditions for its absorbability.

As previously discussed in Chap. 6, assuming different rationality standards for the economic actors yields different implications for the recursivity and absorbability of the economic theory. Correspondingly, this theme can be approached either focussing on the reflexive effects of theories, which rely on the assumption of fully rational economic actors or on those of theories encompassing individuals' bounded rationality.

This study conceives the inquiry of the absorbability of theories of full rationality as a natural first step and is preferred over starting with the observation of the absorption of theories of bounded rationality. This is inspired by different considerations. First, as their denomination suggests, mainstream theories enjoy a broad acceptance, which (with their common axiomatic basis and underlying assumptions) leads to their almost univocal codification when applied to a concrete setting. Adding to that, theories of full rationality interpret the individuals as straightforward utility optimizers. They greatly simplify aims and purposes which steer human decision-making, which is manifestly reductive but still appealing whenever it comes to economic decisions. In principle, mainstream theories can be assessed to be, to some extents, familiar to the people, not for what concerns their content, but for the logic inspiring them.

A further aspect, which sides with first focussing on the absorption of full rational theories, is represented by the ideal normative standards they posit. Although their descriptive validity can be seriously questioned, their usefulness as a normative benchmark for assessing decisional and judgemental quality holds in many cases.

On the contrary, theories of bounded rational behaviour do not enjoy such an ideal normative standard. Despite that, they contemplate a more complex range of aims and purposes which may motivate the acting individual. This adds to their plausibility and their descriptive power, but at the same time increases the variables that should be accounted for.

Moreover, theories of bounded rationality are articulated over a particularly vast panorama of different approaches that frequently ascribe the responsibility for problem-solving and decision-making to very different faculties of the human cognition. Because of their pronounced empirical anchoring, bounded rational theories can be thought of as more realistic in their description of human cognition than neoclassical theories.

Plenty of experimental evidence reveals how theories of unbounded rationality are not always descriptive of the observed economic behaviour. Now, the proof of the absorption of such theories could work as a test of their "viable" normative validity. On the other hand, as theories of bounded rationality have tried to interpret and stylize the systematic violations of theories of full rationality, theory absorption could be seen as an ultimate test of validity of their real descriptive power, as well as of the degree of their acceptance.

The effects of the self-referentiality of economic theories and their absorption can be experimentally analyzed through the observation of how individuals deal with meta-theoretical information in different experimental contexts.

Meta(-theoretical) information (i.e. theoretic information about the experimental situation the experimental subjects face, which aims at adding to the theoretical knowledge of the subjects and not simply to their information)[7] could be communicated to the experimental subjects in the form of "meta-instructions." By "meta-instructions" instructions it is meant, instructions that reveal the theory underlying the experimental situation and/or previous experimental findings. The information contained in the meta-instructions should not overwhelm the cognitive bounded capabilities of the individuals and should be secured by some control questions on the application of the meta-instructions or supplemented by a questionnaire at the end of the experimental sessions.

The experiments have been run over two treatments, with and without meta-instructions. The comparison between the two treatments intends to enhance some conclusions both on the perception of the self-referentiality of economic theories and on their absorption. While a significant difference between the behaviour of the test and of the control groups (i.e. respectively with or without meta-instructions) can reveal the perception of the self-referentiality of the theory communicated by the meta-instructions, the compliance of the experimental subjects with the theory conveys a proof of its absorption.

The experimental hypothesis underlying this study is that meta-theoretical information can support the bounded rational decision process and improve the outcome's efficiency degree in various experimental settings. In the ideal case, the experimental research should cover both individual and interactive decision-making.[8]

Meta-instructions can be also interpreted as an attempt to support the subjective rationality, not in the sense of considering theory as "congealed experience,"[9] according to which, experience could be substituted by theoretic knowledge, but in the sense that if non-optimizing behaviour would simply be a step in a discrete optimization process (as sustained by the "as if" approach), individuals would choose to optimize, if they could; i.e. they would comply with the theory presented in the meta-instructions. It can be argued whether that really happens. In many situations, as for instance in beauty contests, coordination or public good games, common knowledge of the equilibrium does not eliminate strategic uncertainty. In similar settings the outcome of the game cannot be foreseen and the players are mainly concerned with predicting the others' behaviour. In economic interactions where bounded rationality of the others matters, meta-information could be expected to promote the emerging of a sort of meta-rationality, a behavioural-rationality, which transcends full rationality and is superior to it in terms of success in concrete settings.

However, the absorption of meta-instructions is not a trivial question even for unilateral decision making. Supporting the subjective bounded rationality is far from leading to perfect rationality, as corroborated by the resistance of several cognitive biases to debiasing attempts as well as by the evidence on "less is more" heuristics.

[7] The difference between "knowledge" and "information" is here meant as in Sect. 1.2 of Chap. 3.

[8] Except for a trial on the debiasing of the conjunction fallacy through meta-instructions, the evidence discussed in the present study restricts to interactive decision-making.

[9] The expression stems from Morgenstern (1972, p. 707).

The absorption of meta-instructions requires their coherency with the subjective beliefs in order to be trusted as a valid support to the decision. So, even in situations where it has been experimentally shown that actual behaviour differs from the predictions of the rational choice theory, theory absorption could test its normative acceptance and validity.

The compliance with meta-instructions based on theories of bounded rationality could help in testing the real coherency between their underlying assumptions and the mental models of the individuals. The proof of the survival of a theory of bounded rationality to its own acceptance could say more about its cognitive reliability.

In Chap. 6 the self-referentiality of a certain theory has been interpreted as a logical premise for its absorbability, and its perception as one of its necessary condition. In this sense, the perception of the self-referentiality of a theory could be disentangled from its absorbability whenever the experimental evidence accounted for significant differences between the behaviour of the individuals in the absorption and in the control treatment, on the one hand, and between behaviour and theoretical predictions on the other hand.

7.3 Related Experimental Studies

The experimental testing of economic theories at their absorbability level is still at its earliest stages. A research project on the absorbability of satisficing has recently started by the Research Group on Strategic Interaction at the Max Planck Institute for Economics in Jena. The experiments that, to the best of our knowledge, had been conducted in this field, the experimental procedures they adopted and their main findings will be briefly sketched in the following pages. They aim at testing the absorbability and task transferability of the satisficing approach in a portfolio selection task[10] and on exploring whether individuals prefer to conform to principles of bounded rationality rather than to rational choice.[11]

The first study[12] observes the behaviour of individuals in different financial investment tasks with state-specific return rates and with a finite number of states. Such states can be ordinally assessed, associated with different degrees of risk and correspond to different degrees of risk. Different states of nature are equally probable.

Individuals faced with a similar task can be said to behave satisficing if they reveal consistency of return aspirations (i.e. their aspirations refer to non-empty portfolio sets) as well as consistency between their investment decisions and the aspirations they state.

The experiment was articulated over two phases. During the first phase, the satisficing principle of aspirations consistency was enforced by means of a routine. The

[10] These aims are respectively tackled in Güth, Levati, and Ploner (2006) and Fellner, Güth, and Martin (2006a).

[11] Cf. Fellner, Güth, and Martin (2006b).

[12] Cf. Güth, Levati, and Ploner (2006).

respondents were asked to form their subjective return-aspirations as long as they achieved consistency. After that, participants were left free to choose a portfolio and were provided with feedback regarding whether their choice satisfied their return aspirations or not. During the second phase, the respondents could decide whether to further rely on the satisficing routine or to make their portfolio choice freely, while having relaxed, in this case, the requirement of aspiration consistency.

In this setting, evidence for the absorption of the satisficing approach would have been inferable on the basis of the compliance with it. The key chosen to interpret the results was to assume theory absorption to be confirmed when the individuals acted satisficing, i.e. in respect of the two principles stated above. Accordingly, a proof for the absorbability of satisficing would have been during the first phase when individuals opt for a portfolio which meets their aspirations and therefore, revels adherence to the logic underlying the satisficing approach. The respondents who in the first phase absorbed the principles of satisficing were expected in the second phase either to further rely on the satisficing routine or to refuse it but still to comply with the principles of satisficing.

The overall findings seem to confirm the individual ability to absorb the satisficing approach. The data reveal that almost half of the participants who refuse the aid after the enforcement phase still chose satisficing portfolios and that most routine-assisted investment choices are consistent with the aspirations stated. Therefore, the study provides evidence for the possibility of learning the principles of satisficing and at a degree which is dependent on the tasks complexity.

A seemingly disrupting finding is represented by the occurrence of a suboptimal routine among a minority of respondents, according to which, participants did not modify their return-aspirations from period to period.

The second study associates the inquiry of the absorbability of the satisficing approach with some reflections on the universality of the bounded rational approach. This is tackled by testing the prescriptive applicability of aspirations formation and satisficing behaviour with their transferability among similar tasks. For this, the respondents were faced with two investments tasks which differed in their complexity.

The experimental data are analysed by clustering the participants according to their responses into the three categories of unreasonable, potentially satisficing and actually satisficing. These clusters indicate respectively if the stated aspirations do not meet the minimal requirements for consistency, if they define a non-empty set of investment choices, and if they are fulfilled by the investment choice done.

The distribution of the respondents in the different categories confirms that principles of satisficing are not innate in the individuals but need teaching or consulting and learning. The efficacy of learning could however be interpreted in the sense of the absorbability of the satisficing approach. There is some evidence hinting at the high consistency of aspiration formation among tasks, whereas the relative low number of potential and actual satisficers among the respondents suggests caution in this insight.

The third study[13] finally explores whether bounded rational individuals, involved in a simple portfolio choice task, prefer to rely on the satisficing or on the optimizing

[13] Cf. Fellner, Güth, and Martin (2006b).

approach. In an initial phase the participants were familiarized with two routines, each of them enforcing one of the two approaches. Routines were automatically processed on the basis of individual parameters the participants were asked to express. Such parameters expressed down- and upward aspirations and the subjective attitude toward risk. In each period and according to the result of the routine processed, the participants were forced to implement a certain investment. In a second phase the participants were left free to decide whether to opt for one of the two decision modes or to formulate their investment choice freely.

From the analysis of the data it emerges that optimizing was favoured over satisficing by 62% instead of by 38% of choices.

7.4 Some Preparatory Attempts

The indicative results of some trials and pilot experiments, which have been mostly conduced in the classroom, inspire the tentative considerations that will be discussed in this paragraph. It should be premised that neither the size of the sample (which does not offer enough independent observations to corroborate any result), nor the experimental conditions (in a classroom perfect isolation of the subjects cannot be achieved) allow one to trust such results as definitive evidence. In addressing what emerged from these preparatory attempts, it is not intended here to provide facts but merely tendencies or possible developments that could help in sketching the experimental framework for testing theory absorption among bounded rational individuals.

In particular, pilot classroom experiments will be focussed on that deal respectively with (1) the experimental attempt of debiasing the conjunction-effect bias through meta-information,[14] with (2) the role of the theory of integrative negotiation in promoting efficiency in multilateral negotiations, and with (3) a guessing-game with information-feedback and meta-instructions.

As previously put forward, the analysis has been restricted to main stream, hard-core theories that rely on the assumption of full rational economic actors.

As follows, an overview of the experimental procedures adopted for the different attempts will be offered, and the main tendencies that can be subsumed for the observation of their results will be mentioned. The experiments' instructions are available upon request.

7.4.1 An Experimental Attempt of Debiasing the Conjunction-Effect Bias through Meta-Information

To test the absorbability of theoretic principles in non-interactive decision-settings, a task has been focussed on in which typically the conjunction-effect bias occurs.

[14] Referring to the diploma thesis of Pombeni (2005).

This bias consists in the systematic violation of the conjunction rule in subjective probability judgements and is typically ascribed to the representativeness heuristic. This fallacy has been the focus of a considerable amount of experimental studies and has been shown to be quite resistant to debiasing attempts.[15]

A classroom experiment, which strived for the debiasing of the conjunction rule through meta-information, was conducted in 2005 on 30 undergraduate students of the University of Dresden.

The participants had to solve a task involving probability estimation and were divided into two groups, a control and a debiasing one.

The participants assigned to both groups had to solve the same task, which consisted in estimating and ordering the probability of some conjuncted and non-conjuncted events. The only difference between the two treatments was that the participants assigned to the debiasing group received theoretical information in form of meta-instructions containing an explanation of the conjunction rule, while the respondents belonging to the control group were not provided with such a decisional aid.

The evidence from this experiment was that while in the control group the conjunction bias occurred by 75%, its frequency in the debiasing group was reduced to 13.3%. This seems to suggest the absorbability of the conjunction rule and hints at the efficacy of meta-theoretical instructions as a possible debiasing procedure.

7.4.2 A Classroom Experiment on Theory Absorption in Multilateral Integrative Negotiations

Both distributive and integrative approaches to negotiation are in principle not immune to their absorption; their knowledge would inspire outrageous demands to which they would, in all likelihood, no longer hold. Their absorbability decisively depends on the foresight of the contrahents, thus on their rationality and motives, as well as on the institutions which rule the negotiation.

Even if real multilateral negotiations in many cases offer a potential for integrative agreements, negotiators often fail to exploit the possibilities of maximising joint benefits and choose sub-optimal bargaining solutions. This trial aimed at experimentally analysing if meta-information of integrative negotiation theory can be absorbed from the negotiators and promote the settle for optimal integrative solutions. Specifically, the experiment focused on the integrative negotiation technique of logrolling which permits the respondents to solve the negotiation on a win-win basis whenever preferences are asymmetric, by means of the mutual exchange of concessions.[16]

In 2003 a classroom experiment was conducted at the University of Dresden. In the two sessions, 18 undergraduate students were divided into groups of three

[15] For more on the conjunction effect see Sect. 3.1.2 of Chap. 5.

[16] Cf. e.g. Stratmann (1997).

and had to negotiate some issues. The first session resembled the design used by Lehmann-Waffenschmidt and Reina (2003), from which it emerged that under the majority rule the integrative potential of a negotiation does not, as a rule, get exploited.

In the first session, each group had to negotiate over three issues under the majority rule. Free talk was admitted and individual preferences were asymmetric. In the second session, groups were rematched and the respondents were confronted with the more complex task of negotiating six issues. The same rules as in the first session applied. Besides, in the second session participants were provided with theoretical information in the form of a decisional aid explaining the principles of integrative negotiation and logrolling.[17]

The evidence from this trial hints at the absorbability of the principles underlining integrative negotiations: whilst in the first session the integrative solution was chosen by 4 of 6 groups, in the second session, despite of the increased complexity of the negotiation, all groups settle for an integrative solution. However, even after having provided the students with principles of integrative negotiation, the integrative solution that was favoured by all groups was the fair one, yielding the same payoff for each individual, rather than a very optimal one, corresponding to an asymmetric payoff distribution, which might have increased the danger of majority cycles.

7.4.3 An Experimental Guessing-Game in the Classroom with Information Feed-Back and Meta-Instructions

A classroom experiment, which was run in January 2006 among 266 undergraduate students of the University of Dresden, focused on the role of theoretical information and previous empirical evidence in a p-guessing game.[18]

The purpose was to analyse the absorbability of game theoretical principles for inferring the Nash equilibrium and to compare it with the effects of spreading information on empirical evidence. Meta-information, both regarding theoretical principles and empirical results, was expected to promote the emergence of a sort of game-rationality, which is among bounded rational actors superior to the game theoretical prescriptions.

The guessing game represents a setting, in which common knowledge of the equilibrium does not eliminate strategic uncertainty. The outcome of a guessing-game cannot be foreseen, as the players are mainly concerned with predicting the others' choice. More on the guessing game, its theoretic modelling and experimental analysis will be, for convenience, treated later in Sect. 5 of Chap. 7 when discussing one of the core experimental studies of this research.

For this trial the experimental setting of a repeated p-guessing game with $p = 2/3$ was chosen. Within the close interval *[1,100]* the participants had to guess a target number corresponding to the entire number which was closest to the two-third of

[17] For more details on logrolling see Mueller (1997).

[18] More on the evidence of this experiment is presented in the diplomat thesis of Marx (2006).

the mean of all chosen numbers. The three participants with the closest guesses to the target yield a reward of 5 €. Three treatments were run: the 89 participants assigned to the control treatment did not receive any further information except the instructions, while the 87 participants that took part in the data treatment received a decisional aid in the form of feedback information about the empirical results of a previous similar p-guessing game.[19] Finally, in the absorption treatment 89 individuals were provided with theoretical information about the Nash equilibrium and how to derive it through iterated elimination of dominated strategies.

Figure 7.1 summarizes the frequencies of guesses in the different treatments.

The experimental hypotheses were that (1) the target number in the absorption treatment as well as (2) in the data treatment were significantly lower than that in the control treatment. While in the absorption treatment, theory absorption could implement the choice of the Nash equilibrium, in the data treatment, the provided empirical findings could represent an anchor for the participants to make their guess.

The results of t-test indicate that both first and second hypotheses cannot be rejected respective to $p = 0.008$ and $p < 0.001$. When comparing the results of absorption and control treatments, it emerges that the target number in the data treatment is significantly lower $(p < 0.001)$, i.e. closer to the Nash equilibrium, than in the absorption treatment. This seems to hint at the incomplete absorption of the principles of iterated elimination of dominated strategies[20] and can be ascribed to the strategic uncertainty which is unavoidable in guessing games. In this sense, information on empirical evidence might have contributed to reduce strategic uncertainty by providing an anchor for typing a guess.

7.5 On the Absorbability of Guessing Game Theory[21]

Besides, evaluating the trial on a p-guessing game confirms the suitability of this setting for testing theory absorption. The p-guessing game has namely several attractive characteristics for deepening rationality and learning, in that it permits the effects of rationality to disentangle from social preferences such as inequality aversion, fairness or reciprocity, which are, at the same time, very simple to explain, and have a very simple economic interpretation. Therefore, it represents a suitable framework for experimentally testing theory absorption by bounded rational decision-makers.

As follows, after a more detailed introduction on the guessing game, an overview of relevant studies will be offered, including both theoretical and experimental

[19] The empirical evidence provided stems from a previously conduced trial, which had been run under comparable conditions, i.e. as classroom experiment, among undergraduate students of the University of Dresden, and following the same experimental procedure and modalities. A Welch's test proved pooling of the samples to be allowed.

[20] This will be discussed in more details in the next paragraph (Sect.5 of Chap. 7).

[21] The results of this experimental study and the data analysis are discussed in more details in Morone, Sandri, and Uske (2008).

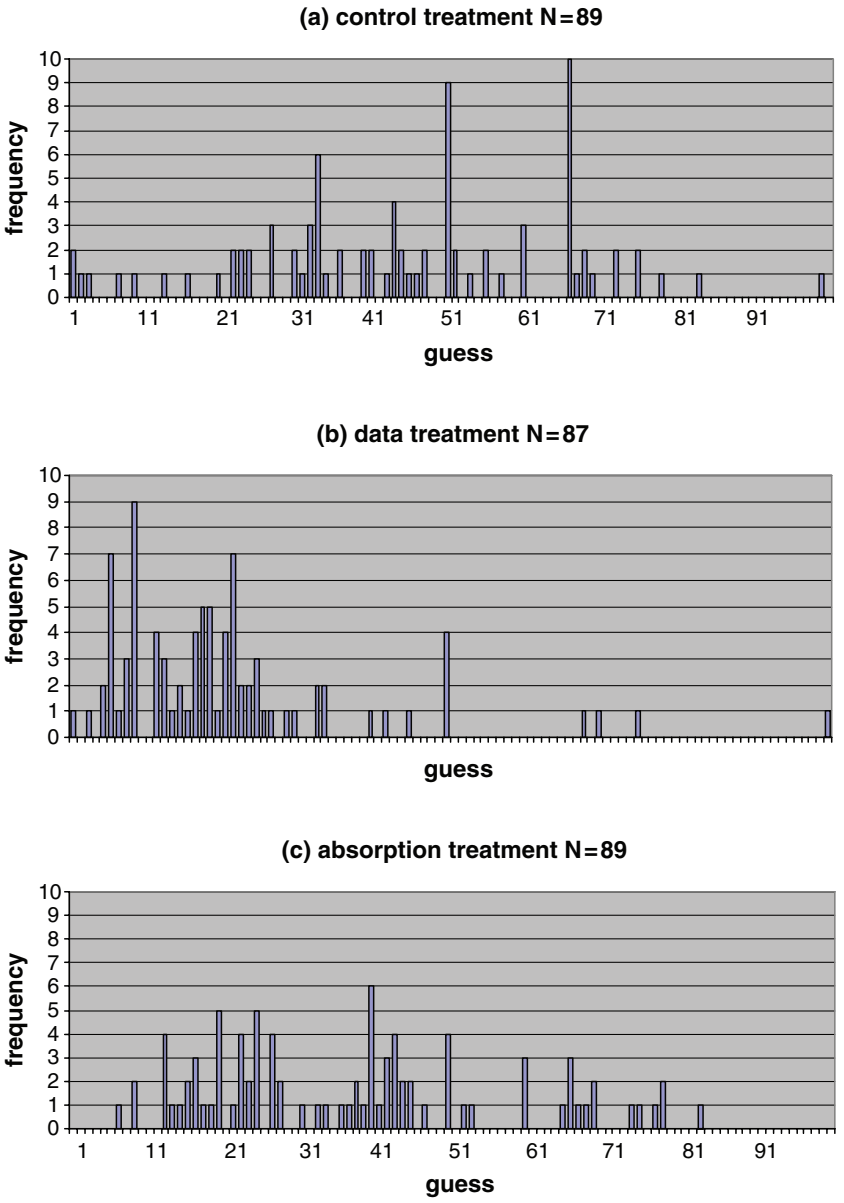

Fig. 7.1 Frequency of guesses in **(a)** the control treatment, **(b)** the data treatment and **(c)** the absorption treatment

contributions. A model for deriving the Nash equilibrium by means of iterated elimination of dominated strategies will be presented and aspects of partial and full theory absorption for guessing games will be related. The experimental design and hypothesis follow, and the main results and some final remarks on them conclude.

7.5.1 On the Guessing Game

The logic underlying the guessing game was first pointed out by Keynes (1936) as a simplification for describing the behaviours of investors in financial markets. As "*most of these [professional investors] are, in fact, largely concerned, not with making superior long-term forecasts [...] but with foreseeing changes in the conventional basis of valuation a short time ahead of the general public*"[22] the logic inspiring the behaviour of professionals on financial markets resembles one of newspaper beauty contests, which ask respondents to guess which one among several faces will be appointed as the prettiest by the majority of the newspaper's readers.

Similarly, in a p-guessing game all $n(\geq 1)$ players $i = 1,\ldots,n$ have to choose simultaneously a number g_i from a closed interval *[L,H]*, with $0 \leq L < H$. With p belonging to *(0,1)* and $d \geq 0$, the wining player is the one whose guess is closest to the target number:[23]

$$g^* = p\left(\sum_{i=1}^{n} \frac{g_i}{n} + d\right) \tag{7.1}$$

Commonly known, iterated elimination of dominated strategies leads to derive the unique equilibrium, as the players are able to conceptualize the equilibrium prediction when they all know that all of them know that all are aware of the principles of elimination of dominated strategies and are capable of deriving its implications.

This illustrates the full absorbability of strategic equilibria in guessing games when the assumption of common knowledge of rationality holds. In a similar setting, partial theory absorption can be illustrated assuming that just a subset of the $n \geq 2$ players knows and is able to apply the principle of iterated elimination of dominated strategies. These players will account for the effect of the own choice on the target number g_i and avoid choosing a dominated strategy. But strategic uncertainty concerning the others' choice questions whether relying on the equilibrium choice is the best response to what the counterparts who are unaware of the theory are going to do.

Experimental evidence on guessing games[24] all reveals convergence to the equilibrium, unlike in other games.[25] Different explanations for what steers convergence toward equilibrium have been proposed. It is still debated whether this relies on

[22] Cf. Keynes (1936).

[23] Cf. Nagel (1995), Duffy and Nagel (1997), Ho, Camerer, and Weigelt (1998) and Weber (2003).

[24] See e.g. Camerer, Ho, and Chong (2001), Camerer (2003a, 2003b), and Nagel (1995, 1999).

[25] For a review, cf. e.g. Roth (1995).

cognitively capturing iterated elimination of dominated strategies rather than on non-cognitive behavioural adaptation, as e.g. the application of individual heuristics or mental representation.[26]

As a possible way of modelling the behavioural patterns that can be observed by guessing games, Nagel (1995) suggested that individuals adjust their guesses by a process of naïve best responding rather than by elimination of dominated strategies. Moreover, Nagel's model posits that for $d = 0$ the individuals take 50 instead of 100 as reference point for their reasoning process, whereby each player assumes herself to be the most sophisticated, i.e. thinks herself to be performing exactly one step of reasoning ahead of all the others. Nagel's model fixes 50 as the reference point from which the individuals start their thinking process, since it assumes level-0 players to randomly choose a number from the interval *[0,100]* and therefore, the mean to equal 50. A level-1 player would then naively best reply, believing everybody else to be level-0, choosing $p \cdot 50$. A level-2 player would think one step ahead of a level-1 player and therefore, guess $p^2 \cdot 50$. In general, a level-k player assumes all the others to be level-(k-1) and accordingly guesses a number equal $p^2 \cdot 50$ Therefore, a player who performs infinite iterative steps of reasoning conceptualises and chooses the Nash equilibrium of 0. The interpretation of the process, which is generated by iterative best responses as converging toward the equilibrium, relies on the implicit assumption that different subjects hold different depths of thinking.

By observing the guessing behaviour in different newspaper and lab beauty-contest experiments[27] subjects can actually be clustered at level-1, level-2, level-3, for then jumping at level-infinity.

Camerer, Ho, and Chong (2001) refuse the idea of behavioural adjustment both by means of sub-game and of trembling-hand refinements toward the Nash equilibrium. They alternatively introduce a model of cognitive hierarchy, according to which a step-k player, perceiving herself to be the most sophisticated, estimates the other players to be Poisson-distributed from step-0 to step-(k-1).

The parameterization of guessing games has been also proved to influence convergence toward the Nash equilibrium: as Güth, Kocher, and Sutter (2002) pointed out, "*interior equilibria trigger more equilibrium-like behaviour than boundary equilibria.*"[28] Because of the individual reluctance to choose the extremes,[29] the heterogeneity of players induces longer deliberation time and higher deviations from the Nash equilibrium, while continuous payment scheme seems to be a force which pushes toward equilibrium convergence. Morone and Morone (2007) discussed Güth et al.'s interior equilibria results in a more general setting and generalised the iterative naïve best replies strategies to a wider class of games, which showed compatibility among Güth et al.'s and Nagel's findings on bounded rational iterative thinking and behaving.

The present experiment exclusively concentrates on first period choices of guessing games with different parameterization, though varying both parameters p and d.

[26] Cf. e.g. Camerer (2003a, 2003b) and Sbriglia (2008).

[27] Cf. Bosch-Domenech, Montalvo, Nagel, and Satorra (2002).

[28] Cf. Güth, Kocher, and Sutter (2002).

[29] Cf. e.g. Rubinstein, Tversky, and Heller (1997).

Hereby, because of the modified parameterization and in particular when subjects are aware of the principle of iterated elimination of dominated strategies, it is focussed on the changes of behaviour rather than the convergence process. In particular, how theoretical prescriptions are related and adapted by subjects on their own as well as others' behaviour in consideration of the supposed bounded rational behaviour of the others will be discussed.

7.5.2 *Iterated Elimination of Dominated Strategies in Guessing Games*

As a natural first step for exploring theory absorption here symmetric guessing games will be exclusively focussed on, which are illustrated by:

$$\Gamma = \left[N, \{G_i\}_{i \in N,} \{u_i\}_{i \in N}\right] \tag{7.2}$$

whereas n is the number of players, g_i the set of strategies of player i satisfying $0 \leq L \leq H$,

$$g_i \in G_i = G_i = [L,H] \forall_i = 1, \ldots, n \tag{7.3}$$

The guessing game chosen relies on a continuous payoff scheme, according to which each player i's payoff function depends on her own deviation from the target number $g^* = p\left(\sum_{i=1}^{n} \frac{g_i}{n} + d\right)$ so that

$$u(g_i) = C - c\left|g_i - p\left(\frac{1}{n}\sum_{i=1}^{n} g_i + d\right)\right| \tag{7.4}$$

whereas $p \in (0,1), d \geq 0, L < g^* < H$ hold. C is a positive monetary endowment, and $c(> 0)$ can be interpreted as a fine that subject i has to pay for deviating form the target number g^*.

It is most common in guessing games[30] that only the player whose guess is closest to the target number g^* yields a positive payoff. Resembling Güth et al. (2002) and Morone and Morone (2007) it is, on the contrary, relied on that a continuous payoff rule, according to which all players win and their deviation from the target number gets punished.

Starting from the trivial case of being sanctioned for deviating from g^* if $n = 1$, it is easy to state that all numbers g_1 belonging to the closed interval *[L,H]* of the only player i that are different from the target $(g_i \neq g^*)$ are given by:

$$g_i^* = \max\left\{\min\left\{\frac{p^* d}{1-p}, H\right\}, L\right\} \tag{7.5}$$

[30] See e.g. Nagel (1995).

For the only player 1 is it easy to recognise g_1^* as the best choice to be made and to compute it. Therefore, informing the player of the principle of iterated elimination of dominated strategies leads her to choose g_1^*. The principle of iterated elimination of dominated strategies can thus be said to be fully absorbable even for boundedly rational decision makers.

The same principle can be applied also increasing the number of players $n \geq 2$. It prescribes for all players $i = 1,\ldots,n$ to eliminate each choice g_i fulfilling

$$g_i < \frac{p\,(n-1)\,L + nd}{n-p} \text{ and } g_i > \frac{p\,(n-1)\,H + nd}{n-p} \tag{7.6}$$

Interestingly, the others are not required to be aware of the principle.

For deriving the unique equilibrium strategies $g_I^* = g^*$ for all players $i = 1,\ldots,n$, i.e.

$$g^* = \max\left\{\min\left\{\frac{p*d}{1-p}, H\right\}, L\right\} \tag{7.7}$$

the following assumptions are required:

1. each of the n players $i = 1,\ldots,n$ eliminates dominated strategies repeatedly and not just once (Principle 1, further P.1);
2. each of the n players $i = 1,\ldots,n$ is aware that all other players know that all other players iteratively eliminate dominated strategies and are aware of all players doing and knowing that (Principle 2, further P.2).

Thus, even if full absorbability of the iterated elimination of dominated strategies is, in principle, (i.e. among rational individuals) granted, exploring partial and full theory absorption of such theoretic principles among bounded rational individuals is a challenging task. Guessing games qualify as good candidates for this analysis. If one of the n players is informed about P.1 and P.2, she cannot conclude much more than that all strategies g_i are dominated (what leaves her choice wide open). On the other hand, informing all of the players about the two principles P.1 and P.2 the immediate choice of g^* should be expected.

7.5.3 The Experimental Design

The experiment was run in three treatments, each of which contained nine successive guessing games with different parameters, as illustrated by Table 7.1.

In the table, p and d stand for the parameters that specify each guessing game. As in Güth et al. (2002), the convergence process is illustrated starting respective from the interval's upper $(L = 100)$ and lower $(H = 100)$ bounds. The first rows refer to what can be eliminated in each successive iteration step-k, while the bottom row considers step-infinity and therefore, specifies the equilibrium strategy for the respective game.

Table 7.1 Nine parameterizations of the experimental guessing games

Elimination steps	$p_i = 1/2$	$d = 0$	$p_i = 1/2$	$d = 25$	$p_i = 1/2$	$d = 50$	$p_i = 1/3$	$d = 0$	$p_i = 1/3$	$d = 25$
k_i in 0	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00
1.00	0.00	50.00	12.50	62.50	25.00	75.00	0.00	33.33	8.33	41.67
2.00	0.00	25.00	18.75	43.75	37.50	62.50	0.00	11.11	11.11	22.22
3.00	0.00	12.50	21.88	34.38	43.75	56.25	0.00	3.70	12.04	15.74
4.00	0.00	0.25	23.44	29.69	46.88	53.13	0.00	1.23	12.35	13.58
5.00	0.00	3.13	24.22	27.34	48.44	51.56	0.00	0.41	12.45	12.80
Infinity	**0.00**	**0.00**	**25.00**	**25.00**	**50.00**	**50.00**	**0.00**	**0.00**	**12.50**	**12.50**
Playing order		7		1		4		5		9

	$p_i = 1/3$	$d = 50$	$p_i = 2/3$	$d = 0$	$p_i = 2/3$	$d = 25$	$p_i = 2/3$	$d = 50$
k_i in 0	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00
1.00	16.67	50.00	0.00	66.67	16.67	83.33	33.33	100.00
2.00	22.22	33.33	0.00	44.44	27.78	72.22	55.56	100.00
3.00	24.07	27.78	0.00	29.63	35.19	64.81	70.37	100.00
4.00	24.69	25.93	0.00	19.75	40.12	59.88	80.25	100.00
5.00	24.90	25.31	0.00	13.17	43.42	56.58	86.83	100.00
Infinity	**25.00**	**25.00**	**0.00**	**0.00**	**50.00**	**50.00**	**100.00**	**100.00**
Playing order		3		2		6		8

In each of the tree treatments, every participant played all nine games in the order specified in Table 7.1 a within subject-design. In each treatment, 32 subjects were divided into groups of 8, yielding four independent observations. The subjects had to guess a number from the interval *[L,H]* whereby as defined above by $u(gi)$, their earnings were inversely proportional to the deviation of their guess from the target number, as defined by $u(gi)$ (cf. Eq. 4).

The first treatment (to which it will henceforth referred as UU-treatment)[31] served as a control treatment in that all the participants received instructions on rules of the game but no information concerning the principles of iterated elimination of dominated strategies.

In the second treatment (the UI- or partial absorption treatment) all subjects were provided with the instructions to the experiment while only half of them were additionally provided with theoretic information regarding the principles of iterated elimination of dominated strategies. Such information on how to derive the equilibrium solution was given in the form of tips about the game and also contained a numerical example for computing the Nash equilibrium.

The tips explained several steps of iterated elimination of dominated strategies and pointed to the Nash equilibrium, whereas e.g. step-1 was illustrated as follows: "*Any number, chosen by all of the group members won't exceed 100. This means, the average won't exceed 100. The average* $+d(d = 50)$ *times q (here q=1/3) is 50. Therefore, the number you should choose, should not exceed 50. If all members of*

[31] The same labelling of the treatments has been used by Morone et al. (2008).

your group realize this, everybody else will choose 50, and therefore the average will be 50. Again, your number should not exceed the average +d times q=1/3, which is 33.33, in order to earn as many points as possible. […]"[32] Both "informed" and "uninformed" subjects (respectively the ones provided with the tips or not) who were assigned to each group knew how many of the participants were informed and how many were not. The common knowledge of the dissemination of the theoretic information intended to put the conditions for testing the principles of iterated elimination of dominated strategies at their partial absorption.

In the third treatment (II- or full absorption treatment) all subjects were informed on how to iteratively eliminate dominated strategies so as to reach the Nash equilibrium and knew that all subjects had received the same information. With that, the third treatment aimed at testing the full absorbability of the guessing game theory.

The participants knew that they would have been randomly matched at the beginning of the experiment and that their group composition would have remained the same for the whole experiment. Control questions on the basic rules of the game ensured the understanding of the experimental instructions (but not of the theoretic information contained in the tips).

After each round, the participants received a feedback on the average of guessed numbers in their group and their own individual payoff per round to control for reputation seeking.[33] Further, changing parameters were adopted to minimise conditioning from the previous choices.[34]

The experiment was run in April 2006 at the experimental laboratory of the Max Planck Institute of Economics in Jena. 96 undergraduate students from Jena University (32 for each session) took part in the experiment, the computerized experiment has been programmed with the software z-Tree,[35] and the ORSEE software[36] was used for participants recruitment. With a show-up fee of 4 € the participants earned in average 8.41 € and the experimental sessions took about 45 min. The experiment's instructions are in appendix.

7.5.4 Experimental Results

Using different guessing games, the experiment aimed at testing whether the principles of iterated elimination of dominated strategies can be cognitively captured rather than non-cognitively adapted. As for bounded rational individuals, absorbability of the equilibrium prediction in the form of perfect compliance may not be granted. The benchmark of perfect strategic interaction has to be compared with the guesses of bounded rational subjects in the three different treatments.

---

[32] The tips are contained in appendix.

[33] Cf. Camerer, Ho, and Chong (2001).

[34] Cf. Ho et al. (1998).

[35] Cf. Fischbacher (1998).

[36] Cf. Greiner (2004).

The concept of thinking depth or of depth of reasoning has been analyzed by many studies and has been typically applied to the analysis of guessing as well as to many other games. It refers to the number of steps of elimination of dominated strategies or, in other words, to the degree of iterated dominance.[37] There is much theoretical and experimental research in exploring the idea of finite depth of reasoning, as illustrated e.g. by the theoretical contributes of Binmore (1987), Selten (1991) and Stahl (1993) and by the experimental studies in Mc Kelvey and Palfrey (1992), Beard and Beil (1994) or Stahl and Wilson (1994).

However, as the consistency of individual choice behaviour over nine guessing games with different parameterizations in terms of stability of revealed thinking depth is unlikely, clustering the subjects according to their depth of reasoning does not represent a suitable analytic procedure for this experiment. Therefore, the analysis of the experimental data will be restricted by using, instead of the concept of depth of reasoning, the absolute deviation from the equilibria to catch the level of application of the principles of iterative eliminating dominated strategies.

Because Nagel (1995) provides evidence for behavioural adjustment toward equilibrium and ascribes it to qualitative learning which is sensitive to changes in the parameters, in this spirit, the present study faces the subjects with different one-period guessing games. Changing parameterizations provides a suitable framing for testing theory absorption in that it prevents and discourages qualitative learning and disentangles the effects of revealing the theoretical principles for deriving the equilibrium.

As follows, the main results on the general effects of providing the individuals with theoretical information will be discussed with consideration to the payoffs, the deliberation times and the choices' deviations from the equilibrium.

7.5.4.1 General Effects of Theory Absorption

Aspects of full theory absorption can be discussed from the comparison between UU- and II-treatments, while partial theory absorption can be analysed considering them in respect to the UI-treatment as well as considering the within-subjects differences in this last treatment.

The experiment provides evidence that revealing the subjects theoretical information about the principles of iterated elimination of dominated strategies yields for (1) smaller equilibria deviations, (2) longer processing time for making a guess, and (3) higher profits.

The choices' deviations from equilibrium convey an indicator for appreciating the absorption of the theory because they can be interpreted as a measure of the acceptance and compliance with it. For a more robust picture of actual behaviour, instead of the comparison of the mean of the nine period choices per subject, the 0.25, 0.50 and 0.75 quantile aggregates of the nine period choices per subject will be focussed on.

[37] Cf. Nagel (1995, p. 107).

Table 7.2 First period results of subjects in the three treatments (Morone et al., 2008)

	Individual choice deviation from the equilibria			Groups' average choice deviation from the equilibria			Time per subject to choose a guess			Individual profit		
Treatment	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
UU*	9.50	16.50	25.00	16.46	18.25	20.48	21.25	32.50	43.50	0.57	21.05	32.85
UI-U$'$	13.25	17.50	25.00	9.84	12.53*	12.60	21.50	27.50	47.00	2.34	15.70	28.85
UI-I^{+}	4.50	12.25	19.75	9.84	12.53	12.60	40.75	57.50'	71.25	3.55	34.10	40.04
II	0.00	3.5^{*+}	15.00	3.66	5.53^{*+}	8.06	23.75	45.00*	63.00	15.27	39.69^{*+}	44.53

*, ', and $^{+}$ mark significant differences to either UU (*), UI-U ($'$) or UI-I ($^{+}$) subjects

Table 7.2 summarizes the first period results in the different treatments and offers an overview of the quantile aggregates of individual and groups' choices deviations from the Nash equilibrium, the deliberation time per subject and the individual profits. Significant differences $(p < 0.05)$ among treatments are marked as well. The labels UI-I and UI-U, which are contained in the table and will, for convenience, be adopted henceforth, respectively refer to the participants to the partial absorption treatment (UI) who were informed (I) about the theoretic principles for deriving the equilibrium and those who were uniformed (U).

First period choices capture choice behaviour which was unaffected by learning and conditioning. Providing theoretic information on how to derive the Nash equilibrium, a modification in the individual choice behaviour was induced, as (see Table 7.2) uninformed subjects participating in the UU-treatment deviated significantly more from the equilibrium than informed subjects in the absorption treatment $(p < 0.01)$.[38]

The same holds for the equilibrium deviations of group averages: it could be observed that the group averages of uninformed (UU) subjects deviated significantly more from equilibrium than those of informed subjects in the full absorption treatment (II) $(p < 0.01)$. Providing all group members with theoretic information on the principles of iterated elimination influenced all group members' behaviour, which thus reduced the group average equilibrium deviation.

Revealing the principles of iterated elimination of dominated strategies induced the subjects to think more deeply about the game and about how to choose because a positive relation between the time needed for making a choice and the thinking effort put in the task of guessing can be reasonably assumed.[39] As shown in Table 7.2, UU subjects needed significantly less time to type in their guesses than II subjects $(p < 0.01)$.

After having a closer look at the payoffs, it could be observed that subjects in the control treatment (II) earned significantly less than subjects in the full absorption treatment (II) $(p < 0.02)$. This is because first period's choices of UU subjects ex-

[38] Whenever not differently specified, all tests have been performed with an asymptotic Wilcoxon signed rank test. Further analysis on the experimental data is in Morone et al. (2008).

[39] Cf. Güth et al. (2002).

hibited a significantly larger variance of the discrepancy of individual choices than groups' averages ($p < 0.002$). Specifically, the absolute variance of UU subjects' choices was 286.74, while the one of II subjects was 85.84. This reveals the immediate impact of providing individuals with theoretical information, which affected the individual choice behaviour from the first period on. Hence, this could be interpreted as a sign of propensity to accept and apply the theory, trusting the others to accept it and apply its prescriptions, as well.

Informing the subjects about the principles of iterated elimination of dominated strategies induced an immediate change in their choice behaviour. In homogenously informed groups (in the II treatment), the subjects self-refer the game theoretical predictions on themselves and suppose the others to do the same. The subjects projected their own choice onto the choice of similarly informed group members. The theoretic principles on how to derive the equilibrium worked in a stabilizing way on the beliefs about the others' choices and increased the trust in the possibility of predicting them with a certain degree of accuracy[40] possibly because a commonly known pattern of reasoning (embodied by P.1 and P.2) had been provided. The advantage of being informed can be ascribed to the regular application and trust in the others to apply the theoretical prescriptions. As informed subjects expected more choices to be close to the equilibrium, the groups' variance of choices decreased leading, therefore, to higher payoffs.

These considerations of the effects of informing the subjects about the guessing game theory hold even in respect to the choices in all periods. As shown by Table 7.3, the choice behaviour over all periods seems to confirm the tendencies revealed by the first period results.

In all of the successive periods, the subjects in the control treatment (UU) revealed a higher choice deviation from the equilibria than II subjects ($p < 0.01$). Groups' averages deviated from equilibrium significantly more in the UU than in the II treatment ($p < 0.001$), too. In all periods the deliberation time was significantly lower among uninformed subjects participating in the UU treatment than among informed subjects in the full absorption treatment ($p < 0.01$). Overall profits of II subjects were higher than those of UU subjects ($p < 0.01$).

Table 7.3 All period results of subjects in the three treatments (Morone et al., 2008)

	Individual choice deviation from the equilibria			Groups' average choice deviation from the equilibria			Time per subject to choose a guess			Individual profit		
Treatment	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
UU*	8.38	19.9	33.00	11.94	18.50	26.91	16.75	24.00	36.00	0.00	24.57	38.20
UI-U′	8.00	17.00*	27.00	10.12	14.58*	20.91	20.00	32.00*	52.25	4.83	29.02*	39.33
UI-U^{+}	5.00	14.75’	25.00	10.12	14.58	20.91	26.00	43.50’	60.00	9.06	26.86	38.82
II	0.00	10.00^{*+}	25.00	7.42	10.96^{*+}	18.20	25.00	43.00^{*+}	64.00	9.17	30.84^{*+}	42.71

Note that *, ’, and $^{+}$, mark significant differences ($p < 0.05$) resp. to UU (*), UI-U (’) or UI-I ($^{+}$) subjects

[40] Cf. Dawes (1988).

7.5.4.2 Equilibrium Behaviour

Focussing on equilibrium choices has been interpreted as a way of disentangling learning effects from immediate theory absorption, i.e. the immediate jump to the equilibrium choice. In this sense providing theoretic information associates overall with an increasing number of equilibrium choices.

As illustrated by Table 7.4, in the first period the subjects in the full absorption treatment (II) chose the equilibrium more often than the subjects both in the partial absorption treatment (UI) and in the control treatment (UU). Both differences are significant with $p < 0.01$. The subjects in all-informed groups performed significantly better than the uniformed subjects $(p = 0.011)$.

In the UI-treatment the number of the first-period equilibrium hits did not differ significantly from their number in the UU-treatment. However, this was revealed to be quite high possibly because of the high number of economics and natural science students among the participants.

A binomial test both on the first and on all successive periods revealed no constant equilibrium choice behaviour of any subject in any of the treatments $(p < 0.001)$. Hence, the principles of iterated elimination of dominated strategies cannot be claimed of having been fully absorbed by the bounded rational experimental subjects. Full absorbability would have required the individuals to comply with the theory's advice. In a satisficing perspective, the compliance with the theory can be translated in the elimination of strategies until further elimination won't produce higher payoffs. At least when all players are known to be informed about the theoretic principles for inferring the Nash equilibrium, it could have been reasonable, even in a satisficing approach, to expect the individuals to jump immediately to the equilibrium from the very first period on.

Still, the significant differences between the subjects' behaviour among treatments prove that the individuals actually perceive the reflexive character of the game's theoretic propositions, which are related to the context faced and induce a modification of behaviour. The individuals modify their behaviour in consideration of the others' expected choices. The others' behaviour is anticipated on the basis of the individual assumptions on the other subjects' limited absorption capabilities and of the doubts on their acceptance of the theoretical advice.

In the experiment the modifications in the individuals' choice behaviour (illustrated e.g. by equilibrium choice deviation) were inversely proportional to the disseminated knowledge: while the average first period choice deviation from equilibrium was 5.53 among subjects in the II treatment, it was 9.97 among UI and 18.7 among UU.

Table 7.4 Percentages of equilibrium choices in the three treatments (Morone et al., 2008)

Period	1	2	3	4	5	6	7	8	9	Total
UU*	15.63	6.25	6.25	34.38	3.13	21.58	12.50	0.00	0.00	**11.11**
UI$'$	12.50	3.13	28.13	9.38	6.25	9.38	6.25	6.25	3.13	**9.38**
II	43.75*’	6.25	59.38	40.63	6.25	28.13	9.38	21.88	12.50	**25.35***’

* and ’ mark significant differences to either UU $(^{*})$ or UI (’) subjects

Analysing choice behaviour over all periods corroborates similar conclusions, as in the full absorption treatment the subjects opted for the equilibrium significantly more often than in the partial absorption treatment ($p < 0.001$) and in the control one ($p = 0.003$).

The common knowledge of the heterogeneous information dissemination (i.e. having informed all subjects in the UI treatment that half of them received theoretic information and half did not) significantly reduced the overall quote of equilibrium hits also in comparison to their number in the UU treatment ($p < 0.001$). The frequency of equilibrium choices in the UI treatment remains however higher among informed (henceforth labelled as UI-I) than among uninformed (UI-U) subjects ($p < 0.001$), adding respective to 21 and 6 equilibrium hits.

Thirty-eight of the 432 choices (8.79%) made by the 48 uninformed subjects equal the equilibrium, while 47.92% of the uninformed subjects never hit the equilibrium. During the nine periods, the 48 informed subjects chose the equilibrium 94 times, i.e. by the 21.76% of cases, while only 33.33% of them never typed in an equilibrium guess.

In consistency with the findings of previous studies,[41] interior equilibria were chosen significantly more often than border equilibria ($p < 0.001$). This provides further evidence for interior solution to be easier to calculate than solutions which lie on the boundaries. The relative frequencies of equilibrium choices in border and interior equilibria games in the different treatments are represented in Fig. 7.2.

The differences in the equilibrium choice behaviour reveal that providing the respondents with theoretical information induces a significant increase of equilibrium choices ($p < 0.001$). Instant changes in the choice behaviour acknowledge the

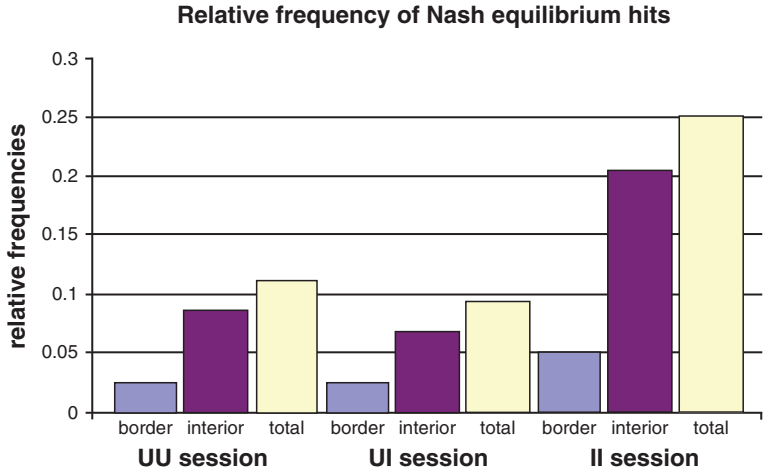

Fig. 7.2 Relative frequency of equilibrium hits in games with interior and border solutions Morone et al. (2008)

[41] Cf. Güth et al. (2002) and Ho et al. (1998).

cognitive capturing of the principles of iterated elimination of dominated strategies, as revealed by the hypothesis of equivalent equilibrium choice behaviour between the II- and the UU-treatment which could be rejected on the 1% level for the first period and on the 5% level over all nine periods.

7.5.4.3 Partial Theory Absorption

The analysis of partial absorption of the guessing game theory is revealed to be a particularly challenging task, since heterogeneous distribution of the theoretic information can raise the strategic uncertainty regarding the behaviour of the others.

The experiment gives evidence first for the definite knowledge about the heterogeneous information structure of the group to induce informed subjects (UI-I), to deviate more from equilibrium than the uniformed subjects (UI-U). Second, it accounts for similar time consumption by both kinds of subjects, which was higher than in the control treatment. Third, the profits in the UI-treatment were similar to those in the UU-treatment and lower than in the II-treatment.

Starting by observing the first period choice behaviour (see Table 7.2), it can be noted that informed UI subjects' choices deviated more from equilibrium than those of the informed II subjects $(p < 0.01)$, while the choices of the uninformed respondents in the UI-treatment were similar to those in the UU-treatment. A further interesting feature of the UI-treatment is that both informed and uniformed subjects acted similarly. Being aware of the presence of subjects who were not instructed on how to derive the equilibrium solution, the informed UI subjects' predictions contemplated quite large equilibrium deviation of the uninformed subjects. The informed subjects adapted their choice to this expectation in the hope of sustaining profits. The uninformed subjects in the UI-treatment acted similarly to the subjects in the UU-treatment and needed the same time to make a guess, while their informed counterparts took significantly more time to deliberate a choice $(p < 0.023)$.

Group averages' deviations from equilibrium were significantly smaller in the UI- than in the UU-treatment because informed UI-subjects typed slightly lower numbers than uninformed UI-subjects (this difference being however non-significant). Consequently, the informed subjects in the UI-treatment earned significantly less than the informed subjects in the II-treatment $(p < 0.03)$.

In summarization, no significant differences could be observed between the first period choices of the uninformed subjects in the UI- and in the UU-treatment, or between those of the uninformed and informed subjects in the UI-treatment. This indicates that subjects possessing superior theoretic information tried, when interacting with less informed individuals, to anticipate their mental limitations and correspondently modified their choice behaviour, but could not avoid profit losses compared to superiorly informed subjects in homogeneous groups.

Considering the choice behaviour of the uninformed UI-subjects during all periods (as shown by Table 7.3), significantly smaller deviations from the equilibrium than among the uniformed UU-subjects can be noted $(p < 0.02)$. This can be ascribed to the awareness of having to interact with superiorly informed sub-

jects and could be derived in absence of a theory to be captured from non-cognitive idea adaptation.[42] Because evidence for this effect can be only found if considering the overall period choices, such a non-cognitive adaptation seems to be based on an improvement of the individual heuristics or changes in the mental representations of outcomes. Overall, however, partial theory absorption lowered the profits which increased the likelihood of deviation from equilibrium because the informed UI-individuals did not reasonably expect their uniformed counterparts to be able to infer the iterative principles of eliminating dominated strategies. Informed UI-subjects adapted their choice behaviour to such expectations so that their choices deviated significantly more from the equilibrium than choices of the informed II-subjects $(p < 0.05)$. Therefore, informed subjects in the UI-treatment can be considered to steer the higher group averages in the UI-in comparison to the II-treatment $(p < 0.012)$. Still, the average absolute deviations from the equilibrium were lower than those of the groups belonging to the control treatment $(p < 0.01)$.

Interestingly, uninformed UI-subjects took more time typing in their choices than UU-subjects $(p < 0.01)$. This indicates that knowledge about the heterogeneous distribution of theoretical information influenced the deliberation time taken by the uninformed subjects, which challenged them to think harder about the game and how to play. While the informed subjects in the UI-treatment needed more time than their uniformed group members $(p < 0.01)$, they required less time than the informed II-subjects $(p < 0.05)$.

In regard to the profits, they were similar within the UI-treatment, whereas the uninformed subjects gained more in the UI- rather than in the UU-treatment $(p < 0.05)$, while the informed UI-subjects yielded significantly lower profits than what II-subjects $(p < 0.05)$.

The informed subjects who were assigned to the partial absorption treatment carried the burden of their groups' uninformed members in the form of profit losses. While the uninformed subjects in heterogeneous groups adapted their choice behaviour without lowering (compared to the uninformed subjects in UU-groups) their earnings, the informed subjects in UI-groups suffered profit losses compared to the subjects interacting in II-groups. This can be alleged to the high strategic uncertainty in the heterogeneous groups regarding the uninformed subjects' behaviour. Higher time consumption and smaller equilibrium deviation of uninformed UI-subjects compared to UU-subjects can be claimed to depend on an essentially non-cognitive behavioural adaption.

7.5.5 Conclusions

The game theoretical predictions for the guessing game were used in this experiment to test them at their absorbability and inform only a part or the totality of the individuals interacting. For disentangling the effects of theory absorption from

[42] "*With no theory to capture, this ought to be a consequence of non-cognitive idea adaptation.*" Cf. Morone et al. (2008, p. 172).

those of learning the experiment comprised nine one-period guessing games with different parameterizations.

As a benchmark for tackling this research aim, a control treatment was run showing that by solely providing the basic rule of the guessing game, a more or less random choice behaviour emerged in a similar setting. On the contrary, in the groups all of whose members were provided with theoretical information on how to derive the guessing games' equilibrium solutions, both the individual guesses and groups' averages were significantly closer to the equilibrium. This could be observed in the very first as well as in all the other periods and hints at the cognitive capturing of the theoretic principles of iterated elimination of dominated strategies and the self-reflection of their advice. Although a significant impact of revealing theoretic information to all group members was observed (in terms of increased frequency of equilibrium hits, longer processing times, lower group average deviation from the equilibrium and higher profits compared to the results of the control treatment), the hypothesis of full theory absorption, in the sense of perfect compliance to the theory's prescriptive, could not be acknowledged.

It can be assumed that the subjects interacting in all-informed groups did not choose to perfectly adhere to the theory in the hope of satisfying their profit aspirations. Assuming limited capabilities in the other group members, the informed subjects adapted their choice behaviour, in this way revealing the cognitive capturing of the theoretic principles with adaptation of their predictions to the capabilities ascribed to the others.

Being aware of the partial dissemination of theoretic information raised strategic uncertainty and diminished trust in the uninformed group members' capabilities. Assuming the same cognitive capturing of the theory by the subjects in the full and in the partial absorption treatments, the informed subjects in the latter adapted the theoretic prescriptions even more because of the presence of uninformed subjects in order to sustain their profits. Being aware of interacting with superiorly informed counterparts induced the uninformed UI-subjects to modify their choice behaviour in an essentially non-cognitive way. In comparison to the uninformed subjects in the control treatment they needed more time to type guesses which deviated less from the equilibrium in all periods and therefore, steered group averages toward the equilibrium.

Therefore, partial theory absorption can be linked to a combination of non-cognitive ideas adaptation by uninformed subjects and cognitive alternations of the captured principles by informed subjects, whereby the degree to which a theory's predictions are adapted depends on the commonly known information dissemination. The trust in the others' capabilities of capturing the principles of iterated elimination of dominated strategies was obviously lower in case of partial rather than full information dissemination. This, adding to the consideration of the bounded rationality of the others, led to revisions of the game's theoretic principles and motivated the observed violation of perfect theory absorption.

The three main findings from this experiment can be summarized as follows:

1. Subjects who received theoretical information about the principles of iterated elimination of dominated strategies revealed smaller deviations from equilibrium

choices, needed longer processing time, and higher profits in the first and in all periods.
2. Theoretic information triggers a higher number of equilibrium choices.
3. "*Definite knowledge about the heterogeneous information structure of the group induces (un-)informed UI-subject to exhibit (lower) higher choice deviations from equilibrium, (higher) similar time consumption, and (similar) lower profits than (UU) II subjects*."[43]

7.6 An Experimental Study on the Absorbability of Herd Behaviour and Informational Cascades Theories[44]

This experiment discusses the absorbability of informational cascades' theory by bounded rational decision-makers. In this insight, it analyses whether providing individuals with theoretic information on informational cascades affects the overall probability of the herding phenomena to occur as well as whether an incorrect cascade can be reversed because of bounded rational adapting of the theory's prescriptive. In particular, the occurrence of informational cascades in an experimental investment task will be observed. The conception which underlies this analysis is that providing the subjects with theoretic information on probability assessment could make them aware of the fragility and idiosyncrasy of informational cascades and thus affect the probability of (erroneous) cascades to occur.

As follows, after a short review of studies on herding and informational cascades, a simple model will provide the theoretical framework for explaining informational cascades. An overview of related experimental evidence will be presented, followed by the design and the main results of this experiment.

7.6.1 Herding and Informational Cascades

Conformity and fluctuations in mass behaviour are frequent features which characterize many social and economic situations.[45] Individuals are influenced by the behaviour of the others, as it can be informative to many extents and promote what has been depicted as "social learning."[46] Trying to learn from the others typically induces imitative behaviour which can under some circumstances be rational even when it implies choosing differently rather than solely relying on one's own information. Despite individual rational behaviour, erroneous, inefficient outcomes can arise on this basis.

---

[43] Cf. Morone et al. (2008, p. 171).

[44] This experiment and its results are discussed in Fiore, Morone, and Sandri (2007).

[45] Cf. e.g, Avery and Zemsky (1998), Çelen and Kariv (2004), Scharfstein and Stein (1990) and Welch (2000).

[46] For surveys, see e.g. Douglas (1996) and Bikhchandani et al. (1998).

The phenomenon of "herding" has been widely analysed in the last decade. The notion of "herd behaviour" refers to the phenomenon according to which people ignore their own private information in order to follow the example of others. Becker (1991) first pointed out this kind of behaviour, whose analysis was then further developed by Banerjee (1992) and Bikhchandani, Hirshleifer, and Welch (1992).[47]

Herding model and informational cascades are interactive, have a clear economic interpretation, are very simple to explain and represent a setting in which the rational action is independent by the subjects' preferences. These features make them particularly attractive for studying rationality and learning, in general, and interesting for experimentally testing theory absorption among bounded rational decision makers.

In addition, as herd behaviour applies to a conspicuous number of social and economic issues, investigating how it can be influenced by theory absorption could yield useful implications, even for defining procedures for preventing erroneous cascades to occur or at least for containing their frequency or dimensioning their consequences.

In many real-life situations, people have to make their decisions sequentially. They can observe the decision taken by previous subjects and often be influenced by that. For example, when having to choose between two restaurants, in absence of other information, people are more prone to prefer the more crowded one. This behaviour is motivated by the assumption, which intuitively sounds rational, that the number of diners reflects the quality of the restaurant. As this kind of logic inspires all subjects, if a few subjects randomly decide to enter the restaurant, then all later subjects will join the queue, which starts a perverse dynamics according to which every new subject joining the queue raises the probability for others to queue up.[48]

Ignoring private information for joining the queue is the basic mechanism of herd behaviour. Assuming rational expectations implies that an agent could compensate her lack of knowledge on the true model of the situation she is confronted with, and thus reach an efficient outcome by drawing on all available information, therefore, encompassing the observation of the others' choices.

In a setting in which individuals act sequentially and have common knowledge of the history of the game,[49] the players will inform their decision-making both to their private information and to the actions of their predecessors, without knowing which information the choices were based upon. As a result, the choices of the individuals will be correlated, even if their personal information and background differ, and mistakes by early decision-makers might be transmitted to latter ones, delaying in all likelihood the settlement for an efficient outcome or even preventing it altogether.

Two simple frameworks for modelling herd behaviour are proposed by Banerjee (1992) and Bikhchandani et al. (1992). They both refer to a setting in which sequen-

[47] Examples of experimental investigations of herding behaviour are among others Allsopp and Hey (2000), Anderson (2001), Cipriani and Guarino (2005), Huck and Oechssler (2000), Hung and Dominitz (2004), Stiehler (2003).

[48] Cf. Becker (1991).

[49] By common we mean that at time t player t knows all the actions taken by previous players.

tially acting agents have to choose a winning action a among a set of alternatives. These two models differ in that while Bikhchandani et al. assumes the wining action a to be an element of the interval $[0,1] \in N$ and all subjects to receive a signal, in Banarjee's model a belongs to the set $[0,1] \in R$. The individuals can be either informed or uninformed, i.e. they can either have received a signal or not.

Further, Bikhchandani et al. (1992) explore the concept of informational cascades. In the process they try to explain not only conformity among agents but also "*rapid and short-lived fluctuations such as fads, fashions, booms and crashes*."[50] They point out that the conformity of followers in a cascade contains no informational value. In this sense, the cascade is fragile because it can be upset by the arrival of new public information (while it cannot be reversed if superior information is not provided). It is also idiosyncratic "*in that random events combined with the choices of the first few players determine the type of behaviour on which individuals herd*."[51]

Based upon the model of Bikhchandani et al. (1992), Willinger and Ziegelmeyer (1998) developed a framework with asymmetric accuracy of private information in order to experimentally investigate the aspect of fragility of informational cascades. Their analysis revels that providing the subjects who decide immediately after the occurrence of a cascade with an additional signal lowered the occurrence of informational cascades and interrupted herding.

7.6.2 A Simple Model of Informational Cascades: A Dichotomy Choice Model

An informational cascade, which occurs when people prefer to follow the behaviour of the others ignoring their own information, can be modelled by the following game.

Consider N players who have to choose sequentially among two options, namely urn B (for black) and urn W (for white). While urn B contains two black and one white balls, urn W has two white balls and a black one. Player 1 draws a ball from a randomly chosen urn and has then to guess from which urn she has pulled the ball. A correct guess will yield her a payoff of 1, a wrong guess will yield a 0. Player 2 who can observe player 1's guess has to pull a ball from the same urn and then to guess. Player 3 observes the choices of both previous players, draws a ball and makes her choice and so on until player N who can type in her guess by having observed the choice of all (N-1) precedent players.

As it is rational of player 1 to guess the black (white) urn when she draws a black (white) ball, player 2 is therefore confronted with one of the following scenarios:

1. she observes player 1 having chosen the black urn, and draws a black ball;
2. she observes player 1 having chosen the black urn, and draws a white ball;

[50] Cf. Bikhchandani, Hirshleifer, and Welch (1992, p. 994).

[51] Cf. Bikhchandani and Sharma (2001, p. 284).

3. she observes player 1 having chosen the white urn, and draws a black ball;
4. she observes player 1 having chosen the white urn, and draws a white ball.

It is therefore rational for player 2 to choose the black urn in the first scenario and to opt for the white urn in the fourth scenario. On the contrary, in the second and third scenarios she should be indifferent to the two urns and choose among them with equal probability.

Under these assumptions and after two individuals have played, the unconditional ex-ante probabilities of an informational cascade choosing the white urn ("White-cascade"), the black urn ("Black-cascade"), as well as the probability of no cascade to occur ("No-cascade") are calculated by Bikhchandani et al. (1992) as follows:

$$\text{White}-\text{cascasde} = \frac{1-p+p^2}{2}; \text{No}-\text{cascasde} = p - p^2;$$
$$\text{Black}-\text{cascade} = \frac{1-p+p^2}{2} \tag{7.8}$$

After an even number of players $(n = 2m)$ have played, they calculate these probabilities to become:

$$\text{White}-\text{cascasde} = \frac{1-(p-p^2)^m}{2}; \text{No}-\text{cascasde} = (p - p^2)^m;$$
$$\text{Black}-\text{cascade} = \frac{1-(p-p^2)^m}{2} \tag{7.9}$$

where p stands for the probability of observing a correct signal. As a result, the bigger p is, i.e. the more accurate the signal is, the sooner an informational cascade can start (see Fig. 7.3).

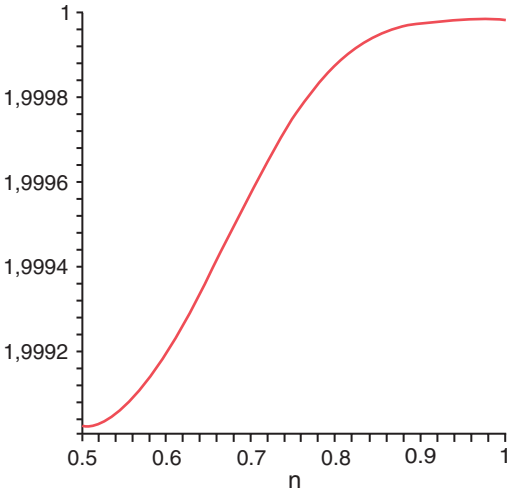

Fig. 7.3 Probability of starting a cascade as a function of p, the correctness of the signal (N = 10)

Further, Bikhchandani, Hirshleifer and Welch calculate the probability of ending up in the correct cascade after two players have chosen, given that the chosen urn is the white one. These are respectively the calculations:

$$\text{White} - \text{cascasde} = \frac{p(p+1)}{2}; \text{No} - \text{cascasde} = p(1-p);$$
$$\text{Black} - \text{cascade} = \frac{(p-2)(p-1)}{2} \tag{7.10}$$

In the general case, i.e. after an even number of players $(n = 2m)$, these probabilities can be generalized as:

$$\text{White} - \text{cascade} = \frac{p(p+1)\left[1-(p-p^2)^m\right]}{2(1-p+p^2)} \tag{7.11}$$
$$\text{No} - \text{cascade} = (p-p^2)^m \tag{7.12}$$
$$\text{Black-cascade} = \frac{(p-2)(p-1)\left[1-(p-p^2)^m\right]}{2(1-p+p^2)} \tag{7.13}$$

Although the probability of observing a correct cascade, illustrated by Eq. 11, increases in p and m, even for very informative signals (i.e. p close to 1), the probability of a wrong cascade remains remarkably high, as shown by Eq. 13.

Figure 7.4 represents the probability of a correct and incorrect cascade for $N = 10$.

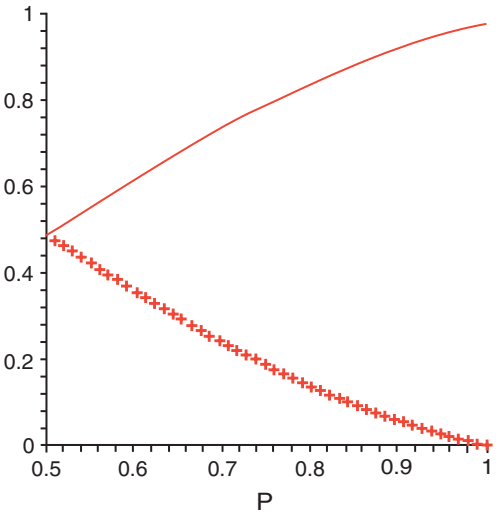

Fig. 7.4 Probability of a correct (continuous line) and incorrect cascade (dotted line) as a function of p, the correctness of signal $(N = 10)$

7.6.3 Experimental Design

The experiment included two treatments: while the control treatment is based on the model in Bikhchandani et al. (1992), in the absorption treatment the participants received theoretical information about informational cascades. More precisely, in the absorption treatment the respondents were provided with an illustration of how to infer the expected value of adoption and rejection in dependency of the accuracy of the private signal and how to deduce the individual optimal decision rule. The absorption treatment informed the experimental hypothesis of the theoretic information provided to prevent incorrect cascades to occur.

The experiment was programmed using the software Z-tree[52] and was run in March 2007 at the laboratory of ESSE at the University of Bari.

Each treatment took about an hour and was made up of 22 periods. Since the first two periods were trials, the final payment was made at the end of each treatment only on the 20 real game-playing periods. The experiment's instructions are contained in appendix.

$N = 10$ subjects participated in each session: each of them sat in front of a PC connected by a net and could neither see the others nor communicate with them. The participants were all undergraduate students in economics, who were not familiar with similar experiments.

In each period, which lasted for about two minutes, the subjects sequentially played in a randomly determined order. A message on their PC screen informed them when it was their turn. The subjects were asked to decide whether to invest in a new product or not, not knowing whether the new product would have been profitable or not. They knew that the profitability of the product was dependent on two equally likely events and that if the product would have been successful $(V = 1)$, each player would have gained 0.5 € in case of having invested, and zero otherwise. The opposite would have been true if the product would not have been successful $(V = 0)$, each player would have gained 0.5 € in case of not having invested (which would have been the right decision), and zero otherwise. In order to exclude losses by the participants, no cost of adopting was considered. The true value of V was exogenously determined in each period but was not revealed to the subjects, who only saw a free-of-charge signal S and knew it had a probability p of 0.75 of being correct.

In each period of the control treatment, the subjects were informed about their own turn to play, all previous guesses, and their own signal S.

In addition to such information, the subjects in the absorption treatment received a decisional aid in the form of tips about the game. It contained theoretical information on how to derive, in dependency of the individual position in the queue, the unconditional ex-ante probabilities of ending up respectively in a correct or in an incorrect cascade.

[52] Fischbacher (1998).

In both treatments the subjects were informed at the end of each period about the true value of *V* and their individual period-payoff. After all periods were played, the subjects were paid and could leave the laboratory. Average earnings were 7 €.

7.6.4 Results

The experimental results enable to discuss some aspects of absorbability of informational cascades' theory by bounded rational decision-makers. The experimental hypothesis which argues whether providing individuals with theoretic information on informational cascades affects the overall probability of herding phenomena and that of their reversal will be in particular discussed focussing on the general effects of theory absorption, comparing the social efficiency of outcomes among treatments and testing for directional learning. The main results, this experiment accounts for, can be discussed as follows.

7.6.4.1 General Effects of Theory Absorption

Considering the benchmark provided by the theory on herding (and informational cascades), it is possible to qualify the individual choices observed in the experiment as rational or irrational, respectively if they are conform to the optimal strategy or not. In the experimental simple set-up, where the two states of nature are equally probable and the private signals identically distributed, the optimal strategy in a Bayesian sense can be defined taking into account the decision of the predecessors and the individual's private information, as doing the count on the previous decisions (the one's own signal being included)[53] and adopting the most chosen option. The adoption of the tie-breaking rule if indifferent has been assessed as rational (optimal) behaviour, both if generating a cascade or not. Further, the case in which "*an imbalance of previous inferred signals causes a person's optimal decision to be inconsistent with his or her private signal*"[54] has been considered as a cascade.

Choice which were not consistent with these rules has been qualified as irrational, whereas it has been distinguished between two subspecies of irrational behaviour, which have been labelled as "signal-keeping" and "not-rationalized." They respective correspond to the cases in which following one's own private signal can provide a somehow logical explanation of the individual's choice and to those in which there is no plausible explanation for it.

According to these criteria the choices of the experimental subjects can be grouped and summarized as shown by Table 7.5.

It can be clearly noted that providing the respondents with theoretical information about the optimal Bayesian strategy yields for higher consistency of behaviour with

[53] Cf. Anderson and Holt (1997).

[54] Cf. Anderson and Holt (1997, p. 851).

Table 7.5 Classification of the individual behaviours observed

	Signal	Rational (Bayes' rule)	Irrational	
			Not rationalized	Signal-keeping
Control Treatment	0,72	0,725	0,15	0,135
Absorption Treatment	0,715	0,935	0,035	0,04

Table 7.6 Perceptual occurrence of informational cascades

	Cascades occurrence (%)	Correct cascades (%)	Wrong cascades (%)
Control Treatment	48.07	24	76
Absorption Treatment	86.66	49.02	50.98

the theoretical predictions and, among the irrational choices, for the lowering of adoption of non-rationalized behaviours.

The higher compliance with the Bayesian optimal strategy provides evidence for the absorption of the theoretical predictions[55].

A further interesting feature which emerges from the experimental data is that while the frequency of optimal behaviour is higher in the absorption than in the control treatment, the overall frequency of signal following is almost the same among treatments (72% in the control treatment versus 71.5% in the absorption one). Therefore, the theoretic information did not affect the propensity of the individuals to rely on their own private information.

The two treatments gave account for a different occurrence of informational cascades. It has been in particular considered, how many informational cascades which could have formed occurred in fact. In this insight, a cascade has been considered as possible to occur whenever the choice between following one's own private information and Bayesian optimization are mutually exclusive. In all of these cases, an informational cascade takes place if the individual ignore her own signal and prefer to herd.

According to these criteria, while in the control treatment 25 of the 52 possible informational cascades formed, in the absorption treatment 51 out of the 59 informational cascades that would have been possible established. This acknowledges for the percentages which are shown by Table 7.6.

Fragility of cascades is higher in the control treatment, in which in particular it never happened that a cascade starts at the beginning of the period and last until the end of it. Instead, in the absorption treatment, informational cascades and herding seem to be more difficult to reverse. In this insight, it should be however considered that the occurrence of correct cascades is higher in the absorption than in the control treatment: while in the control treatment only 9 of the 25 informational cascades which affirmed were correct, in the absorption 25 correct and 26 wrong cascades formed. Table 7.6 summarizes the relative occurrence of correct and wrong cascades in perceptual values.

[55] At a 0.01 significance level.

7.6.4.2 Social Efficiency

For testing if proving the respondents of theoretical information on the optimal decision-rule enhances their profits, improves their winning chances and better therefore the efficiency degree of the outcome, the earnings distribution per position in the queue per treatment can be compared (cf. Fig. 7.5).

We can accept at a 95% significance level that earnings are higher, on average, in the absorption than in the control treatment.

The percentage of winning per position in the queue (cf. Fig. 7.6) which can be considered as a proxy for the individual utility is clearly higher in the absorption than in the control treatment, this difference being significant with $p < 0.043$.

In this sense, providing the individuals with theoretical information on herding behaviour and informational cascades reveals to be a device which can be preferable both from the social and from the individual point of view.

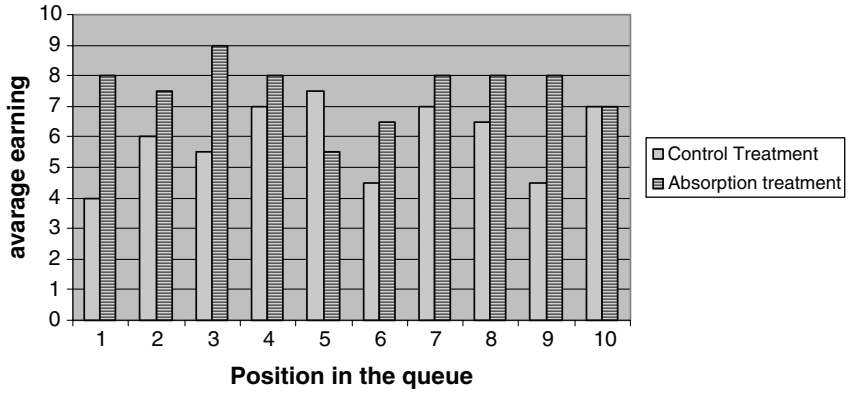

Fig. 7.5 Average earnings per position in the queue per treatment

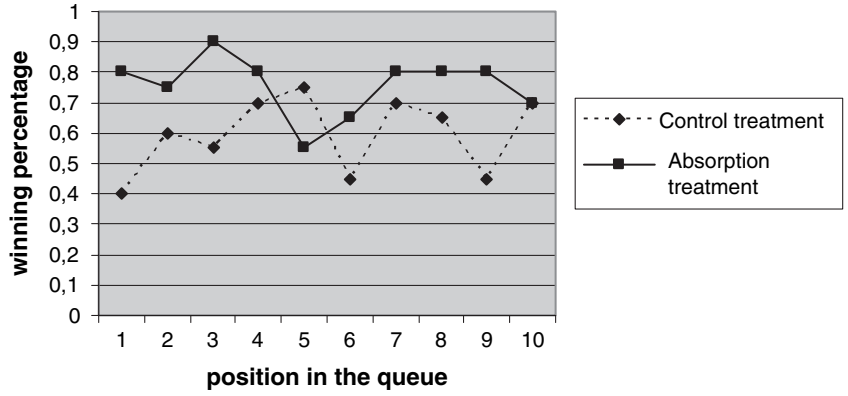

Fig. 7.6 Percentages of winning per position in the queue per treatment

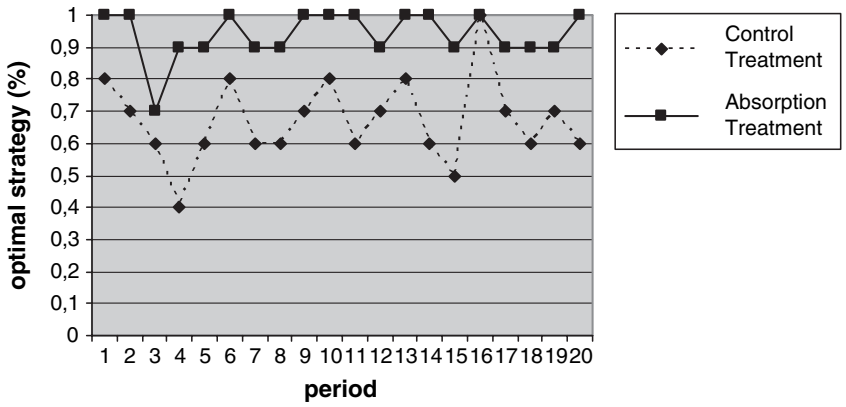

Fig. 7.7 Percentages of optimal strategy per treatment

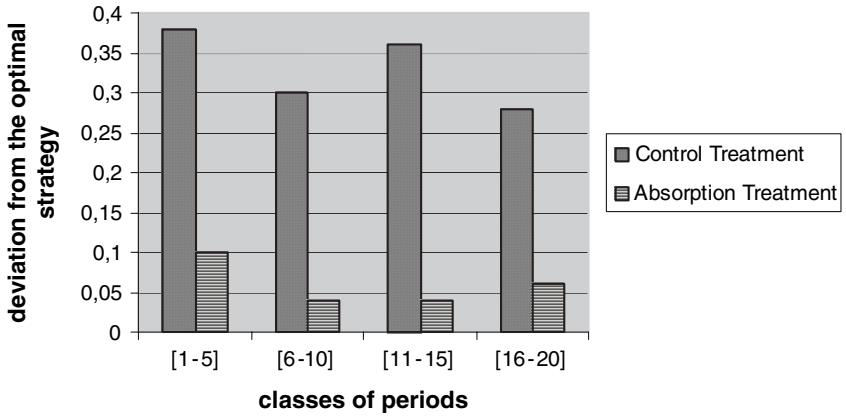

Fig. 7.8 Average percentages of deviant strategies per classes of periods

Figure 7.7 illustrates the percentages with which the winning strategy has been per period adopted in each treatment. At a significance level of 99% higher adoption of the optimal decision-rule implies that the individuals are able to conceptualize the theory and to refer it on the setting they face.

7.6.4.3 Learning

In order to appreciate whether learning effects occurred in the experimental setting the 20 game-playing periods can be divided onto 4 groups. The average percentages of deviations from the optimal strategy per classes of periods (see Fig. 7.8) does seem to corroborate the idea of a marked learning process, neither in the control nor in the absorption treatment. In particular in the control treatment, a slight tendency toward the reduction of deviant choices can be noted.

7.6.5 Conclusions

Herding behaviour and informational cascades refer to cases in which individual rational behaviour may result in a non-optimal strategy at aggregate level. As suggested by Becker (1991), these phenomena deal with situations in which information can be somehow linked with negative externalities. By conforming to the behaviour of the preceding subjects ignoring ones' own private information the behaviour of the others stops up to a certain point to be informative. Some studies on herding argue if "*society may actually be better off by constraining some of the people to use only their own information*."[56] This has been also acknowledged by some empirical evidence.[57] Shifting the perspective, the present study investigates whether providing the individuals acting in settings in which herding and informational cascades are likely to occur with the theoretical principle of Bayesian optimization constitute an alternative mechanism for improving the degree of social efficiency.

The evidence from this experiment seems to corroborate this idea, in that average earnings and the percentage of winning per position in the queue were higher among the responders who received the theoretic information than among those who did not. In this sense, providing the individuals with theoretical information on herding behaviour and informational cascades reveals to be preferable both from the social and from the individual point of view.

The experimental evidence further proves that theoretic information on informational cascades affects the overall probability of herding phenomena and that of their reversal. It could be in particular observed that the behaviour of the respondents who received the theoretic information revealed a higher consistency and compliance with the theoretical prescriptions and yields for lowering the occurrence of non-rationalized behaviour. The higher compliance with the Bayesian optimal strategy corroborates the hypothesis of absorbability of informational cascades' theory, revealing its understanding and acceptance by bounded rational individuals. The theoretic information did not affect the propensity of the individuals to rely on their own private information.

The overall occurrence of informational cascades was higher among informed than among uniformed individuals, whereas the frequency of correct cascades was more than double. Thus providing the individuals with theoretical information on herding behaviour and informational cascades reveals to be a device which improves the occurrence of cascades that can be associated with individually and collectively favourable outcomes. Informational cascades seem however to be less fragile and more difficult to be reversed among individuals who are aware of the theoretical prescriptions.

The experimental results provide, in none of the treatments, significant evidence for marked patterns of directional learning concerning the application of the optimal decision-rule.

---

[56] Bannerjee (1992, p. 798).

[57] Cf. e.g. Fiore and Morone (2007).

7.7 Concluding Remarks on the Experimental Examination

What emerged from the experimental examination of the self-referentiality of economic theories and their absorption confirms so far the intuition that the investigation should include a large spectrum of experiments. The experimental evidence corroborates the thesis that absorbability of economic theories is, in principle, always possible and that its concrete modality are context specific and depends on the complex mechanisms ruling bounded rational cognition and decision-making.

In this dissertation the effects of the self-referentiality of economic theories and their absorption have been experimentally analyzed through the observation of how individuals deal with meta-theoretical information in different experimental contexts. The experiments which have been conduced and presented in the framework of this dissertation focussed on the absorbability of economic theories of full rationality.

The effects of providing bounded rational individuals with meta-theoretical information of full rationality have been in particular observed in the experimental settings of p-guessing games and of informational cascades. Both settings have several attractive characteristics for testing theory absorption among bounded rational decision-makers. They namely permit the effects of rationality to disentangle from social preferences and have at the same time a very simple economic interpretation. Further, both the guessing game and herding behaviour represent interactive settings in which the individuals, in order to achieve a satisficing result, have to anticipate or interpret the others' behaviour and in which individually optimal behaving does not per se ensure success.

In general, the two experiments provide evidence that, although a significant impact of revealing theoretical information to the experimental subjects was observed, the hypothesis of full and complete theory absorption cannot be acknowledged as setting transcendent. Instead, the experimental findings corroborate the idea that individuals reflect the theoretical statements on themselves and on the situation they face and adapt them to their own bounded rationality as well as to the supposed bounded rationality and boundedly rational absorption capabilities of the others. In interactive settings, similar to the ones experimentally tested, strategic uncertainty about the behaviour of the others could be revealed to play a central role in orienting individual behaviour and to definitely inspire the modalities of theory absorption by bounded rational decision-makers.

Conclusion

The phenomenon of referring is pervasive and regards all fields of human thought and activity, so much so that it appears to be an inescapable basis of all that can be thought, conceptualized and expressed. Reflexivity can affect the dynamics of social and economic systems at different levels because it concerns the social reality per se, but it also represents an essential feature of the social sciences and social research.

The focussing on the recursivity and self-referentiality of economic theories intends to investigate the effects the knowledge and acceptance of a theory among the acting individuals in an economic system who provide the dynamics and development of that system.

For this reason the present analysis has deepened the modalities by means of which real economic actors perceive the recursive character of economic theorizing, the conditions under which economic theories affect in a self-referential way the economic actors' behaviour and the possibility of empirically testing the self-referentiality of economic theories.

In doing that it has relied on the notion of "theory absorption," whose implications have been discussed considering the boundaries that are posited to the subjective rationality of the economic actors. A theory is said to be fully absorbable whenever its own acceptance by all of the individuals belonging to a certain population does not question its predictive validity. This accounts for strategic equilibria and can be related to the logic underlying convergence of behaviour and the stability of equilibrium outcomes.

While assuming the economic actors to be perfectly rational, a theory of rational choice will be completely and universally absorbed. Allowing for the bounded rationality of the individuals requires that a certain margin be left for the individual who is adapting to the theory's prescriptive.

Among the requisites for a theory to be absorbed are, in particular, its understanding by the individuals and its compatibility with their subjective beliefs and mental representations. A theory is absorbed by an individual if that individual internalizes it in her own mental models and chooses to act according to its logical content. In interactive contexts, theory absorption will also be strongly related to the supposed mental models of the others. These have been considered to be elaborated

by means of introspection. Depending on the number of individuals – from one to all – who follow its prescriptions and are satisfied with the result, it can be distinguished among unilaterally-, partially-, and fully-absorbable theories, in particular.

Striving toward a better understanding of the mechanisms on which theory absorption relies has been interpreted as one of the possible approaches for defining bounded rational expectations, bounded rational best replies and formulating more realistic economic forecasts.

The effects of the self-referentiality of economic theories and their absorption have been experimentally analyzed by observing how individuals deal with meta-theoretical information, i.e. theoretical information about the experimental situation faced by the respondents. For this purpose, two experiments have been conducted focussing respectively on the absorbability of the guessing-game theory and the herding behaviour theory.

The experiments provided evidence that, although a significant impact of revealing theoretical information to the experimental subjects was observed, the hypothesis of full theory absorption (in the sense of perfect compliance to the theory's prescriptive) could not be acknowledged. Instead, the experimental findings corroborate the idea that individuals reflect the theoretical statements on themselves and on the situation they face and adapt them to their own bounded rationality as well as to the supposed bounded rationality and boundedly rational absorption capabilities of the others. In interactive settings, like the ones experimentally tested, strategic uncertainty about the behaviour of the others could be revealed to play a central role in orienting individual behaviour and to decisively inspire the modalities of theory absorption by bounded rational decision-makers.

As potentially any economic theory can yield recursive effects and be absorbed for the resolution of a concrete problem, the analysis of the self-referentiality of economic theories and their absorption can be approached either focussing on the reflexive effects of theories, which rely on the assumption of fully rational economic actors, or on those of theories encompassing individuals' bounded rationality. This study conceived the inquiry of the absorbability of theories of full rationality as a natural first step, and favoured it above beginning from observing the absorption of theories of bounded rationality. Testing the absorbability of theoretical principles of bounded rational behaviour could constitute an interesting development of the analysis.

Besides tackling the pure speculative intent in exploring the psychology of choice by testing if theories are accepted by bounded rational decision-makers and if their predictive content survives such test, this research program could yield several useful applications.

It could for example enhance the development of effective decision support systems, teaching methods and training procedures which enjoy a broad acceptance among real (bounded rational) decision-makers. A better understanding of the modalities by which a theory can be applied for the resolution of a practical problem could permit the development of effective debiasing procedures which would enable some mental blockades that are responsible for biased judgements and decisions to be overcome.

Furthermore, the findings on the absorbability of theoretical prescriptions could be applied to policy advising. On their basis of these findings, advice, that is fully workable (thus absorbable) by at most bounded rational actors, could be inferred. Hereby, absorbable advice should insist on the mechanisms which rule human cognition and guide decision-making and should acknowledge both the boundedness of its addressee and that of the behaviours on which it ought to be ruled.

Appendix

1 Instructions to the Experiment on the Absorbability of Guessing Game Theory

Welcome to this experiment! Thank you for your participation.

You are member in a group, consisting of 8 persons. All of your group members have these same instructions you have.

In each round, each person of your group has to choose a number between 0 and 100 (0 and 100 are possible as well). You can choose each number you like. Please note, that numbers with more than 2 decimals are excluded. The chosen numbers of your group members will remain unknown to you.

You can earn points in each round. The points you earn depend on how close your number is to a target number. The closer your number to a target number, the higher your payoff in points.

You can calculate you payoff in points this way:

target number
Earned points per round = 50–2, 5 · | your chosen number-q · (group average + d) | distance

At the beginning of each round all group members choose a number simultaneously. The target number (modified group average) is the average of the numbers your groups' choice, added by a constant d, all multiplied by q.

The values you need can be calculated as follows:

1.1 Example

x_1 = your chosen number
x_2 = the number chosen by the 2nd group member
x_3 = the number chosen by the 3rd group member

...
x_8 = the number chosen by the 8th group member

The *average* is determined by:

$$\frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8}{8}$$

The *target number* is determined by: q^* (*average* + d)
while q and d vary in every round and will be shown to you on the PC screen.
The target number might be bigger or smaller than your chosen number All what matters for the calculation of your payoff is the distance between your chosen number and the target number, that's why we take the absolute value of this difference.

Therefore the distance is always positive and determined by:
|your chosen number-target number|
Your payoff per round is: 50 points – $2,5 \cdot$ distance

If the distance time is 2.5 bigger than 50 (it means, your earnings should be negative) you will still receive 0 points. You might not earn additional points but you will never loose points.

At the end of each round you will receive information about your chosen number, your earnings and the average of all chosen numbers of your group. Since q and d vary in every round, please after each round choose your number again/please after each round think again about the number you want to choose. After you played 9 rounds, you'll receive recapitulating information about your payoffs. The transaction course is 1 point = 1 Cent. Payoffs are accumulated over all rounds and paid in cash and privately at the end of the experiment.

[*Instructions for treatment 3*]

Remember, all of your group members have the same information like you. All know as well that all participants received the same information.

[*In treatment 1 and 2 respective all and half of the participants received following "additional tips"*].

1.2 Additional Tips

1. To earn as many points as possible, you have to guess which number results from the calculation "average of the numbers your groups' choice added by d, all multiplied by q". Choose then this number.
2. Any number, chosen by all of the group members won't exceed 100. This means, the average of all chosen numbers won't exceed 100. The average added by d (here, e.g. d = 50) equals 150. Multiply then this number (group average + d) times q (here, e.g. q = 1/3) and you get 50. Therefore the number you should choose, should not exceed 50.

3. If all members of your group will realize this, everybody else will choose a number which does not exceed 50 and therefore the average will not exceed 50. The average plus d = 50 (that is 100) multiplied by q = 1/3 is 33.33. Therefore the number you should choose, should not exceed 33.33.
4. If all members of your group will realize this, everybody else will choose a number which does not exceed 33.33 and therefore the average will not exceed 33.33. The average plus d = 50 (that is 83.33) multiplied by q = 1/3 is 27.77. Therefore the number you should choose, should not exceed 27.77 and so on.
5. If you and every group member will think this way, everybody will realize that the number that will be chosen will not exceed 25, so that the best everybody can do is to choose 25. All would then choose 25 ensuring themselves the maximum payoff of 0.5*e* per round.

[*Instructions for treatment 1*]

All of your group members received as you those additional tips. They know as well that all group members received those tips.

[*Instruction for half of participants in treatment 2*]

Four people (you included) in your group received those additional tips. The other four people are just aware of the basic instructions, but not of those additional tips. All people in your group know as well, that there are four members who received some additional tips and four who did not.

[*Instruction for the remaining half of participants in treatment 2*]

All of your group members received as you those instructions. In addition, four of them received some advices (tips) for playing this game. All people in your group know as well, that there are four members who received some additional tips and four who did not.

Please answer some control questions on your PC screen before starting the experiment. This ensures the understanding of the rules of this experiment. Please remain seated during all the time. If you have questions, please raise your hand. Your question will be answered privately. Please, remain seated after the experiment and wait for further instructions. Thanks!

1. Assume, you chose 24, the other group members chose 17, 66, 100, 91, 73, 82, and respectively 13.
   a. Calculate the average.
   b. What is the target number, if d = 0 and q = 2/3?
   c. How much is the distance to your chosen number?
2. Assume, you chose 72, and the average of numbers the other group members chose is 78.
   a. Calculate the average, inclusive your chosen number.
   b. What is the target number, if d = 0 and q = 2/3?
   c. How much would you earn if d = 25 and q = 1/3?

3. Assume, in the previous round the parameters were $d = 0$ and $q = 1/3$. Now, in this round $d = 50$ and $q = 1/3$. Which of the following sentences is then correct?

   a. Group average added by d becomes higher.
   b. The target number gets smaller.
   c. Nothing can be said.

2 Instructions to the Experiment on the Absorbability of Informational Cascades' Theory

Welcome to this experiment! Thank you for your participation.

This is an experiment in the economics of decision-making.

Although you will have to use a computer, the experiment is extremely simple.

If you follow the instructions and make careful decisions, you may earn a considerable amount of money which will be paid to you cash at the end of the experiment.

Your earnings will depend partly on your decisions and partly on chance.

Think that you have to act like an entrepreneur who has to decide either to produce a new product or not.

Two situations will be equally possible: either you will sell all of the products you produce or none of them.

In each of the 20 periods you are going to play, the computer will determine the result of your sales (if you will sell everything or nothing). The sales result will be the same for all of people playing this game and will be determined period after period.

Each time you will make the right choice, you will earn 0.5€, while if you take the wrong decision, you will earn nothing. This is summarized in the following table:

	Choice: to invest	Choice: not to invest
Sales go well	0.5€	0€
Sales go bad	0€	0.5€

You will play with other subjects and you will be asked to make your choice in a certain order which will be randomly determined and will be different in each of the 20 periods.

Further, in order to make you choice, you will receive two different kinds of information:

The first one is private: think for example that you commissioned a survey on the success of your new product. The agency which did the survey ensures their results to be correct with a probability of 75%.

In the experiment, the computer will give you this information, in form of a signal which can be equal 1 or 0.

As explained in the following table, a signal equals 1 tells you that sales will go well in 75% of the cases while they will go bad in 25% of the cases.

A signal equals 0 tells you, on the contrary, that sales will go well in the 25% of the cases and that they will go bad in the 75%.

	Signal = 1	Signal = 0
Sales go well	75%	25%
Sales go bad	25%	75%

The second information is available to all the participants (it is "public") and consists in knowing the choices of the subjects who played before you.

Therefore, the number of information you will receive depends on the order with which you will be asked to play. Assume, for example, that you are the fifth subject to play: in this case you know your signal and the choices of the four preceding players. If you are the sixth player, you will know your signal and the choices of the five preceding players and so on.

After receiving this information, you will have 10 s to decide, either to invest or not (the time that remains you for typing your choice will be displayed in the upper-right corner of your pc-screen).

After all of the players will have made their choice, you will be informed about the right decision and about your profit.

You will play in this way 20 times, added by two initial trials. After having played all periods, the profit you gained in the game (which excludes the profit you earn in the two initial trials) will be paid you cash. You will be then allowed to leave the laboratory.

Please, do not talk to anyone during the experiment. We ask everyone to remain silent until the end of the experiment. If you have questions about the game, you can ask the experimenters during the two trials by raising your hand.

Your participation in the experiment and any information about your earnings will be kept strictly confidential.

Good luck!

[*In the absorption treatment all of the participants received following "additional tips"*].

2.1 Additional Tips

Think of the game carefully:

If you are the first player to move and the signal you receive is "1", you know that in 75% of the cases sales will go well, that is, the probability that sales will go well, given the signal 1, is 0.75. Therefore, if you choose to invest, your chances to win stays 75%. Since this is more than 50% (i.e. more than 0,50 probability), if the signal you receive is "1", you better choose to invest.

However, if you are the first player and the signal you receive is "0", you know that in 25% of the cases sales will go well, that is, that is, the probability that sales will go well, given the signal "0", is 0.25. Therefore, if you choose to invest, your

chances to win stay 25%. Since this is less than 50% (i.e. less than 0,50 probability), if the signal you receive is "0", you better choose **not** to invest.

If you are not the first player, the first player will think the same and therefore you can expect him/her:

- to invest, if his/her signal is "1";
- not to invest, if his/her signal is "0".

The 2nd, the 3rd or the 4th player will think the same, so that you can expect each of them:

- to invest, if his/her signal is "1";
- not to invest, if his/her signal is "0".

Assume you are the 5th player, in addition to the signal you receive, you can observe what the players before you have chosen.

In particular, you can be faced with one of the following scenarios:

- all of the 4 players before you have chosen to invest;
- 3 of them have chosen to invest, only one of them not to invest;
- 2 of them have chosen to invest, the other 2 not to invest;
- only one of them has chosen to invest, the other 3 not to invest;
- they have all chosen not to invest.

E.g. if all the 4 players before you have chosen to invest, they all probably received a signal equal "1".

Therefore, if you receive a signal equal "1", you better choose to invest, too.
BUT, what would you do, if you receive a signal equal "0"?
You can think that, if the signal was "1" in 4 of the 5 cases, and "0" just in your one case,
the **probability that given all signals sales will go well** can be calculated this way:

$$\frac{1}{5}\cdot 0.75+\frac{1}{5}\cdot 0.75+\frac{1}{5}\cdot 0.75+\frac{1}{5}\cdot 0.75+\frac{1}{5}\cdot 0.25$$

This is because, in the first fifth of cases (for the 1st out of 5 players) the probability that given the signal sales will go well is 0.75, so that it counts for $\frac{1}{5}\cdot 0.75$;

For the 2nd fifth of cases(for the 2nd out of 5 players) the probability that given the signal sales will go well is further 0.75. That adds to the previous probability and makes in total: $\frac{1}{5}\cdot 0.75+\frac{1}{5}\cdot 0.75$;

The same adds for the 3rd player (that is $\frac{1}{5}\cdot 0.75+\frac{1}{5}\cdot 0.75+\frac{1}{5}\cdot 0.75$) and the 4th player (that is $\frac{1}{5}\cdot 0.75+\frac{1}{5}\cdot 0.75+\frac{1}{5}\cdot 0.75+\frac{1}{5}\cdot 0.75$)

In the fifth case (your case) your signal, which equals "0", tells you that the probability for the sales to go well is 0,25, so that $\left(\frac{1}{5}\cdot 0.25\right)$ adds.

Overall, the **probability that given all signals sales will go well** is:

$$\frac{1}{5}\cdot 0.75+\frac{1}{5}\cdot 0.75+\frac{1}{5}\cdot 0.75+\frac{1}{5}\cdot 0.75+\frac{1}{5}\cdot 0.25=0.65>0.50$$

Since this probability is bigger than 0.50, your chances to win if you invest, despite of your own signal, stays higher than 50%.

Therefore, you can better choose to invest, even if your own signal is "0".

Imagine now that you are the 6th player and that 3 players before you have chosen to invest and 2 not to invest.

In this case, if your signal is 1, you better choose to invest, since the probability that given all signals sales will go well is bigger than 0.50,

Whilst if your signal is 0, the probability that given all signals sales will go well is exactly 0.50, since $\frac{1}{5} \cdot 0.75 + \frac{1}{5} \cdot 0.75 + \frac{1}{5} \cdot 0.75 + \frac{1}{5} \cdot 0.25 + \frac{1}{5} \cdot 0.25 + \frac{1}{5} \cdot 0.25 = 0.50$

This means, both if you choose to invest or not, you have the same chances to win. You can therefore choose at random, as tossing a coin.

From all that you can think of a **simple rule for playing this game**:

1. if an equal number of players before you have chosen to invest or not to invest (e.g. 3 invested and 3 not), than follow your signal;
2. if among your predecessors there is one more player that have chosen to invest (this is e.g. the case when 4 players invested and 3 not) and your signal is 1, then choose to invest. Respective, if one more player have chosen not to invest, and your signal is 0, then choose not to invest;
3. invest, regardless of your signal, if among your predecessors two more players before you have chosen to invest (this is e.g. the case when 5 players invested and 3 not). Respective, if two more players have chosen not to invest, do not invest, regardless of your signal.

Literature

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature Review, 4*, 165–178.

Agnoli, F., & Krantz, D. H. (1989). Suppressing natural heuristics by formal instruction: The case of the conjunction fallacy. *Cognitive Psychology, 15*(4), 439–449.

Albert, H. (Ed.). (1972). *Theorie und realität*, Tübingen: J. C. B. Mohr (Paul Siebeck).

Allais, M. (1953). Les comportement de l'homme rationell devant le risque: Critique des postulats at axiomes de l'ecole americane. *Econometrica, 21*, 503–46.

Allsopp, L., & Hey, J. D. (2000). Two experiments to test a model of herd behaviour. *Experimental Economics, 3*, 121–136.

Anderson, L. R. (2001). Payoff effects in information cascade experiments. *Economic Inquiry, 39*(4), 609–615.

Anderson, L. R., & Holt, C. A. (1997). Information cascades in the laboratory. *American Economic Review, 87*, 847–862.

Arrow, K. J. (1982). Risk perception in psychology and economics. *Economic Inquiry, 20*, 1–9.

Avery, C., & Zemsky, P. (1998). Multidimensional uncertainty and herd behavior in financial markets. *The American Economic Review, 88*(4), 724–748.

Bannerjee, A. V. (1992). A simple model of herd behaviour. *Quarterly Journal of Economics, 107*, 797–817.

Bartlett, S. J. (1987). Varieties of self-reference. In S. J. Bartlett, & P. Suber (Eds.) *Self-reference. Reflections on reflexivity* (pp. 5–28). Dordrecht: Martinus Nijhoff Publishers.

Bartlett, S. J. (1992). *Reflexivity. A source-book in self-reference*. North-Holland: Elsevier Science Publishers B.V.

Bartlett, S. J., & Suber, P. (Eds.). (1987). *Self-reference. Reflections on reflexivity*. Dordrecht: Martinus Nijhoff Publishers.

Barwise, J., & Etchemendy, J. (1987). *The Liar – An essay on truth and circularity*. New York: Oxford University Press.

Bateson, G. (1972). The logical categories of learning and communication. In G. Bateson (1972). *Steps to an ecology of mind*. New York: Ballantine.

Bateson, G. (1991). A social scientist views the emotions. In R. Donaldson (Ed.) *Gregory Bateson: Further steps to an ecology of mind*. New York: Harper & Row.

Beard, R. T., & Beil, R. (1994). Do People Rely on the Self-interested Maximization of Others?: An Experimental Test. *Management Science*, Vol. 40, No. 2 (Feb., 1994), 252–262.

Becker, G. S. (1991). A note on restaurant pricing and other examples of social influences on price. *Journal of Political Economy, 99*(5), 1109–1116.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity and social history. *Games and Economic Behavior, 10*, 122–142.

Berger, P. L., & Luckmann, T. (1966). *The social construction of reality: A treatise in the sociology of knowledge*. Garden City, NY: Anchor Books.

Berninghaus, S., Güth, W., & Kliemt, H. (2003a). From teleology to evolution: Bridging the gap between rationality and adaptation in social explanation. *Journal of Evolutionary Economics, 13*(4), 385–410.

Berninghaus, S., Güth, W., & Kliemt, H. (2003b). Reflections on equilibrium: Ideal rationality and analytic decomposition of games. *Homo Oeconomicus, 20*, 257–302.

Bernoulli, D. (1738). Exposition of a new theory on the measurement of risk. *Econometrica, 22*, 23–36.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy, 100*, 992–1026.

Bikhchandani, S., & Sharma, S. (2001). Herd behavior in financial markets. *IMF Staff Papers, 47*(3), 279–310.

Binmore, K. (1987). Modeling Rational Players, Part I. Economics and Philosophy, October 1987, 3, pp. 179–214.

Bolander, T. (2002). Self-reference and logic. *Phi News, 1*, 9–44.

Bolander, T. (2003). *Logical theories for agent introspection*. PhD thesis, Informatics and Mathematical Modelling (IMM), Technical University of Denmark.

Bolle, F. (1979). *Das problem des optimalen Stoppens: modelle und experimente*. Frankfurt: Peter Lang.

Bolton, G. E. (1998). Bargaining and dilemma games: From laboratory data towards theoretical synthesis. *Experimental Economics, 1*, 257–281.

Bombach, G. (1962). Über die möglichkeit wirtschaftlicher voraussagen. *Kyklos, XV*, 29–67.

Bosch-Domenech, A., Montalvo, J. G., Nagel, R., & Satorra, A. (2002). One, two, (three), infinity…: Newspaper and lab beauty-contests experiments. *The American Economic Review, 92*(5), 1687–1701.

Bosse, L. (1957). Über die möglichkeiten und den Nutzen von kurzfristigen Wirtschaftsprognosen. *Weltwirtschaftliches Archiv, 79*, 65–83.

Buck, R. C. (1963). Reflexive predictions. *Philosophy of Science, 30*(4), 359–369.

Bush, R., & Mosteller, F. (1955). *Stochastic models of learning*. New York: Wiley.

Calandro, J., Jr. (2004). Reflexivity, business cycles, and the new economy. *The Quarterly Journal of Austrian Economics, 7*(3), 45–69.

Camerer, C. (1995). Individual decision making. In J. H. Kagel, & Roth, A. (Eds.), *The handbook of experimental economics* (pp. 587–616). Princeton: Princeton University Press.

Camerer, C. F. (2003a). *Behavioral game theory*. New York: Russel Sage Foundation; Princeton, NJ: Princeton University Press.

Camerer, C. F. (2003b). Behavioural studies of strategic thinking in games. *TRENDS in Cognitive Sciences, 7*(5), 225–231.

Camerer, C. F., Ho, T.-H., & Chong, J.-K. (2001). Sophisticated experience-weighted attraction learning and strategic thinking in repeated games. *Journal of Economic Theory, 104*, 137–188.

Campbell, D. T. (1969). Reforms as experiments. *American Psychologist, 24*, 409–429.

Carey, S. (1985). Conceptual differences between children and adults. *Mind & Language, 3*(3), 167–181.

Carnap, R. (1942). *Introduction to semantics*, Cambridge, MA: Harvard University Press.

Carnap, R. (1956). *Meaning and necessity*. Chicago: University of Chicago Press.

Cass, D., & Shell, K. (1983). Do sunspots matter? *Journal of Political Economy, 91*(2), 193–227.

Cassel, S. (2000). *Politikberatung und Politikerberatung – Eine institutionenökonomische Analyse der wissenschaftlichen Beratung der Wirtschaftspolitik*. Bern: Verlag Paul Haupt.

Çelen, B., & Kariv, S. (2004). Distinguish information cascades from herd behavior in laboratory. *American Economic Review, 94*(3), 484–498.

Chase, V. M., Hertwig, R., & Gigerenzer, G. (1998). Visions of rationality. *Trends in Cognitive Sciences, 2*(6), 206–214.

Chew, S. H., & McCrimmon, K. R. (1979). *Alpha-nu choice theory: An axiomatization of expected utility*. Working paper, 669. University of British Columbia Faculty of Commerce.

Church, A. (1951). The need for abstract entities in semantic analysis. *Proceedings of the American Academy of Arts and Sciences, 80*, 100–112.

Cipriani, M., & Guarino, A. (2005). Herd behavior in a laboratory financial market. *American Economic Review, American Economic Association, 95*(5), 427–1443.

Coase, R. (1974). The lighthouse in economics. *Journal of Law and Economics, 17*(2), 357–376.

Cole, W. G. (1989). *Understanding Bayesian reasoning via graphical displays*. Paper presented in CHI'89 Proceedings, Austin, TX: ACM Press.

Copeland, T., Koller, T., & Murin, J. (2000). *Valuation*, New York: Wiley.

Cournot, A. (1838). *Recherches sur les principes mathématiques de la théorie des richesses*, Paris.

Cyert, R. M., & March, J. G. (1963). *A behavioral theory of the firm*. Englewood Cliffs, N.J.: Prentice-Hall.

Cyert, R. M., Simon, H. A., & Trow, D. B. (1956). Observation of a business decision. *Journal of Business, 29*, 237–248.

Dacey, R. (1976). Theory absorption and the testability of economic theory. *Zeitschrift für Nationalökonomie, 36*, 247–267.

Dacey, R. (1981). Some implications of 'theory absorption' for economic theory and the economics of information. In J. C. Pitt (Ed.), *Philosophy in economics*, Dordrecht: D. Reidel Publishing Company.

Davidson, D. (2001). *Inquiries into truth and interpretation* (2nd ed.) New York: Oxford University Press.

Davidson, D., & Marschak, J. (1959). Experimental test of a stochastic decision theory. In Measurement: definitions and theories. Churchman, C. W., & Ratoosh, P. (eds). New York: Wiley. 233–69

Davidson, D., Suppes, P., & Siegel, S. (1957). Decision making: an experimental approach. Stanford, Calif.: Stanford University Press.

Davis, J. B., & Klaes, M. (2003). Reflexivity: Curse or cure? *Journal of Economic Methodology, 10*(3), 329–352.

Dawes, R. (1988). *Rational choice in an uncertain world*. San Diego: Harcourt, Brace, Jovanovich.

Demandt, A. (2001). *Ungeschehene Geschichte. Ein Traktat über die Frage: Was wäre geschehen, wenn…?*, 3. erw. Aufl., Vandenhoeck und Ruprecht, Kleine Reihe V & R 4022, Göttingen.

Dennett, D. (1987). *The intentional stance* (Reprint ed.). Cambridge, MA: MIT Press.

Douglas, G. (1996). What have we learn from social learning? *European Economic Review, 40*, 617–628

Duffy, J., & Nagel, R. (1997). On the robustness of behaviour in experimental 'beauty contest' games. *The Economic Journal, 107*(445), 1684–1700.

Duncker, K. (1935). On problem solving (L. S. Lees, Trans.). *Psychological Monographs, 58* (original work published 1935).

Edwards, W. (1953a). Experiments on economic decision-making in gambling situations. Abstract. *Econometrica*, 21:349–50.

Edwards, W. (1953b). Probability preferences in gambling. *American Journal of Psychology*, 66:349–64.

Egidi, M. (1992). Organizational learning, problem solving and the division of labour. In H. Simon, M. Egidi, R. Marris, R. Viale (1992). *Economics, bounded rationality and the cognitive revolution* (pp. 148–173), UK: Edward Elgar Publishing Company.

Egidi, M. (2003). Decomposition patterns in problem solving. *Social Science Research Network Electronic Library*. Accepted Paper Series.

Egidi, M. (2005). From bounded rationality to behavioral economics. *EconWPA*, Experimental 0507002.

Egidi, M., & Narduzzo, A. (1997). The emergence of path-dependent behaviors in cooperative contexts. *International Journal of Industrial Organization, 5*, 677–709.

Einhorn, H. J. (1982). Learning for experimence and suboptimal rules in decision-making. In D. Kahnemann, P. Slovic, A. Tversky (Eds.), *Judgement under uncertainty* (268–283). Cambridge: Cambridge University Press.

Einstein, A. (1905, March). On a heuristic point of view concerning the generation and transformation of light (Concerning the generation and transformation of light from a heuristic point of view). *Annalen der Physik, 17*, 132–148.

Elveton, R. O. (Ed.) (2000). *The phenomenology of husserl: Selected critical readings*. Seattle: Noesis Press.

Fellner, G., Güth, W., & Martin, E. (2006a). *Satisficing or optimizing? – An experimental study*. Papers on Strategic Interaction, No. 11–2006, Max Planck Institut für Ökonomik, Jena.

Fellner, G., Güth, W., & Martin, E. (2006b). *Task transcending satisficing – An experimental study*. Papers on Strategic Interaction, No. 09–2006, Max Planck Institut für Ökonomik, Jena.

Ferguson, N. (Ed.). (1999). *Virtuelle Geschichte. Historische Alternative im 20. Jahrhundert* (German Trans.). Darmstadt: Primus.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanfort, CA: Stanford University Press.

Fiore, A., & Morone, A. (2007). *A simple note on informational cascades*. Economics Discussion Papers, 2007–21, Kiel Institute for the World Economy.

Fiore, A., Morone, A., & Sandri, S. (2007). *On the absorbability of herd behaviour and informational cascades: An experimental analysis*. Dresden Discussion Paper in Economics, No. 13/07.

Fischbacher, U. (1998). *Z-tree: A toolbox for readymade economic experiments*. University of Zurich.

Fischer, H. R. (Ed.). (1995). *Die Wirklichkeit des Konstruktivisms – Zur Auseinandersetzung um ein neues Paradigma* (pp. 47–71). Heidelberg: Carl-Auer-Systeme-Verlag.

Fischoff, B. (1982a). For those condemned to study the past: Heuristics and biases in hindsight. In D. Kahnemann, P. Slovic, A. Tversky (Eds.), *Judgement under uncertainty* (pp. 335–352). Cambridge: Cambridge University Press.

Fischoff, B. (1982b). Debiasing. In D. Kahnemann, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty* (pp. 422–442). Cambridge: Cambridge University Press.

Franz, W. (2000). Wirtschaftspolitische Beratung. *Perspektiven der Wirtschaftspolitik*, Band 1, Heft 1, 53–71.

Frederick, S. (2003). Time preference and personal identity. In G. Loewenstein, D. Read, & R. Baumeister (Eds.). (2003). *Time and decision: Psychological perspectives in intertemporal choice*. New York. Russel Sage.

Frey, B. S. (2000). Was bewirkt die Volkswirtschaftslehre? *Perspektiven der Wirtschaftspolitik*, Band 1, Heft 1, 5–33.

Frey, D. (1986). Recent research on selective exposure to information. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 41–80). NY: Academic Press.

Friedman, D., & Sunder, S. (1994). *Experimental methods, a primer for economists*. Cambridge: Cambridge University Press.

Friedman, M. (1953). *Essays in positive economics*. Chicago: Chicago University Press.

Friedman, M., & Savage, L. J. (1952). The expected utility hypothesis and the measurability of utility. *Journal of Political Economy, 60*, 6.

Fulda, E., Lehmann-Waffenschmidt, M., & Schwerin, J. (1998). Zwische Zufall und Notwendigkeit – Zur Kontingenz ökonomischer Prozesse aus theoretischer und historischer Sicht. In G. Wegner, & J. Wieland (Ed.), *Formelle und informelle Institutionen. Genese, Interaktion und Wandel*, Marburg: Metropolis.

Geanakoplos, J. (1989). *Game theory, partitions, and applications to speculation and consensus*. Cowles Foundation Discussion Paper, 914, Yale University.

Gebotys, R. J., & Claxton-Oldfield, S. P. (1989). Errors in the quantification of uncertainty: a product of heuristics or minimal probability knowledge base? *Applied Cognitive Psychology, 3*, 237–250.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*, 650–669.

Gigerenzer, G., & Goldstein, D. G. (2002).Models of ecological rationality: The recognition heuristic. *Psychological Review, 109*(1), 75–90.

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 513–525.

Gigerenzer, G., & Selten, R. (Eds.). (2001). *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT Press.

Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In G. Gigerenzer, P. Todd, & the ABC Research Group (Eds.). (1999). *Simple heuristics that make us smart* (pp. 3–34). Oxford: Oxford University Press.

Gigerenzer, G., Todd, P., & the ABC Research Group (Eds.). (1999). *Simple heuristics that make us smart*. Oxford: Oxford University Press.

Gilbert, D. T. (2002). Inferential Correction. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases* (pp. 167–184). New York: Cambridge University Press.

Greiner, B. (2004). The online recruitment system orsee 2.0 – A guide for the organization of experiments in economics. Working Paper Series in Economics 10, University of Cologne, Department of Economics.

Grether, D. M., & Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *American Economic Review, 69*, 623–638.

Groener, R. (Ed.). (1983). *Methods of heuristics*. Hillsdale, NJ: Erlsbaum.

Grünbaum, A. (1956). Historical determinism, social activism and predictions in the social sciences. *British Journal for the Philosophy of Science, 7*(27), 236–240.

Grunberg, E., & Modigliani, F. (1954). The predictability of social events. *Journal of olitical Economy, 62*, 465–478.

Gul, F. (1991). A theory of disappointment in decision-making under uncertainty. *Econometrica, 59*(3), 667–686.

Güth, W. (2000). Boundedly rational decision emergence – A general perspective and some selective illustrations. *Journal of Economic Psychology, 21*, 433–458.

Güth, W. (2006). *Satisficing in portfolio selection – Theoretical aspects and experimental tests*. Discussion Papers on Strategic Interaction, Max Planck Institute for Economics, Jena.

Güth, W., & Kliemt, H. (2000). *From full to bounded rationality – The limits of unlimited rationality*. ZiF – Mitteilungen Special 2000, Zentrum für interdisziplinäre Forschung der Universität Bielefeld, 1–15.

Güth, W., & Kliemt, H. (2004a). Bounded rationality and theory absorption. *Homo Oeconomicus, 21*(3/4), 251–540.

Güth, W., & Kliemt, H. (2004b). Perfect or bounded rationality?: Some facts, speculations and proposals. *Analyse und Kritik: Zeitschrift für Sozialtheorie, 26*(2), 364–381.

Güth, W., Kocher, M., & Sutter, M. (2002). Experimental'beauty contests' with homogeneous and heterogeneous players and with interior and boundary equilibria. *Economic Letters, 74*, 219–228.

Güth, W., Levati, M. V., & Ploner, M. (2006). *Is satisficing absorbable? – An experimental study*. Papers on Strategic Interaction, No. 09–2006, Max Planck Institut für Ökonomik, Jena, 2006, forthcoming in Journal of Behavioral Finance.

Güth, W., Rolf Schmittberger & Bernd Schwarze (1982). 'An Experimental Analysis of Ultimatum Bargaining', *Journal of Economic Behavior and Organization*, *3*, 367–388

Haberlandt, K. (1998). *Human memory: Exploration and application*, Boston: Allyn & Bacon.

Handlbauer, G. (1997). Self-reference in individual and social decision-making. *Journal of Institutional and Theoretical Economics (JITE), 153*, 762.

Hands, D. (2001). *Reflections without rules: Economic methodology and contemporary science theory*. Cambridge: Cambridge University Press.

Hayek, F. (1970). Can we still avoid inflation? In R. Ebeling (Ed.), *The austrian theory of the trade cycle and other essays*. Auburn: Ludwig von Mises Institute.

Hazlitt, H. (1996). *Economics in one lesson*. San Francisco: Laissez Faire Books

Hedgeworth (1981). *Mathematical Psychics* (reprint of 1961), London.

Heisenberg, W. (1927). Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik, 43*, 172–198.

Heisenberg, W. (2000). *Physik und philosophie* (6th ed.), (1st ed., 1959), Stuttgart: Hirzel Verlag.

Hempel, C. G. (1960). Inductive inconsistencies. *Synthese, 12*, 439–469.

Hempel, C. G. (1962). Deductive-nomological vs. statistical explanation. In H. Feigl, & G. Maxwell (Eds.), *Minnesota studies in the philosophy of science* (Vol. III, pp. 98–169). Minneapolis: University of Minnesota Press.

Hey, J. D. (1991). *Experiments in economics*. Oxford: Blackwell.

Hey, J. D. (1994). *Experimental economics*. Heidelberg: Physica Verlag.

Higgins, E. T. (1996). Knowledge activation: accessibility, applicability, and salience. In E. T. Higgins, A. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133–168). New York: Guilford Press.

Hintikka, J., & Pietarinen, J. (1966). Semantic information and inductive logic. In J. Hintikka, & P. Suppes (Eds.), *Aspects of inductive logic* (pp. 96–112), Amsterdam: North Holland.

Hintikka, J., & Suppes, P. (Eds.). (1966). *Aspects of inductive logic*. Amsterdam: North Holland.

Ho, T. H., Camerer, C., & Weigelt, K. (1998). Iterated dominance and iterated best response in experimental 'p-beauty contests'. *The American Economic Review, 88*(4), 947–969.

Holmes, O. W. (1980). *Ralph Waldo Emerson*. New York: Chelsea House Pub.

Hoyenga, K., & Hoyenga, K. (1988). *Psychobiology: The neuron and behaviour*. Pacific Grove, CA: Brooks/Cole Publishing.

Huck, S., & Oechssler, J. (2000). Informational cascades in the laboratory: Do they occur for the right reasons? *Journal of Economic Psychology, 21*(6), 661–671.

Hung, A., & Dominitz, J. (2004). Homogeneous actions and hetergeneous beliefs: Experimental evidence on the formation of information cascades. *Econometric Society 2004 North American Winter Meetings*.

Kagel, J. H. & Roth, A. (Eds.). (1995). *The handbook of experimental economics*. Princeton: Princeton University Press.

Kahneman, D. (2002, December). *Maps of bounded rationality: A perspective on intuitive judgement and choice*. Nobel Prize Lecture.

Kahnemann, D. (2003, December). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 1449–1475.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics & biases: The psychology of intuitive judgment* (pp. 49–81). New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185*, 1124–1130.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(19, 33), 263–291.

Kahneman, D., & Tversky, A. (1983). Extension versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293–315.

Kahneman, D., & Tversky, A. (Eds.). (2000). *Choices, values and frames*. New York: Cambridge University Press and the Russell Sage Foundation.

Kahnemann, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty*. Cambridge: Cambridge University Press.

Keynes, J. M. (1936). *The general theory of employment, investment and money*. London: Macmillan.

Kincheloe, J. (1991). *Teachers as researchers: Qualitative inquiry as a path of empowerment*. London: Falmer Press.

Kliemt, H. (2001). Rationality and reality. *Kyklos, 54*, 309–316.

Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy, 72*, 690–712.

Krippendorff, K. (1986). Web dictionary of cybernetics and systems. *A Dictionary of Cybernetics*. Retrieved March 21, 2006, from http://pespmc1.vub.ac.be/Asc/Kripp.html

Lacey, A. R. (1996). *A dictionary of philosophy* (3rd ed.). London, UK: Routledge.

Lehmann-Waffenschmidt, M. (1990). Predictability of economic processes and the morgenstern paradox. *Schweizerische Zeitschrift für Volkswirtschaft und Statistik, 2*, 147–160.

Lehmann-Waffenschmidt, M. (1996). Limitations of social forecasting. *Wissenschaftliche Zeitschrift der Technischen Universität Dresden, 45*(4), 43–47.

Lehmann-Waffenschmidt, M. (2002). Neuer Fokus Viabilität. Zur Bedeutung des (radikalen) Konstruktivismus für die Wirtschaftswissenschaften und die Nachhaltigkeitsdebatte. *Ökologisch Wirtschaften, 6*, 23–26.

Lehmann-Waffenschmidt, M. (2006a). Konstruktivismus und Evolutorische Ökonomik. In G. Rusch, (Ed.), *Konstruktivistische Ökonomik* (pp. 27–54). Marburg: Metropolis-Verlag.

Lehmann-Waffenschmidt, M. (2006b). *Self-referential optimal delays when reactions are delayed*. Dresden Discussion Paper Series in Economics. Technical University of Dresden.

Lehmann-Waffenschmidt, M., & Reina, L. (2003). *Coalition formation in multilateral negotiations with a potential for logrolling: An experimental analysis of negotiators' cognition processes*. Dresden Discussion Papers Series in Economics, No. 17, Dresden University of Technology.

Levi, I. (1967a). *Gambling with truth*, New York: Alfred Knopf.

Levi, I. (1967b). Information and inference. *Synthese, 17*, 369–391.

Lichtenstein, S., & Slovic, P. (1971). Reversal of preferences between bids and choices in gambling decisions. *Journal of Experimental Psychology, 89*, 46–55.

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes, 65*, 272–292.

Loomes, G., & Sudgen, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal, 92*, 805–825.

Lorenz, K. (1973). *Die Rückseite des Spiegels*. München: R. Piper & Co. Verlag.

Luger, F. G. (Ed.). (1995). *Computation and intelligence: Collected readings*. Cambridge, MA: AAAI Press/MIT Press.

Luhmann, N. (1984). *Soziale Systeme. Grundriss einer allgemeinen Theorie*. Frankfurt: Surhkamp.

Mackie, J. L. (1973). *Truth, probability and paradox*. Oxford: Oxford University Press.

Mackinnon, L. A. K. (2003). *The Social Construction of Economic Man: The Genesis, Spread, Impact and Institutionalisation of Economic Ideas.* Doctoral Thesis submitted at the University of Queensland in March 2006 (available at http://search.arrow.edu.au/09.09.2008).

March, J. G. (1994). *A primer on decision making*. New York: Free Press.

Margert, A. W. (1929). Morgenstern on the methodology of economic forecasting. *Journal of Political Economy, 37*(3), 312–339.

Mark, T. K., & Eysenck, M. W. (2005). *Cognotove psychology: A student's handbook*. UK: Psychology Press.

Martin, R. L. (1967). Toward a solution to the liar paradox. *The Philosophical Review 76*, 279–311.

Martin, R. L. (1992). On non-translational semantics. In S. J. Bartlett (Ed.), *Reflexivity. A sourcebook in self-reference*. North-Holland: Elsevier Science Publishers B. V.

Marx, R. (March, 2006). *Die Prognostizierbarkeit des wirtschaftlichen Verhaltens – eine theoretische Analyse aus radikalkonstruktivistischer Perspektive und eine experimentelle Fallstudie zu Selbstreferentialität von Prognosen und Theorien*. Dissertation submitted for a diploma at the Technische Universität Dresden under the supervision of Prof. Dr. M. Lehmann-Waffenschmidt.

Maturana, H. R., & Varela, F. J. (1973). Autopoiesis: The organization of the living. In H. R. Maturana, & F. G. Varela (Eds.), *Autopoiesis and cognition*. Dordrecht, Netherlands: Reidel.

Maturana, H. R., & Varela, F. G. (1980). *Autopoiesis and cognition*. Dordrecht, Netherlands: Reidel.

Maturana, H. R., & Varela, F. J. (1987). *The tree of knowledge: The biological roots of human understanding*. Boston: Shambhala.

May, K. (1954). Intransitivity, utility, and the aggregation of preference patterns. *Econometrica*, 22:1–13.

McKelvey, R., & Palfrey, T. (1992). An Experimental Study of the Centipede Game. *Econometrica*, July 1992, *60*(4), 803–36.

McClelland, J. L., & Rumelhart, D. E. (Eds.). (1986). *Parallel distributed processing*. Cambridge, MA: MIT Press.

McNeil, B. J., Pauker, S. G., Sox, H. C., & Tversky, A. (1982). On the elicitation of preferences for alternative therapies. *New England Journal of Medicine, 306*, 1259–1262.

Meier, A., & Slembeck, T. (1994). *Wirtschaftspolitik. Ein kognitiv-evolutionärer Ansatz*. München Wien: Oldenbourg Verlag.
Mendelson, E. (1997). *Introduction to mathematical logic* (4th ed.). London: Chapman & Hall.
Merriam Webster's Unabridged Dictionary (2000).
Merton, R. K. (1936). The unanticipated consequences of purposive social action. *American Sociological Review, 1*(6), 894–904.
Mistri, M. (1997). Changing preferences and cognitive processes. In R. Viale, (Ed.), *Cognitive economics*, LaSCoMES Series, I/1997.
Morgenstern, O. (1928). *Wirtschaftsprognose, eine Untersuchung ihrer Voraussetzungen und Möglichkeiten*. Vienna: Julius Springer Verlag.
Morgenstern, O. (1935). Vollkommene Voraussicht und wirtschaftliches Gleichgewicht. *Zeitschrift für Nationalökonomie, 6*, 337–357.
Morgenstern, O. (1972). Descriptive, predictive and normative theory, *Kyklos, XXV*, 699–714.
Morgenstern, O., & Schwödiauer, G. (1976). Competition and collusion in bilateral markets, *Zeitschrift für Nationalökonomie, 36*, 217–245.
Morone, A., & Morone, P. (2007). *Guessing game and people behaviours: What can we learn?* Mimeo.
Morone, A., Sandri, S., & Uske, T. (2008). On the absorbability of the guessing game theory – A theoretical and experimental analysis. In A. Innocenti, & P. Sbriglia (Eds.), *Games, rationality and behaviour. Essays on behavioural game theory and experiments* (pp. 161–181). Houndmills: Palgrave McMillan.
Mosteller, F., & Nogee, P. (1951). An experimental measurement of utility. *Journal of Political Economy*, 59:371–404.
Moutier, S., & Houdé, O. (2003). Judgement under uncertainty and conjunction fallacy inibition training. *Thinking and Reasoning, 9*(3), 185–201.
Mueller, D. C. (Ed.). (1997). *Perspectives on public choice: A handbook*. Cambridge: Cambridge University Press.
Mumma, G. H., & Wilson, S. B. (1995). Procedural debiasing of primacy/anchoring effects in clinical-like judgements. *Journal of Clinical Psychology, 51*(6), 841–853.
Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review, 85*(5), 1313–1326.
Nagel, R. (1999). A survey on experimental guessing games: A study of bounded rationality and learning. In D. V. Budescu,, I. Erev, & R. Zwick, *Games and human behavior* (pp. 105–42). London: Lawrence Erlbaum.
Nash, J. (1951). Non-cooperative games. *Annals of Mathematics, 54*, 286–295.
Neisser, U. (1976). *Cognition and reality. Principles and implications of cognitive psychology*, San Francisco: Freemen.
Nelson, R. R., & Winter, S. (1982). *An evolutionary theory of economic change*. Cambridge, MA: The Belknap Press of the Harvard University Press.
Newell, A., & Simon, H. A. (1972). *Human problem solving*, Englewood Cliff: Prentice Hall.
Niiluoto, I., & Tuomela, R. (1973). *Theoretical concepts and hypothetico-inductive inference*. Dordrecht: Reidel.
Peirce, C. S. (1931–1958). In Hartshorne, C., Weiss, P., & Burks, A. (Eds.), *Collected papers I–VIII.* Cambridge, MA: Harvard University Press.
Pelikan, P., & Wagner, G. (Eds.). (2003). *The evolutionary analysis of economic policy*. Cheltenham, UK; Northampton, MA: Edward Elgar.
Pies, I. (2000). *Ordnungspolitik in der Demokratie: ein ökonomischer Ansatz diskursiver Politikberatung*. Tübingen: Mohr Siebeck.
Pohl, R., & Hell, W. (1996). No reduction in hindsight bias after complete information and repeated testing, *Organizational Behavior on Human Decision Processes, 67*(1), 49–58.
Polya, G. (1954). *Mathematics and plausible reasoning. Vol. 1: Induction and analogy in mathematics* [Gigerenzer/Todd (1999)]. Princeton, NJ: Princeton University Press.

Pombeni, F. (February, 2005). *Debiasing-Problematik, Methoden und eine experimentelle Fallstudie*. Dissertation submitted for a diploma at the Technische Universität Dresden under the supervision of Prof. Dr. M. Lehmann-Waffenschmidt.

Popper, K. (1957). *The poverty of historicism*. New York: Harper and Row.

Popper, K. (1967). *Unended quest: An intellectual autobiography*. London: Fontana.

Preston, M. G., & Baratta, P. (1948). An experimental study of the auction value of an uncertain outcome. *American Journal of Psychology*, 61:183–93.

Priest, G. (1987). Unstable solutions to the liar paradox. In S. J. Bartlett,, P. Suber (Eds.), *Self-reference. Reflections on reflexivity*. Dordrecht: Martinus Nijhoff Publishers.

Pruitt, D. G. (1981). *Negotiation behavior*. New York: Academic Press.

Quine, W. V. (1962). Paradox. In S. J. Bartlett (Ed.), *Reflexivity. A source-book in self-reference* [Reprinted form *Scientific American, 20*(4), 84–96.]. North-Holland: Elsevier Science Publishers B. V.

Rabin, M. (1998). Psychology and economics. *Journal of Economic Literature, 36*(1), 11–46.

Raiffa, H. (1982). *The art and science of negotiation*. Cambridge: Harvard University Press.

Rakow, T., Harvey, N., & Finer, S. (2003). Improving calibration without training: The Role of task information. *Applied Cognitive Psychology, 17*, 419–441.

Reichenbach, H. (1947). *In elements of symbolic logic*. New York: The Free Press.

Rosenthal, R. (1998). Covert communication in classrooms, clinics, and courtrooms. *Eye on Psi Chi, 3*(1), 18–22.

Rosenthal, R., & Jacobson, L. (1968/1992). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development*. New York: Irvington Publishers.

Roshwald, M. (1955). Value-judgements in the social sciences. *British Journal for the Philosophy of Science, 6*(23), 186–208.

Roth, A. E. (1995). Bargaining experiments. In J. H. Kagel, & A. Roth, (Eds.). *The handbook of experimental economics*. Princeton: Princeton University Press.

Roth, A., & Erev, I. (1995). Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior, 8*(Special Issue: Nobel Symposium), 164–212.

Roth, G. (1985a). Die Konstruktivität des Gehirns: Der Kenntnisstand der Hirnforschung. In H. R. Fischer (Ed.), *Die Wirklichkeit des KOnstruktivisms – Zur Auseinandersetzung um ein neues Paradigma* (pp. 47–71). Heidelberg: Carl-Auer-Systeme-Verlag.

Roth, G. (1985b). Die Selbstreferentialität des Gehirns und die Prinzipien der Gestaltwahrnehmung, *Gestalt Theory, 7*, 228–244.

Rothstein, B. (2005). *Social traps and the problem of trust*. Cambridge: Cambridge University Press.

Roy, M. C., & Lerch, F. J. (1996). Overcoming ineffective mental representations in base-rate problems. *Information System Research, 7*(2), 16–25.

Rubinstein, A., Tversky, A., & Heller, D. (1997). Naive strategies in competitive games. In Albers, W., Güth, W., Hammerstein, P., Moldovanu, B., Van Damme, E., Selten, R. (Eds.), *Understanding strategic interaction: Essays in honor of Reinhard Selten* (pp. 394–402). Heidelberg: Springer.

Rusch, G. (Ed.). (1999). *Wissen und Wirklichkeit – Beiträge zum Konstruktivismus – Eine Hommage an Hernst von Glasersfeld*, Heidelberg: Carl-Auer-Systeme Verlag.

Russell, B. (1903). *Principles of mathematics*. Cambridge: Cambridge University Press.

Russell, B. (1940). *An inquiry into meaning and truth*. London: Allen & Unwin.

Russo, J. E., & Schoemacher, P. J. H. (1992, Winter). Managing overconfidence, *Sloan Management Review*, 7–17.

Salewski, M. (1999). *Was wäre wenn. Alternativ und Parallelgeschichte: Brücken zwischen Phantasie und Wirklichkeit*. Steiner: Stuttgart.

Sanna, L. J., & Schwarz, N. (2004). Integrating temporal biases: The interplay of focal thoughts and accessibility experiences. *Psychological Science, 15*, 474–481.

Sanna, L. J., Stocker, S. L., & Schwarz, N. (2002). When debiasing backfires: Accessible content and accessibility experiences in debiasing hindsight. *Journal of Experimental Psychology, Learning, Memory and Cognition, 28*(3), 497–502.

Sapir, E. (1929). The status of linguistics as a science. In E. Sapir, & D. G. Mandelbaum (Eds.), *Selected writings of edward sapir in language, culture, and personality*. Berkeley: University of California Press.

Sapir, E., & Mandelbaum, D. G. (Eds.). (1986). *Selected writings of edward sapir in language, culture, and personality*. Berkeley: University of California Press.

Sbriglia, P. (2008). Revealing the depth of reasoning in p-beauty contest games, *Experimental Economics, 11*(2), 107–202.

Scharfstein, D. S., & Stein, J. C. (1990). Herd behaviour and investment. *American Economic Review, 80*, 465–479.

Scheutz, M. (1995). *Ist das der Titel eines Buchs? Selbstreferenz neu analysiert*. Dissertationen der Universität Wien, Band 13, WUV-Universitätsverlag.

Schmidt, S. J. (1987). *Der Diskurs des Radikalen Konstruktivismus*. Frankfurt am Main: Suhrkamp Verlag.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: 1. Detection, search, and attention. *Psychological Review, 84*, 1–66.

Schrödinger, E. (1935, November). Die gegenwärtige Situation in der Quantenmechanik. *Naturwissenschaften*.

Schumpeter, J. A. (1954). *History of Economic Analysis* [Revised arch 1996], USA: Oxford University Press.

Schwegler, H. (1999). Wissenschaft als Sprachspiel. In G. Rusch (Ed.), *Wissen und Wirklichkeit – Beiträge zum Konstruktivismus – Eine Hommage an Hernst von Glasersfeld* (pp. 17–34), Heidelberg: Carl-Auer-Systeme Verlag.

Scott, R. W. (1995). *Institutions and organizations*. Thousand Oaks, CA: Sage.

Searle, J. (1995). *The construction of social reality*. NewYork: Free Press.

Sebenius, J. K. (1992). Negotiation analysis: A characterization and review. *Management Science, 38*(1), 18–38.

Selten, R. (1991). Anticipatory Learning in Two-Person Games, in Selten, R. (Ed.), *Game equilibrium models* I. Berlin: Springer Verlag, 1991, pp. 98–154.

Selten, R. (1999). *What is Bounded Rationality?* Paper prepared for the Dahlem Conference, *Mimeo*.

Shiller, R. (2001, May). *Bubbles human judgment and expert opinion*. Cowles Foundation Discussion Paper No. 1303.

Simon, H. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69*, 99–118.

Simon, H. (1956a). Rational choice and the structure of the environment. In H. Simon, M. Egidi, R. Marris, & R. Viale (Eds.), *Economics, bounded rationality and the cognitive revolution*, Northampton, MA: Edward Elgar Publishing Company.

Simon, H. (1956b). Rational choice and the structure of the environments. *Psychological Review, 63*, 129–138.

Simon, H. (1957). *Models of man*. New York: Wiley.

Simon, H. (1977). *Models of discovery and other topics in the methods of science*. Boston: D. D. Riedle Publishing Company.

Simon, H. (1981). *The sciences of artificial*. Cambridge, MA: MIT Press.

Simon, H. (1982). *Models of bounded rationality*. Cambridge, MA: MIT Press.

Simon, H. (1988). Scientific discovery as problem solving. In H. E. Simon, M. Egidi, R. Marris, & R. Viale (Eds.), *Economics, bounded rationality and the cognitive revolution*. Northampton, MA: Edward Elgar Publishing Company.

Simon, H. (1990). Invariants of human behaviour. *Annual Review of Psychology, 41*, 1–19.

Simon, H. (1992). Introductory comments. In H. Simon, M. Egidi, R. Marris, & R. Viale (Eds.), *Economics, bounded rationality and the cognitive revolution* (pp. 3–7). Northampton, MA: Edward Elgar Publishing Company.

Simon, H., Hayes, J. R. (1976). The understanding process: problem isomorphs. *Cognotive Psychology, 8*, 165–190.

Sipser, M. (2006). *Introduction to the theory of computation*. Boston, MA: Thompson Course Technology Division of Thompson Learning, Inc.

Slembeck, T. (2003). Ideologies, beliefs and economic advice – A cognitive-evolutionary view on economic policy-making. In P. Pelikan, & G. Wagner (Eds.), *The evolutionary analysis of economic policy*, Cheltenham, UK; Northampton, MA, USA: Edward Elgar.

Smith, K. K., & Crandall, S. D. (1984). Exploring collective emotions. *American Behavioral Scientist, 27*, 813–828.

Smith, K. K., Simmon, V. M., & Thames, T. B. (1989). 'Fix the Women': An intervention into an organizational conflict based on parallel process thinking. *Journal of Applied Behavioral Science, 25*, 11–29.

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

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

Smullyan, R. (1991). *Godel's incompleteness theorems*. New York: Oxford University Press.

Smullyan, R. M. (1984). Chameleonic languages. *Synthese, 60*(2), 201–224.

Solé, R. V., & Bascompte, J. (2006). *Self organization in complex ecosystems*. Princeton, NJ: Princeton University Press.

Soros, G. (1994). *The Alchemy of finance – Reading the mind of the market*. New Work: Wiley.

Stahl, D. O. (1993). The Evolution of Smart Players. *Games and Economic Behavior*, October 1993, *5*(4), 604–17.

Stahl, D. O., & Wilson, P. W. (1994). Experimental Evidence on Players' Models of Other Players. *Journal of Economic Behavior and Organization*, December 1994, 25(3), pp. 309–27.

Stahl, D. O. (1995). Boundedly rational rule learning in a guesssing game. *Games and Economic Behavior, 16*, 303–330.

Stanovich, K. E., & West, R. F. (2002). Individual differences in reasoning: Implications for the rationality debate. In T. Gilovich, D.Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 421–440). Cambridge, UK: Cambridge University Press.

Stapel, D. A., Reicher, S. D., & Spears, R. (1995). Contextual determinants of strategic choice: Some moderators of the availability bias. *European Journal of Social Psychology, 25*, 141–158.

Steier, F. (1991a). Introduction: Research as reflexivity, self-reflexivity as social process. In F. Steier (Ed.), *Research as reflexivity* (pp. 1–11). California: Sage Publications Ltd.

Steier, F. (1991b). Reflexivity and methodology: An ecological constructionism. In F. Steier, (Ed.), *Research as reflexivity*(pp. 163–185). California: Sage Publications Ltd.

Stiehler, A. (2003). *Do individuals recognize cascade behavior of others? – An experimental study*. Discussion Paper, 02, Max Planck Institute for Research into Economic Systems, Jena.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy, 69*, 213–225.

Stratmann, T. (1997). Logrolling, In D. C. Mueller, (Ed.), *Perspectives on public choice: A handbook*. Cambridge, UK: Cambridge University Press.

Suber, P. (1987b). Varieties of self-reference. In S. J. Bartlett, & P. Suber (Eds.), *Self-reference. Reflections on reflexivity*. Dordrecht: Martinus Nijhoff Publishers.

Suber, P. (1989). The reflexivity of change: The case of language norms. *Journal of Speculative Philosophy, 3*(2), 100–129.

Suber, P. (1990). *The paradox of self-amendment: A study of logic, law, omnipotence, and change*. New York: Peter Lang Publishing.

Tamborini, R. (1997). Constructivism. A pattern of human knowledge for economics. In R. Viale, (Ed.), *Cognitive economics*. LaSCoMES Series, I/1997.

Thomas, W. I., & Thomas, D. (1928). *The child in America*. New York: Knopf.

Thurston, L. L. (1931). The indifference function. *Journals of Social Psychology*, 2:39-67.

Tietzel, M. (1989). Prognoselogic, oder: Warum Prognostiker irren dürfen – On the logic of economic forecasting. *Jahrbuch für Nationalökonomie und Statistik, 206*(6), 246–262.

Todd, P. M., & Miller, G. F. (1999). From pride and prejudice to persuasion: Satisficing in mate search. In G. Gigerenzer, P. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 287–308). Oxford: Oxford University Press.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgement under uncertainty. In M. Fischbein, (Ed.), *Progress in social psychology*. Hillsdale, NJ: Erlbaum.

Tversky, A., & Kahnemann, D. (1982). Judgement under uncertainty: Heuristics and biases. In D. Kahnemann, P. Slovic, A. Tversky (Eds.), *Judgement under uncertainty* (pp. 3–21). Cambridge: Cambridge University Press.

Van Fraassen, B. C. (1970, October). Inference and self-reference. *Synthese, 21*(3–4).

Venn (1966). *The logic of change* (4th ed.). New York: New York Chelsea Publishing Co.

von Foerster, H. (1992). Entdecken oder Erfinden – Wie lässt sich Verstehen verstehen? In G. Heinz, & H. Meier, (Eds.), *Einführung in den Konstruktivismus* (pp. 41–88), 7. Aufl. (1. Aufl. 1992).

Von Foerster, H., & Von Glasersfeld, E. (1999). *Wie wir uns erfinden. Eine Autobiographie des radikalen Konstruktivismus*. Heidelberg: Carl Auer.

Von Glasersfeld, E. (1996). *Radikaler Konstruktivismus – Ideen, Ergebnisse, Probleme* [Radical constructivism. A way of knowing and learning, London: The Falmer Press, 1995]. Frankfurt am Main: Suhrkamp.

Von Humboldt, W. (1945). *Über das vergleichende Sprachstudium in Beziehung auf die verschiedenen Epochen der Sprachentwicklung*. Leipzig: Felix Meiner.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Düsseldorf: Verlag Wirtschaft und Finanzen.

Watzlawick, P. (Ed.). (1981). *Die erfundene Wirklichkeit – Was wissen wir, was wir zu wissen glauben? – Beiträge zum Konstuktivismus*. München: R.Piper & Co. Verlag.

Weber, R. A. (2003). Learning with no feedback in a competitive guessing game. *Games and Economic Behavior, 44*, 134–144.

Wehner, B. (1995). *Die Logik der Politik und das Elend der Ökonomie*. Darmstadt.

Weick, K. E. (1995). *Sensemaking in organisations*. Thousand Oaks, CA: Sage.

Weiss, P. (1992). The theory of type. In S. J. Bartlett (Ed.), *Reflexivity. A source-book in self-reference*. North-Holland: Elsevier Science Publishers B. V.

Welch, I. (2000). Herding among security analysts. *Journal of Financial Economics, 58*(3), 369–396.

Wellman, H. M. (1990). *The child's theory of mind*, Cambridge, MA: MIT Press.

Whewell (1987). Self-reference and meaning in the natural language. In S. J. Bartlett, & P. Suber (Eds.), *Self-reference. Reflections on reflexivity*. Dordrecht: Martinus Nijhoff Publishers.

Whitehead, A. N., Russell, B. (1910, 1912, 1913). *Principia Mathematica* [2nd ed., 1925 (Vol. 1), 1927 (Vols 2, 3). Abridged as: *Principia Mathematica to *56*, abridged, Cambridge University Press, 1962]. Cambridge: Cambridge University Press.

Whorf, B., & Carroll, J. (Eds.) (1964). *Language, thought, and reality: Selected writings of benjamin lee whorf*. Cambridge, MA: MIT Press.

Willinger, M., & Ziegelmeyer, A. (1998). Are more informed agents able to shatter information cascades in the lab? In P. Cohendet, P. Llerena, H. Stahn, & G. Umbhauer (Eds.), *The economics of networks: Interaction and behaviours* (pp. 291–305), Heidelberg: Springer.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103–128.

Winrich, J. S. (1984, December). Self-reference and the incomplete structure of neoclassical economics. *Journal of Economic Issues, 18*(4).

Wittgenstein, L. (1953). *Philosophische Untersuchungen*. Bibliothek Suhrkamp.

Wittgenstein, L. (1999). *Tractatus logico-philosophicus. Tagebücher 1914–1916. Philosophische Untersuchungen* (Vol. 1). Frankfurt am Main: Suhrkamp.

Woolgar (1992). Some remarks about positionism: A peply to collins and yearly. In A. Pickering, (Ed.), *Science as practice and culture* (pp. 327–342). Chicago: University of Chicago.