The Behavioral Economics of Social Interaction

Incentives, Intrinsic Motivation, and Value Learning

Blaž Remic

The Behavioral Economics of Social Interaction:

Incentives, Intrinsic Motivation, and Value Learning

De gedragseconomie van sociale interactie:
Prikkels, intrinsieke motivatie en leren van waarden

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. A.L. Bredenoord

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on
Friday the 20th of May 2022 at 13:00 hrs.

by
Blaž Remic
born in Ljubljana, Slovenia.

Erasmus University Rotterdam

Doctoral Committee:

Promotor:	Prof.dr. A. Klamer
Other members:	Prof.dr. P. Lewis
	Prof.dr. E.M. Sent
	Dr. C.W. Handke
Copromotor:	Dr. E. Dekker

"At the present moment economic interest and discussion are in one of their periodic swings away from the more philosophical aspects of the subject, in the direction of the empirical and the concrete. Expressions of weariness and impatience with methodology and speculation and all generalities are the familiar note. We are urged to be 'scientific' in the manner of the laboratory sciences, to devote ourselves to the observation of 'facts,' and to eschew generalisation and all assertions outside the realm of empirical verification. Such movements come and go. The balance between observation and analysis will always be a shifting one. The philosophical interest never dies out and will always come into its own–and be overdone in its turn. Not for any long period, certainly, can any science which deals with human conduct and social policy remain aloof from the broad and difficult but unescapable problems connected with the nature of value and its relation to reality and the methods by which both are tested and known. The great names in the history of economic thought are to a remarkable extent prominent also in the history of moral science and of logic, and it is no more probable than from the standpoint of economics it is desirable, that this condition of affairs will be greatly changed in the future."

Frank H. Knight, Economic Psychology and the Value Problem (1925)

Contents

Acknowledgements

To me, recognition that many more people besides myself have been involved in the process of producing this thesis is neither simply a nod to the truism about the role of specialization and division of labor in the modern industrial society (well, thank you Steve Jobs for this computer!), nor a mere gesture of academic decorum. It rather means naming the components of the extended mind that in an utterly literal sense created this thesis.

First and foremost, my supervisors. The role of Erwin Dekker cannot be overstated. Erwin, I consider your part to be without a substitute. Were it not for you, I would most probably not have written a thesis. Not in this, not in any other form. Simple as that. Arjo Klamer represented the quiet background presence and an ongoing inspiration, but also active interventions at the right moments. Arjo, it's an honor to be part of your tribe.

Academic seminars and conference sessions are both critical for the well-functioning republic of science. Portions of this thesis have been presented to engaging audiences at the conferences organized by ESHET, WINIR, EAEPE, and INEM. I benefited from the intellectual input as well as from the sense of belonging and collegiality brought about by each challenging question. Roland Fritz and Nils Goldschmidt invited me to contribute to the immensely stimulating conference on contextual economics at the University of Siegen, which enabled me to develop one of the chapters into a publication in the Journal of Contextual Economics. I much appreciate the generous gesture by Mark McAdam to proofread the final version of the paper. Econ and Culture seminar at the ESHCC has been something of an intellectual home base throughout the last years. I wish to thank the regular seminar participants, particularly Hans Abbing, Carolina Dalla Chiesa, Aldo do Carmo, Christian Handke, Ellen Loots, Valeria Morea, Lucy Petrova, and especially Anna

Mignosa for the remark after my very first presentation, which planted a seed that in time grew into one of the pillars of this thesis. At the Erasmus Institute for Philosophy and Economics, Jack Vromen gave valuable feedback and suggestions during my presentation at their student seminar.

Besides my wonderful co-authors Erwin and Carolina Dalla Chiesa, yet another group of people had a direct influence on the contents of this thesis. I have benefitted enormously from the unpaid labor of a number of anonymous journal referees. It takes a village to write a paper, and this thesis would have looked substantially different without their valuable challenge and input.

I am extremely grateful to Elbert Visser and Jan Stroeve for hosting me at three critical times in the writing process, when I was lucky enough to be able to take a much-needed break from the reality of domestic family life during the pandemic and enjoy a writing retreat in their bed and breakfast.

Finally, my family. I wish to thank my parents Vilma and Martin, as well as Erika and Peter, for all the help and support. Oliver and Emil made sure that my home office was never dull. Building a quiet sanctuary of study and thought will have to wait for later periods in life. I dedicate this thesis to my wife Elizabeth. Thank you for all your love and understanding.

Preface

Three vignettes from the life of a musical novice

It is a hot summer evening in a beautiful setting of an open auditorium in Perugia, Italy. The great jazz pianist Keith Jarrett is performing with his legendary trio. Not long into the second song, Keith stops playing abruptly and asks for a microphone. "Here we go," I say to myself, and my friends and I chuckle as we brace ourselves for a lecture from an artist notorious for his zero tolerance of any disturbances coming from the audience. Indeed, he calls out and openly shames and humiliates a poor fellow in the third row for daring to reach for his camera and take a picture in the middle of the song. "If you want to take pictures," he says with an annoyed tone in his voice, "go out and take pictures of the countryside! Music is not about snapshots! Music is about continuous flow!"

Outside a beautiful morning. Inside a room packed with budding music students. A famous pianist is giving a masterclass. "What are the three most important ingredients of jazz?" he asks. As we look at each other, he starts to lay it out for us. "Blues…." A round of nods. Of course, 'the feel' and pure emotion that can't be codified and taught, what else! "Bebop…" Another round of nodding. Of course, technical virtuosity on the instrument and a fluent command of the musical language, what else! A silent moment of tension follows, filled with expectations of what the

third element is. Finally, he cuts the silence. "…and surprise." The master grins as he delivers his punchline.

The evening is dark and cold, and I am sitting in my dimly lit music room, alternating between practicing, listening to music, and watching interviews with my musical heroes. Brian Eno is talking about how things come about. Contrary to what many people seem to think, he says, Beethoven did not first come up with his string quartets in his head, and then wrote them down. Musician is more like a gardener than an architect, planting a seed and then tending to it to see what it will grow into. Eno also says that new artforms are product of scenes, not of individuals. Groups and social settings have a way of generating and developing ideas that transcend any individual effort. Our obsession with the individual genius prevents us to understand how society actually works.

Each of these three vignettes contains a lesson that has through the years had a deep sustaining influence on my thought. Not only in my previous musical life, but to the extent that they keep on shaping the way I think about economics and social science. As will hopefully become clear to the reader, the arguments in the following chapters are all in one way or another informed by these few basic insights.

The first one is about my interest in processes rather than discrete states, in dynamic rather than static theories. The second one is about the discovery, innovation, creativity, and emergent aspects of human action, rather than any notions of efficiency or maximization. And the third one has informed my interest in social intelligence rather than in the characteristics of individual ability. The last one also means that academic work is very much attractive to me for this reason. When working on this thesis I experienced lonely individual work, no doubt about

that. But I have even more to thank to the collective intelligence, sparks of insight that come from others and that evolve and grow into a whole that I could never claim to be utterly mine. The proof is in the pudding, they say. And I believe that the work that has led to this thesis, with all the people that have been involved in shaping it, stands as evidence of this process in its most fulfilling form.

Introduction

The relationship between economics and psychology has been a contested subject among economists. The standard story about the interaction between the two disciplines is the 'in-out-and back in' story (Bruni & Sugden, 2007; Sent, 2004; Hands, 2010). It usually starts something like the following: there was once a period in which economic reasoning was very much informed by psychological considerations about the role of human sensations, and introspection was a viable method to derive insights about the nature of economic action. But as the world was becoming more and more rational and bureaucratized, and economics gradually became *the* science of studying this world, such ostensibly pre-scientific thinking was abandoned. The decisive step in the separation of economics and psychology is often attributed to Vilfredo Pareto. He described two types of actions: logical, which is in the domain of the rational, and non-logical, which consists of what is left. As Bruni and Sugden (2007) explain, "for Pareto, an action is logical if and only if it is the result of valid instrumental reasoning from objectively true premises" (p. 155). Pareto declared only this kind of instrumental reasoning to be the domain of economics. He dispelled the leftover aspects of human action to other disciplines, particularly sociology and psychology, while economics reigned in the queendom of rational action. Other giants of economic thought, such as Frank Knight and Paul Samuelson, followed the suit and for a while, so it seemed, economics was firmly established as a pure science of choice and allocation, free from any 'non-logical' psychological or sociological influences.

After a while, however, cracks started to appear in the structure of purely rational action, and scholars slowly started to fill those cracks with insights from psychology. And thus eventually, so the standard narrative goes, psychology found its way back into economics in the form of the new subdiscipline of behavioral

economics that today gets recognized by the Nobel committee, features prominently in all the top economics departments and journals as well as in popular imagination, and became a serious force in the world of policy making. What's more, psychological insights have been firmly recognized to be part of the mainstream approach to economic analysis (e.g., Chetty, 2015). In this thesis, I take such state of affairs as given, which requires us to modify what is to be explained. The relevant question is not *whether* psychology is in or out. The relevant question is *what kind* of psychology found its way into economics.

The return of psychological considerations into economic analysis during the last half century is characterized by the introduction of two distinct kinds of psychology. This has resulted in two very different types of behavioral economics. The prevalent one–or Type I, as I will call it here–is what is today known as *the* behavioral economics. Its development can be traced to two main influences. On the one hand, it grew out of the work by Herbert Simon on bounded rationality, and on the other, out of the research on behavioral decision making (Heukelom, 2014). The latter research program originated in the work by organization and decision theorists in the 1950's. It brought in the emphasis on "[measuring] experimentally which decision subjects make with respect to [an] objective stimulus" (Heukelom, 2014, p. 62). As such it offered the possibility of experimentally testing the assumptions of rational choice theory. Many of the most prominent behavioral economists today–such as Collin Camerer, for example–were originally trained in behavioral decision research. Bounded rationality, the second prominent aspect of Type I behavioral economics, has a central status because the standard assumptions of rationality are deemed unrealistic (Kahneman, 2003). As Simon (1955) emphasized, people's decision-making abilities are bounded by limited cognitive capacities and limited access to information, which both prevent the agents to do what the standard economic theory assumes them to do: to maximize. The increased realisticness of the Type I behavioral economics is said to lie in the ability of the approach to provide descriptive accounts that are closer to how agents are actually observed to behave and choose in experimental settings.

In my characterization, Type I behavioral economics encompasses both 'old' (corresponding to the Simonian line of influence) and 'new' (corresponding to the

experimental influence of the behavioral decision making research) behavioral economics as Sent (2004) defined them. Despite the differences between these two, they are, in essence, both primarily interested in correct descriptions of behavior, and they share the view that the neoclassical approach is problematic precisely because it is descriptively false. Type I behavioral economics is thus an approach that seeks to identify anomalies in the applications of the standard economic theory[1]. But this kind of critical stance, in turn, also makes it a perfect bedfellow for neoclassical economics, which becomes especially relevant once behavioral economists start adopting an openly normative stance (Camerer et al., 2003; Thaler & Sunstein, 2003). Following that, Type I behavioral economics is not anymore only a way to demonstrate anomalies; it is a way to suggest interventions that aim to correct certain aspects of behavior and change the outcomes.

This jump to explicitly normative implications of the Type I analysis has been recently increasingly problematized (Rizzo & Whitman, 2020; Sugden, 2018). Gal and Rucker (2018) trace its roots to the change in conception about the nature of the work in behavioral economics that happened somewhen between the early 1980's and the late 1990's. Consider prospect theory, for example. Kahneman and Tversky (1979) are explicit in stating that prospect theory offers "a useful framework for the *descriptive* analysis of choice under risk" (p. 289, emphasis added). But, as Gal and Rucker (2018) point out, in time it got to be used as an *explanatory* account. This has led to circularities in accounting for the observed behaviors, where "for example, the endowment effect, described shortly, is cited as evidence of loss aversion, and loss aversion is cited as an explanation for the endowment effect" (Gal & Rucker, 2018, p. 499). Critics point out that proponents of the Type I behavioral economics have not been able to provide convincing justifications for this jump from descriptive to explanatory. And the experimental research itself has come under scrutiny during the recent replication crisis.

[1] Note that for years Richard Thaler was writing a famous section in the Journal of Economic Perspectives called *Anomalies* where he reported "successful searches for disconfirming evidence" (Thaler, 1987, p. 198).

Parallel to these developments, another type of psychology was influencing the work of economists. It resulted in a different type of behavioral economics–let us call it Type II–that does not start from the premise that rational choice theory is descriptively false and thus we need a correction, but from the premise that rational instrumental action does not lead to human flourishing. Rather than being interested in the behavioral patterns and choices people make, this second type of behavioral economics seeks to account for the subjective experiences that people have while acting. It has more to do with subjective well-being, general satisfaction, and quality of life. Tibor Scitovsky was the economist that first prominently imported this kind of psychology into economics. For him, work in purely instrumental terms, as a means to satisfy one's given desires through consumption, strives for comfort. But since comfort means a lack of surprises and novelty, it also means the absence of pleasure. The very pursuit of preference satisfaction thus results in decreased welfare. It leads to an undesired situation that Scitovsky famously called 'the joyless economy' (Scitovsky, 1992). The view that rational preference satisfaction is by definition dull and unstimulating, and thus has direct negative consequences on economic development and growth by stifling creativity, is a view far removed from the normative ideal of rational maximization sought for by Type I behavioral economists. It comes as no surprise that Scitovsky was never part of any contemporary groups that were working on the Type I behavioral economics.

In general, the central feature of the Type II behavioral economics is a conceptual distinction between two fundamentally different kinds of motivations: intrinsic and extrinsic (Frey, 1997). Proponents of the Type II behavioral economics argue that incentives and other extrinsic motivators are potentially harmful to the psychological wellbeing of the agent because they negatively affect the agent's sense of autonomy. The common feature of the Type II approaches is thus the opposition to instrumentality as a basis of economic activity. Type I is still instrumental, albeit with limited cognitive capacities and limited information; Type II, on the contrary, wants to abolish instrumentality altogether. Thus, rather than advocating for a more realistic descriptive accounts of decision making, Type II behavioral economics advocates for a change in focus towards the study of motivation for economic action.

The distinction between the two types of behavioral economics that I described thus far maps well to the distinction that Boulding (1969) drew between economic and heroic ethics. The former is the ethics of calculation and instrumentally rational conduct. It is true that Type I behavioral economics seeks to identify the cracks in the image of the rational economic man; but the essence of the economic action is still instrumental, despite people making cognitive mistakes. The focus of Type I behavioral economics is on the assessment of the outcomes of the actual choice process in the experimental settings as compared to the outcomes that would follow from the rational choice analysis. Moreover, what Thaler and Sunstein (2008) called an Econ, referring to someone that possesses "complete information, unlimited cognitive abilities, and complete self-control" (p. 5), is the *normative* ideal of the whole enterprise. It is better to be omniscient than to make calculating mistakes. Type I is thus firmly in the domain of economic ethics.

In contrast, Type II behavioral economics is characterized by heroic ethics, which is the expressive ethics of 'doing your own thing' without regard to any considerations of costs and benefits. The normative ideal here is an agent that acts unconstrained by external influences such as rewards or punishments, only following her own autonomous impulses and acting out of her own perception of personal identity. In this framework, being an Econ cannot be the ideal, because identifying and acting as an Econ means that one obeys the analysis of costs and benefits as assessed through the available external incentive structure. Within the framework of heroic ethics, being and acting like an Econ is in fact undesirable. For one, it has negative reputational effects, because it reflects poorly on whoever is labeled to be one. But it may also directly lead to lower levels of psychological wellbeing. Rather, the ideal is one of performing an activity for the sake of itself, without regard to external incentive structures and often even at the expense of severe personal costs (Loewenstein, 1999).

The distinction I am drawing between the two types of behavioral economics also implies different dimensions of choice that matter for the economic analysis. Type I emphasizes the perspective on choice of the external observer, such as experimentalist or the policy maker, who can objectively assess both the problem at hand and the costs and benefits linked to alternative choice options. Thus, the

relevant level where choice is assessed is this policy maker's rationality, and the normative concerns are mostly directed at this external entity (also called the 'planner', or the 'choice architect') who has the ability to intervene into the decision process and guide the fallible individual towards an optimal outcome. Contrary to that, Type II sees the level of policy maker's rationality not as normative but as negative. It emphasizes choice as a deeply individual expression of values, identity, and authenticity. The relevant level of choice is thus individual psyche. This has important policy implications, because the success of a policy intervention cannot be assessed based only on positing or observing some set of economic outcomes. It has to take into account the autonomy and authenticity of the individuals and their subjective wellbeing.

Both polarizing views have strengths but also important limitations. Type I offers a rather straightforward way of deriving testable hypotheses, and a set of analytical tools that are readily available to the practicing economist. But it also raises important concerns over paternalism and the neglect of the perspective of the individual actors. Type II offers a perspective on individual autonomy that many people find intuitively attractive. But, as Boltanski and Chiapello (2005) have argued, the emphasis on individual autonomy and authenticity has led to many negative unintended consequences on the societal level, such as increased anxiety brought about by the flexibilization of work that was motivated precisely by the calls for more personal freedom and authenticity. Most importantly, both types of behavioral economics are characterized by a decisive turn inwards, into the realm of the individual mind as the locus of cognitive activity. In Type I this is demonstrated by the focus on cognitive biases and on social preferences as expressions of individual utility functions. And in Type II it is demonstrated by the conceptualization of the social environment as external factor that has potential corrupting effects on the individual.

The overarching theme of the following chapters is that a yet different type of behavioral economics is both possible and needed. A behavioral economics that does not focus exclusively on the workings of individual minds but takes seriously the fact that people are embedded in a social world. Taking this embeddedness seriously means that we need to consider the deeply entangled nature of the

relationship between individuals and their social and institutional environments. This is a perspective that resides in the realm of social interactions, between the two extremes of Type I's instrumental and Type II's expressive perspectives. It echoes Kenneth Boulding's memorable quote that "economic man is a clod, heroic man is a fool, but somewhere between the clod and the fool, human man, if the expression may be pardoned, steers his tottering way" (Boulding, 1969, p. 10). People living in between these two extremes act and interact through and with the help of institutional and social environment. In figuring out what the right thing to do is, they are guided not only by their inner impulses, or responding to some external standard of rational action, but primarily by a constant process of discovery and learning–through interacting with each other and with their environment–about how to appropriately interpret and evaluate the situation. To study this, we need not behavioral economics of individual cognition or of individual psychological wellbeing. We need *the behavioral economics of social interaction.* An approach that centers on the study of intersubjective meaning and builds on an insight from the recent cognitive science that individual minds and their environments are epistemically and ontologically entangled.

In the first chapter I set the stage by addressing the problem of the effectiveness of incentives. The idea that incentives matter is rightfully considered one of the cornerstones of economic thinking. But different strands of economic research, including the psychologically informed ones mentioned above, have demonstrated that incentives cannot be treated as some objective natural force that directs human behavior. The chapter examines the obvious but somewhat neglected peculiarity that while some incentives are felt as powerful reasons to alter actions, other incentives have little, or even counterproductive effects. I demonstrate that social world plays a crucial role in the conception of incentives because incentives become meaningful in relation to the social settings, social roles, and institutional practices that people engage in as part of their ordinary business of life. The relative meanings of costs and benefits are not objective facts of the social world but are arrived at and acquired intersubjectively through processes of social interaction within particular social settings. In this regard, incentives may backfire when their meaning is unclear or contested by another competing meaning.

When accounting for the potential ineffectiveness of incentives, economists mostly evoke the psychologically informed literature on intrinsic motivation. In Chapter 2 I review this literature and argue that the concept of intrinsic motivation has been used by economists in inconsistent ways because the underlying theories of intrinsic motivation, imported into economics from psychology, are competing and mutually exclusive despite employing the same terminology. I demonstrate that these differences have important implications for the empirical work and incentive-based policy interventions. The relationship between psychology and economics gets complicated when we consider that different psychological theories may imply fundamentally different visions of human psychology. This means that the standard argument for importing psychological insights into economics, which is that psychology increases realisticness of economic theory, is not as straightforward as it is often assumed by economic practitioners.

Chapter 3 further develops this insight through the analysis of the concept of ecological rationality, which has recently been advanced as an alternative to the conception of bounded rationality based on cognitive biases. In the chapter, I argue for the importance of understanding the underlying differences in both psychological and economic approaches that get combined as part of the efforts of behavioral economists to explain market outcomes. This chapter also provides a first step towards the new type of behavioral economics. We come across a new way of thinking about psychology and cognition that enables us to depart from the strictly individualistic and internalist perspective of the theories encountered so far. This type of psychology combines especially well with the economic approaches that emphasize the role of institutions in economic life. Environment starts to acquire cognitive properties on its own.

In Chapter 4 I develop this idea into a novel approach for conceptualizing environment in economics. I argue that the traditional conception of environment as a type of constraint on individual action has been limiting because it neglects the important interdependencies that arise as people interact with their environments. The chapter employs insights from the recent cognitive science on socially extended mind to demonstrate how the study of economics can benefit from reconceptualizing environment not as a constraint on individual action but as a resource for

constituting socially extended cognitive processes. By enabling people to tap into the knowledge embodied in institutional practices, these resources play a key role in making possible particular economic actions, types of reasoning, and the creative discovery of new features of the environment, or new potential actions within that environment, that further fuel the learning process about the possibilities for action and interaction.

The themes of incentives, motivation, interactions between the individuals and their environments, and learning through the discovery of new potential actions, are weaved together in Chapter 5 with the help of Agnes Callard's account of aspiration and value acquisition (Callard, 2018). The chapter starts with an observation that psychological literature on intrinsic motivation only offers a psychological mechanism of why incentives may be ineffective, but it falls short of explaining why people engage in activities in the first place, why they value one activity over another, and how those values come about. I argue that value learning must be seen as the central process. This transcends both Type I's emphasis on rational normative preferences, and Type II's emphasis on intrinsic reasons for action. The chapter also demonstrates that the exclusive focus on choice, be it policy maker's or individual, is misleading, because it presupposes that whoever is making the choice possesses the evaluative apparatus that enables them to make a relevant judgment. Rather, we should turn our analytical focus on the study of processes and activities in between the extremes of economic and heroic ethics; and in between the extremes of omniscient policy makers and autonomous expressive individuals, where people continuously strive to resolve value conflicts and learn new values through acting and interacting in, and with, their social and institutional environments.

1

Incentives Matter, But What Do They Mean? Understanding the Meaning of Market Coordination*

(with Erwin Dekker and Carolina Dalla Chiesa)

1. Introduction

The idea that incentives matter has become a near mantra in economics. However, there is something peculiar about the idea that incentives matter, an oddity that gets little attention in the discipline. Why are some incentives felt as very powerful reasons to alter actions at the same time that other incentives hardly manage to produce any effect at all and yet other incentives actually have counterproductive effects? Why do social norms sometimes act as a strong incentive not to cheat, while in other situations social norms seem powerless to prevent cheating? And why are monetary incentives often a powerful stimulus for a certain type of behavior, yet at other times they actually work as a deterrent? This chapter argues that the start of an answer to these questions is present in current literature at the intersection of sociology and economics.

In this growing body of research, both economists and sociologists see the fundamental problem of the social sciences as the question of how coordination between individuals with different objectives and interests can come about. Rather than thinking in objective terms about incentives, they study how coordination

* This chapter has been published as "Incentives matter, but what do they mean? Understanding the meaning of market coordination." *Review of Political Economy,* 2020, Vol. 32, 2, 163-179.

comes about through processes of meaning-making. These studies analyze how social environments and embedded incentives come to be understood in particular ways. This work disputes the existence of 'hard' objective incentives, and instead see only incentives as understood by the actor. In empirical approaches within economics, particularly the institutional branch, and in pragmatic approaches to decision-making in sociology, we see a convergence not only in the type of problems studied and the empirical strategies pursued, but also in their underlying theoretical and methodological approach. It shows important similarities with the mid-twentieth-century work of Alfred Schutz and his social phenomenology, and identifies various methodological strategies as embracing a common 'Schutzian' answer to the problem of mutual coordination. Lastly, we explain why this convergence is a fruitful ground for a shared approach to study economic interaction and human interaction more broadly.

The empirical approaches we are talking about comprise work in institutional and Austrian economics on the one hand and work in the tradition of pragmatic (economic) sociology on the other. In these approaches, coordination among individuals with subjective understandings of the situation is the central theoretical problem, and empirical work focuses on the emergence of intersubjective understandings through communication, signaling, conventions, institutions, and formal or informal rules. The flip side of that focus is how shared understandings break down, are upset, and change. Although this might sound highly abstract, there is an important difference between these approaches, on the one hand, and alternative approaches in economics and sociology. This difference is, first and foremost, that everyday understandings of the social world are taken as the starting point in these alternative approaches, something labeled as 'the pragmatic approach' in sociology (Thévenot, 2001). This distinguishes them from more structuralist approaches in sociology, such as Marxist and social network analysis traditions, and the rational choice approach in economics, based on costs and constraints.

Based on this pragmatic methodology, studies often focus on everyday understandings of the social world. They are particularly attentive to the open-ended nature of human action, and the potential for coordination and discoordination in social interaction. Scholars involved in these traditions seek to study how

intersubjective understandings coordinate individual and collective plans, and how particular (powerful) agents seek to change these understandings to their advantage. In contrast to structuralist perspectives, which emphasize the close ended nature of, and the determinism in, the existing structures, the alternative approach pays close attention to interpersonal meaning structures that facilitate decision-making processes.

Nevertheless, we recognize there are differences between the individual authors within and between these approaches, and there remain important differences between the more sociologically and more economically minded authors. Most notably, they disagree over the extent to which individuals are able to shape intersubjective understandings, and the agency ascribed to these intersubjective structures. But such disagreements should not obscure the fundamental agreement and shared conception of the problem of coordination in human interaction, or the need to study coordination processes through the formation of shared understandings.

This chapter provides three contributions. First, it shows the nature of the convergence between recent work in economic sociology and economics. Second, it demonstrates that this convergence has one of its theoretical roots in the intersubjective social phenomenology of Alfred Schutz, whose influence tends to be unacknowledged. The work of Schutz will help clarify the distinctiveness of the recent convergence, and show how different elements of different alternative approaches might form a coherent research program. Third, this chapter applies these insights to critique the way economists tend to think about incentives as objective stimuli for certain types of behavior. It demonstrates how the empirical work discussed provides an alternative way to think about incentives as intersubjectively understood reasons for performing a particular action. This will enrich the theory of incentives, which do not matter objectively, independent of their meaning, but matter precisely because they have meaning to actors.

The chapter proceeds as follows. Section Two identifies four Schutzian building blocks that allow us to detect shared elements in the different empirical approaches. This provides hints for a renewed theoretical and methodological justification of some recent empirical work. Section Three is devoted to a discussion

of these approaches in sociology, and Section Four examines these approaches in economics. Section Five explains why this convergence is not merely an interesting opportunity for cross disciplinary learning but also an important correction to prevalent ideas in sociology and economics. In particular, it addresses what economists can learn from these alternative approaches for studying incentives.

2. Four Schutzian Building Blocks

Alfred Schutz was one of several thinkers who brought sociology and phenomenology together by combining the contributions of Max Weber and Edmund Husserl (Wagner, 1983). For Schutz, the social world is a world of meaning in which individuals coordinate their actions with those of others. In order for this to be possible, individuals enact ways of understanding the world around them. What sets Schutz's approach apart from many other sociological approaches is that he embraces Husserl's phenomenology to discuss the problem of intersubjectivity in the realm of the social world. He unveils the everyday life structures through the notion of subjectively meaningful actions as his starting point. These experience-based understandings are the subject of his first book *Der sinnhafte Aufbau der sozialen Welt.* Our purpose here is not to give an overview of Schutz's intellectual project, which is covered by Wagner (1983) and Prendergast (1986)[1], but rather to identify four central tenets in his work.[2] The following sections identify similarities between elements of his work and recent empirical approaches in sociology and economics.

The first building block, and most central notion in his work, is that of *Verstehen*, which lies at the core of everyday understandings of the world. Schutz adopted this interpretive methodological position from his predecessors, especially Max Weber, but critically expanded on it. Verstehen, for him, is relevant on multiple

[1] These two authors disagree on the connection between Schutz and Austrian school of economics. Augier (1999) correctly observes that Wagner downplays the importance of this connection in Schutz's intellectual history.

[2] We draw primarily on the English translation of his seminal book (Schutz, 1967) and the collection of essays published in Schutz (1962).

levels. First, there is the problem for the actor in understanding the actions of others and drawing out his own plan or project. Schutz thinks of individuals as coordinating their own actions with the potential actions of others in mind because interpretations of the social world of individuals are subjective. Second, there is the problem for the outside observer, who is similarly engaged in a project of Verstehen in order to figure out what individuals have intended or meant with their action. This also brings up the epistemological problem of any social science project: how much can we ultimately know about the plans and motivations of others?

In order to deal with both the first and the second problem, Schutz argues we have a need for the Weberian notion of ideal types, which he expands by including the everyday life perspective and providing a way for phenomenologically based research. This leads us to the second building block, a discussion of how actors make sense of the (potential) actions of others and the social world by drawing on ideal-types reasoning. This is particularly true when we move beyond the interaction of a very small number of individuals, which Schutz sought to do. Since actors cannot form subjective interpretations of each individual's action, they rely on types of actions, or ideal-typical motivation, of which the profit-oriented businessman is but one example. These ideal types are not fixed but depend on the relevant action-situation of the actor. Correspondingly, the scientific observer faces a similar epistemological problem. If we are analyzing the real estate market, we might want to differentiate between certain ideal-typical actors such as the financier, the real estate agent and the homeowner, while if we are analyzing the economy as a whole, we might feel the need to form higher-order ideal types of a greater degree of anonymity to arrive at relatively concise explanations.[3]

The third Schutzian building block, perhaps the most important for the purpose of this chapter, is how interpersonal coordination comes about. A central aspect of Schutz's theory of human action is the idea that individuals form plans or projects about their actions, and within these projects, they have to take into account

[3] Machlup (1936) applied the Schutzian distinction between different degrees of anonymity of ideal types to the study of economic activity, showing that problems involving different ideal types must also involve different methodological approaches.

what they expect others to do. This process of project formation[4] is the coordinative and potentially disruptive process of social interaction. As is well known, coordination is a central concept in modern economics, in particular in game theory, but Schutz's primary concern is how shared understandings of particular situations come about because only when there exists a shared understanding of the situation are individuals likely to form correct expectations about the behavior of others. It is here that his emphasis on intersubjectivity and the way in which situations become socially constructed, or built-up, becomes central. In modern terms, one might say that institutions help foster shared understandings by providing what Lachmann (1971) has called 'guideposts' for the behavior of various individuals. But as we will see below, many empirical studies in economic sociology and economics have paid particular interest to situations involving great uncertainty, where coordination is anything but given.

From interaction among individuals emerges a pattern of coordination. Schutz calls these patterns 'domains of relevance' or 'distinct provinces of meaning', which is the fourth building block.[5] For Schutz, the social world is not a coherent whole but instead a set of somewhat autonomous orders that overlap in complex ways. The provinces of meaning create what in modern sociology is called 'spheres' or 'logics of interaction', which are internally coherent but might conflict with other spheres or logics (Boltanski & Thévenot, 2006; Thornton et al., 2012). The strong point of Schutz is that he explicitly distances himself from the idea that these provinces are somehow ontologically different. Instead, the boundaries between them come about through the meaning-making process of human interaction, and hence new situations can be conflictual in part because it is not clear how they should be classified by the actors.

[4] Schutz's work on project formation and emphasis on time originates in the work of Henri Bergson.

[5] In developing this concept Schutz draws on William James who analyzed reality as consisting of several sub-universes, "each with its own special and separate style of existence" (James, 1890, p. 291).

These four building blocks are, perhaps surprisingly, present in important strands of the contemporary empirical literature in economics and (economic) sociology. Sometimes Schutz's legacy is implicitly acknowledged, while in other instances his themes have been rediscovered quite independently from his work. His work, however, provides a good framework for interpreting the convergence between economics and sociology. This chapter is not about the history of ideas, so we are not primarily concerned with how these contemporary authors came to adopt their approach; instead, we demonstrate how they converge around these four building blocks.

3. Recent Economic Sociology

A prominent description of economic sociology is that the field seeks to combine into a unified analysis economic interests and social relations (Swedberg, 2003). This is done in structuralist ways, where little attention is paid to meaning-making and more to underlying social and economic differences, and it is done in less structuralist ways, where attention is directed to actor perspectives and processes of meaning-making (in the extreme by such approaches as ethnomethodology, which relies completely on qualitative actor perspectives). It is important to observe that Schutz cannot simply be placed within one camp. Although his work utilizes the methodology of Verstehen, the goal is to explain social coordination and social structures.

Within economic sociology, the approach of Viviana Zelizer contains similarities with that of Schutz. Zelizer (2004) has developed the notion of *circuits of commerce*, which are best thought of as distinct provinces of meaning that structure the market and provide guidance for the behavior of individuals. Even though 'Zelizer circuits' (Collins, 2004) do not explicitly refer back to a phenomenological root, the two central arguments of her work–the existence of multiple currencies that are distinct and not perfectly exchangeable, and the idea that different markets give rise to different subjective experiences–provide a way to trace back Schutzian building blocks. Zelizer (2005) details how the valuation of homemaking labor is subject to a plethora of negotiations that intertwine love and money, and how the logic of both structures the way rewards and incentives are

perceived. Her earlier empirical work details the way in which life and death, or the loss of a close relative, came to be valued differently through the rise of the markets for life insurance (Zelizer, 1978).

In related work, Velthuis (2005) describes the symbolic meaning of prices and the intricate way in which artistic and economic ways of valuing a new work of art are negotiated in the front and back room of contemporary art galleries. He demonstrates that gallerists do not simply face incentives to increase or lower prices but, instead, that price symbolizes possible paths of action in ways understood to market insiders and negotiated between buyers, dealers, and artists (Velthuis, 2004). These works rely on actors' understandings, focus on the process of coordination of plans, the associated language, and the resulting structures of meaning.

One might be tempted to argue that this is only to be expected in these somewhat more marginal markets. But recent explorations of the world of finance show that the same is true of central capitalist markets. Knorr Cetina and Bruegger (2002) contend that global financial markets work on the basis of a 'temporal coordination'. This notion, inspired by Schutz's work, is used to discuss the intersubjectivity that agents in global markets develop during the process of doing their work, especially currency trading in global investment banks. Based on empirical findings, the authors argue that if markets coordinate, it is because there is spatial–temporal synchronicity that allows for intersubjective relationships between traders. Their research delves into the work of participants who are geographically distant and disengaged from local settings but bound together by global microstructures. Knorr Cetina and Bruegger (2002) contend that these "patterns of relatedness and coordination … are global in scope but microsocial in character" (p. 907). Also, in the realm of temporal coordination, Abolafia (2001) produces one of the first works that provide an in-depth look at subcultures of Wall-Street, showing how agents negotiate tensions between short- and long-term plan coordination, and how the temptation toward excess spurs market activity.

The issue of interaction in financial markets is widely discussed in economic sociology. Based on ethnographic fieldwork, Smith (2012) shows that narratives mold prices in financial markets since agents work in highly ambiguous environments. Thus, narratives would provide "meaningful, ordered and unified accounts of how

particular events unfold" (Smith, 2012, p. 141). Preda (2012) argues that foreign currency traders in uncertain environments, where the price is set in the process of opening offers and receiving counteroffers, adapt their decisions contingently, not depending on previous decisions about which price is optimal. Ethnographic work, such as Zaloom (2006) and Ho (2009), show how financial markets work from the perspective of daily life. These studies all show the rich symbolic nature of markets, where actors attempt to make sense of the world around them and the actions of others. Prices do not act as simple incentives in these markets but are instead created in the process and given meaning in the narratives that become dominant.[6]

The process of creation is shown to be even more central in financial markets that prioritize the notion of performativity. This concept is used to describe the process by which actors adopt models or concepts that were originally conceived (in scientific discourse) to describe their actions. Holmes (2013) studies the way in which actions of central bankers are informed by concepts and models derived from economic theory. And especially in the ethnographic work of Miyazaki (2006) we find an interesting way in which performativity leads to coordination. His argument is that a particular set of (rational) trading strategies are possible in the first place because of a 'faith' in efficient market assumptions, which entail the notion of an anticipated future, projected and managed to correspond to theoretical constructs. Thus, we have here, in Schutzian terms, an interesting way in which ideal types are enacted by traders who expect other traders to behave like the ideal type. As such, it becomes an important coordinating structure, but only because of mutual understanding and only within a confined domain.

The notion of performativity, as used by these authors, builds on the view that coordination comes about as a result of the familiarity of the actors with the theory that is behind their actions. Morgenstern and Schwödiauer (1976), however, showed that, especially in the cases that involve small numbers of actors, attaining

[6] This ongoing sense-making is crucial since, as James Buchanan has argued, the act of choice is not based on "[maximizing] utilities described in independently-existing functions," and therefore "the potential participants do not know until they enter the process what their own choices will be" (Buchanan, 1982, p. 5).

a stable equilibrium may actually get upset when the knowledge of the theory is 'absorbed' by the actors. In that case, some of them might try to trick others into thinking that they are following some other theory. This would, in turn, result in a change in the observations that others are using for their predictions, and, consequently, the predicted equilibrium would break down.

There are two other domains worth highlighting. In the work of Boltanski and Thévenot (2006) six different 'worlds of justification' are delineated. These are distinct and partially overlapping and conflicting provinces of meaning. For example, their notion of justifications found in the inspired world, and how these interact with those given in the market world, provide excellent case studies for how different understandings of the world give rise to differing understandings and to only partially coordinated plans, and how discoordination and conflict arise. Also, they note how particular signals that are interpreted as powerful incentives in one domain (i.e., the critical praise of peers) can be far less powerful in another domain. David Stark (2009) has extended this framework to analyze conflicts within organizations. And, from a somewhat different background, the institutional logics literature has analyzed conflict between different 'logics' operating within organizations, communities, or societies (Thornton et al., 2012).

Jens Beckert emphasized how goods and other artifacts come to be valued in the modern economy (Beckert & Aspers, 2011; see also Karpik, 2010; Lamont, 2012). This literature examines the question of uncertainty and ambiguity regarding how things get valued, and how the value of goods gets justified, perceived and actively constructed. In more recent work, Beckert (2016) has extended this to how the uncertain future is valued. Drawing on Schutz's work, in which actors convey information and expectations about the future, entrepreneurship comes to play an important role as the quintessential economic practice that transforms uncertainty into potential actions and opportunities. It starts from the perspective of the actor, and the possibility of discoordination stemming from uncertain future projects and shows the process of coordination in an uncertain world. In Beckert's work, we thus find the first elements of Verstehen as well as that of project formation.

4. Recent Economics

While it is perhaps to be expected that Schutzian elements are present in contemporary economic sociology, his work has not been explicitly acknowledged in economics. Nonetheless, we aim to demonstrate in this section that recent economic approaches contain plenty of Schutzian building blocks. And precisely because this is not fully recognized, there is great potential for more engagement between sociology and economics.

New institutional economics has been influenced by game theory, a theory of strategic interaction. This has led to discussions over the nature of institutions in game-theoretic terms. Some have argued that institutions are best understood as constraints on individual behavior (e.g., North, 1990). This approach still relies on an objective notion of costs and constraints. However, a more interesting approach suggests that institutions are equilibrium outcomes of repeated games (Greif & Kingston, 2011). Multiple equilibria are possible, and through coordination, individuals settle on a particular equilibrium that turns into a norm or a rule. These norms are self-enforcing to the extent that individuals have an interest in following the norm given that they expect others to do the same. The seminal example is driving on the right side of the road. A particular institutional arrangement is but one of several possible solutions that can emerge based on the same objective factors. As Thomas Schelling (1960) puts it, "[p]eople can often concert their intentions or expectations with others if each knows that the other is trying to do the same" (p. 57). He describes the mechanism as involving focal points, which he explains are "clue[s] for coordinating behavior." He illustrates this with an example of a husband and a wife trying to find each other in the department store. There is no 'right' or dominant strategy, to solve this coordination problem; it critically depends on what they "expect [each other] to expect to be expected to do" (Schelling, 1960, p. 57). From this, they may be able to choose the right place to meet. It is telling that Schelling uses an example of a married couple and not a pair of completely random strangers, since it is crucial for the successful coordination that participants know

each other or share some common beliefs that they can both rely on when identifying the appropriate, or salient, focal point (Mehta et al., 1994; Sugden, 1989).[7]

The idea that shared beliefs play a role in equilibrium selection is central in the work of Greif (1994, 2006). His comparative study of the medieval societies of Genoese merchants and Maghribi traders assumes that cultural beliefs, that is "ideas and thoughts common to several individuals that govern interaction" (Greif, 1994, p. 915), play an important role in shaping economic outcomes by contributing to the path dependence of the emergent institutional arrangements. Greif shows that individualistic and collectivist cultural heritages of the Genoese and Maghribis, respectively, shaped the different expectations that merchants in these societies held with respect to retaliation for the cheating behavior. This, in turn, affected the institutional solutions that developed to deal with these agency problems. Greif's game-theoretic analysis depends on the meaning structures provided by a particular cultural and historical context, since the development of institutions as stable equilibria that guide behavior through the alignment of incentives relies on shared belief systems in the society.[8] Other economists have extended this approach to study organizations and political institutions (see Greif & Kingston, 2011, Sections 5.2-5.3).

Similarly, Aoki evokes the concept of societal rules, which he defines as "commonly cognized, salient patterns of the ways in which societal games are recursively expected to be played" (Aoki, 2011, p. 23). They differ from formal rules

[7] Another famous Schelling example is two people trying to meet in New York City. While experimental results showed that majority of participants indeed succeeded in meeting each other by choosing the information booth at Grand Central Station at 12 o'clock noon, the fact that this, as Schelling points out, "may reflect the location of the sample in New Haven, Connecticut" (Schelling, 1960, p. 55n) suggests precisely the presence of a certain shared knowledge among the sample population.

[8] "In situations in which an institution generates behavior, the knowledge and information that are compressed into the institutionalized rules enable and guide individuals, despite their limited perception, knowledge, and computational ability, to act in a manner that leads to behavior and reflects the constraints on admissible beliefs and behavior that the game-theoretic equilibrium analysis captures" (Greif, 2006, p. 126).

in that they are recursively observed. In order to be effective, these observed rules need to be shared. Aoki argues that while individuals learn and form beliefs by recognizing patterns in their environment, real shared knowledge comes in the form of cultural heritage. He demonstrates this by comparing different institutional solutions to irrigation problems in two villages in Japan and Korea, showing that cultural factors had a key impact on the equilibrium selection.

Bates et al. (1998) describe a unified methodology for combining interpretive work with the rational choice approach. Their programmatic statement comes in the form of what they call analytic narratives. The group of researchers involved in this program characterize themselves as being part of a "critical trend among a subset of rational choice theorists who have been trying to integrate interpretive and rationalist accounts" (Bates et al., 2000, pp. 697–698). At the core of their approach is a combination of a narrative and historical approach with rational choice theory and game theory, where the case studies and narrative techniques are employed in order to "understand the actors' preferences, their perceptions, their evaluation of alternatives, the information they possess, the expectations they form, the strategies they adopt, and the constraints that limit their actions" (Bates et al., 1998, p. 11), which is then used to construct a game-theoretic account. Several Schutzian building blocks can be recognized here: there is a reliance on the method of Verstehen, the central question is how mutual coordination can happen, and, although generally not explicit, there is often a reliance on ideal types in analyzing the actions of different groups.

There is also some empirical work on how entrepreneurs actively seek to change established understandings of particular goods, which builds on the analytical narrative approach. Shared meaning is in these studies not taken as a background against which economic actors make decisions, but rather as a malleable foreground through which actors try to change the possible range of projects which can be (legitimately) undertaken. In his case study of surrogate motherhood, Pavel Kuchař has studied how entrepreneurs played an active role in altering the understanding of contested commodities. His work demonstrates how the illegitimate practice of 'selling babies' was transformed into an accepted market for 'renting wombs' (Kuchař, 2016). His analysis shows how what was originally both legally and socially

perceived as an illegitimate activity gets transformed into an accepted—or even honorable—practice. He has extended this work to suggest that market exchange, more generally, builds on accepted understandings of particular artifacts, which are transformed by entrepreneurs to create new market categories through exemplary goods (Dekker & Kuchař, 2016, 2017a).[9]

Even the new institutional approach, building on the work of Douglass North and treating institutions as rules rather than equilibria, recognizes that cultural understandings cannot be ignored. North draws attention to the idea that "subjective perceptions of the actors are not just culturally derived but are continually being modified by experience that is filtered through the existing (culturally determined) mental constructs" (North, 1990, p. 138). He develops this in work with Arthur Denzau on shared mental models that serve as an aid to overcome uncertainty (Denzau & North, 1994). Although North makes only limited attempts to explain how such shared mental models emerge, his approach, too, recognizes the importance of the meaning of incentives by emphasizing the role of the diverse intersubjectively shared frameworks of mental models for interpreting the environment. While his subsequent work does not put the questions of meaning at the center of interest, his recognition of the role of meaning structures has opened the door for many later researchers, as we demonstrate below.

More explicitly Schutzian, and more explicitly concerned with the study of the emergence of shared mental models, is the recent empirical work in the Austrian tradition. Don Lavoie (2011) has argued that economists have wrongfully restricted themselves to price coordination only. His work has sought to explore other types of coordination. He did so through studies of entrepreneurs, which he called the interpretive agents seeking to develop new ways of understanding the world: "profit opportunities are not so much like road signs to which we assign an automatic meaning as they are like difficult texts in need of a sustained effort of interpretation" (Lavoie, 2015, p. 59).

[9] These exemplary goods have affinities with Weberian ideal types (Dekker, 2016, p. 107).

Such understanding led Virgil Storr (2004) to study different entrepreneurial spirits in the Bahamas. He demonstrates that a particular way of understanding entrepreneurship based on the narrative of the pirate is central to the local entrepreneurial spirit, which makes entrepreneurs see a range of opportunities and ways of making a deal that are very different from traditional notions of market entrepreneurship. The pirate, Storr argues, is a kind of ideal-typical entrepreneur, the reliance on which structures later actions. In more recent work he has studied the entrepreneurship of local community leaders in the aftermath of hurricane Katrina (Storr et al., 2015; Storr & Chamlee-Wright, 2010). Emily Chamlee-Wright (1997) similarly started out by studying the entrepreneurial culture in a non-standard context, in her case women in Ghana. Her methodology is based on interviews through which she seeks to explore how the actor's perspectives of opportunities are structured by cultural frames. Culture within these works serves as an interpretive lens through which entrepreneurs perceive the world, and through which coordination in markets is made possible. Furthermore, this work emphasizes that shared cultural frames are even more important in the absence of formal institutional structures, and thus it acts as a complement to some of the work in institutional economics discussed above.

A slightly different approach that relies on the use of ideal types is the work of Roger Koppl (2002) on 'big players'. He explicitly develops Schutz's notion of the levels of anonymity on which social actors rely. He argues that in most markets there will be a relatively high degree of anonymity, and thus the actors can form their expectations based on anonymous ideal types. But in markets with a few big firms (oligopoly), or in markets with one big player such as the central bank, economic actors will develop quite sophisticated interpretations of the likely actions of some other actors, since in forming their expectations, small actors will have to take into account the power of the big players to make idiosyncratic moves. For Koppl, the ability and willingness of certain, usually big, economic agents to act discretionary as opposed to following a set of rules results in a distortion of expectations among other agents, and in attempts by the latter to align as close as possible to the actions of the big player. He applies this to financial markets and the elaborate attempts to

predict the likely course of action, typically regarding the interest rate, of the central bank.

Koppl also raises the possibility that expectations need not be identical to be coordinated. In such a case, coordination may be sustained even as the mental models are not shared but differ. Such 'false mental models' may thus nevertheless lead to a situation where "the players are oriented to different visions of the future and yet neither party is ever disappointed" (Koppl, 2002, p. 91). While such coordination remains a possibility, the reality of the social world nevertheless is predominately based on the shared models due to a mix of invariant human biological traits and 'universal' social principles emerging from human interaction.

The insight emerging from these different economic studies is that coordination in markets depends on shared frames of reference as much as on prices. There are no simple incentives out there; actions arise from shared understandings of what is likely to happen and what is expected of others. In the case of contested commodities, the buying and selling of a certain artifact is regarded improper and hence the 'incentive' to do so ignored. In other instances, entrepreneurs discover new opportunities and create incentives for others to follow. In yet others, there is no single signal guiding the way, since the success of a projected action depends on what others will do, and hence mutual coordination is crucial. Even more than in sociological studies, economists emphasize ways in which shared understandings emerge and are transformed. In line with Beckert (2016), they demonstrate that the future is not merely uncertain—it is actively shaped through market coordination to overcome that uncertainty. The one building block virtually absent from these studies is that of different provinces of meaning.

5. The Meaning of Incentives, a Constructive Research Program

Above we demonstrated how a number of Schutzian themes are present in recent empirical work in both economics and economic sociology. It has been acknowledged for some time now that economics seems to be moving away from theory toward more applied or empirical work (Hamermesh, 2013), and this might foster new avenues for exchange with neighboring disciplines. The exchange with psychology in

the form of behavioral economics is a famous example of such an exchange. Meaning in this particular exchange is largely ignored in favor of understanding underlying psychological mechanisms. This need not be the case, as two of us have argued elsewhere, since other combinations of psychology and economics are possible (Dekker & Remic, 2019) that leave more space for a focus on meaning. But incentives dominate the economic literature. As one of the popular books on the subject has it: "An incentive is simply a means of urging people to do more of a good thing and less of a bad thing" (Levitt & Dubner, 2005, p. 17). What could be simpler?

It is the assumption of the simplicity of incentives that is undercut in the empirical literature we have explored. By showing how shared meaning structures coordinate the actions of individuals it becomes clear that it is the meaning of incentives, and not their inherent or natural force, that gives them power. There is nothing natural about the symbolic value of certain photoshoots that makes them prestigious and hence so attractive that they do not require a payment, as in the study by Mears (2011). Nor is there anything inherent in lower prices that make them suspicious, yet in particular settings such as primary art markets, lower prices might function as a reason not to buy (Velthuis, 2005). Similar effects can be observed in the examples cited by Frey and Jegen (2001) on intrinsic motivation, where the introduction of monetary compensation works as a deterrent rather than an attractor for certain behavior. A classic example is the way in which the amount of blood donated drops after payment for it because the prosocial meaning of the act is undermined. Another example is the way in which parents interpret a fine for picking up their children late from the daycare center. Without this fine more parents were on time, as they considered this their duty; the fine was interpreted as a price for being late which removed the duty of being on time (see Gneezy & Rustichini, 2000).

This could amount to little more than a note of caution: beware of the simple use of (monetary) incentives. The Schutzian insight is that the problem goes deeper than it appears at first sight. Since particular signals are interpreted in relation to the existing intersubjective meaning structures, the real problem facing the individual is how to interpret a particular signal. This is a cognitive or knowledge problem, not merely a problem of fixing the incentives. Foss and Garzarelli (2007)

drive that point home in an article critical of the way mainstream economics deals with incentives:

> "Mainstream conceptions of institutions such as firms and markets ignore the positive cognitive role that such institutions play, that is, their ability to coordinate different expectations through time is downplayed, and all attention is focused on how these institutions may align incentives" (p. 795).

In other words, by thinking in terms of incentives we take for granted that these can be meaningfully interpreted by actors in the first place, and we ignore the institutional structure (in economic language), or the shared meaning (in Schutzian language), that makes such interpretation possible in the first place. Our first conclusion is that understanding the meaning structure within which incentives operate is essential to understanding what effects they will have.

In the previous sections we have seen that in many empirical studies a methodology of ideal types is implicitly or explicitly developed in order to make sense of the actions of others. This was explicitly done in Koppl's (2002) study of big players in financial markets, the interpretation of whose actions involve great cognitive exertion by many of the other players in the market, all the while they are being content to accept a fairly simple representation of the other (small) players in the markets. It was explicit in work on the performativity of rational-actor models in financial markets by Miyazaki. In the economic histories we mentioned we also saw that particular groups (as well as individuals) are modeled (often in a game-theoretic setting) as being of a certain type with a stylized motive. This allows economic historians to analyze the various coalitions seeking to arrive at a beneficial outcome. Their effort is interpretive in the sense that they try to model the decision situation for the historical actors, but they utilize an implicit methodology of ideal types to analyze large-scale historical developments.

Although not all studies do this explicitly, it does open up the possibility for more sophisticated thinking about incentives and institutional change. Many economic histories show how conflict situations are reinterpreted so that cooperative

coalitions can be built, which support new institutions that shape the future expected behavior of the actors (obviously this can also happen the other way around so that others are no longer regarded as potential coalition partners). On a smaller scale, Elinor Ostrom (1990) studied this issue for communities seeking to overcome common-pool resource problems. This underlines the fact that thinking about ideal types, and the mutual understanding of actors, shapes the (potential) outcomes of a situation. Non-cooperative situations can turn into cooperative ones. This point is emphasized within the economic sociology literature, which often demonstrates the intricate ways that competition and cooperation go together in social interactions, as well as the importance of certain social logics with clearly defined roles (Steiner, 2010; Thornton et al., 2012). This leads to our second conclusion: incentives are inextricably linked to the roles of different actors, which can be made intelligible through the methodology of ideal types.

The third building block is that of the coordination process that happens through project formation. Economics has a long tradition of studying entrepreneurship in a variety of settings and, as we demonstrated above, increasingly in non-market settings. Within the Schutzian framework, every action based on a project or plan has entrepreneurial aspects. But, when shared meaning structures are absent or disrupted, empirical studies of entrepreneurial behavior and the resulting coordination or conflicts can be highly illuminating of how individuals understand the incentives and opportunities in the world around them. Recent studies in the sociology of finance have highlighted how institutions help shape expectations about highly uncertain futures, and Beckert (2016) is a deeply Schutzian project, as demonstrated above. Increasingly, economic sociologists have studied how uncertainty is reduced through a variety of judgment devices (Karpik, 2010), and other types of coordination which act as guideposts in uncertain terrain.

Economists typically focus on prices as the primary institution providing guidance for future economic action, but we have highlighted how they increasingly are paying attention to other institutions that help structure expectations about the future. Again, this work is in part stimulated by game-theory, where prices are not as central as elsewhere in economics. Moreover, Storr and Chamlee-Wright's (2010) work on post-disaster recovery shows that entrepreneurship can be social and

coordination can take place through announced actions by leaders. That work also shows how particular individuals are important in shaping the future, a point that also comes out clearly in Koppl's (2002) work on big players, such as central banks. This leads us to our third conclusion, which is that through the study of entrepreneurship we can grasp how different actors understand and imagine possible futures of the world differently. In the open-ended situations that entrepreneurs face, we can study how they interpret the limited cues and signals about the future and base their projects on these. Limited coordination or discoordination are likely to occur. Empirical studies are crucial here since the way uncertainty is reduced, or judgments are formed, differs in different markets.

The fourth building block we identified in the second section–multiple provinces of meaning–is reflected in the empirical and conceptual literature in economic sociology. It underlies the theory of the circuits of commerce, work on the different worlds of justification by Boltanski and Thévenot (2006), and the literature on institutional logics. It is also reflected in empirical work in economic sociology, whose purpose is often to identify the particular province of meaning that structures a particular market. Within economics, however, it is harder to find explicit instances where authors draw on the idea of different provinces of meaning. The exception is a paper on surrogate motherhood by Kuchař (2016), which explicitly deals with the different logics attached to the idea of motherhood and to the exchange on markets. While nobody is in favor of selling and buying babies, reconciling these different logics occurs when the idea emerges that not the baby is bought and the mother receives money for it in exchange, but instead the womb is rented out, for which the surrogate mother can be compensated.

Empirical work can be done in economics on related issues such as 'contested commodities' (Radin, 1996). These are goods that, traditionally, have not been exchanged on markets, or whose commodity status is contested. For these goods it is clear that competing understandings of the good are involved, linked to different provinces of meanings and institutional structures. This is more relevant because markets are being designed for some of these contested commodities, such as kidneys and school enrollment. These mechanisms, which often mimic market mechanisms (although only partially), are used for the distribution of particular scarce goods for

which normal market exchange is not acceptable. There is awareness that this project runs into what Alvin Roth (2007) has called 'moral repugnance', but there is a poor understanding that this moral repugnance is not a natural repugnance but instead a competing set of meanings associated with particular artifacts. Just like monetary incentives might upset existing motivations, so here market-like distribution will upset existing meanings and associations. If we wish to make these markets function well, or to criticize them intelligently, an understanding of the multiple provinces of meaning is necessary. The fourth conclusion we draw from this is that markets are typically their own province of meaning which is (potentially) in conflict with other social processes containing rival and complementary meanings. To understand why markets and (monetary) incentives are sometimes not accepted, we need to understand these interrelations.

6. Conclusion

This chapter demonstrates the convergence of recent empirical approaches in economics and economic sociology along Schutzian lines. It identifies four Schutzian building blocks: Verstehen, methodology of ideal types, coordination of projected plans, and provinces of meanings. Except for the fourth, they all appear in recent economic work, especially in new institutional and Austrian economics; all appear in recent work in economic sociology.

This does not mean that there is a coherent or unified new approach; however, there is a promising avenue for future research at the intersection of economic sociology and these branches of economics. We have shown that this can be a constructive research program that can contribute to a central issue in economics–how incentives work. The approaches discussed illuminate: (i) why certain incentives are perceived as powerful reasons for action, while others are mostly ignored; (ii) why incentives are typically tied to certain social roles that can be identified through ideal-type analysis; (iii) why situations of high uncertainty are useful in studying how actors make sense of the world and how an uncertain future is understood by actors; and (iv) why monetary rewards and market exchange sometime provide the wrong type of incentive.

To pursue this line of research, it is valuable to recognize the important contribution of Schutz. Elements of his work can be recognized within recent empirical work in both disciplines, and his work can serve as a theoretical and methodological foundation for these new approaches. His interpretative social science, focusing on mutual coordination by many individuals operating in divergent provinces of meaning, is not only appropriate for the modern plural world but can also help us analyze small-scale (micro) interactions and large-scale (macro) interactions. It also combines elements of individual choice and subjective valuation common to economics, with notions of norms, shared (sub)cultures, and different domains in society, that are found in sociology.

Incentives indeed matter, but they do not matter in and of themselves. We have to pay attention to the general dynamics of intersubjective meaning that enable economic actors to interpret signals. The way a particular signal is interpreted, so that it comes to be understood as an incentive or a disincentive for action, is crucial for understanding social interaction. Within this research program, incentives are not objective facts of the social world; they are thing understood by actors.

2

Three Accounts of Intrinsic Motivation in Economics: A Pragmatic Choice?*

1. Introduction

In recent decades, one of the cornerstone concepts in the literature at the intersection of economics and psychology has been intrinsic motivation. The concept of intrinsic motivation was developed in psychology as part of a critique of behaviorism and psychologists generally define it as the motivation for activities that are not means to some further end but *an end in itself* (e.g., Deci, 1971). Contrary to behaviorism's straightforward positive relationship between reward and performance,[1] the idea of intrinsic motivation is built around the hypothesis that when the activity is pursued as an end in itself rewards may have hidden costs and therefore be ineffective–or even deterrent–for the subsequent performance (Lepper & Greene, 1978). For economists, this hypothesis represents an especially intriguing puzzle since it seemingly violates the law of supply by implying that an increase in price may lead to less quantity supplied. The classic example is blood donation, where the claim is that introducing monetary compensation will result in less donations (Titmuss, 1970; Mellström & Johannesson, 2008). Some of the empirical applications of this

* This chapter has been published as "Three accounts of intrinsic motivation in economics: a pragmatic choice? *Journal of Economic Methodology,* forthcoming.

[1] Rather than *reward*, behaviorists employ a technical term *positive reinforcement,* which refers to any stimulus that increases the probability of a response. Conversely, negative reinforcement–akin to punishment–refers to a stimulus that decreases the probability of a response (Skinner, 1953; for a discussion see McCullers, 1978, pp. 12–16).

hypothesis that helped to firmly establish it in the economic literature can be found in the analysis of tax compliance (Pommerehne & Frey, 1992), environmental morale and public virtue (Frey & Oberholzer-Gee, 1997), principal-agent relations (Bénabou & Tirole, 2003), prosocial behavior (Bénabou & Tirole, 2006), artistic production (Throsby, 2000), work in the care sector and public services (Nelson, 2006), and within the discussion on piece-rate pay vs. fixed wages (Gneezy & Rustichini, 2000b). Apart from accounting for negative effects of incentives, intrinsic motivation can also have a role as a sorting mechanism for workers (Bohnet & Oberholzer-Gee, 2002; Heyes, 2005; Prendergast, 2008; for a criticism, see Nelson & Folbre, 2006).

While intrinsic motivation has been recently increasingly discussed in economics (notable examples are Le Grand, 2010; Bowles & Polanía-Reyes, 2012; Luttmer & Singhal, 2014; Festré & Garrouste, 2015; Bénabou et al., 2018; Besley & Ghatak, 2018; Sugden, 2018; Dellavigna & Pope, 2018), a proper systematic account of the theoretical underpinnings is still missing. Empirical work often proceeds based on ad hoc applications that only superficially delineate the underlying psychological mechanisms. Ashraf et al. (2014), for example, code work motivation by hair stylists as intrinsic if the subjects report they are doing it to "make people look nice" and extrinsic if their reason is to "make money" (p. 4). This reflects the impression shared by many economists that intrinsic motivation refers to motivation that is negatively correlated to (or independent of) financial incentives. However, this is at odds with the above-mentioned general definition from psychology. Both making people look nice and making money may be seen as separate pursuits with respect to the activity of cutting hair. This implies that in both cases the activity is not performed *as an end in itself*. Furthermore, the conceptual issues are sometimes entirely sidestepped. For example, DeVaro et al. (2017) do not provide a single definition of intrinsic motivation and do not cite any foundational research; they proceed with their empirical analysis based on a tacit assumption that the meaning and significance of the term intrinsic motivation is self-evident and uncontested.

The aim of this chapter is twofold. In Section 2 I will first demonstrate that economists who imported the concept of intrinsic motivation from psychology into economics have been–often unknowingly–drawing on competing underlying psychological theories. In psychology, the counterintuitive hypothesis that extrinsic

incentives might have detrimental effects was in the early 1970's developed, tested, and published independently by three research groups (Deci, 1971; Kruglanski et al., 1971; Lepper et al., 1973). This work initially culminated in a joint edited volume *The Hidden Cost of Reward* (Lepper & Greene, 1978) with contributions from all three groups. After that, however, their research programs diverged. The volume remains one of the key references for the detrimental effects of incentives, but what gets overlooked is that the three groups that contributed to it in fact offer distinct theories of intrinsic motivation.[2] As a result there is not only a fundamental dispute with behaviorists about the correct theorizing and interpretation of the empirical findings concerning the relationship between rewards and subsequent performance; the research program of intrinsic motivation is divided internally as well. As Kruglanski et al. (2018) assert, "despite *identical nomenclature* [our models] refer to *entirely different explananda*" (p. 167, emphasis added). The economic literature mostly neglects this fact, and hence has imported a contested notion rather than a unified concept. Section 3 discusses what is at stake when considering empirical work and incentive-based policy interventions.

The second aim of the chapter is to argue (in Section 4) that the discussion about the different accounts of intrinsic motivation has important consequences for the recent methodological debate concerning the relationship between neoclassical and behavioral approaches in economics. Most accounts of behavioral economics assume a relatively straightforward marriage of behavioral insights and rational choice theory. Sometimes this is even made explicit, and a 'pragmatic' stance is adopted suggesting that more realistic (i.e., psychologically plausible) behavioral assumptions can be applied (or added-on) whenever they prove necessary (Chetty, 2015; Angner, 2019). Given that within the pragmatic approach psychological

[2] Two of these three psychological accounts have been employed by economists and will be discussed later. The third one stems from the initial work by Kruglanski et al. (1971) and has recently been restated as a *means-ends fusion* model of intrinsic motivation (Kruglanski et al., 2018; see also Harackiewicz & Sansone, 2000). While this approach has recently received some attention in the management literature (e.g., Foss & Lindenberg, 2011; Woolley & Fishbach, 2018) there has not yet been any systematic attempt to adopt it within economics. Therefore, it will not be further considered in this chapter.

plausibility explicitly corresponds to realisticness of the assumptions, this chapter argues that a pragmatic stance towards choosing among different psychological insights when doing economics is often misguided because of the fundamentally different psychological underpinnings of the resulting integrated accounts. I demonstrate that such pragmatic application of intrinsic motivation insights to the economic analysis has resulted in a curious case of importing competing, contradictory and mutually exclusive theories of intrinsic motivation from psychology into economics. It gave rise to multiple accounts of the same phenomenon. Arguably, the pragmatic attitude is in danger of ending up as a case of a particularly bad kind of ad hocness, leading to results that are closer to a set of descriptions than to real explanation.

2. Three Accounts of Intrinsic Motivation in Economics

Within economics we can distinguish three distinct explanatory accounts of the detrimental effects of incentives on intrinsic motivation: (i) *the motivational account* (Frey, 1997); (ii) *the signaling account* (Bénabou & Tirole, 2003, 2006); and (iii) *the allocational account* (Holmström & Milgrom, 1991). They differ with respect to the underlying causal mechanism, but the full extent of their systematic differences is poorly, if at all, recognized in the literature. For example, Bowles and Polanía-Reyes (2012) include (i) and (ii) in their comprehensive paper surveying the literature addressing the problem of substitutability of incentives and social preferences. But they nevertheless discuss the topic of intrinsic motivation exclusively in terms of the psychological mechanism underlying (i).[3] Although Kreps (1997) and Festré and Garrouste (2015) include (iii) as an alternative account of the problem of the hidden costs of rewards, and Fehr et al. (2001) discuss it explicitly in terms of intrinsic

[3] They justify their decision to do so by arguing that "it is likely that [in (ii)] more than one mechanism is at work" (Bowles & Polanía-Reyes, 2012, p. 398). Furthermore, they conceptualize (ii) in strictly informational and strategic terms without recognizing that the underlying logic of this account stems from a distinct psychological theory of intrinsic motivation that is different from the one they mention in relation to (i).

motivation, this approach has been somewhat lost in the recent scholarly work as a distinct way of thinking about the problem of intrinsic motivation with its own psychological underpinnings. We will now look at (i) to (iii) in turn.

2.1 The Motivational Account: Are Your Psychological Needs Met?

The motivational account is based on the motivation crowding theory (Frey, 1997; Frey & Jegen, 2001) that posits a possible downward sloping supply curve for effort by introducing the crowding out effect as an additional force working in the opposite direction than the relative price effect. In other words, an external intervention in the form of a reward may crowd out intrinsic motivation and thus result in less rather than more effort.

The concept of intrinsic motivation was for the first time explicitly introduced into economics by Stroebe and Frey (1982). They follow psychologists in characterizing intrinsic motivation as a motivation for performing an activity when "there is no reward except the activity itself" (Stroebe & Frey, 1982, p. 121). In such a case, they argue, motivation is directed to the satisfaction of a psychological need. Crucially, as they point out, the self-interest postulate is not contradicted and thus the utility calculus essentially remains the same. What gets tweaked is the content of the personal utility function that in the motivational account acquires an additional expression in terms of the satisfaction of psychological needs, in particular the inner need for autonomy. In the motivational account, intrinsic motivation is modelled as a particular type of incentive for achieving a desired psychological state (Frey & Jegen, 2001, p. 590). The state of being intrinsically motivated is associated with improved well-being. And because improved well-being, in turn, is associated with higher productivity and willingness to exert effort, fluctuations in intrinsic motivation have a clear economic relevance.

The motivational account is often portrayed as a critique of the standard economic account based on the narrow notion of a rational economic agent. However, Frey's earlier work reveals that he conceived it not as a critique but as an upgrade in the spirit of Becker's economic imperialism (Romaniuc, 2017). In his initial foray into psychologically informed economics, Frey explicitly asserts that economic man and psychological man share the same utilitarian roots, both "assumed to respond

systematically to positive and negative incentives" (Stroebe & Frey, 1980, p. 120), and that both disciplines would largely agree that "man tries to behave rationally, i.e., chooses the action alternatives which are likely to be associated with the highest overall utility" (*ibid.,* p. 127). Stroebe and Frey (1980) point out that Becker's economic approach to human behavior is a largely content-less theory and argue that psychology provides the needed empirical content that makes it possible to develop testable hypotheses that are based on more than monetary income alone.

To fully make sense of this model, we therefore need to understand the psychological theory underlying the motivational account: the self-determination theory (henceforth SDT; Ryan & Deci, 2017). The theory proposes that people have an inner need for autonomy and self-determination. This need is met when they perform an activity autonomously without a sense of being controlled from the outside: namely, when they are intrinsically, rather than extrinsically, motivated. SDT builds on the previously developed notion of the 'perceived locus of causality' (DeCharms, 1968), according to which it matters whether the reward is perceived as external (that is, controlling), or internal (that is, when the person perceives herself as the origin of her behavior). The presence of a reward that is perceived as controlling will negatively affect the sense of autonomy and shift the locus of control to the outside, thus substituting extrinsic for intrinsic motivation. Such crowding out is not categorical, however, but happens on a scale representing a continuum between autonomously motivated and controlled behaviors (Ryan & Deci, 2017, p. 14).

Autonomy, referring to "the need to self-regulate one's experiences and actions" (Ryan & Deci, 2017, p. 10), is in SDT conceptualized as one of the three basic innate psychological needs, the other two being the need for *competence* ("to feel effectance and mastery"; *ibid.*) and the need for *relatedness* ("feeling socially connected"; *ibid.*).[4] The motivation directed at fulfilling them is labeled intrinsic

[4] This reflects the fact that SDT consists of a set of 'mini-theories', one of those being *cognitive evaluation theory*, which focuses on the issues of autonomy and control (see Ryan & Deci,

motivation. The crucial part of this theory–and also the core of its critique of behaviorism–is that fulfilment of the three needs cannot be elicited by external stimuli. Even more, *any attempt to do so will be harmful* because it will undermine intrinsic motivation and thus prevent the needs for autonomy, competence and relatedness to be met. Importantly, this is not to say that rewards have an *immediate* detrimental effect. As Ryan and Deci (2017) emphasize, "the scientific problem here is specifically their impact on the maintenance of intrinsically motivated behavior over time" (p. 127). The experimental evidence suggests that when a reward is first introduced and then taken away, the initially intrinsically motivated effort does not return to its before-treatment baseline.[5] In other words, rewards lead to long-term corrupting effects. The shift of the perceived locus of control, brought about by extrinsic incentives, thus causes long-lasting damage to the psychological well-being of the individual.

It is important to stress that the three psychological needs of autonomy, competence and relatedness are in SDT understood as "nutrients that are essential for growth, integrity, and well-being" and, as such, are "*objective* phenomena in that their deprivation or satisfaction has clear and measurable functional effects, effects that obtain regardless of one's subjective goals or values" (Ryan & Deci, 2017, p. 10,

2017). Cognitive evaluation theory was the first one to be formulated within this research program (e.g., Deci, 1975), and is what Frey primarily based his account on. While autonomy alone would arguably be enough to present the core of the motivational account, it is instructive to include a discussion of the other two needs as well, since the recent work within this account has been drawing on all three of them. Competence has been emphasized by economists discussing craftsmanship and artistic creativity (e.g., Klamer, 2016). Frey himself has been incorporating the need for relatedness in his more recent work on happiness (e.g., Frey & Stutzer, 2002; Frey & Gallus, 2012), but it arguably also applies well to the situations governed by social norms. Empirical work in economics based on SDT sometimes incorporates all three needs (e.g., Cnossen et al., 2019; Nikolova & Cnossen, 2020).

[5] In a classic demonstration of this effect participants in two groups are assigned a task of solving puzzles (Deci, 1971). The experiment consists of three rounds, the only difference being that the treatment group receives a performance-contingent reward in the second round (but is told that there is no more money available for the third round). The results show that participants in the treatment group demonstrate lower levels of motivation in the third round compared to the control group.

emphasis in original). These effects are akin to the effect that vitamin C deprivation will have on the body regardless of whether one believes in it or not. Furthermore, this functional relationship is an adaptive trait of the organism. As Ryan and Deci (2017) emphasize, "the assumption of SDT is not that social-contextual events 'cause' intrinsic motivation–on the contrary, intrinsic motivation is understood as an evolved and inherent human propensity. The ultimate causes of intrinsic motivation lie in the selective advantages this propensity yielded in human prehistory" (p. 124). These advantages have to do with the inherent satisfaction brought about by an intrinsically motivated activity, which contributes to human learning and overall flourishing. Some economically relevant examples are experimentation and voluntary action (driven by the need for autonomy) that brings about innovative ideas; sustained play and persistence (driven by the need for competence) that have an important role in learning various skills and getting things done; and forming and sustaining communities (due to the need for relatedness) that is crucial for cooperation.

This discussion enables us now to provide a more detailed account of why we should expect voluntary blood donations to fall after an extrinsic reward has been introduced. The motivational account predicts that an extrinsic incentive will have measurable negative psychological effects on the individual person by depriving her of the ability to act autonomously, that is, out of intrinsic motivation. The psychological need for autonomy will not be met. But since, as we saw earlier, crowding out is not categorical, the extent of it will be determined by the strength of the extrinsic incentive. This implies that a large enough incentive will completely crowd out intrinsic motivation to donate blood out of an inner impulse to act altruistically. Once the intrinsic motivation to donate blood is crowded out, blood donation would be completely dependent on the extrinsic incentive. In the short term, this could even mean that an adequate total supply of blood for transfusions might be secured by offering a high enough monetary reward. However, the motivational account also implies that such crowding out of intrinsic motivation will have important long-term costs in the form of negative effects on the psychological well-being of donors.

The motivational account thus puts at the center the question of personal psychological well-being, specifically the well-being that stems from the individual ability to act in a self-determined way. It is therefore not surprising that in terms of the explananda more specifically the motivational account prominently seeks to explain work morale and virtuous, moral, or prosocial behavior (Frey, 1997), which are the areas where the supply of activity by the agents is considered to be positively related to the ability to act out of inner impulses. The scope of the account is thus very much tied to the psychology of individual actors and the more general societal regularities are explained to the extent that they can be considered as aggregates of individual psychological reactions.

2.2 The Signaling Account: You Are What You Do

The signaling account models the interactive process in which actions are potentially perceived by others–or, in the case of self-signaling, by the agent himself–as displaying intrinsic motivation. It thus translates a motivational problem into a signaling problem. Intrinsic motivation is modelled as an incentive to achieve a desired reputational effect: being intrinsically motivated is a salient signal that people try to convey either by refusing to accept rewards or by diminishing the effort in the presence of rewards.

In the earlier incarnation of this account, Bénabou and Tirole (2003) developed a model of crowding out of intrinsic motivation based on the principal-agent framework with asymmetric information, where the principal has some private information about either the task or the ability of the agent to perform that task. Rather than relying on direct mapping from stimulus (i.e., incentive) to response (i.e., crowding out of intrinsic motivation), Bénabou and Tirole model the situation in strategic terms based on the information that is available and conveyed at any given moment. The agent is assumed to respond and adapt to what happens. In such a situation, explicit incentives will signal possible 'bad news' to the agent: perhaps the task is boring; or perhaps the principal does not trust him. This will change the agent's information structure about his own abilities and, consequently, undermine his self-confidence. The underlying mechanism is based on an assumption that "naturally, the agent will undertake the task only if he has sufficient confidence in

his own ability to succeed, and in the project's net return" (Bénabou & Tirole, 2003, p. 491). However, recasting intrinsic motivation in terms of a rational calculation about the probability of success is the Achilles heel of this initial attempt, for it seemingly implies that what makes one intrinsically motivated is the increased chance of an extrinsic reward.

The signaling account was given its mature shape in Bénabou and Tirole (2006). They gave up on their initial attempt to model intrinsic motivation directly. Instead, they build on the assumption that people buy self-esteem by signaling some desired social trait, such as being intrinsically, rather than extrinsically, motivated to do something good for others. Crucially, the *state* of being intrinsically motivated does not play the decisive role in the deterring effects of incentives. The key assumption of the signaling account is the fundamental uncertainty with respect to the underlying motivations. This uncertainty applies to both the audience's ability to observe the true motivation of the agent, as well as to the agent's ability to directly know his or her own underlying motivations. Therefore, intrinsic motivation is not just an *unobservable*; it is also an *unknown*. As such, it always has to be inferred from actions. Thus, as Bénabou and Tirole emphasize, what matters in their model is that the "agents *value being perceived, or perceiving themselves,* as having high [intrinsic motivation]" (2006, p. 1657, emphasis added). This valuation–and not the motivational state as such–is what in the model determines the outcomes. Crowding out potentially happens due to reputational concerns regarding the inferences other people (or oneself, in the case of self-signaling) might draw about the supposed motivation for a particular act.

To capture that, Bénabou and Tirole propose a trichotomy of intrinsic, extrinsic, and reputational motivation. The latter relates to "what a person's behavior says about him or her, which will depend on the informational and economic context, including what others are doing" (Bénabou & Tirole, 2006, p. 1658). While the person in the model derives marginal benefit from all three factors in the mix, only the effort as the sum of the three motivations can be observed, not the particular mix. Therefore, the practical problem of strategic interaction is one of signal-extraction, where "rewards act like an increase in the noise-to-signal ratio, or even reverse the sign of the signal" (*ibid.*, p. 1654). Incentives create noise in the inference

about the underlying motivation because the audience will question whether the person did something for money rather than out of intrinsic motivation. This will, in turn, reduce the incentive to perform the activity in the presence of explicit incentives. As Bénabou and Tirole argue, "the presence of extrinsic incentives spoils the reputational value of good deeds, creating doubt about the extent to which they were performed for the incentives rather than for themselves" (*ibid.*, p. 1654). Thus, payments for blood donations do not have direct corrupting psychological influence, as the motivational account would argue. What might lower the overall amount of blood donated is that payments remove the possibility for signaling virtue through blood donation. The person wants to be perceived as intrinsically motivated to donate blood, but because in the presence of a reward she cannot be sure this is indeed the right reason, she will go further than necessary to increase the chances of convincing others or herself and thus achieving the salient effect. This will result in reducing donations when the person starts to worry that getting rewarded will make her look as if she is doing it for the money.

The signaling account of intrinsic motivation in economics is a formalization of the *overjustification hypothesis* first put forward by psychologists Lepper, Greene and Nisbett (1973). This approach explains intrinsic motivation in negative terms: in the absence of clear extrinsic rewards to account for given behavior, the person will attribute it to intrinsic motivation. Conversely, if rewards are clear and psychologically sufficient, behavior will be attributed to extrinsic motivation. The rather awkward name of the hypothesis refers to a situation where behavior initially attributed to intrinsic motivation will feel 'overjustified' once a reward is introduced. The presence of a reward will cause the person to reevaluate her engagement in the activity; and since now there is a clear link between activity and reward, the person will infer that the activity could not have been that intrinsically interesting after all. This will prompt her to reduce the effort when the reward is withdrawn again. Motivation is thus conceptualized as a kind of knowledge (Kruglanski, 1978, p. 24). For example, when children are given a reward for drawing–which is presumed to be intrinsically interesting–they will start to see the activity as a way to get the reward. This will lead to lower effort once the award is not presented, because children now simply have a different understanding of the reasons to do the activity. Important

for our discussion is that the attributional mechanism does not depend on any particular nature of intervention: it is not about money as such. As Lepper, Greene, and Nisbett (1973) argue, "an overjustification effect is predicted for any situation which results in an extrinsic attribution where previously intrinsic interest was the only salient attribution" (p. 130).

Theoretically, this psychological approach is a combination of self-perception theory (Bem, 1972) and attribution theory (Kelley, 1967). According to the former, "individuals come to 'know' their own attitudes, emotions, and other internal states partially by inferring them from observations of their own overt behavior and/or the circumstances in which this behavior occurs" (Bem, 1972, p. 2). In other words, we cannot know ourselves directly, but only through indirect inferences. Attribution theory is a cognitive theory concerned with the processes of inference in general (Kruglanski, 1978, pp. 24–25), but with regard to intrinsic motivation it argues that people explain behavior–their own or that by others–by attributing it to either internal or external reasons. Internal attribution means that one thinks of oneself as being the kind of person that behaves in the observed way; and external attribution means that the causes of behavior are attributed to some situational factors. To be intrinsically motivated to donate blood then means that one perceives oneself to simply be the kind of person that donates blood without having to be paid for it. However, self-perception theory says that that can never be directly known. In the presence of an extrinsic reward, it will not be possible to clearly infer that one indeed is such a person, for even in the case of strong self-signaling it is impossible to be sure–perhaps the person is simply repressing the fact that she is doing it just for the money. Therefore, also in the case of self-signaling it is only reputation that is observable. In this sense, self-signaling is actually a model of self-deception (Mijovic-Prelec & Prelec, 2010) and not a model of the underlying motivation.[6] The combination of self-perception and attribution theories reveals the dynamic between,

[6] Gold (2019) provides a similar observation with respect to the (in)ability of the signaling approach to explain the claims about the corrupting effects of commodification: signaling explanations "do not involve any change in valuations" (p. 173).

on the one hand, our inability to know ourselves and, on the other, the ways in which we nevertheless get to define ourselves by the actions we perform. In other words, actions serve as signals about the possible reasons and motivations for behavior. It is now obvious that this theory remains closer to behaviorism than the motivational account.

These signals are important for social interaction, since they enable the agents–individuals, but potentially also organizations–to convey and infer information that guide further action and allow coordination. Rather than what happens at the individual psychological level, of interest here are the consequences that reading other people's actions and responding to them has for social interactions. The scope of the signaling account is thus rather indistinct, extending from the micro perspective of the principal-agent interactions to the larger-scale institutionalized rules that structure the incentives and serve as sources of information for strategizing about the possible actions of others and how certain actions will be perceived in the context of interaction. Illustrative examples thus encompass any situation where it is desirable to be perceived as doing something out of intrinsic motivation, such as contribution to public goods, donating to non-profit or charity, being altruistic, reciprocating a friendly action, refraining from imposing negative externalities on others, and more (see Bénabou & Tirole, 2006). Importantly, the scope is determined by the socio-cultural meaning and desirability of these acts, and not by inner psychological drives.

2.3 The Allocational Account: What Are You Really Rewarded For?

Contrary to the first two, the allocational account developed by Holmström and Milgrom (1991) is not explicitly about the concept of intrinsic motivation. However, the fact that it does effectively represent a distinctive account of it has not gone unnoticed among economists (see, e.g., Fehr et al., 2001). The account focuses on the actor's allocation of effort among the various aspects of the task to be performed. In the model, these aspects are called dimensions. The key factor is that the different dimensions are, for various reasons, differently incentivized. Intrinsic motivation can

thus be modeled as a motivation for activity that is either not explicitly incentivized or is incentivized less, relative to other dimensions of the same task.

For example, the job of a university professor consists of dimensions such as research, teaching, administrative tasks, and networking with stakeholders outside of academia. On top of that, those dimensions can be further divided into subdimensions. Research activity consists of publications, generating new ideas, and serving the scientific community by doing peer review. However, it may be that it is harder or more costly to measure and evaluate the performance on some of these dimensions. In such a case it is likely that incentives will be tied to the dimension where performance standards can be more easily specified and monitored. However, when the employer starts to explicitly incentivize a particular dimension, employees will start to focus on it at the expense of the others. The provision of explicit incentives for a measurable dimension will raise the opportunity costs of exerting the effort on the dimension where performance is hard to measure: by engaging in the latter, the employee is foregoing the additional money he could be earning by focusing solely on the former. He will then reallocate the effort to the measurable dimension in order to maximize the financial benefits. Consequently, the motivation to exert effort on the non-measurable dimension will appear to be crowded out by some extrinsic incentive. If professors are evaluated and rewarded primarily based on their publication record or their administrative effort, this may come at the expense of their engagement in teaching.

Such reasoning, however, may still seem like the standard microeconomic logic regarding the substitutability of inputs. What nonetheless makes this model specifically about intrinsic motivation is the assumption–made explicit by Holmström and Milgrom in a footnote only–that agents "are motivated to [...] supply [some] inputs even without incentive pay" (1991, p. 32n9). In other words, the absence of explicit incentives for a particular dimension does not mean that the agent will not exert herself. However, that effort can get completely reallocated to another dimension if that dimension is clearly incentivized. Holmström and Milgrom reason that it may, therefore, be better to reduce explicit incentives for measurable dimensions if the desired outcome is some effort on the unmeasurable component–which is analogous to reasoning that it is better not to introduce extrinsic incentives

lest intrinsic motivation would get crowded out. However, the point is not that incentives as such are harmful. According to the allocational account, some form of incentive contract is always needed, since Holmström and Milgrom assume that the unincentivized supply will not reach the optimum level. But rather than performance pay, the optimal incentive contract in such cases would pay a fixed wage, because the latter does not provide explicit incentives for any particular dimension and thus avoids crowding out.

As with the other two accounts, it helps at this point to examine the underlying psychological theory implied by the allocational account: the theoretical take on intrinsic motivation as developed within the tradition of behaviorism in psychology.[7] It must be said from the onset that Holmström and Milgrom do not build directly on this theory *qua* the psychological theory underlying their account. Nevertheless, its main points are tacitly accepted.

For our purposes, the most important aspect of the behavioristic approach to intrinsic motivation is that it does not see intrinsically motivated behavior as functionally different from extrinsically motivated behavior. They both follow the same logic of responding to some set of incentives to achieve a desired goal. However, the intrinsic/extrinsic dichotomy is not understood in terms of incentives being internal or external to the person but as internal or external to the activity (Dickinson, 1989). We can illustrate this distinction by considering a puzzle from Rheinberg and Engeser (2018, p. 591): "Why [do] some top-earning football and tennis players give up the game altogether when they retire, whereas some former professional skiers and world cup surfers continue to practice their sports enthusiastically, even without the prospect of material rewards?" Intrinsic motivation is here associated with the incentives related to the activity itself. Aside from the factors external to the activity, such as competition, fame, or money, the activity of surfing as such seems to be more engaging than playing football is. This relates to the part of the allocational account that assumes people will to some extent engage in the activity even without incentive pay. However, as pointed out earlier,

[7] The psychological literature refers to this approach as *behavioral* approach. In order to avoid a possible confusion with behavioral economics I adopt the term *behavioristic* instead.

such intrinsic aspects of the activity are–on their own–not enough to provide the optimal amount of effort. After all, it is not surprising that top surfers train the hardest and develop the most when the activity is part of a competitive context with clear financial rewards, and not after they retire from the competitive career.

For proponents of the behavioristic approach, the two motivations being functionally the same means that potential crowding out effects "lose much of their philosophical importance," yet "they remain empirically interesting" (Dickinson, 1989, p. 12). This empirical interest lies in the challenge the empirical findings present to researchers: to keep on looking for the *hidden* incentive effects that are actually at work in a given case, rather than simply focusing on seemingly *obvious* rewards. As Cameron and Pierce (1994) state, "intrinsically motivated behavior is simply behavior for which appropriate controlling stimuli have yet to be specified" (p. 364). In the language of behaviorism: the reward that is offered may not be the actual reinforcer affecting the behavior. To recall the difference between rewards and reinforcers: a reinforcer is something that positively affects the behavior; a reward, on the other hand, is only assumed to have such a function, without the direct link being really shown (Cameron & Pierce, 1994, p. 364). Rewarding the teachers based on students' results on the standardized tests may be meant to incentivize their teaching efforts, but in reality, the real reinforcer at work is the mechanism that rewards teaching to the test. Therefore, it may be misguided to focus only on links between rewards and behavior, without a proper understanding of the nature of the task: the interplay between its many dimensions and incentives (perhaps yet unknown) that are present in the situation. This pertains all the more to complex and multifaceted forms of behavior–which behaviors high on intrinsic motivation are usually assumed to be–since behaviorists argue that in such cases it becomes increasingly difficult to identify the sources of control (Dickinson, 1989).

The allocational dynamic is thus not explained by referring to competing types of motivation but with respect to competing reinforcers. One clear implication is that the allocational account does not see activities such as blood donation as being primarily intrinsically motivated in the first place. If so, it would probably mean that one is motivated by the activity of sticking a needle into one's vein and experiencing pumping out the blood. For the allocational account, everything that

is outside of such activity is extrinsic by definition. Altruism, for example, is clearly an extrinsic aspect of an activity, which is at odds with the fact that altruistic acts are often being portrayed as intrinsically motivated. Once we recognize this, we see why the most important question is not about the intrinsic or extrinsic motivation but about the relative effects of competing incentives that lead to different allocations of effort.

This question, however, crucially depends on how the actor perceives the nature of the activity, since perception will define the relevant dimensions of the task. Consider that the actor in our exemplary case of blood donation perceives the task as consisting of the following two dimensions: (a) *to incur costs in terms of pain and foregone work*; and (b) *to benefit unspecified others.* Those two dimensions can be pursued simultaneously. But since the reward is meant to provide compensation for a costly activity, and because the benefit to others is much harder to measure than the individual cost, the reward will reallocate the attention from (b) to (a), because the latter is what is effectively incentivized. The fact that most countries have a compensation scheme in place for blood donors suggests that this mechanism is rather plausible. However, it also means that the dimension to benefit others will receive proportionally less attention as the reward gets larger.

3. The Differences Between Accounts Matter for Empirical Work and for Incentive based Interventions

The preceding elaboration now enables us to more carefully examine what the differences between the approaches imply for economic research and for the design of incentive-based interventions. Perhaps the most obvious observation is that the three accounts employ different notions of intrinsicness. In line with the underlying psychological theories, intrinsic motivation is conceptualized either in terms of motivation for rewards internal to the person (motivational account); the absence of extrinsic rewards (signaling account); or rewards internal to the task (allocational account). This is not only of theoretical significance but has especially important implications for empirical research, where appropriate and successful operationalization of intrinsic motivation depends on which account is adopted. When asking respondents to reflect on the reasons for their actions it matters a great

deal what notion of intrinsicness researchers employ when coding and sorting the responses. Similarly, to explain experimental results, the domain of behavior for which intrinsic motivation might be plausibly supposed, differs crucially.

The second observation is that the three accounts imply different visions for the nature of the economic problem that they are addressing. It matters whether the economic problem is how to balance the incentives in order to achieve the optimal level of well-being (motivational account), to achieve the optimal level of social reputation (signaling account), or to achieve an optimal level of a multi-dimensional task performance (allocational account). Organizations or governments seeking to avoid crowding out through appropriate rewards or incentive schemes need to know which theory is the most appropriate, in order to properly function.

The feature of the motivational account that has probably proved to be the most attractive for practical applications is that it builds on a psychological account that theorizes psychological well-being. As such, it has an advantage of incorporating a wider notion of welfare than only pecuniary payoff. As pointed out above, it is a version of Gary Becker's economic approach to human behavior. Becker refrained from giving his notion of stable preferences any objective form (see Emmett, 2006), but Frey's account does precisely that. The utility brought about by the state of being intrinsically motivated is conceptualized in terms of the satisfaction of objective psychological needs. This is especially attractive for the study and interventions in the workplace, since it helps to disentangle the monetary compensation of workers and the actual feelings people have while working. It upsets the traditional notion that sufficiently high monetary compensation can buy any kind of hardships at work. The reason is that while compensation may indeed have short term benefits, it will nevertheless lead to long-term negative psychological effects by crowding out the intrinsic motivation. The needs-based motivational account implies an emphasis on the research into well-being. Not surprisingly, following this direction has led Frey to recently engage with the economics of happiness as a natural extension of this account (see Frey, 2018).

The focus in the signaling account on strategic interaction has great merit and value for analysis of any situation where bargaining or concerns over reputation play a decisive role. While the motivational account has trouble accounting for the

adaptations people make in response to the changes in the social desirability of particular activities, the signaling account in this case provides more precision when describing the exact mechanism of the change. It takes 'being perceived as intrinsically motivated' as a desired outcome of the signaling process and then shows the dynamics of the decision making and adaptation when incentives interfere with the signals. It is not that people seek to signal something that is objectively related to their own individual psychological well-being; they seek to signal whatever is socially desirable, which adds a distinctly social component to the individual preferences. The agent is self-centered, but the preferences that enter his strategy calculus are completely socially constructed and entirely conditional on the preferences of others. Because the model is not directly about intrinsic motivation, the signaling account implies further research into social norms and values that drive what is considered to be a desirable trait, and how the relative social status affects those values. For policymakers or managers seeking to promote more intrinsically motivated behavior, simply raising the status of that activity might be sufficient.

The allocational account operationalizes intrinsic motivation in terms of the changing effort along different dimensions of the task. Thus, it forces us to think about the possible dimensions that we might be overlooking and to think about how incentives affect the performance along those dimensions. Contrary to the other two, the model does not directly require an additional type of motivation. This seemingly makes it immune to some of the issues that come along when extra psychological or reputational factors are added to the analysis. However, according to the allocational account, incentives only affect the marginal utility of effort, not the fundamental decision to engage in the activity. The reasons why people do what they do are not directly linked to the incentive schemes. There seems to be a fundamental element of purpose to the activity that does not depend on economic considerations. This implies, perhaps somewhat surprisingly, that the allocational account offers a fundamental modification in which work is no longer simply a cost. However, while the motivational account explicitly assumes that without the activity the person would be unhappy; and the signaling account assumes that acquiring social prestige in the form of being perceived as intrinsically motivated is central to living in the social world; the underlying assumption of the importance of the unincentivized

supply of activity remains unexamined by the allocational account. While the latter is most closely aligned with the standard notion of utility in economics, this omission leaves a gap when it comes to explaining behavior in the first place. This chapter provides some clue by linking allocational account with the behavioristic account of intrinsic motivation. Our discussion suggests that in order to better understand the workings of incentives, the allocational account implies further research into how the actor perceives and interprets the nature of the activity.

The third observation that follows from our discussion is that contrary to what many economists think, the notion of intrinsic motivation does not simply refer to motivation negatively correlated to (or independent of) financial incentives. While much experimental work is indeed performed by using money as a proxy for extrinsic motivation, a more careful reading of the literature reveals that it is not about money as such. The motivational account is about any reward that feels controlling; the signaling account is about any reward that negatively affects the reputation; and the allocational account is about any reward that targets a particular measurable dimension of the task. Empirical research thus faces a much more complex challenge of identifying and tracing these incentive effects, be it that they are dependent on the individual psychology of the actor, on the social environment, or on the actor's interpretive scheme about the task.

Last but not least, our fourth observation is that all three accounts point to a clear economic relevance of the crowding out effects. This leads to–perhaps somewhat ironically–a pronounced economic motive by policymakers and employers to intervene. The desire to formulate advice for interventions probably explains a large portion of the interest by economists in intrinsic motivation. By clarifying the conceptual confusion, this chapter provides a more solid ground for developing policy proposals and workplace modifications, since our discussion shows that differences in the accounts imply fundamentally different nature of the proposed interventions. However, since the three accounts are built on different competing psychological theories of the same concept, the question that arguably matters greatly for policy makers and designers of incentives is: which theory is the right one?

4. Methodological implications: Is it really a matter of pragmatic choice?

So far, the aim of our discussion has been to provide clarification and better understanding of the competing accounts of intrinsic motivation in economics. At this point one may ask why can't all these theories be equally right? Why can't it be that different people in different contexts simply act according to different accounts? Aren't all three accounts just additional tools available to practicing economists to choose from as they see fit? In this section, I want to demonstrate that the fact that there exist different theoretical accounts of the same concept has deeper consequences for the recent methodological debate concerning the relationship between neoclassical and behavioral approaches in economics.

Angner (2019) argues that a new synthesis has taken place between neoclassical and behavioral economics. His claim is built on an observation that economists are increasingly adopting a "more pragmatic, policy-oriented perspective" whereby behavioral factors are to be incorporated in the analysis "to the extent that they improve empirical predictions and policy decisions" (Chetty, 2015, p. 1). However, contrary to the view held by many economists that such a synthesis of behavioral and neoclassical economics makes behavioral economics obsolete as a distinct research pursuit, Angner argues that the nature of the new synthesis is behavioral and not neoclassical. Economists, such as Chetty, who advocate for the pragmatic approach, are only able to legitimately do so by assuming a methodological position specific to behavioral economics that "it is legitimate and sometimes necessary to incorporate behavioral factors in economic models" (Angner, 2019, p. 200). The pragmatic inclusion of behavioral factors is meant "to increase the explanatory and predictive power of economic theory on the margin" (Angner, 2019, p. 199) by providing a set of realistic (or psychologically plausible) assumptions, where realistic is to be understood as "consistent with the best available psychology" (Angner & Loewenstein, 2012, p. 642).

The choice to use the word synthesis is of course not arbitrary. It is an explicit allusion to the neoclassical synthesis, as well as to the neo-Darwinian synthesis, which have both been considered to be examples of explanatory unification (Kitcher, 1981; Mäki, 2001). Therefore, we can assume that the new synthesis has

similar unificatory aspirations. It might perhaps be obvious that the new synthesis is meant to unify neoclassical and behavioral economics in order to increase explanatory power. However, Angner's insistence on pragmatic attitude towards the use of assumptions raises important questions with regard to the nature of the proposed unification. To use the terminology developed by Mäki (2001), unification can assume two forms: derivational and ontological. In the case of the latter, what matters is that more is explained by using a smaller number of assumptions. In general, the tendency of neoclassical economics to favor theoretical consistency over many other things (sometimes in silly ways) reflects a high degree of derivational unification. Ontological unification, on the other hand, refers to the discovery that phenomena to be explained share the same underlying principles. Systematic application of psychological insights to derive behavioral microfoundations for macro theories would be an example of such an attempt. The new synthesis is clearly not a case of derivational unification, since the pragmatic attitude towards the use of behavioral and neoclassical assumptions as tools suggests that the goal of unification is not to decrease the number of axioms, but rather to expand the available toolbox. Some problems do not need behavioral factors and they can be accounted for with the use traditional tools; others will be explained in a better way by incorporating psychology. For similar reasons, the new synthesis also cannot be categorized as an example of ontological unification, since behavioral factors are not assumed to play a role everywhere. Pragmatism makes it hard to assess the nature and degree of unification implied by the new synthesis.

It may be argued, however, that this is precisely what the new synthesis is about; that it is so pragmatic as to only care about the internal consistency of the use of a concept in a given paper. However, offering such a large degree of freedom of choice to the practicing scientist when explaining the social world raises the problem of ad hocness of explanations. Indeed, one desirable property of a theory of explanation is that "it should be objective–what counts as an explanation should not depend on the idiosyncrasies and changing tastes of scientists and historical periods" (Friedman, 1974, p. 14). Furthermore, as the classic example of Ptolemaic epicycles teaches us, increased predictive power is not enough to accept an ad hoc solution (Forster & Sober, 1994). Granted, the new synthesis is not pragmatic to this extent

and it surely does not imply that 'anything goes'. After all, behavioral factors are supposed to be realistic. However, the discussion in this chapter shows that one major problem with this position is that the question of what is considered realistic– what for the pragmatic practitioner counts as *the best available psychology*–is not as straightforward as it may seem. Psychological theories are competing not only in terms of which one best explains behavior. They are also competing in terms of their fundamental view of human psychology. Ranking may thus be severely complicated. Do we have direct access to our motivation status, as SDT suggests? Or do we always have to infer it from our actions, as is suggested by the self-perception theory? It is hard to square these issues with the prescription to employ the best available psychological theory, because the choice here automatically entails a choice between two or more rather different underlying visions of human psychology. The pragmatic approach has a hard time dealing with competing psychological approaches of the same concept. The discussion in this chapter should remind us that by ignoring the underlying differences there is a danger of sliding into ad hocness of the worst kind: only behavioral description, rather than explanation. The new behavioral synthesis can indeed increase explanatory power only if we assume that its psychological part is unified.

Thus, Angner's view that the new synthesis is behavioral in nature suffers from the same problem that he identifies with respect to Chetty's take on the pragmatic approach: they both assume certain answers to fundamental questions about the foundations of behavioral economics. Chetty assumes the answers that are the same as the one's provided by behavioral economists. And Angner assumes that the clause about the best available psychology secures coherent and uniformly plausible psychological foundations. This highlights the problem of thinking about behavioral economics simply as an instance of importing psychological insights into economics to provide it with more realistic (i.e., psychologically plausible) assumptions. A synthesis that is pragmatic to the extent that it encompasses any model that combines economics with any psychological account can hardly be thought as explanatory. It is closer to being a set of descriptive accounts. More emphasis should be given to the fact that psychology is a plural discipline and that we thus have to be careful when commenting on seemingly similar psychological

theories that have been imported into economics as part of a coherent behavioral economics approach.

5. Conclusion

This chapter has argued that the concept of intrinsic motivation–imported into economics in attempts to explain why incentives may sometimes backfire–has been used in economics in inconsistent ways. We identified three distinct accounts of intrinsic motivation in economics and demonstrated how they differ in terms of the underlying psychological theories they employ, and in the explanatory mechanisms they provide for the potential detrimental effects of incentives. Understanding these differences helps clearing out the conceptual confusion that surrounds the concept of intrinsic motivation used in economics.

Economists are increasingly adopting a view that integration of psychological insights into economics is a pragmatic issue, guided by the nature of the puzzle that empirical data present to us. According to this view, models based on psychological insights are simply another tool in the economist's toolbox. However, it gets complicated when we consider that the underlying psychological theories might not only be at odds with economics but competing within psychology as well. It is not only that we may wonder how to choose the appropriate tool. If a certain psychological model is adopted, the problem arises that this particular combination of economics and psychology may be fundamentally incompatible with combinations that rely on other psychological models. Arguably, the new behavioral synthesis cannot be based on fundamentally different psychological theories.

Intrinsic motivation provides a good exemplary case for demonstrating what is at stake, because the psychological underpinnings employed by the economic accounts may be either fundamentally at odds with the subjectivist economic approach (such as in the case of the motivational account, which is based on an objective notion of human needs), they may lead to a reconceptualization of the nature of the individual in economics (as is the case in the signaling account that turns individual into a seeker of social prestige and self-esteem), or they may enable economists to bypass any psychological issues (such as in the allocational account,

which in terms of psychology remains an empty shell). In light of this, the pragmatic choice may lose much of its straightforwardness.

3

Two types of ecological rationality: or how to best combine psychology and economics*

(with Erwin Dekker)

1. Introduction

Ecological rationality is a concept used by an increasing number of economists as an alternative to mainstream behavioral economics, in particular to the so-called heuristics-and-biases program rooted in the pioneering work of Kahneman and Tversky. The two main proponents of ecological rationality are Gerd Gigerenzer and Vernon Smith. Both Gigerenzer and Smith refer to each other's work repeatedly to suggest that they are talking about the same concept (Berg & Gigerenzer, 2010, p. 149; Gigerenzer, 2015, pp. 115–116; Smith, 2003, p. 469), and they both contributed to several of each other's edited volumes (Gigerenzer & Selten, 2001; Plott & Smith, 2008). And both authors claim Herbert Simon as an important precursor to their work, although they put different weights on this claim (Gigerenzer et al., 1999, p. 14; Smith, 1991, p. 877). We would thus expect that "ecological rationality" provides a coherent alternative to the "heuristics-and-biases" (H&B) program and the associated idea of bounded rationality (Kahneman, 2003). And indeed Davis (2011) has argued that Smith adopted and extended Gigerenzer's ecological rationality.[1] This chapter will, on the contrary, argue that the ecological rationality promoted by

* This chapter has been published as "Two types of ecological rationality: or how to best combine psychology and economics." *Journal of Economic Methodology,* 2019, Vol. 26, 4, 291-306.

[1] Smith (2003) himself, however, points to Norman (2002) as the initial source for his constructivist-ecological distinction.

Gerd Gigerenzer is radically different from the ecological rationality proposed by Vernon Smith. In fact, it is safe to say that the meanings of ecological rationality diverge as much from each other as they do from the H&B program.[2] This chapter sets out to resolve this conceptual confusion by demonstrating the differences between the ecological rationality of the Gerd Gigerenzer type (ER1) and ecological rationality of the Vernon Smith type (ER2). We will also argue that these differences can be best understood as resulting from two different ways of combining psychology and economics, which give rise to different conceptions of rationality and most importantly different experimental methodologies.

The resulting systematic analysis of ER1 and ER2 will also allow us to show different ways in which psychology can be employed in economics. This will undercut the simple claim that the H&B program represents the natural or best way to integrate the two disciplines (see also Hands, 2010, p. 645).

Vernon Smith in his Nobel lecture describes ecological rationality as "an undesigned ecological system that emerges out of cultural and biological evolutionary processes: home grown principles of action, norms, traditions, and 'morality'" (Smith, 2003, pp. 469–470). Gerd Gigerenzer argues that "ecological rationality refers to the study of how cognitive strategies exploit the representation and structure of information in the environment to make reasonable judgments and decisions" (Gigerenzer, 2000, p. 57). From these two statements, it should be fairly clear that they are talking about quite different research programs, which not only differ in what they study–Gigerenzer mostly cognitive strategies (heuristics), Smith mostly institutional environments–but also in what they seek to explain. Gigerenzer is much closer to psychology and modern decision sciences (including the H&B program), which attempt to explain individual choices and decisions, whereas Smith is primarily interested in social systems and processes of social interaction. For Gigerenzer, the central puzzle is how individuals manage to achieve their tasks given their limited cognitive abilities. For Vernon Smith, the central puzzle is much closer

[2] The difference between ER and H&B is examined by Rich (2016).

to that of the other Smith, Adam: how can socially beneficial results emerge from actions which are self-centered and based on limited knowledge?

That being said, they share a common rival research program: the H&B program (see Grüne-Yanoff et al., 2014). This might explain why they have sometimes presented themselves as allies, but we will demonstrate it is an unholy alliance. H&B has become widely accepted as the standard way of integrating psychology and economics (Angner, 2019). The H&B program presents itself as a serious challenge to the neoclassical picture of the rational economic man, and argues that a serious reconsideration of rationality is necessary, since individuals are only boundedly rational (Camerer et al., 2004; Mullainathan & Thaler, 2015; Tversky & Kahneman, 1974). The proponents of ecological rationality do not seek to challenge this claim of bounded rationality. They fully accept it. In fact, they sometimes go even further in emphasizing the cognitive limitations of individuals. However, they argue that in the interaction with their (social) environment individuals are nonetheless able to make reasonably good, or as good as rational, decisions, because they are able to use the environment to their advantage through cues or institutional features of that environment.

This chapter will proceed as follows. In the first section, we provide a broad outline of the different notions of ecological rationality as understood by Gigerenzer and Vernon Smith. In what follows we will focus our attention on these two prominent authors, with occasional reference to their co-authors, since the work of Gigerenzer and Smith provides the clearest contrast between the two types of ER. This has the virtue of making the conceptual contrast clear but does not imply that there could be no convergence or attempts to bridge between them (i.e., Petracca, 2017). In the second and third section, we will analyze the different psychological theories that lie behind the two types of ecological rationality, we will occasionally contrast this with the cognitivist psychology that underpins the H&B program. In section four we will then provide an analytical overview of the differences between the notions of ecological rationality–the different conceptualizations of the cognitive abilities of the individual and the structure of the environment–as employed in the two accounts.

2. Delineating two types of ecological rationality

To understand the differences between the two types of ecological rationality it is important to understand that they represent different marriages of psychology or cognitive science and economics.[3] For Gigerenzer's program, this is fairly clear, as he acknowledges the importance of Brunswik and Simon for his work. For Vernon Smith matters are less clear, since thus far the dominant view has been that he tries to keep psychology out of economics. Don Ross (2014b) recently put forward that Vernon Smith's theoretical account is virtually devoid of psychology apart from allowing that the latent individual cognitive processes may be included in the choices individuals make. Similarly, Davis (2011) argues that while Smith's view is that "interaction between individuals somehow produces or reinforces individuality" by the discovery of preferences through processes of trial-and-error learning and adaptation, his abstract conceptions of socioeconomic institutions make it "hard to see how they could be associated with any particular conception of the individual" (Davis, 2011, p. 154), and consequently, to connect it with a psychological theory. And finally, our argument is at odds with Gul and Pesendorfer (2010), who have argued for a 'mindless economics' and whose perspective is sometimes associated with that of the experimental tradition of Vernon Smith and Charles Plott (Ross, 2014a, p. 200). We demonstrate below that the underlying account based on interaction, learning, and adaptation is itself a distinct psychological perspective.

Let us start with ER1 as conceptualized by Gigerenzer. He defines ecological rationality as: "the study of how cognitive strategies exploit the representation and structure of information in the environment to make reasonable judgments and decisions" (Gigerenzer, 2000, p. 57). His focus is not on the individual and his cognitive capacities, which is the central concern in cognitivist psychology and the H&B program, nor on the underlying cognitive structure of the environment central in Smith's work. Rather, it is on the heuristics which operate as an intermediary between the individual and the natural environment. It is tempting to think of

[3] Psychology and cognitive science have considerable overlap if one looks at it from the perspective of cognitivist psychology. We do not explore the differences between the two here.

heuristics as capturing the cognitive structure of the environment, but as Gigerenzer recently reaffirmed, ecological rationality "means functionality, not veridicality" (Chater et al., 2018, p. 800). As a consequence, heuristics should be thought of as tools to function in the environment, not parts of the environment. To achieve such functioning, the individual relies on strategies, which range from relying on emotions, categorizations, and recognition, to the use of adaptive heuristics. And the latter especially are studied in the 'fast-and-frugal-heuristics' program (Gigerenzer et al., 1999). But whereas in the H&B program the idea is that heuristics lead to systematic underperformance (as compared to some idealized rationality), Gigerenzer argues that these heuristics are at the very heart of how individuals perform successfully.

Heuristics exploit structures in the environment to facilitate choices in real-world situations. The importance of this real-world aspect is repeatedly stressed by Gigerenzer, and it is reflected in the type of tasks which his subjects are asked to perform: predict the winner of Wimbledon, recognize the names of cities in Switzerland, or predict the outcomes of elections (all mentioned in Gigerenzer & Goldstein, 2011). In these examples, Gigerenzer makes it clear that individuals rely on simple heuristics: who is the first tennis player that comes to mind, what is the most famous city in Germany, or which brand do I recognize. That is, they make smart use of the information they already have in order to make (logically) unwarranted inferences. These nonetheless often prove to work well and, as Gigerenzer likes to stress, frequently even better than more sophisticated cognitive strategies.

The cognitive abilities of the actor receive little attention in Gigerenzer's research program. He has no interest in developing a full picture of the human mind and its internal working, something which is far more prominent in the H&B program, as evidenced by Kahneman's discussion of System I and System II thinking (Kahneman & Frederick, 2002). In Kahneman's theory biases result from cognitive structures in the human brain, in Gigerenzer's work they stem from the flawed use of heuristics. Consequently, we find Gigerenzer and others working in this tradition studying messy real-world environments, not the clean laboratory settings which characterize the H&B program. The structure of task environments is conceptualized as a natural environment that contains informational cues that allow individuals to

use their heuristics. In the examples above these are formed by the everyday experiential knowledge of individuals and the way in which the environment is structured: what is central or focal?

At this point, we are ready to contrast the notion of ER1 with ER2 as defined by Vernon Smith. He defines ecological rationality as "an un-designed ecological system that emerges out of cultural and biological evolutionary processes: home grown principles of action, norms, traditions, and 'morality'" (Smith, 2003). This is somewhat puzzling, since ecological rationality is defined as a property of the system, rather than a set of strategies or a field of study as in Gigerenzer's definition. Smith proposes to clarify matters with a contrast between ecological rationality and constructivist rationality. He argues that this distinction is a reformulation of Simon's distinction between objective and subjective rationality, the former meaning rationality from the experimenter's point of view, and the latter rationality "given the perceptual and evaluational premises of the subject" (Simon, 1956, p. 271).[4] Hence, Smith argues that ecological rationality simply develops subjective rationality as Simon defined it (Smith, 2008, pp. 39–40, 176–177).

However, that hardly gets at the difference between Smith and Gigerenzer, since both might equally argue that their interest is not in optimal decisions, but rather in the strategies of the subject. What is more, in the H&B program the Simonian distinction is equally useful to distinguish between the behavioral prescription of rational choice theory and the biases that determine the actual choices of subjects. Therefore, we have to dig a little deeper. The first experiment that Vernon Smith performed was a classroom experiment in which subjects were free to talk and walk around. Its purpose was not to test a particular response to a choice setting but to observe exchange behavior as participants learn from experience. While the subjects were given particular behavioral constraints about minimal selling prices and maximum buying prices, they were apart from these restrictions left free to do what they desired. The focus was on the exchange process and learning within

[4] Simon first developed this distinction in his early book *Administrative Behavior* (Simon, 1997, p. 324).

the constraints by using these constraints to one's advantage (Smith, 1962). In later experiments, the institutional market setting was refined and varied, but the basic idea remained the same: the design of experiments resembled particular market settings. The outcome Vernon Smith was interested in was not individual choices but instead emergent market prices, suggesting that for him the interesting feature was the way in which institutional settings influence behavioral outcomes and learning (Smith, 1967). What emerges is rational because there is a convergence to price level predicted by neoclassical market theory, not because every individual makes a rational choice. This puts him at odds with the H&B tradition, although both perform lab experiments (Smith, 2008, Chapter 7; see also Svorenčík, 2016).

Smith, therefore, had a very different notion of the relevant environment, which had to match some stylized market institution, rather than a particular choice situation. The subjects in the experiment were performing a semi-open task, such as trading. This semi-open task was facilitated by the task-environment, which was nonetheless closer to an economic model situation than a real-world task. Smith, too, is interested in the extent to which market participants are capable of performing these tasks, but more importantly, he wanted to test under what institutional conditions the predictions about equilibrium prices held. When he limited the amount of information provided to the experimental subjects, it was not to purify the experiment but instead to see if with this more limited information prices still converged to a predicted equilibrium level. And when he varied the environment in other ways, it was to test how robust the outcome of price convergence was to changes in the environment. Vernon Smith accepted that individuals had limited capabilities but was interested to see when the institutional environment would still lead to rational outcomes on the market level (Smith, 2008).

This means that Vernon Smith thinks of ecological rationality as rationality that results from the interaction of individuals under different institutional settings, moral norms, and conventions. It is a property of the institutional setting, the system, rather than of individual choices. Like Gigerenzer, he is little interested in the cognitive structure of the human brain, but unlike him, he is more interested in the rules that are part of the environment. And, unlike Gigerenzer, Smith believes that the (experimental) environment has to be varied to investigate the consequences

of different institutions and norms. This means that he is not restricted to 'real-world settings', and–more importantly–that he does not highlight 'informational cues' as essential to the structure of the environment but instead focuses on a variety of stylized settings and the 'institutional rules and norms', which make up the structure of the environment.

We have now arrived at the point where we know enough about the two different approaches to ecological rationality that we can explore their psychological underpinnings. In section four we will then return more systematically to the differences between the two approaches.

3. Gigerenzer and Brunswik

While important references to Simon's bounded rationality feature prominently in Gigerenzer's approach, there is an even more important influence from Egon Brunswik. While Gigerenzer himself constantly refers to and praises Brunswik's work in his writings, he nevertheless seems to suggest that the influence does not go much beyond the level of inspiration, except for Brunswik's idea on probability which he discusses at length (Gigerenzer, 2000; Gigerenzer et al., 1999). However, in this section, we demonstrate that, on the contrary, the approach of Gigerenzer is thoroughly Brunswikian. So much so, that we can hardly make sense of the program if we do not understand the underlying functionalist psychology of Brunswik.

Before we discuss Brunswik's particular approach it is worth briefly revisiting the difference between cognitivist and functionalist psychology. Functionalism is primarily associated with the American pragmatists: William James and John Dewey. It is less interested in the structure of the mind and more in its capabilities, how well it functions in its environment (Levin, 2016). Functionalism's main explanatory purpose is adaptive behavior as expressed in action. As such, it occupies a position between behaviorism, which neglects cognitive psychology altogether, and cognitivism, which seeks to study the structure of the mind. This focus on the structure of the mind within cognitivism includes both conscious and non-conscious forms of reasoning. One might say that Gigerenzer, as far as he is interested in cognition, emphasizes the non-conscious forms.

The contrast between functionalism and cognitivism is drawn well by Brunswik[5] when he states: "Both historically and systematically psychology has forgotten that it is a science of organism-environment relationships, and has become a science of the organism" (Brunswik, 2001, pp. 300–301).[6] He seeks to resolve this by his own brand of functionalist psychology, which he dubs *probabilistic functionalism*, which brings together (i) his view that perception of environmental cues by the observer is probabilistic in nature, and (ii) functionalism's emphasis on achievement (getting the task done). Brunswik's work builds on the smaller European strand of functionalism based on the notion that "any given stimulus … will be perceived differently when placed against a different contextual background" (Leary, 1987, p. 118). The classic example would be the difference between a wink and a blink, which as stimuli might be identical but have nevertheless a radically different meaning. In such case, the challenge for the subject is to act correctly on the cues present in the environment.

In the following, we will address what we identify as four central topics in Brunswik's theory and elaborate how they each map onto the approach of Gigerenzer's research program. Those topics are (i) achievement (as opposed to knowing) as the key goal for the individual; (ii) the environment as a probabilistic texture of cues; (iii) the central place for perception; and (iv) the critique of the standard experimental design.

3.1 Achievement

In Brunswik's view, the key problem of psychology is not that of knowing, but rather that of the adjustment of the organism to a complex environment. The primary goal of the study is to understand how the subject achieves her task, rather than whether

[5] Egon Brunswik came of age in the post WWI Vienna and later emigrated to the United States where, in the years before his premature and tragic death, he fully developed his own branch of functionalist psychology. For a bio-graphical account see Leary (1987) and Gigerenzer's contribution to Brunswik (2001).

[6] Here and below, we refer to the collection of Brunswik's writings from 2001 rather than the original papers.

she understands it. This difference becomes clear when we think about a simple everyday task, such as climbing the stairs. For Brunswik, the important criterion is whether we *manage* to climb the stairs, not whether we *know* how to climb the stairs. The cognitive problem thus extends beyond what the mind knows and should include the full scope of achievement problems (Brunswik, 2001, pp. 300–312). This is also of crucial importance for economics, since if the emphasis is on achievement, it might turn out that there are multiple ways to achieve a particular task. Cognition is one means toward achievement, but achievement can also be reached by making use of cues within the environment. The study of these different strategies became the focus of Brunswik's scientific efforts. As we saw, the most important of such strategies that Gigerenzer develops is the use of heuristics. The very characterization of heuristics as 'fast and frugal' tools for decision making implies that the goal is not any form of certainty or perfect knowledge but simply a satisfactory level of correspondence of the inferences to the real-world problem. Therefore, it is functioning–and not knowing–that is essential to Brunswik's and Gigerenzer's accounts of human decision-making.

3.2 The environment as a probabilistic texture of cues

Brunswik has a peculiar way of thinking about the environment. Firstly, it is important to realize that he thinks strongly in terms of natural environments, rather than social or institutional environments. Secondly, in Brunswik's terminology cues function vicariously, meaning that they derive their usefulness from their relations and inter-substitutability with other cues in the environment. This is well illustrated by Brunswik's experiment where he had the subject followed around the campus by an assistant who asked the subject to estimate sizes of various objects (Brunswik, 2001, pp. 68–105). The results thus obtained demonstrated that we are much better at inferring size of objects than the results of the lab experiments that isolate objects from their environments would have us believe. The height of a building was not just estimated based on the number of stories it had but also whether it was big compared to nearby trees and other buildings. For Brunswik this means that cues should not be isolated from their environment. Brunswik refers to this as the

texturedness of environments (Brunswik, 2001, pp. 17–35). A final element of this theory of cues is that they might be somewhat incoherent, and thus redundancy plays an important role. When asked to estimate the weight of an object we have to weigh several cues against one another: it looks big but feels light. Exactly which cue gets utilized for inference will remain to a large extent uncertain. Thus, environments rich in cues providing an excess of information will enable more accurate inference. This leads him to propose that the perception of the environmental texture of cues is probabilistic in nature, a point that is of great importance for Gigerenzer.

One might think that the emphasis on the probabilistic environment also means that Gigerenzer would be heavily interested in artificial environments, and indeed much of his other research deals with the foundation of statistics. Nonetheless, his interest regarding ecological rationality is predominately about real-world environments and representation. A good example is his paper on whether children can solve Bayesian problems, which varies the way in which frequencies are presented to children (Zhu & Gigerenzer, 2006). His paper on the recognition heuristic formulates the heuristic as an abstract decision rule but argues that recognition itself is based on personal memory and recognition of a person's natural environment "before entering the laboratory" (Gigerenzer & Goldstein, 2011, p. 101).

Brunswik's theory also diverges in one sense from that of Gigerenzer. Whereas Brunswik focused much of his efforts on a theory of perception of the environment, Gigerenzer thinks that heuristics mediate between the individual and the environment.[7] But, as in the work of Brunswik, we find in Gigerenzer's approach the study of the utilization of environmental cues to be the prime factor in successful decision making. And in his critique of the conjunction fallacy, the so-called Linda-problem, Gigerenzer relies on the probabilistic texture of cues (Hertwig & Gigerenzer, 1999).

[7] As a consequence, these heuristics can be transferred from one environment to another and as such provide a stable factor that we can study.

3.3 Perception is key

In order to address the problem of the inability to provide a full causal account of cognitive achievement, Brunswik distinguishes between two cognitive processes: intuitive perception and analytical reasoning, which are fit for tasks that require estimation with some degree of uncertainty and certainty, respectively. In an experiment meant to illustrate this point, Brunswik assigned half of the respondents to infer the size of a pole 'intuitively', and the other half by triangulating (Brunswik, 2001, pp. 260–271). While the former resulted in the normal distribution of error (nobody being excessively off the mark), the latter resulted in half of the group being precise with the other half being very wrong. In his words, this demonstrated "one of the pitfalls of reasoning, namely, the going off in the wrong direction by being right about something else" (Brunswik, 2001, p. 261). Thus, Brunswik concludes that what more successfully guides behavior is 'intuitive' perception, not reasoning and thinking. The emphasis on intuition should not be overdrawn, it is really perception based on cues that does the work. A book title like Gigerenzer's *Gut Feelings* (2007) also suggests that intuition is central. But for Gigerenzer, too, it is the process of perception of cues in the environment that matters, not pure intuition.

This view maps almost perfectly to Gigerenzer's notion of variance-bias trade-off, which implies that heuristic reasoning can indeed be biased but will also be characterized by the low variance of results, thus providing more robust and homogenous outcomes. As such, in contrast to the data-hungry complex models that decrease bias at the expense of increased variance, it might even result in a lower total error–a phenomenon he calls the less-is-more effect (Gigerenzer & Gaissmaier, 2011).

3.4 Representative design of experiments

As will be clear by now, Brunswik was a proponent of experiments in real-world settings. Human perception functions best when it can rely on a variety of cues, rather than in a clean experimental set-up. Furthermore, in order for the experiment to be valid, he introduced an additional requirement: not only the sample of the participants has to be representative (as is usual in experimental social sciences),

but also the conditions themselves. This did not mean a variety of lab experiments but instead a variety of real-world environments, since in each different situation the availability of cues will be different. Gibson (1950), who started along Brunswikian lines and later developed his highly influential ecological psychology, has argued that, in order to account for the fact that "there is literally no such thing as a perception of space without the perception of a continuous background surface" (p. 6), experiments need to be performed 'outdoors' and not in the context-less lab. After all, as Hogarth (2005) points out, abstract experimental tasks test only abstract theory, while the more important step should be to test the theory in the situations that are representative of the real economy. This is well reflected in Gigerenzer's work on the recognition heuristic (Gigerenzer & Goldstein, 2011).

To put it in terms of ecological rationality, the individual acts ecologically rational to the extent that she uses the right heuristics adapted to various real-world environments. If we want to test organismic achievement in reading environmental cues, conditions must mimic those of the environments in which the organisms typically operate. Brunswik was a fierce critic of the clinical design of psychological experiments, because they either remove or radically simplify the environment in terms of cues provided.

The most striking contrast between the H&B experiments that are often based on game-theoretic and choice-theoretic situations, and the 'dirty' ER1 experiments, is their approach to experimental design. But now that we understand the underlying functionalist psychology of Brunswik, we are able to better understand the rationale for Gigerenzer's focus on real-world tasks rather than abstract choice situations. For him, the functioning of heuristics is always dependent on particular choice environments. For this reason, the tasks in the experiments are so specific: name the capital of Germany; catch a ball; which has more cholesterol cake or pie (Gigerenzer, 2008)?

There is, however, one curious aspect which is shared by the program of Brunswik and Gigerenzer, and which makes it (somewhat) less suited for the social sciences. They both think of the environment in mostly naturalist terms. Their theories largely avoid the problems of symbolic processing, and hence efforts by humans to make the world more (easily) navigable are left out of the picture they

paint. It is one thing to infer the size of a building, but quite another to strategize about the prices of copper or trust a stranger to deliver what promised. It is those type of issues that are central in ER2.

4. Vernon Smith and distributed cognition

In contrast to both cognitivism and functionalism, Vernon Smith focuses on properties of the system rather than the cognitive capabilities of the individual or her achievements. Therefore, he is interested in more open-ended problems, where the outcome emerges out of interactions of individuals and their (institutional) environment. By an institutional environment, Smith means typically a market setting in a laboratory. He does not study different actual market settings; instead, he is changing the institutional rules within the laboratory market to allow for different types of interaction–to allow for resales, for example. Therefore, it is tempting to believe that psychology plays no significant role in Smith's account, and that he simply builds on a longer tradition in economics which has shunned psychology altogether. As such, we might be tempted to categorize Smith's work as what Gul and Pesendorfer (2010) have described as 'mindless economics'. Although later in his career Vernon Smith does call for a fitting psychological account for his program, he has not developed anything substantial in that direction (Smith, 1991, p. 880). Ross (2014b) has suggested that the approach of Vernon Smith and other market experimenters allows for psychological factors to be included in, rather than identified with (as, for example, in H&B approach), the processes that generate choice data. But for Ross, too, it remains a desideratum, not a developed program.

We argue that the research program of Vernon Smith is, on the contrary, highly compatible with an existing approach in psychology called *distributed cognition*, and hence not at all mindless or anti-psychological. This psychological underpinning is not only relevant for the theory to provide a complete alternative to both the H&B and ER1 programs but also for us to provide a full understanding of ecological rationality in the sense that Smith uses the term. Distributed cognition is a more recent approach that developed as a critique of cognitivism and the associated focus on the study of the individual mind. It challenges, somewhat similar

to functionalism, the idea that human cognition can be separated from the environment in which it is situated (Hutchins, 1995). But it goes a step further and emphasizes not the structure of the individual mind, but rather the cognitive properties of the interaction of the individual and the environment. This might still sound rather abstract, but those familiar with Hayek's views on the working of the price system will quickly recognize that in his account prices are an essential part of the cognitive system that allows rational action to take place (Hayek, 1937, 1945). As such, Hayek's work is a precursor to this type of thinking about cognitive properties and can be usefully reinterpreted in the context of distributed cognition.

Distributed cognition is associated with a number of related approaches called *situated*, *embodied*, or *grounded* cognition (Anderson, 2003; Barsalou, 2008; Robbins & Aydede, 2009) where the common trait can be found in the emphasis on the contextual and interactionist understanding of the mind in its environment. Wilson's (2002) review article provides an excellent introduction to the developing field and the outstanding issues that researchers are grappling with, and it is not our intention to arbitrate between the various approaches here. What crucially differentiates distributed cognition from functionalism is its focus on the system, rather than the individual (Rogers, 2006). So, when the system selects or rewards a particular type of behavior, we can also say the system has certain cognitive properties that differ from the cognitive properties of its elements (Menary, 2006). In the following, we will address some key points of the distributed cognition approach that will enable us to show why it fits so well as a psychological account of Vernon Smith's theory.

4.1 Situatedness, distribution, and interaction

In the distributionist approach, the situated nature of activity is the starting point. Although not quite as pragmatic as the functionalist approach, and hence less interested in 'mere achievement', the emphasis here, too, is on the fact that cognition is always related to particular tasks, material environments, and institutional settings (Osbeck & Nersessian, 2014, p. 89ff.). But this approach is far less individualistic and highlights the distributed nature of the cognitive processes, where new cognitive properties emerge from the interaction among the elements of the

system (Hollan et al., 2000; Hutchins, 2014).[8] This interaction can take the form of learning, usage of tools, imitation of others, or responding to incentives. Hutchins' (1995) seminal example is the study of the navigation of a large ship, where many individuals perform separate tasks and use specific tools without direct central control. Together they are involved in the joint cognitive process of navigating the ship that is more than the sum of those individual contributions.

It is clear that the market experiments of Vernon Smith allow for interaction with others where learning takes place, such as in the repeated trading games. But even within one trading round, the market experiments provide information to the subjects: is my offer rejected or accepted? Is the price I am asking too high? At what price have other exchanges taken place? In Smith's own words, his work studies the "interactive experience in social and economic institutions" (Smith, 1991, p. 878). In doing so he is constructing an experimental account of the utilization of the Hayekian socially distributed knowledge. The experimental setting is so designed that it replicates essential institutional features of the market economy but not the complete causal texture that Brunswik and Gigerenzer are interested in (Smith, 1994).

4.2 Off-loading and cognitive scaffolding

Where the first point emphasizes social interaction between people, and interaction between people and their environments, the distributed cognition approach also emphasizes the cognitive properties of the environment. To describe this, they argue that individuals 'off-load' cognition to the environment by using and modifying the environment (Hollan et al., 2000). This alters fundamentally how we think of cognition, since there is no longer a single (internal) representation. Instead, the distributed cognition approach starts from the idea that the environment itself is a cognitive structure that has been shaped by previous interpretations and interventions. In the literature, these are called external representations (Zhang,

[8] Hutchins argues that additional cognitive flexibility is made possible through interaction: "A phenomenon that is entirely missed by research paradigms that, for good reasons, intentionally limit the methods subjects may use to perform a task" (Hutchins, 1995, p. 289).

1991). Streets have names, houses numbers, and cars are equipped with navigational systems. These external representations are themselves part of interpretive structures, sometimes overlapping, sometimes at odds with one another. The individual 'plugs in' to this cognitive system when he is able to 'read' these external representations. But the individual also contributes to these cognitive structures by numbering his house, for example. The focus is thus not on internal representations, but rather on the cognitive structures of social systems. Rather than being just an external memory aid, external representations are the central feature of the cognitive system (Kirsh, 2010).

The way to think about this is to think not about individual trades, or individuals maximizing their utility, but rather to think about the price system. Hayek has argued that we can use prices as shorthand for the availability of a particular resource. An increase in the price is a signal that it has become scarcer, yet we do not need to know the cause of it, or all the uses and different markets in which the resource is used more generally. Hayek turns this into a more general point about the cognitive structure of the environment: "We make constant use of formulas, symbols and rules whose meaning we do not understand and through the use of which we avail ourselves of the assistance of knowledge which individually we do not possess" (Hayek, 1945, p. 528).[9] This system is the cognitive structure of the environment, and individuals merely plug in and rely on the system. In Vernon Smith's work, we see a similar emphasis on 'the price system' rather than individual actions (Smith, 2015).

In a somewhat extreme form, this view has also been labeled the extended mind hypothesis, where the crucial emphasis is precisely on seeing the coupling of the individual and the external artifact as a new cognitive entity in its own right (Clark & Chalmers, 1998). Vernon Smith certainly does not completely abandon an individualist methodology but does recognize the cognitive properties of the environment. John Davis has suggested that a distributive cognition view might

[9] Hayek, on who Vernon Smith draws so heavily, was himself engaged in serious psychological inquiries. He saw psychology as interactionist and his book *The Sensory Order* anticipated later development in neurology and artificial intelligence (Hayek, 1952; Steele, 2002).

entail a more serious revision of individualism in economics. He argues that a genuine distributionist perspective entails we can no longer stick to a notion of independent individuals (Davis, 2016, p. 28). This would mean that the internalist-externalist distinction would dissolve and that we would instead study distributed socio-cognitive structures consisting of multiple interrelated individuals and a cognitively rich environment (see also Davis, 2010). We recognize the importance of these implications, but in this chapter, we will not explore such a distributionist perspective further since it does not fit Vernon Smith's notion of ecological rationality. It is, however, worth noting that such an attempt has been undertaken by Bardone (2011; see also Petracca, 2017).

Ross (2014b) has highlighted a similar insight by evoking the notion of *cognitive scaffolding* from the cognitive philosopher Andy Clark (see also Davis, 2010). Cognitive scaffolding is the process by which human beings use and transform their environment to their cognitive needs. We thus develop ways in which to navigate our environment, say a particular path through the grocery store to avoid missing essentials and to not overspend, or we transform our environment to make it more navigable, say through the use of signs at an airport. This is a far cry from the naturalistic way of thinking about the environment in the work of Brunswik and Gigerenzer, but it has interesting parallels to the concept of choice architecture in the H&B program.[10]

4.3 Environmental pressures

In her overview article on situated and other related types of cognition, Wilson (2002) emphasizes the importance of time pressure for shaping cognition, but we can make the point somewhat more general here: environments exert pressure. Tasks have to be performed and decisions cannot be postponed forever, so the resources to make decisions and to perform tasks are limited. As Wilson puts it succinctly: "When situations demand fast and continuously evolving responses, there may simply not be time to build up a full-blown mental model of the environment, from which to

[10] For an attempt to bridge some of these differences see Arnau, Ayala, and Sturm (2014).

derive a plan of action" (Wilson, 2002, p. 628). Distributed cognition provides an answer to this in terms of redundancy of the capabilities of the elements in the cognitive system (Rogers & Ellis, 1994, p. 123). When navigating a ship, people performing particular tasks have knowledge also of some of the others, and this redundancy is crucial for the functioning of the system and thus the emergence of the larger-scale cognitive capacities (Hutchins, 1995).

That is, so to say, one side of the coin. The other side is that environments select for certain types of behavior. The extent to which they do so is a contested issue in economics. If one takes this perspective to its extreme, the environment contains the rationality, and that is indeed what some economists have suggested. Famous are the experiments by Gode and Sunder (1993) which simulate a market in which the only market rationality required of agents–or achievable for their simulated 'zero-intelligence' traders–is that they not bid beyond their budget constraints. They explicitly suggest that the institutional structure of markets is a substitute for individual rationality. That argument has origins in work by Chicago economist Armen Alchian who argued that "there may have been no motivated individual adapting but, instead, only environmental adopting" (Alchian, 1950, p. 214). In fact, Vernon Smith was criticized by others that the way his initial experiments were set up was so institutionally constrained that convergence to the equilibrium price was inevitable.

In psychology, this is matched by those who argue that the environment does all the cognitive work, and in terms of the methodology it leads to a pure form of behaviorism. Vernon Smith sometimes comes quite close to adopting a view of rationality that places it entirely in the environment. In a particularly strong statement to this effect, he states: "The current manager does not know about opportunity cost or even why the policy is what it is; only that he learned it from the last manager. He is an instrument of the 'law' of one price in a market" (Smith, 1991, p. 892).

The tension between the rationality of the individual and the rationality of the environment fact runs even deeper, since in the scaffolding view above, the individual becomes rational through the use of the environment, and rationality is thus to be located on the level of the individual. In the environmental pressure view,

on the other hand, rationality emerges in the aggregate; in markets, for example, through the forces of competition. Vernon Smith seems to toggle between these two notions. On the one hand, he suggests that individuals learn to make rational decisions by relying on human institutions and practices. But he also emphasizes that reason is a limited guide, and that the process of evolution is necessary "to serve the process of selection" (Smith, 2008, p. 38).

The two views need not be mutually exclusive, and might, in fact, be reconciled or complement one another. Smith recognizes that there is more mediation between the environment and the individual than the selection-only view suggests. This provides scope for the study of the interaction between the individual and the environment that is central to the distributed cognition perspective. The environment is thus important for two reasons: for its cognitive properties (the feedback it provides) and for the selection pressures it exerts. This means that this approach will be less voluntaristic than both the Brunswikian and the H&B approaches–the individual is more constrained by the situation, and acts 'in accordance' with the situation. Consequently, this type of psychology aligns more easily with institutional economics than with accounts derived from rational actor models.

In this section, we have demonstrated how central aspects of the distributed cognition approach in psychology map to the market experimental approach of Vernon Smith and others. In particular, the focus on the joint system of individuals and the environment, rather than the individual in relation to her environment, distinguishes this approach from those of Brunswik and of cognitivism.

5. Different combinations of economics and psychology

We are now in a position to put the different combinations of psychology and economics together. We have provided a conceptual overview in which we have contrasted the two approaches discussed with the H&B program. The overview and our analysis so far should have made it clear that the question whether economics needs psychology is rather misguided. Instead, we should ask what type of psychology should be combined with economic analysis. Just like economics is

characterized by different approaches, this is true in psychology as well. Consequently, we should gain a better understanding of the different approaches in psychology, and how to best set up a conversation between economics and psychology, instead of mindlessly adopting a particular type of psychology. For this chapter, it is of crucial importance to see that the concept of ecological rationality, which has been suggested as an alternative to the H&B program, is a result of two different combinations of psychology and economics. And hence, it gives rise to two fundamentally different understandings of the term.

	Distributed cognition & Smith	**Brunswikian functionalism & Gigerenzer**	**Cognitivism & 'heuristics and biases'**
Level of analysis	Systems	Individual choices and tasks	Individuals' choices
Cognitive problem	Rationality of the system	Achievement	Knowing
Guiding behavior	Rule-following	Perception	Reasoning
Normative benchmark	Competitive market outcomes	Good enough/Better than	Rational choice theory
Type of environment	Institutional context	Real-world complexity	Information patterns
Guiding metaphor	Interaction and emergence	Fast and frugal heuristics	Cognitive biases

Table 1: Three Combinations of Economics and Psychology

Gigerenzer's use of ecological rationality stems from a functionalist tradition, which, although critical of cognitivism, accepts the latter's methodological focus on choices

and acts of individuals. This means that, like cognitivists, we end up with an individualistic perspective on human cognition, although that of Gigerenzer is far more focused on the ability to perform specific tasks than on the cognitive ability to make more abstract decisions. Vernon Smith's idea of ecological rationality is a radical departure from this perspective since it does not locate rationality in the individual mind, but rather conceives of it as a property of a system: a combination of individuals and environment. Contrary to Davis (2011) who interprets Smith's account in terms of institutional environments enabling individuals to be ecologically rational, we can say that particular market outcomes are rational despite the fact that individual choices are not. Since there is an opportunity for learning, imitation, and there are feedback mechanisms in the system, individual errors are corrected.

What Gigerenzer is particularly interested in is how individuals use heuristics to perform specific tasks in what he calls the natural environment. As he states, ecological rationality is the ability "to exploit the structure of the information in natural environments" (Gigerenzer et al., 1999, p. 24). This means that he is less interested in social tasks, although he occasionally touches on them, and has very little to say about the transformation of environments by humans, or what we might call the cultural, that is non-natural, part of the environment. The heuristics that humans use have evolved in natural environments and hence rely on cues in natural environments, but this does not mean that the heuristics themselves can be regarded as 'cultural'. As far as the recognition heuristic relies on cultural aspects of the experienced environment, those cultural aspects are treated as 'natural' parts of the environment.

That is different for Vernon Smith, whose approach is far less focused on the natural environment and more on the (constructed) institutional environment. So, when he discusses the challenges of psychology to economics, he emphasizes that risk-preference is dependent on the institutional context, so that it will differ depending on different economic settings; and that some psychological effects, such as the endowment effect, tend to disappear when the institutional pressure is sufficiently strong (Plott & Smith, 2008). But most importantly, Vernon Smith does not try to locate the rationality on the level of the individual. His early classroom experiments tested the convergence of the observed prices in his experiment to the

predicted market equilibrium price, and rationality was the outcome of a process of interactions in a system guided by certain rules, not a property of the acts of the individual. For him, rationality is to be found at the level of the system, like it is for the modern theorists of distributed cognition.

Just like in a democracy we do not think it is important that everybody holds the same or the right view but that through discussion and deliberation we will reach a well-informed decision at the collective level, so Vernon Smith argues that rationality emerges as a property of systems. As much should be clear from his definition of ecological rationality, which talks of 'the system', but our analysis of distributed cognition makes us better able to understand what that means. On the contrary, for Gigerenzer, ecological rationality is a property of individual decisions, which occur through heuristics. It shares the focus on individual decisions with cognitivism, but contrary to cognitivism, it is not interested in whether the subject can reason her way toward the superior alternative, but rather whether she reaches, or achieves, this decision. Its criterion is far more pragmatic than that of the cognitivists, who are interested in the way in which a particular choice is made.

This directly impacts how the three programs think about normative issues, and what standard they use to evaluate choices: the normative benchmark. This benchmark is more or less shared between Vernon Smith and the H&B program. The benchmark for them is the rational choice as laid out by neoclassical economics, with the exception that Vernon Smith, like we argued, is less interested in individual choices and more in aggregate outcomes. So, the benchmark he ends up using are the equilibrium outcomes of market models based on rational choice theory. This is different from Gigerenzer, who has toyed with the idea that heuristic decision-making leads to better than rational choices but in more recent years has emphasized that heuristics are superior for a certain domain of tasks in which the solution is either intractable, there is an estimation error because of small number of occurrences, or, finally, because the problem is ill-specified. Gigerenzer believes that most human choices suffer from at least one of these three problems. The relevant benchmark for him, therefore, cannot be rational choice theory, and instead, he proposes that given strategies perform good enough, or better than the alternatives (Gigerenzer & Todd, 2012; see also Hands, 2014). The normative benchmark is not that of optimization

in some abstract sense but whether the functioning is achieved, or the task completed. Implicit in this notion is an evolutionary account in which strategies have to be good enough to survive, but not necessarily optimal.[11]

This completes our contrast of the two types of ecological rationality. ER1, part of Gigerenzer's fast-and-frugal heuristics approach, is based on the picking up of cues in the environment, which are inputs for heuristics used to arrive at quick decisions. The individual is 'ecologically rational' to the extent that the strategies, the heuristics, used are well adapted to her environment. ER2, part of the market-experimental approach developed by Smith, Plott and others, is based on institutional constraints and social and cultural norms, which help the individual to navigate his social environment. The system is 'ecologically rational' to the extent that it facilitates this navigation by the individual by means of embedded norms, learning or feedback mechanisms, and leads to outcomes that are efficient in the aggregate. Both the institutional rules and in particular the social norms emerge in the process of social interaction and represent the cognitive content that is off-loaded to the environment.

6. Conclusion

It is tempting to accept the standard narrative that modern behavioral economics is the reintroduction of psychology into economics. However, that narrative relies heavily on the idea that there is one economics and one psychology. In this chapter, we demonstrate that this narrative is severely complicated when we look at the two conceptions of ecological rationality, which are offered as alternatives to modern behavioral economics and accounts of bounded rationality. As we have demonstrated, these present two alternative combinations of economics and psychology. The methodological question for economists, therefore, shifts from a concern over whether economics needs psychology to what type of integration of the two fields is desirable and fruitful. This chapter does not aim to provide guidelines

[11] For a more radical position see Cosmides and Tooby (1994).

for what the best integration might be, but it does show the choices and stakes involved.

Firstly, we have demonstrated that ecological rationality does not present a uniform challenge to the idea of bounded rationality. Instead, there are two different conceptions of ecological rationality, which represent different ways of integrating psychology and economics. What we have called ER1, ecological rationality as promoted by the FFH-program associated with Gerd Gigerenzer, is a merger between the functionalist psychology of Brunswik and an individualist economics focused on specific real-world tasks. What we have called ER2, ecological rationality as promoted by the market-experimental program associated with Vernon Smith, is a merger between the psychological approach situated cognition and a more institutional economics. We have repeatedly contrasted them with the H&B approach which represents a merger between cognitivist psychology and the individualist economics associated with rational choice theory.

Secondly, we have argued that these different combinations of psychology and economics have important consequences for relevant methodological practices. They impact the level of analysis, the way in which experiments are conducted, the perceived central problem to be explained, and the way in which central concepts, such as the individual and the environment, are conceptualized. To move forward the fruitful discussion between these different programs, it is important to realize that such deep methodological differences exist. Psychology is and will remain relevant to economics, as is by now broadly accepted. But importing parts of another discipline, or even merging the two, should not and cannot occur without a good understanding of which goods we are importing.

4

Environment as a Resource, not a Constraint*

1. Introduction

In their manifesto for contextual economics, Goldschmidt, Grimmer-Solem, and Zweynert (2016) argue that *raison d'être* of contextual approach is that mainstream economics has been too narrow in its scope by relying on an isolated view of individuals and their actions. As such, it has neglected the importance of the wider social, institutional, and historical context in which those individuals are embedded. Many might be tempted to agree with this statement immediately. However, it is actually hard to maintain that contemporary orthodoxy ignores the study of how environment affects the economic outcomes. To the contrary, the study of institutions has been fully incorporated into the standard economic framework (e.g., Williamson, 2000) and the orthodoxy has 'pragmatically' accepted behavioral economics' emphasis on framing and choice architecture (Chetty, 2015). For many economists, doing contextual economics may hardly mean much else than doing more of what has already been done. On this view, addressing contextual issues is purely a matter of collecting more and better data. Does this mean, however, that there is no distinct place for a self-consciously contextual approach to economics?

Answering this question requires us to delve deeper into the underlying conceptualization of context in economics. In this chapter I argue that the existing

* This chapter has been published as "Environment as a resource, not a constraint," *Journal of Contextual Economics-Schmollers Jahrbuch,* forthcoming.

mainstream approaches that integrate contextual issues into the economic analysis share a particular perspective: context matters because it shapes behavioral outcomes by imposing various internal and external constraints on individual action. Such perspective, however, is arguably not the only possible one. The central methodological question legitimizing a distinct contextual approach to economics should thus not be *whether* context matters but *how* and *why* it matters. The aim of this chapter is, first, to examine and refine the 'context matters' dictum, and then to suggest a distinct theoretical ground for contextual economics that differentiates it from the existing approaches that analyze the environment as a type of constraint.

The central argument of this chapter builds on the assertion by Sturn (2016) that contextual economics should abandon the notion of economizing within "the orderly world of pure economic transactions" as its central analytical concern (p. 80). Instead, it should move toward embracing the analysis of "complex, messy situations" (*ibid.,* p. 80). Sturn argues that the exclusive focus on economizing has resulted in the neglect of certain types of contexts where scarcity is not the defining feature but where various non-scarcity interdependencies result in dynamics of increasing returns. This view, however, still conceives of the environment as having essentially a passive role. To move away from that, I suggest that contextual economics should rather focus on yet another type of interdependencies where the environment becomes an active part in cognitive processes extending beyond the individual mind. I draw on recent developments in cognitive science based on rejecting the view of cognition as being limited exclusively to the individual brain in favor of understanding cognition as stemming from interactions between the individual and the environment. Hence, the discussion in this chapter also contributes to the literature on the possible ways of combining economics with psychology and cognitive science.

We proceed as follows. Section 2 discusses how context has been conceptualized and approached within both neoclassical and behavioral economics and identifies a lacuna to be filled by contextual economics. Section 3 addresses three common-sense conventions implicit in the standard approaches. These conventions create a pull towards conceptualizing the environment as a constraint. With regard to each I propose alternative ways forward for contextual economics. Informed by

this discussion, Section 4 develops a reconsideration of the role of the environment: rather than a type of constraint, it should be instead seen as a resource. Section 5 concludes.

2. Circumventing the Perceived Danger of Relativism: Environment as a Constraint

For many economists, the idea that contextual factors matter for the economic analysis carries a seed of relativism: how are we to do economics as a science if every economic action depends on the given context? The standard answer to this dilemma circumvents the problem altogether by focusing on the pure logic of choice: a context-less logic based on the relationships between abstract entities. In contrast to historians or sociologists, to which such perspective more often than not resembles a particularly limited understanding of the social and economic world, for economists the context-independency of the analysis is a feature and not a bug. It shields the analysis from dangers of contextual relativism.

Combining the pure logic of choice with the most widely accepted definition of economics–economics as the study of efficient allocation of scarce resources among competing ends–the standard analysis treats contextual factors as additional constraining factors on the efficiency of allocation. In other words, the relationship between the individual agent (or firm) and her context resembles the way the maze structures and influences the movements of a mouse trying to reach a snack in the middle of it. Such analysis is in economics mostly cast in terms of constrained maximization and associated with "the economic approach to human behavior" (Becker, 1976). This approach has been incredibly successful in generating powerful accounts of human behavior within a wide range of contextual settings. But maximization must in the first place, of course, assume something that is being maximized. Most generally, what is maximized is the utility that the individual gains from satisfying her preferences in the most optimal way possible. Stigler and Becker's (1977) (in)famous assumption that preferences are given and stable across

population[1] has turned the focus exclusively on the analysis of the means for achieving the desired goals, delegating the questions about the nature of the wants to other disciplines. The individual in the standard economics carefully plans her actions in order to get the most out of what is possible, given the constraints she is facing.

By concentrating on the properties of this individual and isolated maximizing entity, the neoclassical perspective has been notoriously a-social and a-cultural. It has prominently neglected the role of institutions in economic and social life. However, the importance of economizing remains paramount also for institutionally minded economists (Sturn, 2016). For example, in Williamson's (2000) institutions-based model of multi-level economizing, higher levels are imposing constraints on the levels immediately below. From level zero that consists of the mechanisms of the mind, to level four where the everyday continuous marginal adaptation in the allocation of resources takes place. Characteristically, the hierarchy is linear and the influence unidirectional[2]: mechanisms of the mind constrain the

[1] It must be emphasized that in this framework changes in preferences (or tastes; Stigler and Becker use these terms interchangeably) are to be distinguished from ordinary changes in demand for various market goods. Stigler and Becker conceive of preferences as something fundamental: not as preferences over ordinary goods and services but over "basic pleasures" or "final objects of choice" (Becker, 1976, pp. 145–146). In their framework, economic agents do not *consume* ordinary goods, but *produce* these final objects of choice by using market goods as inputs. Thus, "the consumer's demand for these market goods is a *derived demand* analogous to the derived demand by a firm for any factor of production" (*ibid.,* p. 134, emphasis added). The demand for different market goods may indeed differ across population, but such differences do not reflect differences in preferences but differences in "*the constraints of time, consumer knowledge and inter-household differences in consumption efficiency*" *(ibid.,* p. 145, emphasis added). Cf. also the discussion in Vanberg (1994, pp. 27–28) where he distinguishes between *ordinary* and *fundamental* preferences in Becker's framework; and the more general discussion in Kirzner (1992, pp. 41–46) on the difference between *induced* and *underlying* variables.

[2] Although Williamson's schema in principle allows for feedbacks on every level, they do not enter the analysis. His justification for this omission is unclear, but it appears that he assumes that on the time scale that is relevant to his analysis feedbacks do not take place. As he puts it, "although, in the fullness of time, the system is fully interconnected, I mainly neglect these feedbacks" (Williamson, 2000, p. 569).

culture, culture constrains institutions, institutions constrain the governance, governance constrains the allocation of resources. The purpose of the processes on each level is explicitly tied to solving the optimization problem ('getting it right') within those constraints. In this framework, the environment is 'the rules of the game' that structure the incentives and transactions costs. In turn, incentives and transactions costs constrain and shape behavior and are thus considered–assuming, of course, that people are "maximizers [that] maximize, *always and everywhere*" (Leeson, 2020, p. 146, emphasis in original)–as determining factors for the economic and behavioral outcomes. In this sense, such analysis of institutions stands solidly in the neoclassical tradition of spotlighting relative costs and benefits as the relevant determinants of behavior. However, as Sturn (2016) points out, the focus on economizing within constraints has led economists to ignore certain types of contexts where the relationship between individual agents and their environment is not as straightforward as the standard theory suggests.

A parallel development that forced economists to squarely face the problems of context has been the arrival of behavioral economics. The attempts of behavioral economists to open up the black box of preferences have corrected some of the aspects of the standard approach by forcefully showing that revealed preferences are context dependent. However, by conceptualizing contextual and environmental influences as exogenous factors, behavioral economics takes a decisive turn inwards. It focuses on the psychological and cognitive mechanisms inside the head of the individual, which may lead to biased actions when influenced by the external context. In contrast to Williamson (2000) who delegates the possible influences of the mechanisms of the mind only to the pre-cultural level zero, psychological constraints are in behavioral analysis continuously present on the level of everyday decision making. Importantly, these psychological constraints are dependent on the context in which the decision making takes place. Behavioral economists analyze the different contexts in terms of different frames within which a decision problem can be presented (Tversky & Kahneman, 1981). For example, the way the choice among various snacks is presented in the supermarket (e.g., by placing them in different order) will likely affect what the customer will buy. Such examples of individual inconsistency of choice within different contexts are then used to argue that the

environment affects preferences simply by framing the decision problem differently. And since behavioral economists consider stability of preferences to be one of the normative hallmarks of rationality, it can be thus said that, on this view, environment imposes a constraint on rational thinking. Framing effects count as evidence of the context-induced irrational behavior.

Another step away from the isolating view–and one where correction of the standard neoclassical theory has gone into the direction of acknowledging a deeply social nature of human behavior–has been recently described as 'the strand two behavioral economics' (Hoff & Stiglitz, 2016). This second wave of behavioral economic analysis aims to break the focus on the individual psychological mechanisms and to branch out into more social aspects of decision making. Rather than inherent psychological cognitive biases, it is concepts such as identity and culture which gain prominence in the analysis. However, the turn inwards remains, because strand two behavioral economics follows the conventions of cognitive psychology by conceptualizing culture, social identities, narratives, and other social factors as various mental models existing in the mind of the individual. Context is thought of as a factor that triggers particular mental models which individual then "[uses] to process information and interpret the world" (Hoff & Stiglitz, 2016, p. 36). Via the concept of mental models, contextual and social factors become additional arguments in the individual utility function (see Section 5.7 in Hoff and Stiglitz 2016). Despite the repeated claims of centrality of social interaction for their argument, this over-reliance on cognitive psychology nevertheless makes strand two behavioral economics vulnerable to the objection raised by Davis (2011), who argues that "rather than *contextualize individuality* … [these approaches] *internalize sociality* by giving the utility function an unmistakably social dimension" (p. 69, emphasis added).

The preceding overview helps us to refine the case for a differentiated approach to contextual economics. Again, our criticism is not that mainstream economics ignores the context. Our main objection is that it arrives at it through the individual. It conceptualizes the environment as a particular constraining factor that has a role in shaping individual behavior. The main constitutive methodological question for contextual economics as a distinct analytical pursuit can thus be stated

as follows: how should contextual economics go about contextualizing individuality without reducing it to internalization of sociality? In order to better answer this question, we will in the following section address in more detail three common-sense assumptions–implicit in the standard approaches–that create a strong pull towards the conceptualization of the environment as a constraint. I argue that these assumptions stand in the way of a more fully contextual approach to economic analysis. Resolving the problems associated with them will point us toward a more distinct theoretical foundations for contextual economics.

3. Three Problematic Conventions Underlying the Constraint-based Views, and How to Overcome Them

Apart from offering powerful analytical tools, conceptualizing the environment as a constraint on the individual action appears to most economists to be a matter of common-sense. However, it is actually not at all obvious why it should be so. In this section I argue that it appears so straightforward because economists are pulled into the constraint-based view by tacitly employing one or more of the following 'common-sense' conventions: (i) the obvious first step in the analysis is to separate and isolate the environmental variables from the individual action; (ii) veridical perception–as reflected in a disinterested representation of a true state of the world–is the gold standard that characterizes the normative ideal; and (iii) the individual brain, resembling an input/output machine, is the natural locus of information processing. Granted, all three conventions are related and overlapping. After all, they all represent a pull towards the same direction of conceptualizing the environment as a type of constraint. Yet, they relate to different aspects of thinking about contextual problems: the first one relates to the analytical procedure, the second one to the ontology of the environment, and the third one to the nature of cognition. I will problematize each of them in turn and discuss how the proposed alternatives can help us in reconsidering the central subject matter of contextual economics.

3.1 Convention 1: The Analysis Starts by Separating and Isolating the Variables

The first problem to be emphasized has to do with the fact that in all of the above-mentioned approaches the individual agent and her environment are assumed to be, at least in principle, analytically separable and independent. The analysis starts with some form of optimization calculus on the level of the individual. Environment is broken down into a set of constraining variables, which are then added to the calculus where they are conceptualized as co-determinants of behavior in terms of their constraining properties. On this view, environment is an *add-on* to whatever is considered a rational economic action. Thus, the resulting calculus can be simplistically represented as: *rational action + environmental constraints + cognitive limitations + narratives + …* In general, all the elements that complicate the clean and frictionless fundamental picture. This 'rational + x' perspective can be most clearly recognized in the standard economic theory, where the assumption of rationality defines the very nature of the individual agent.

Behavioral economics, however, may seem to be anything but a 'rational + x' account, since it was developed primarily through showing the inadequacy of the rationality assumption. Yet, the explicit appeal by behavioral economists to the normativity of the rational choice theory (see, e.g., Tversky & Kahneman, 1986) reveals the underlying relatedness to the neoclassical analysis. Let us again consider the framing effects as an example. Behavioral economists today typically talk about choice architecture, which refers to the particular structure of the environment that is explicitly designed and presented to bring about certain behavioral patterns. Choice architecture can be designed to exploit or to cancel out a particular cognitive bias that is inherent to the psychology of the agent. On a closer look, however, behavioral economists appeal to an *inner rational agent* that serves as an indicator of the individual's true preferences against which the outcomes observed in the behavioral experiments are assessed (Infante et al., 2016). This inner agent

corresponds to the neoclassical image of a fully rational individual, which makes behavioral economics in its essence a 'rational + x' account.[3]

In neoclassical economics, the important implication of the assertion of analytical separability is that the economic action can in principle be decontextualized. Much of the explanatory power of the theory derives from the claim of being context-free and thus readily available to be applied to any specifiable context. Behavioral economists, on the other hand, have showed that behavior is always context-dependent–choice in the supermarket may be rational under one frame and irrational under another, but it will inevitably have to be presented within some frame. However, decontextualization is relevant in behavioral economics as well. Behavioral analysis implies that there exists an abstract essence of the problem–knowable to the observing experimentalist–which can be framed differently in order to measure the framing effects. For example, in the famous 'Linda problem' (Tversky & Kahneman, 1983), different frames may lead the subjects to choose one option or the other. However, what really (normatively) matters are the underlying logical relationships based on the probability calculus.[4] In other words, framing of the problem is imposed on the individual reasoning to divert it away from its logical path.

This discussion touches upon the perennial discussion about the agency/structure dualism in the social sciences, which revolves around the relationship between "active agents and … constraining social structures" (Loyal, 2012, p. 1). However, Collins (2004) argues that this debate mostly "confuses the distinction of micro/macro with the distinction between what is active and what is not" (p. 5), which mistakenly leads the subsequent discussion to be about what it is that propels the activity forward. In standard economics, this is evident by the centrality of the maximization assumption that accounts for the 'energy' propelling

[3] For a more exhaustive discussion of framing in behavioral economics, see Rizzo and Whitman (2020, pp. 69–75).

[4] For a critical discussion of the Linda problem, which emphasizes that people actually use contextual cues to their advantage rather than being merely misled by them, see Hertwig and Gigerenzer (1999).

forward the individual action. But, as Collins (2004) points out, this energy is always about the "processes of real human beings doing something in a situation" (p. 5). Resembling the point made by John Dewey that "[living] in a world means … [to] live in a series of situations" where "the conceptions of situation and of interaction [between individuals and objects and other persons] are inseparable from each other" (Dewey, 1938, p. 43), Collins proposes to start with the situation–rather than the individual–and to focus on the interactions that constitute and characterize it. As he argues, "we get more by starting with the situation and developing the individual, than by starting with individuals; and we get emphatically more than by the usual route of skipping from the individual to the action or cognition that ostensibly belongs to him or her and bypassing the situation entirely" (Collins, 2004, p. 3).[5]

For contextual economics, the move away from the analytical focus on the properties of an individual within the exogenously determined context, to the focus on the situation and the interactions within that situation, promises a viable way forward in overcoming the internalized and constraint-based view of the environment. It points to the importance of changing the unit of analysis. We should study not only the structural characteristics of institutional and social environments but the various social practices, relationships, and interactions that define them.

3.2 Convention 2: Veridical Perception Is the Benchmark

The second convention relates to the first one in that they both involve existence of a one-and-only objective and invariant point of reference. However, rather than stemming from a common-sense intuition about analytical separability of individual

[5] See also Popper (2013[1945]) on situational logic: '"Psychological 'part of the explanation is very often trivial, as compared with the detailed determination of [individual's] action by what we may call the logic of the situation" (p. 308). Ostrom (2010) on the centrality of action situations for the institutional analysis. And Smith and Wilson (2019), who recently raised a similar objection when arguing that their "narrative" approach to experiments is unlike framing, because narratives provide the context in which the decision becomes meaningful. In other words, decontextualized problem is meaningless and cannot be used as a benchmark for the evaluation of framing effects.

action and external influences, the second convention has to do with how perception is generally understood. A view to which many economists tacitly subscribe is vividly captured in the following quote: "We see *correctly*–'veridically' is the official term– to the extent that we manage to emulate the All-Seeing Eye. To the extent that we miss out on this, we live in a state of blindness. To the extent that we deviate from it, we dwell in a state of illusion. Each equally embarrassing" (Koenderink, 2014, p. 2, emphasis in original). This implies that there exists an objective picture of the environment, where objectivity refers to the state of affairs as seen by the all-seeing eye. Furthermore, this objective camera-like picture of the environment defines the benchmark of unbiased perception that forms a basis for unbiased rational action. This pulls the focus of economists into identifying the conditions that cause individuals to fail to meet the benchmark.

The all-seeing eye assumption is rather explicit in behavioral economics, evident by the statements such as that we are "blind to the obvious" and by the frequent analogies with visual illusions (Chater et al., 2018; Kahneman, 2011). Thus, the problem to be explained is the agents' errors in the perception of some obvious and invariant features of their situation (Tversky & Kahneman, 1986). Crucially, these invariant features come to resemble the axioms of the rational choice theory. For example, economic agents are in the framework of behavioral economics considered boundedly rational (or sometimes even outright irrational) if they attach utility to the relative gains and losses, rather than to the absolute levels of wealth as should be the case in the expected utility framework. But, as Rizzo and Whitman (2020) ask, "why shouldn't *expected utility agents* be considered limited or bounded because they are *not* sensitive to gains and losses (except to the extent that they affect absolute levels of wealth)" (p. 32, emphasis in original)? It does not obviously follow from the experimental observations that there is some mechanism at work that limits the perception and rationality. Felin et al. (2017) point out that behavioral economists, following both Herbert Simon as well as Kahneman and Tversky, treat the environment as having "a true, actual nature … which can be learned over time" (Felin et al., 2017, p. 1044). The environment is thus supposed to have an objective and invariant existence that can be defined and measured in absolute terms. Consequently, this objective state of the world as perceived by an

assumed all-seeing eye is in behavioral economics adopted as a normative benchmark for the unbiased perception. Thus, it is central for the evaluation of rationality of actions based on such perception.[6]

But, as Felin et al. (2017) argue, "perception *necessarily* originates from a perspective, or point of view" (p. 1049, emphasis added). Consequently, the environmental cues do not exist independently of the observer. They are, therefore, also not simply given triggers for particular mental models, as proposed by strand two behavioral economics. In order to perceive contextual cues as relevant, agents first need to conceive of them as cues in the first place. This becomes clear when applied to the issue of entrepreneurship. As Lavoie (2015) argues in his discussion on the discovery of entrepreneurial profit opportunities, "entrepreneurship is not a matter of opening one's eyes, of switching on one's attentiveness; it requires directing one's gaze" (p. 59). Perceiving an opportunity as an opportunity means interpreting it as such. And context plays a crucial role in the process of interpretation "because it gives meaning to outcomes" (Smith & Wilson, 2019, p. 159). However, the interpretation does not happen *ex nihilo.* The entrepreneur interprets the profit opportunities by relying on a specific pre-existing cultural framework that enables her to direct her attention. Similarly, Boltanski and Thévenot (2006) claim that "subjects do not constitute the meaning of a scene by the gazes they bring to bear on it" (p. 144) but that what we perceive mostly gets to us pre-interpreted. In their view, people have no choice but to perceive things in certain ways or notice what is relevant in a situation, because interpretation is directed by the particular province of meaning (Schutz, 1962) that the agent relies upon in a given situation.

The emphasis that we simultaneously live and act in different worlds, each governed by its own distinct logic, provides a clear alternative to the all-seeing eye

[6] Gerd Gigerenzer claims that one of the main points that differentiates his proposed alternative to the standard behavioral economics is that in his account heuristics are "functional, not veridical" (Chater et al., 2018, p. 801). However, the emphasis on functionality does not rule out the constraint-based view of the environment.

perspective.[7] Rather than assuming that the problem is how to access the invariant objective reality blocked by a variety of interfering factors–which creates a strong pull toward the constraint-based view–the starting problem of this perspective is to determine the particular interpretive framework that agents are relying upon in making sense of their situation, which influences the very meaning of rational action within a particular logic.[8] Researchers cannot rely on their own common-sense notions of what matters in particular situations (Smith & Wilson, 2019). An opportunity for contextual economics thus lies in utilizing the 'multiple worlds' perspective in order to bring the analytical focus to the interpretive dynamics taking place between and within the plurality of different provinces of meaning that economic agents inhabit.

3.3 Convention 3: Cognition Is a Matter of the Mental Processes in the Individual Head

When discussing the possible reasons for why philosophers and economists are often drawn to the idea of an inner rational agent who is affected by the 'alien' (to him or her) constraining psychological forces, Sugden (2018, p. 66) speculates that this may be due to the tacit acceptance of a particular view of human mind and cognition. According to this view–consistent with the so-called cognitivist paradigm in cognitive science[9]–cognition refers to the processes of symbolic manipulation of the internal

[7] Variations of this perspective are found in Schutz (1962) on multiple realities, Thévenot (2001) on pragmatic regimes, Boltanski and Thévenot (2006) on worlds of justification, and Thornton et al. (2012) on institutional logics.

[8] The institutional logics literature makes it explicit that it is not only about institutions as constraints: "Institutional logics represent frames of reference that condition actors' choices for sense-making, the vocabulary they use to motivate action, and their sense of self and identity. The principles, practices, and symbols of each institutional order *differentially shape how reasoning takes place and how rationality is perceived and experienced*" (Thornton et al., 2012, p. 2, emphasis added).

[9] For the discussion of the core tenets of cognitivism, including its "deliberate decision to de-emphasize certain factors … [including] … the contribution of historical and cultural factors, and the role of the background context in which particular actions or thoughts occur," see Núñez et al. (2019, p. 783).

representations of reality, where these processes are ontologically linked exclusively to what happens in the brain (Newen et al., 2018). In line with this internalist and representationalist view of cognition as a software running on the hardware of the brain, economists usually conceive of the individual agent as an input-output machine that takes the input received from the environment through the senses and combined with the existing preferences, transforms it into plans of action which then direct the preference-satisfying behavior (Ross, 2014a). Psychological realism claimed by behavioral economists comes from separating mistakes and true preferences, but the basic image remains the same: cognitive biases are a case of bad software (Lecouteux, 2016).

The internalist view can be contrasted with the view of cognition as extending into the world rather than being confined to the processes in the head. The so-called extended mind hypothesis (Clark & Chalmers, 1998) is part of a family of related approaches in recent cognitive science that emerged out of the 'immanent critique' of the brain-bound and representationalist view of cognition within the cognitivist paradigm (for overviews, see Wilson, 2002; Robbins & Aydede, 2009) Newen et al., 2018).[10] The central argument of the extended mind hypothesis is the so-called *parity principle*: "If, as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process" (Clark & Chalmers, 1998, p. 8). The classic examples of such extended cognitive processes include paper-and-pencil calculations,

[10] This family of approaches is not entirely coherent and researchers working within the paradigm vary considerably in their commitment to the relative strength of the claims. The perspective is thus far from settled. However, the approaches share a common thread of thinking of cognition as not bound to the mental processes in the head. At the moment, the whole family of approaches is gathered most comprehensively under the umbrella term *4E cognition*, standing for *embodied, embedded, extended,* and *enacted* (see Newen et al., 2018). Acknowledging the emphasis on embodiment, the nuances in the level of extendedness, and the question whether extended mind hypothesis nevertheless still assumes the center of cognition to be in the head, this chapter adopts the term *extended mind* as a colloquial term to refer to the whole paradigm.

where the cognitive process of calculation cannot be separated from the external artefact; and relying on notes for the retrieval of information, where it is argued that the process is essentially the same whether I engage with either the memories stored in my brain or in the notebook.

However, the idea of extended cognitive processes does not apply only to various physical tools and technologies as the vehicles of extension. It applies also to other minds we interact with, institutions, and social practices. This is captured by the notion of *the socially extended mind* (Gallagher, 2013), which suggests that cognition should be looked at "in terms of activities and processes, such as problem-solving, decision-making, judgment, etc." (Petracca & Gallagher, 2020, p. 7). On this view, cognition is extended when the individual agent 'couples with' an institution or practice. In the center of this approach are the so-called *cognitive institutions* that "consist of those practices, rules and structures that have been instituted for cognitive purposes (such as making judgments, making decisions and solving problems)" and without which "specific classes of cognitive processes would simply not exist" (Slaby & Gallagher, 2015, pp. 33–34; for an application to economics, see Gallagher et al., 2019; Petracca & Gallagher, 2020). A typical example is the legal system, which is not just a set of rules governing our actions but a set of institutions that make legal reasoning possible in the first place. By engaging with these institutions through various practices–such as contracting, court procedures, and argumentation–I 'plug into' the system and so extend my cognitive problem-solving ability.[11] In stark contrast to the cognitivist account that conceives of cognition primarily in terms of computational processes, cognition in the (socially) extended

[11] "Contracts are institutions that embody conceptual schemas that, in turn, contribute to and shape our cognitive processes. They are not only the product of certain cognitive exercises but are also used as tools to accomplish certain aims, to reinforce certain behaviors, and to solve certain problems. Institutions of property, contract, rights, and law not only guide our thinking about social arrangements, for example, or about what we can and cannot do, but *allow us to think in ways that were not possible without such institutions.* Insofar as we cognitively engage with such tools and institutions we extend and transform our cognitive processes" (Gallagher, 2013, p. 6, emphasis added).

mind approach refers to a process of active engagement with the (social) environment.

This points to the pragmatic roots of the extended mind approach. For pragmatists like Dewey, cognition is a form of action in the environment (Gallagher, 2009), which implies a more direct engagement with the world than the cognitivist Cartesian vision based purely on inner mental processes. Accordingly, the individual agent perceives the environment directly by way of the potential actions that the environment makes possible. That is, in terms of *affordances* (Gibson, 2015).[12] This suggests that not only various physical tools and technologies but also other people, social and institutional structures, and abstract categories (Felin et al., 2016) can offer possibilities for interaction and engagement as a direct part of the extended problem-solving cognitive process (Petracca & Gallagher, 2020). The pragmatic emphasis on action, environmental affordances, and direct perception also means that much of what we do does not involve any contemplation about the aspects of the environment we are interacting with. As I am finishing this paragraph, I am not thinking about the keyboard under my fingers; I write. When I present a paper at a seminar as part of the development of my ideas, my activity is not the result of processing the mental representations of the academic conventions and the institutional structure of the university; I simply discuss with my peers, which is a result of directly perceiving, and acting on, the social affordances made possible by both the institutional (and social) environment of the university department and the practice of a university seminar.

There are two related implications of accepting cognition as extending into the world, which point to a promising way forward for contextual economics. First, the perspective implies a different kind of interdependencies: individual-environment cognitive couplings. Due to the non-computational and non-representational nature of the (socially) extended cognitive process, the dynamics of such couplings is simply intractable from the perspective of the standard framework. Contextual economics

[12] Pezzulo and Cisek (2016) define an affordance as a "potential action that is made possible to an agent by the environment around it" (p. 415).

could explicitly study the situations and contexts where direct interactions between the individual and the environment are central. The second take-away from accepting this position is that cognitive processes do not refer only to computational processes. For economics, this means that the act of economizing should be rethought. For what is economic action if it is not some kind of calculation of costs and benefits, if it is not maximization within a set of constraints? The following section offers a reconsideration of the role of the environment that utilizes the discussed alternatives to the unit of analysis, to perception, and to cognition, in order to propose a view of the environment not as a constraint but as a resource.

4. 'Context Matters' Reconsidered: Environment as a Resource

In the Section 1 we defined the methodological challenge of contextual economics to be how to contextualize individuality without relying on the reductive internalization–characteristic of social utility function–of environmental and social constraints. In this section we reconsider the 'context matters' dictum in light of this challenge. The aim is to shift the emphasis in the understanding of the role of the environment from a passive constraint on individual action to an active factor in constitution and carrying out of cognitive processes. The argument develops through three stages that progress from conceptualizing the environment as a resource of potential actions; through acknowledging the (often distributed) knowledge embodied in its various components; to, finally, emphasizing environment as providing resources for constituting socially extended cognitive processes.

4.1 Environment Is an Expanding Opportunity Set of Potential Actions

Our discussion in the previous section revealed the important role of the environment in making possible the various affordances–or potential actions–with which individual agents can directly engage in the process of (social) interaction. This is the view of environment as an affordance landscape (Pezzulo & Cisek, 2016), which "can be physical, but also social and cultural" (Gallagher, 2017, p. 174). To

emphasize the economically relevant dimension of choice among various options, I will call such a conception of the environment *an opportunity set of potential actions.*

This opportunity set is, however, far from static or given. Developmental researchers studying human ability to interact with each other have argued that "infants gradually learn about their world through learning the interactive potential of aspects of this world. They learn what they can do with it and how it reacts, that is, what happens as a result of their actions. They come to perceive their world in terms of potential actions" (Carpendale et al., 2016, p. 195). It probably comes as no surprise that discovery and learning would feature heavily in developmental studies, but for us this helps to emphasize an important additional point about the nature of the environment as an opportunity set of potential actions: it is ever *expanding,* based on learning and discovery that happens as agents are acting and interacting within it. When we are acting and interacting in the world, more importantly than to simply provide a set of constraints, the environment serves as a learning ground for discovering potential actions it affords. Thus, the expansion of the opportunity set does not happen as a result of an exogenous change in the structure of the environment but as a result of learning and discovery of the interacting potentialities it offers.

However, being good in social practices and interactions goes beyond only knowing the correct code or script. This means that the learning process is not about uncovering the rules for a more successful interaction but about developing expertise in what Dreyfus (2014) referred to as *skillful coping:* the state of being absorbed in the interactive situation. In the well-known model of skill acquisition, Dreyfus and Dreyfus (2005) argue that rules-based systems cannot capture expertise, "since expertise is based on the making of immediate, unreflective situational responses" (p. 779). Learning to directly perceive the social affordances that are tied to specific situations, and how to exploit them, requires the agent to acquire a certain level of expertise in the (social) practice. Following context-free rules is a sign that the agent has not yet gained understanding of the relevant context and is thus a characteristic

behavior of a novice.[13] This adds an interesting perspective on the characterization of the standard neoclassical economic agent, where the ability to stick to context-free rules is a hallmark of rationality. Therefore, institutions analyzed as a set of rules will only give a very partial answer. Granted, a novice needs rules because she does not yet know what is relevant. In this sense, rules are important on the basic level because they help to stabilize the interactions in a specific context (Linson et al., 2018). But an important implication for economics is that learning the rules alone does not uncover the relevant incentives, because only when the agent masters the skills that enable her to perceive what the environment affords will she be able to fully understand the activity. And such understanding is a prerequisite for interpreting the meaning of the relevant incentives in a given situation.

Besides not being static and given, the environment as an expanding opportunity set of potential actions also does not have any objective existence in the sense of being perceivable by an all-seeing eye. As emphasized above, the pragmatic emphasis on skillful coping when dealing with the world brings attention to the continuous process of learning and developing new ways of engaging with the world. But rather than learning as a process of uncovering true and objective underlying features of the environment, this implies learning as a creative process of discovery of previously non-existent opportunities. As Felin et al. (2016) argue, "uses are not ontological properties of a resource per se, but rather are *attributions of specific actors*, to the extent *actors perceive* resources–as affordances–by means of potential uses that such resources enable" (p. 138, emphasis added). This aspect is obviously important for the study of entrepreneurship (Koppl et al., 2015), particularly as related to the emergence of latent markets (Cazzolla Gatti et al., 2020), since the perception of the entrepreneurial opportunities and the ability to act on them are highly dependent on the level of skill the actor possesses. But it is also highly relevant for our understanding of the organizational dynamics.

[13] "Normally the instruction process begins with the instructor decomposing the task environment into context-free features that the beginner can recognize without the desired skill. The beginner is then given rules for determining actions on the basis of these features, like a computer following a program" (Dreyfus & Dreyfus, 2005, p. 782).

To see this more clearly, we first need to reconceptualize the idea of work from maximizing some objective function to perfecting one's craft (Klamer, 2016). For example, a librarian working within organizational and institutional structure of a library is continuously acquiring particular knowledge of time and place (Hayek, 1945) and gradually becoming more and more competent at her job. As she is perfecting her craft of librarianship, she is also both discovering and expanding the scope of possible actions and interactions within the practice and within what the library as an institutional and organizational environment makes possible, which is leading her to further discover what it means to be a good librarian. But crucially, this process has important spillovers for the organization as a whole. Expanding the opportunity set of potential actions means opening up new lines of work, which in turn creates a positive feedback loop of further specialization and division of labor. Organizational development is thus directly linked to the ability of the organizational infrastructure to afford active engagement and skill development.

We do not only use the environment, however, but also actively create and shape it to serve our cognitive needs. In this sense, institutions are resources that have a particular mode of being produced, maintained, and reproduced (Dekker & Kuchař, 2020), which has a dynamic that is quite different from the standard thinking about the diminishing returns from a fixed factor. The central role of action and interaction in developing and sustaining the institutional resources makes such resources shared goods (Klamer, 2016). The counterintuitive characteristic of such goods–for example, friendships or conversations–is that as you use them, you actually have more of them. Such increasing returns dynamic is central to understanding the environment in an active way as an expanding opportunity set of potential actions.

4.2 Environment Serves as an External Resource of Embodied Knowledge

Skillful action and interaction are not all there is to the active engagement between individual and the environment. The extended cognitive processes crucially rely on "external (and conventional) cognitive schemas and rules … provided by the … institution itself" (Gallagher, 2020, p. 214). These conventions are resources of

embodied knowledge that emerge in the workings of an institution. Let us consider an example. As I walk into a bookstore, I am faced with the problem of how to buy. Simply contemplating the thousands of available titles, potentially on different media, can be an overwhelming experience. However, the spatial layout of the shop, various sorts of lists and rankings, and different product categories that establish and qualify relationships and groupings, all help in guiding me towards solving the problem. As such, they are all vital part of the practice of buying. Context clearly matters because it provides the proper cues, established through previous practice, that help me to solve the problem by simplifying the cognitive load.

This aspect of the environment is explicit in the idea of prices aggregating and communicating the knowledge dispersed throughout the economy, where the knowledge of the relative scarcities and desirability of goods is distilled to a single number (Hayek, 1945). But viable knowledge can also be embodied in various cultural practices, such as the practice of people forming a queue (Hutchins, 2014). Forming a queue goes beyond a mere array of people. A queue clearly indicates the point of service, lets everyone know who comes next, who came before whom, and how long it will approximately take to get to the service. It is not simply an external source of information but a participatory device that actively helps in solving a coordination problem. In a similar sense, a book review is a judgment device that plays a vital role in solving the uncertainty problem in the market for books by serving as a guidepost that helps orient the actions of consumers (Karpik, 2010). Dekker and Kuchař (2020) argue that these judgment devices and cognitive practices are *instruments of interpretation*: institutional elements that are particularly important because they enable interpretation that is "needed to transform [the] institutional sources of information into knowledge that can guide action" (p. 31). As such, they are pivotal in making institutions effective in Lachmann's (1971) sense: as signposts that orient individual plans and form expectations about the future actions of others. On this view, the environment consists of social and institutional resources that we may draw on when we need them to tell us what to do.

However, interpretation having such a central role in the process suggests two problems that need to be addressed. The first is how and why do people understand the devices and practices in particular ways in the first place. Why does

a queue embody the order of arrival? And how do we tell when to ask for a friend's recommendation for a good book and when to consult a critical consensus among the experts? This implies that there exists a more or less tacit (intersubjective) understanding of cultural practices, institutions, and judgment devices. In other words, the world mostly comes to us pre-interpreted and ready for use. This underlies the importance of the processes of socialization and initiation (Gallagher, 2020; see also Smith and Wilson (2019) on the role of *maturation* in the formation of moral sentiments). As Zawidzki (2013) argues, the general relevance of such processes is that they homogenize us as members of the community of interpreters by providing us with the common interpretive framework within which the shared interpretation instruments can indeed serve as vehicles of extended cognitive processes and have coordinating powers.

While stable meanings are important prerequisites for institutions to be successful as coordinating devices, interpretive frameworks, however, are never fixed. The second problem is thus that there is always a latent possibility for circumventing the norms that homogenize the social world, or for applying a different interpretive framework. This means that situations that shake the stability of the accepted meanings will be especially interesting topics of study. In such conflictual situations, the different interpretive frameworks that are employed by the interacting agents are revealed and a reconciliation is required in order to solve the problem (Boltanski & Thévenot, 2006). Importantly, the solution to a problem does not depend on any notion of veridicality with respect to the interpretive framework. Rather, it resembles a game-theoretic convergence: actors must simply reach a common agreement on what the right framework is. In such moments, the important question is not only how people choose the proper framework but what interpretive frameworks are at their disposal. This suggests that another important potential research topic for contextual economics is a study of the unequal distribution of access to available interpretive frameworks.

4.3 Environment Has an Active Role in the Cognitive Processes

As discussed above, the socially extended mind perspective suggests that cognition should be understood in terms of various interactive problem-solving processes that are constituted as the individual agents act and interact within the so-called cognitive institutions. We thus move away from the 'rational + x' model (where the environment is conceptualized as a variable in the individual maximizing calculus) to the idea of the environment and the individual agent as being coupled and entangled to such extent that a particular cognitive process can only take place through this coupling. However, it is not only that institutions have a constitutive role in cognition. New institutional resources are constantly emerging through these actions and interactions. This means that the individual-environment coupling is bi-directional: individual actions and interactions shape the various rules and practices, which, in turn, shape the subsequent actions and interactions. For example, a music chart, which serves as a judgment device for the individual buyer making a purchasing decision, is being simultaneously transformed in the process. The chart thus reflects and embodies decision-making processes of all the individuals that engaged with it up to the present moment.

There are two interesting issues that appear as a consequence of this bi-directional influence taking place through interaction. One is methodological, the other conceptual. Concerning the former, the general challenge of extended cognition approaches based on the individual-environment couplings is that, because the "agent and environment exert continuous and mutual causal influence on each other … agent and environment cannot be modeled as separate systems. They are instead best modeled as a single extended cognitive system" (Kiverstein, 2018, p. 4). Similarly, Ross (2014a) rejects the "descriptive individualism" on the grounds that "our economically interesting preferences … are generated in and by the social and material marketplaces where we interact" (p. 311). And Davis (2016) suggests that, since individuals are constituted (individuated) through their relations to others and their environment, this may call for a reconsideration of the relevant unit of analysis. Individuals and their environment should be perhaps studied jointly. This raises a practical issue of how to actually go about that. The discussion in this chapter suggests that one possible answer to this–and a viable strategy for contextual

economic research–is to shift the focus from studying the mechanics of individual actions and decisions to studying situated social practices in which individuals engage when pursuing their cognitive goals. The concept of social practices implies both individual actions and particular situations in which they take place, yet it is not reductionistic in terms of getting to the environment through the individual. The elements of the environment–such as conventions, judgment devices, and other instruments of interpretation–play within a practice perspective an active role in the analysis and are not subordinated to the individual choice calculus.

The relevant conceptual issue follows from recognizing that the continuous bi-directional transformation implies a central role of action and interaction in the process. With regard to that, Petracca and Gallagher (2020) point out that "once we acknowledge the centrality of social interactions and of the dynamical notion of constitution, institutions are no longer understood, as in Denzau and North and as in Clark, as structures that merely constrain and enable individual actions" (p. 16).[14] But how should institutions be understood instead? The answer suggested by the literature on cognitive institutions is to shift the understanding of institutions from *shared mental models*[15] to *shared mental processes.*[16] This conceptual change shifts

[14] The reference to Clark in this quote is (mostly) about his landmark book *Being There* (Clark, 1997) where Clark, influenced by the conversations he had with Douglass North, develops the notion of *scaffolding institutions*. This notion, however, is in its essence still about institutions as constraints.

[15] "Institutions are the rules of the game of a society and consist of formal and informal constraints constructed to order interpersonal relationships. The mental models are the internal representations that individual cognitive systems create to interpret the environment; the institutions are the external (to the mind) mechanisms individuals create to structure and order the environment" (Denzau & North, 1994, p. 4).

[16] "If we think of the mind not as a repository of propositional attitudes and information, or in terms of internal belief-desire psychology, but as a dynamic process involved in solving problems and controlling behavior and action–in dialectical, transformative relations with the environment–then we extend our cognitive reach by engaging with tools, technologies, but also with institutions. We create these institutions via our own (shared) mental processes, or we inherit them as products constituted in mental processes already accomplished by others. We then engage with these institutions–and in doing so, participate with others–to do further

the focus from analyzing the effects of the institutional structure on the individual agent to understanding of the dynamics of the knowledge embodied in the various institutional elements as the result of continuous actions and interactions.

The answer to our initial methodological question of how to properly contextualize individuality thus suggests two ways forward: on the one hand, contextual economics would benefit from studying situated social practices and other resources for extended cognitive processes as the relevant units of analysis; and on the other hand, the way forward is opened up by conceptualizing institutions as shared mental processes. In both cases, at the center are not the mechanics of the individual action and decision-making but the interactive aspects of the relationship between the individual and the environment. And in both cases, it is required that we shift our understanding of environment from a constraint to a resource.

5. Conclusion

In this chapter I have argued that taking contextualism in economics seriously entails going beyond viewing the environment as another constraining variable in the individual maximizing agent's decision-making calculus. We cannot understand the social world by analyzing the context into individuals' heads. A proper contextualization of individuality implies a move away from the analytical focus on stable preferences as the benchmark for rational action, to the focus on stable situations providing sense-making settings for meaningful action and interaction. Such move enables a much broader understanding of social dynamics that take place as agents interact with each other and with their environment, engage in various problem-solving social practices, and build relationships that define their roles within the processes of social interaction. However, in order to make this move, we have to understand cognition not as an internal process performed by individual brains but as a process that extends across interacting individuals, elements of their

cognitive work. These socially established institutions sometimes constitute, sometimes facilitate, and sometimes impede, but in each case enable and shape our cognitive interactions with other people" (Gallagher, 2013, p. 7).

environment, and the various practices that shape action and interaction. This chapter has demonstrated that context matters because the environment is a resource for extended and shared cognitive processes.

Seeing the environment as having a constitutive role in cognitive processes, rather than as simply being an add-on to the individual action, has important consequences not only for the understanding of interactive and organizational dynamics on the micro scale but also for understanding of the developments on the macro level of society and culture. As the society gets more complex, the opportunity set of potential actions afforded by the social environment expands as well. A complex society involves a broad variety of interactive situations that afford a large number of possible individual-environment cognitive couplings. Contextual economics, as envisioned in this chapter, contributes to the understanding of the dynamics on both micro and macro scales and provides a framework for a comprehensive study of complex social processes.

This new vision of contextual economics transcends the standard institutional critique of neoclassical economics by incorporating the continuous bi-directional transformative influence between the individual and institutional environment; and it transcends the psychological critique by showing how individual psychology and cognition are entangled with the environment through situated social practices and interactions. Contextual economics thus conceived also raises a bigger question of whether we have to give up some of our methodological commitments, such as methodological individualism or the explanatory focus on optimization or rule-following. The conceptual shift from institutions as shared mental models to institutions as shared mental processes entails the impossibility of treating agents and environments as separate systems. However, if the environment is not only helping individuals to achieve their goals or providing the information and rules that they use in their planning but is playing a constitutive role in the cognitive problem-solving processes, then a methodological caution about the relevant unit of analysis is warranted. Thus, the first step when doing contextual economics might be to take seriously the possibility that the properties of an individual agent–be it actual or representative–might not be the appropriate starting point of the analysis.

5

Incentives, Proleptic Reasons, and Intrinsic Motivation: Value Learning as a Discovery Procedure

1. Introduction

Incentives matter. However, as the preceding chapters have demonstrated, human action does not always conform to the logic of this premise. As Tyler Cowen points out, "put up some money and hire someone to make you diet [...] won't make you *want* to lose weight, and building up that desire is usually the only real long-term solution" (Cowen, 2007, p. 3, emphasis in original). Incentives may well lead one to stick to the diet in the short term, but it does not in and of itself teach one to value healthy diet as an activity worthy of being pursued for the sake of itself. This may be an important reason why incentives often prove to be unsuccessful over longer periods. If people understand dieting as a means to avoid financial punishment, then perhaps it simply becomes easier to save money in some other way. This issue has important consequences not only for our understanding of human behavior but also for any incentive-based policy intervention. For example, in the context of educational policy it matters if giving children incentives gets them, on the one hand, to read more books but, on the other hand, fails to teach them to value reading as an activity worthy of being pursued for the sake of itself.

Besides being potentially ineffective, incentives can also have an outright negative effect on the outcomes. To account for such incentive failures, economists often rely on psychological literature on intrinsic vs. extrinsic motivation that explains the underlying mechanism in terms of basic psychological needs, one of

them being autonomy (Frey & Jegen, 2001; Ryan & Deci, 2017). The fulfilment of these basic psychological needs is crucial for the psychological well-being of the organism. According to this literature, incentives, as extrinsic motivators, have a direct negative effect on the perceived sense of autonomy. Extrinsic motivation is perceived as controlling, and incentives backfire because of the organism's natural negative reaction to control.[1] As an often-referenced anecdote reported by Deci and Flaste (1995) goes, when you reward a child for practicing the violin, the child's intrinsic motivation to play the violin will be crowded out by the incentives perceived as controlling. The child starts practicing only easy pieces and starts to become stressed-out by the prospect of failing to get the reward. Therefore, despite the incentives being introduced to support the child in her practice, they may have a negative effect on motivation and may eventually even result in child quitting playing the violin altogether.

But something is missing in this explanation. It does not explain why the child is playing the violin in the first place. If the goal is to fulfill the innate psychological need for autonomy, there may be other ways for achieving this goal. Why is the child playing violin and not juggling, or doing any other autonomy-supporting activity? Socialization is by many deemed suspicious because it suggests that social conditioning impairs the true inner self. But people are indeed creatures of their social, historical, and geographical context. One may find ski jumping intrinsically satisfying, but this activity would be impossible to be even conceived of were one to be born in a seventeenth century Middle Eastern peasant family. Psychology provides rich accounts of the underlying experience and mechanisms of

[1] It is thus not surprising that these authors explicitly motivate their research program in terms of a contribution to the project of human liberation: "Our overriding, sociopolitical interest is examining the possibilities and obstacles for human freedom. In our thinking, this pertains not only to social, political, and economic structures, but also to internal psychological structures that reflect and anchor the external ones. It is our hope that, by engaging in a serious investigation of motivational issues, we can make some small contribution toward the larger goal of human freedom" (Deci & Ryan, 1985, p. vii).

valuation, but it fails to help us understand why any *particular* value, how they are acquired, and where they come from.

In this chapter I argue that instead of looking within, we should rather look outside to our social and institutional environments and how we interact with them in the process of valuation. My intent is not to argue against the insights from psychology but to demonstrate that the neighboring disciplines of sociology and philosophy offer rich accounts of learning and acquiring new values that can help economists to better understand why and under what conditions people choose to engage in one activity rather than another. Furthermore, this chapter will provide a new way of accounting for the formation of long-term commitments (as opposed to a short-term change of behavior) and thus contribute to the understanding of the workings of incentives and incentive-based policy interventions. Understanding the formation of such commitments has important implications for thinking about welfare consequences of interventionist policies designed to achieve some desired behavioral goals.

The chapter proceeds as follows. In section two I will present the outline of the proposed theory and its basic building blocks. Section three will address the main features and implications that follow from this new theoretical approach. These will be brought together and applied in section four where we will use the theory to reconcile the two critical approaches to what has become the predominant view of the welfare implications of behavioral findings. In the conclusion I will draw some further implications of the new approach to argue that the approaches built on preferences and choice are often misguided in accounting for human motivation.

2. Value acquisition: building blocks of a theory

2.1 Frank Knight and his legacy

Economists mostly accept that economics as a discipline is about exploring different ways to allocate scarce means among given ends. Where they differ is in recognizing the limits of such a method. One of the most remarkable expositions on this issue comes from the early Chicago economist Frank Knight. After establishing that scientific economics is indeed about rational economizing, he continues by asking the

question of "how far life is rational, how far its problems reduce to the form of using given means to achieve given ends," to which he gives a perhaps atypical answer for an economist: "not very far" (Knight, 1935, p. 105). The recognition of the stark limitations of the method of his own scientific discipline is interesting in itself; but for our purposes what is truly remarkable is how he continues:

> "The scientific view of life is a limited and partial view; life is at bottom an exploration in the field of values, an attempt to discover values, rather than on the basis of knowledge of them to produce and enjoy them to the greatest possible extent. We strive to 'know ourselves,' to find out our real wants, more than to get what we want" (Knight, 1935, p. 105).

Knight demonstrates the limitations of economic science by arguing that problems of economizing are but a tiny fraction of what life is actually about. Rather than striving to maximize utility through preference satisfaction within the constraints we are facing, we strive to develop new and better preferences, to acquire new and better values[2]. However, despite encompassing most of what life is about, value acquisition for Knight does not belong in the domain of economics because it is not a rational process. What he seems to suggest, though, is that economists should be humble with respect to the applicability of their methods to the study of human behavior in general.

This lesson was not absorbed by later generations of Chicago economists. Gary Becker became the poster child of economics imperialism with his approach that treats preferences as given and aims to analyze any changes in behavior based exclusively on the perceived changes in relative prices (Stigler & Becker, 1977). The so-called 'economic approach to human behavior' (Becker, 1976) not only eschews value formation from the domain of economics (Knight also did that); it altogether

[2] On the point that, for Knight, the problem is not only a change in preferences but indeed an acquisition of qualitatively 'better' ones, see Knight (1935, p. 101).

bypasses the problems of accounting for value change by proclaiming it irrelevant for explaining human behavior. Economists can explain any kind of individual behavior simply by treating values, tastes and preferences as given and stable across population. According to this view, life is at heart indeed about maximization.

Parallel to these imperialistic developments, the legacy of Frank Knight was revived by his student James Buchanan whose essay *Natural and Artifactual Man* (Buchanan, 1979) grew out of a referee report for the above-mentioned famous paper by Stigler and Becker.[3] Rather than assuming the man to be a rational maximizer, Buchanan acknowledged an aspirational nature of man. Where for Stigler and Becker addiction to music is simply a rational reaction to the increasing marginal ability of listening to produce a desired state of music appreciation, Buchanan sees it in terms of a struggle between what the person is and what she imagines herself being capable of becoming. And while Stigler and Becker cannot explain why increasing music appreciation would induce a switch from rock to classical, for example, Buchanan's artifactual man is able to imagine changes in tastes, rank them as better or worse, and so, given the liberty to do that, "become the man he wants to become" (Buchanan, 1979, p. 259). Contrary to both Knight and Becker, Buchanan sees the study of preference change as a necessary component of economics. As he warns us, "the utilitarian origins of nineteenth-century political economy may have come to haunt us and to do us great damage" (Buchanan, 1979, p. 250). The inquiry into value acquisition should indeed have a prominent place in economics.

Buchanan's notion of an artifactual man who aspires to change his preferences assumes that this man can evaluate *ex ante* the prospects of changing values. As suggested by the ability to rank the prospective changes, this man is said to have preferences for preferences he wants to hold. When discussing the issues of intertemporal choice, Brennan and Buchanan (1985) point out that "some futures must be deemed better than others, and choices in the present will tend to reflect these preferences" (p. 71). This is further complicated by the fact that one does not know if one's future self will endorse the preferences that the present self deems more

[3] This story is reported in Lewis and Dold (2020).

desirable. In Buchanan's framework this dilemma is resolved by assuming that those new preferences can be fully known and grasped in the present, and so any future self that potentially fails to endorse them will do so based on a lack of adequate reflection (Brennan & Buchanan, 1985, pp. 71–73). Buchanan's artifactual man is thus characterized by the ability to pull himself up by his own bootstraps (on this see also Lewis & Dold, 2020).

There is also a more subtle point to be made with respect to Buchanan's normative assumptions. Buchanan rejects instrumental defense of liberty, according to which people want to be free to maximize some form of utility. Rather, he proclaims that "man wants liberty to become the man he wants to become." But in characterizing the nature of aspiration, he follows Knight in assuming that central to it is a "tendency to want better things, to become a better man" (Buchanan, 1979, p. 251). Given that aspiration is in his framework guided by second order preferences, this means that people naturally aspire to become qualitatively better than they are. And crucially, what is 'better' and what is 'worse' seems to have in this framework an unambiguous moral content: when granted liberty, people will demonstrate that deep down they are not evil but good and cooperative.[4] Perhaps somewhat ironically for Buchanan, and against his own assertion in the same paper that "the economist […] has no justification for building his theories on romantic notions of man that will not stand empirical test" (Buchanan, 1979, p. 255), this reveals a rather romanticized view of human nature.

2.2 Learning the meaning of the activity

So, we have established that the change in values and preferences is what needs to be explained, but that it can't be explained by referring to yet another instrumental relationship based on preferences about preferences. We need a different way of

[4] This is not unlike the view from the literature on intrinsic motivation, where violence *by definition* cannot be part of autonomous intrinsically motivated behavior. To the contrary, a truly free person is basically inherently good. 'Bad' motivations are a result of introjection and moral disengagement, not real intrinsic motivation. Violence and doing harm to others is need thwarting and cannot promote true autonomy (see Ryan & Deci, 2017, pp. 635–644).

accounting for the change. Means-ends instrumentalism is often contrasted with intrinsic motivation, or motivation to engage in the activity as an end in itself. For example, the child might practice the instrument in order to get a reward, but she might practice simply because she finds playing to be an enjoyable activity. According to psychologists, enjoyment is one of the key experiences that characterize intrinsic motivation (see Ryan & Deci, 2017, p. 193). And in economics, Tibor Scitovsky argued that enjoyment, and not preference satisfaction, should be at the center of welfare considerations (Scitovsky, 1992; 1986). But how do people get to the point at which something feels 'intrinsically enjoyable'? How so that the child may feel enjoyment while practicing the violin? Sociologist Howard Becker argued that deriving pleasure and joy from performing the activity does not depend on any internal traits of the person but is chiefly a function of learning how to enjoy it with the help of an experienced mentor. As he put it,

> "the presence of a given kind of behavior is the result of a sequence of social experiences during which the person acquires a conception of the meaning of the behavior, and perceptions and judgments of objects and situations, all of which make the activity possible and desirable. *The motivation or disposition to engage in the activity is built up in the course of learning to engage in it and does not antedate this learning process.* For such view it is not necessary to identify those 'traits' which 'cause' the behavior. Instead, the problem becomes one of describing the set of changes in the person's conception of the activity and of the experience it provides for him. […] On completion of this process he is willing and able to [engage in the activity] for pleasure. He has learned, in short, to answer 'Yes' to the question: 'Is it fun?'" (Becker, 1953, pp. 235, 242, emphasis added).

Enjoyment thus arises from the learning process, from the process of *acquiring a conception of the meaning of the behavior*, as Becker puts it. This is a big step forward from simply observing that some activities feel intrinsically motivating to some people. Becker's account demonstrates that such observation must be preceded

by a process of learning that enables the individual to experience it in a way that feels enjoyable, or autonomous, or intrinsically motivating. And, perhaps most importantly, it shows that this learning is distinctly social in nature. It suggests that the conception of any behavior that feels intrinsically motivating must first come from the outside, so to speak. In this sense, the question of enjoyment and intrinsic motivation is in fact a secondary question. The primary question, at least for a social scientist, is the availability of these external resources for learning. Is there a suitable role model in my environment from whom I can learn to enjoy and appreciate certain activity? If we want to answer why the child is playing the violin, and not doing karate or juggling, we need to account for the process that led the child to be able to find playing the violin enjoyable in the first place.

Drawing on sociological insights requires a caveat with respect to the role of the social structures. For instance, it has been recently argued by Hayes (2020) that the work of Pierre Bourdieu on habitus and practice offers a viable alternative to the exclusive focus on individual cognition usually found in the work of (behavioral) economists. This is a welcoming contribution, because it emphasizes the aspects of action and decision-making that are not cognitive in the sense of individual psychology but are cognitive in a more extended sense of including the various social structures into the cognitive process. Social structure has an important effect on what and how practices are performed by the individual. However, we need to be careful not to get carried away too far to the side of structural determinism and claim that *all that matters* is the social structure. My position here (as well as in the previous chapter) is that individual agency matters, but that this agency cannot be conceptualized as isolated from its environment. The important lesson that economists can learn from sociology is that, in studying motivation, we should indeed pay much more attention to practices that people engage in, how learning happens within those practices, and how that learning enables the conception of the meaning of the activity. Incentives matter to the extent that people find them meaningful (Dekker et al., 2020).

2.3 Aspiration and proleptic rationality

A framework that helps us reconcile the issues raised so far has been recently proposed by Agnes Callard in her book *Aspiration: The agency of becoming* (Callard, 2018). She builds on a distinction between 'valuing something' and 'believing that something is valuable.' It is much easier for us to believe that something is valuable than to in fact value it. This is because "in order to value something, we must engage with it in a way that takes time, effort, and practice" (Callard, 2018, p. 117). Callard calls this engagement an act of *aspiration* or *value acquisition*. For example, one may believe that children are valuable, but one will only be able to fully grasp this value by engaging in an ongoing practice of child rearing. Only such engagement enables the eventual discovery and acquisition of the value of having children. However, this does not mean that a person without children cannot in any meaningful way make any rational decisions about whether to become a parent or not. As discussed above, Knight argued that there are limitations to the science of economics because the process of acquiring and discovering new values is in fact not a rational process and it is thus outside the scope of economic thinking. Callard, on the contrary, argues that aspirational process of striving for new values is indeed a rational process, albeit of a different kind. In the process of aspiring to acquire a specific value we act in a rational way given that we actually do not yet have the values or preferences that we strive to acquire. Callard labels this kind of rationality *proleptic rationality*.

Proleptic rationality is based on proleptic reasons, which are "defective variants of the reasons [the aspiring agents] will come to grasp fully at the end of their transformations" (Callard, 2018). Callard's account of aspiration is thus built on a completely different foundations than Buchanan's. While Buchanan assumes that the aspiring agent relies in her aspirational pursuits on the choices based on higher-level ordering of preferences about preferences, Callard is clear that "proleptic reasons are not internal reasons—they cannot be arrived at by sound deliberation from what the agent already cares about. Instead, they reflect the possibility of rationally coming to care about something new" (Callard, 2018). The idea of proleptic rationality emphasizes that values are not only things we can reason *from*. By acting on proleptic reasons we can also reason *towards* them.

But if these proleptic reasons do not come from within the person, from the values that the person holds, or from some second-level ordering about what values one should acquire, where do they come from? Callard argues that "on an aspirational account of self-creation, the creator does not determine, choose, or shape the created self; rather, she *looks up to, imitates, and seeks to become* the created self. *The source of normativity lies at the end of the process* rather than at the beginning" (Callard, 2018, p. 13, emphasis added). Sources of proleptic reasons can thus be found in mentors and role models whom we look up to and imitate; activities and institutionalized practices that help us to structure and model our actions; and even the process of competition that provides feedback to our actions, such as in the case of the proverbial watercooler conversation where we bounce our ideas off each other and, in the process, discover their value. The bottom line is that for Callard valuation is a social process and not a matter of individual cognitive activity.

3. Learning Values

3.1 The case of weakness of will

For economists in general, the notion of aspiration is probably the most relevant in the context of intertemporal choice, specifically when this choice is prone to be undermined by time-inconsistent behavior. Consider the example of a New Year's resolution to start eating healthier. In this case, the present self makes an intertemporal choice on behalf of the future self. However, as time goes by and the individual gets confronted by opportunities for unhealthy but attractive food choices, the initial choice may lose its power and the individual may succumb to temptation. This is the ancient problem of akrasia, or weakness of will, which is commonly described in terms of 'acting against one's better judgment'. The central economic problem here is that a person will reveal different preferences at different points in time. As behavioral economists have demonstrated, preferences are context-dependent, which presents a particular challenge to standard welfare economics where an outcome is thought to be welfare promoting to the extent that people's preferences are satisfied.

The developments in behavioral economics have recently given rise to an increased interest in the normative implications of this research. Time-inconsistency, or the disagreement between the present and the future self about what action to take, is in this normative version of behavioral economics usually framed as a problem of rationality. This is because failing to act in one's best interest is thought to be a consequence of the individual being either ill-informed or having her preferences distorted in some way (Hausman, 2012). So, if the individual seems to act impulsively and against some well-considered judgment made in the past, this means that there is either something wrong with the available information, or preferences got distorted by some behavioral bias (what Hausman refers to as a decision-making flaw), such as weakness of will. Revealed preferences are thus said to be context dependent. The key feature of these accounts is that the degree of distortion of the revealed preferences is established in comparison to the benchmark of decontextualized preferences as conceptualized in the standard rational choice theory. Thaler and Sunstein (2008) define a bad decision as a decision that people "would not have made if they had paid full attention and possessed complete information, unlimited cognitive abilities, and complete self-control" (p. 5). Compared to this benchmark, weakness of will is clearly a failure. It is a consequence of cognitive limitations that prevent one from acting rationally by exercising self-control.

However, as Reinhard Selten has pointed out, cognitive limitations are not the only possible factors that shape the behavior. It may well be that "the lack of complete control over behavior is not due to cognitive bounds of behavior, but rather to motivational ones" (Selten, 2002, p. 12). One may be fully convinced that sticking to the New Year's resolution is rational both in the sense that it is consistent and that it maximizes health outcomes; and yet still accept a piece of cake and a glass of cognac at the end of a fine dinner in a good company. The issue is thus not rationality but motivation. Similarly, the effort of the violin student does not drop after the incentives have been introduced because of the inability to act rationally, but because introducing the incentives has taught her something new about the activity (i.e., that what is rewarded is simply playing as such, not necessarily learning new and challenging pieces), which in turn affected her motivation. As Howard

Becker would put it, a new conception of the meaning of the activity has been acquired.

Contrary to viewing weakness of will as a decision-making flaw that violates some benchmark of rational choice, Agnes Callard has argued that akrasia is a consequence of the conflict between several evaluative systems within the individual, which she calls intrinsic conflict. For example, I can at the same time wrestle with two competing desires: one to be a good husband by spending the evening talking and watching a movie with my wife, and another to be a good husband by finishing the work that will bring in the necessary income to support the family. These two ways of conceptualizing what being a good husband means belong to two different evaluative frameworks. One built around the logic of companionship and emotional support, and the other around the logic of contributing to the material security of the family. In this regard, it is not possible to say whether I commit a decision-making flaw if I opt for an evening on the sofa rather than in my home office chair. What is possible, though, is to say whether I aspire to act from one of the two.

When akrasia happens, we may indeed experience it as something that goes against our better judgment. However, contrary to the all-things considered judgments, Callard argues that we cannot deliberate on all the reasons and grade them. This is because the relevant value has not been acquired yet. In my case, I aspire to be a good husband by having a somewhat vague idea of what being a good husband feels like. This idea is facilitated by the proleptic reasons that I have available in my environment. But I do not yet grasp this value fully, because being married is an ongoing process of discovering what it means to be a good husband. Callard's central insight is that persevering in moments when values clash is guided by proleptic reasons. We stick to the activity because we aspire to possess a certain value, not because we already possess that value (and then fail to act on it). If we would already fully possess that value, the moment of doubt would not have arisen in the first place. It would be meaningless to doubt. Doubt has a meaning because there is an underlying clash in valuations. What Cowen (2018) calls 'stubborn attachments' must be stubborn precisely because there is a clash in values. Therefore, given that we want to stick to the plan, it is of the utmost importance in such

moments that we can follow the proleptic reasons, and not follow the actual reasons dictated on the spot by whatever our current preferences seem to be.

However, sometimes I will act based on one evaluative framework despite me aspiring to act from another. In other words, my context-dependent reasons to go and work will differ from my proleptic reasons to be a supportive husband. And yet, this does not mean that I 'irrationally' acted against my dominant value, only that I did not act according to what I proleptically aspire to be my dominant value. The fact that I am an aspirant means that I am unable to unambiguously rank the context-dependent and proleptic reasons for acting one way or another. Weakness of will is thus a result of a not-yet-fully-grasped value. In general, resolving intrinsic conflict is a matter of grasping the value and thus establishing a clear relative status of values. This, however, is not a matter of will and deliberation but of learning the value through an ongoing practice and engagement, where my grasp of the value is shaped by my role models, ways of doing things, stories, but also by the course of events.

3.2 Value learning as a discovery procedure

So far, we established that the way people grasp the value of the activity is an important factor when we want to understand why people do what they do, because simply observing that something gets done is not enough. If the problem of incentives is that they do not teach the value of the activity, this suggests that even if incentives can be useful in overcoming the problems of self-control, or the problems of motivation for things such as education where benefits are harder to observe in the short term, they are unsuccessful in creating long-term commitments based on genuinely held values about the worthiness of these pursuits. We need to investigate the underlying learning processes of value acquisition.

I want to emphasize learning because in the existing economic accounts of changes in preferences and values learning has been largely undertheorized. The account of learning the meaning of the activity discussed above gives us an insight into the process that has been in economic literature discussed in terms of cultivation of tastes (McCain, 1979, 1981) or endogenous preferences (Bowles, 1998). These economic accounts are important contributions to our understanding of the general

conditions under which values, tastes, and preferences change. However, the accounts of learning that they offer are rather thin. Since the emphasis is on formal modeling of the general mechanisms, the learning component is often taken for granted. Consider, for example, how Bowles and Polanía-Reyes (2012) talk about the possible long-term changes in taste for cuisine after living in Rome for a period of time:

> "Which case it is–state-dependent preferences or endogenous preferences–would be revealed by what we will eat back in Bogota or Santa Fe. If we go back to arepas or potatoes, then our taste for pasta was state-dependent. If we remain pastaphiles, then our preferences have endogenously changed" (Bowles & Polanía-Reyes, 2012, p. 375).

Here, the emphasis is on discovering behavioral discrepancies and for this purpose it suffices that learning is simply assumed to be some black box function of living in Italy. But the danger is that these approaches may lead one to think of learning simply as an automatic and passive process of being exposed to something. Compare this to how Agnes Callard talks about a similar case:

> "Given the expertise and work involved, it is implausible that anything but the earliest stages of such transformation can be explained through fully external factors. For instance, the fact that someone found himself, for incidental reasons, in the exceptional gastronomic environment of Osaka, Japan, might be the beginning of the story. Those experiences could ignite a spark of interest, but then something more would be needed to drive someone's systematic development of that initial spark into a full-fledged passion. The 'something more' in question is unlikely to be a value to which he was antecedently committed, from which a passionate interest in culinary excellence could be derived" (Callard, 2018, p. 207).

Note how Callard emphasizes the *work* and *expertise* that are involved in acquiring a new value. This aspect is absent in the Bowles' account of endogenous preferences. While, in Bowles' framework, context-dependency means attributing the main role to the working of the individual mind reacting to the changing circumstances, endogenous changes in preferences are often attributed to the corrupt influences of 'the system'. In other words, endogenous preferences seem to be a consequence of passively copying and internalizing some aspects of the surrounding environment. However, the reason for preferences change is not that the individual is simply exposed to the effects of the social, institutional, or cultural environment. To the contrary, these effects, as Callard argues, will tend to be rather minuscule if not accompanied by aspiration and ongoing practice. In this sense, changes in preferences and values require interaction between the individual mind and the environment, where none of them is a dominant factor.

Expertise and work are closely linked also because learning a new skill is not simply a matter of executing a learned script. In a well-known account of expertise discussed in the previous chapter, Dreyfus and Dreyfus (2005) have argued that simply following a set of rules is a sign of a novice, whether these rules are conscious or unconscious. Expertise must be built up through practice and is characterized by flexibility in the application of skills, a quality of action that Dreyfus (2014) calls skillful coping. Crucially, these skills are often tacit and cannot be codified. Therefore, becoming an expert in any activity implies picking up these tacit skills through engagement in an ongoing practice, with a principal emphasis on mentorship, imitation, and trial and error as sources of proleptic reasons.

Learning can also be facilitated by competition. Hayek famously argued that competition is a discovery procedure because it helps us to discover relative values of things by enabling market prices to emerge (Hayek, 2002). But competition not only enables us to discover what people are willing to pay for some product; through competitive process we also discover new product categories, ways of doing things, and relating to others. Dekker and Kuchař (2016) demonstrate that competition enables the discovery of new qualities of goods where these qualities then serve as exemplars around which market activity can be coordinated. For example, in discussing the case of Starbucks they emphasize that the outcome of the competitive

process was not simply that Starbucks managed to lower its input costs, but that "the product and the way it came to be understood by consumers, and hence how it was valued was the key aspect that changed" (Dekker & Kuchař, 2016, p. 246). Consequently, they argue, the process of competition enables consumers to learn about new alternatives. This learning about new qualities will, in turn, expand the available sources of proleptic reasons. It will, for example, enable people to aspire to become a coffee connoisseur.

Common to the accounts of learning discussed in this section is that learning is not simply a matter of information processing by an individual mind (as in cognitivist psychology) but happens through active engagement with the resources available in the (social) environment. As extensively discussed in the previous chapter, such resources present themselves to an agent in form of what Gibson (2015) called affordances. An affordance is "a potential action that is made possible to an agent by the environment around it" (Pezzulo & Cisek, 2016, p. 415). For example, a chair affords sitting. But if needed, it can also afford standing. Crucially, the meanings of affordances for the agent are not objectively preexisting. They depend on the type of embodiment (a chair affords something else to a bird than to a human) and on the context (am I looking for a place to rest, or do I want to reach the box on the top shelf?). But they also depend on the parallel developments in the wider environment that may create new economic possibilities, the process that Cazzolla Gatti et al. (2020) describe using a biological evolutionary metaphor of the opening of new economic niches. For example, a metal rod was equally capable of receiving radio signals 1000 years ago, but that niche was not opened yet and so the affordance of receiving radio transmissions was not being able to be perceived as such. Learning is what creates an affordance, both in terms of learning from others and in terms of a creative discovery of new alternatives made possible by the evolving situation.

This also means that value acquisition encompasses discovery in the sense of genuine creation. It is often assumed, as the discussion in this chapter has demonstrated, that discovery of values is a matter of uncovering one's 'true self', be it in the sense of unconstrained intrinsic motives, latent preferences, or second-order preferences about who we want to become. In contrast, I argue that answering the 'why' question of human behavior is a creative act. Changes in behavior are a result

of the changes in the conception of the meaning of a behavior. And these changes in conception are the result of a learning process. It should also be added that any newly grasped value can be contested at any time. The ongoing emergence of new alternatives that the environment affords will inevitably make sure that the process of valuation is also ongoing and that values are never completely fixed and stable. But this also means that it matters greatly what kind of environment we interact with. A 'creative' environment with lots of new niches and possibilities opening all the time, will afford more opportunities for proleptically learning new values than a stale environment with few sources of proleptic reasons (Koppl et al., 2015).

The emphasis on learning as an ongoing process implies that one must sometimes act even though the value may not be fully acquired yet. As Callard (2018, p. 176) puts it, "the learning of a value takes time, and sometimes the decision has to be made now." Therefore, short term changes in behavior that resemble weakness of will are explained by not yet having a full grasp of a value. The individual has not yet fully learned to perform the activity for the sake of itself, or for the sake of pleasure, or any other aspirational goal. Understanding the process of grasping a new value is crucial. Without it, we cannot explain why one activity, and not another.

3.3 Value conflict, not choice: learning to live with multiple values

The issues discussed so far are in the scholarly literature predominately in the domain of decision theorists that study the conditions under which a choice might be more or less optimal in terms of bringing about a more or less optimal outcome. Economists are mostly attracted to the work of decision theorists because economics is said to be the science of choice. Economists study and assess the costs and benefits of a potential action as inputs into the calculation that aims to show what the optimal choice would be in given circumstances. And in normative terms, the debate is often focused on the issue of autonomous choice on the one hand, and paternalistic interventions into the choice process on the other. But how is a science of choice possible if a value has not been learned and there are multiple evaluating frameworks available? Callard explicitly argues against the decision theoretic approach to the study of value acquisition by pointing out that "the problem is that because one (or

both) of the options promises a substantial change in preferences, the agent doesn't have a single, stable set of preferences that could provide the input for the decision procedure" (2018, pp. 41–42). The focus on choice misses an important aspect of this situation.

The implications of such simultaneously existing multiple evaluating frameworks have been explored under the banner of heterarchy, defined by Dekker and Kuchař (2017b, p. 1) as "an order with more than one governing principle." These principles are competing and conflicting, without one clear dominating principle on which to decide. Thus, the notion of heterarchy emphasizes the conflict between several mutually incompatible but internally well-ordered modes of valuation. It shifts the focus from the challenge of making an optimal choice to the challenge of resolving value conflicts. For example, in a case of a family firm, heterarchy involves competing values such as the ones that govern family and competitive market. A family member is at the same time an employee, and while family logic would suggest that each of the children is equally deserving, market logic requires that they get paid based on their marginal product. This can naturally result in lots of tensions among the family members. The ongoing process of negotiating interactions between the different conflicting logics aims at resolving some of the tensions by enabling family members to discover and learn ways to make trade-offs between, for example, the logic of profit and the logic of family loyalty. In other words, the challenge is to make commensurable what is essentially incommensurable. Crucially, this is not a frictionless or 'rational' process, because there is no clear recipe. It is a process that "requires effort" (Dekker & Kuchař, 2017, p. 2). This effort entails discovering and learning how to meaningfully compare the overlapping modes of valuation. Callard (2018, p. 4) agrees: "Grasping new values is work." The outcome of this process is an expertise in negotiating different value systems, be it on the level of an individual negotiating between different simultaneously existing values but also on the level of organizations as in our family firm example.

This discussion also involves a paradox. On the one hand, commensurability is generally thought of as freedom-promoting because it makes possible to assess the opportunity costs and thus easily change the course of action. For example, this

would be the case when I switch from violin to juggling seamlessly and still entirely meet my inner need for autonomy. But on the other hand, incommensurability (the fact that I can't simply switch at will) also opens new opportunities by making certain activities possible. As Cass Sunstein argues,

> "Incommensurability is not just a barrier to action; it also makes possible certain valuable human relations. Without incommensurability, our understanding of what a friendship is, or what it means to be a parent, or what a beach is, would be compromised badly. So, we ought not to think of the existence of incommensurability as simply an obstacle. It is also freedom-promoting, in the sense that it makes possible certain valuable human connections and relationships" (Sunstein, 1994, p. 1667).

This seemingly paradoxical relationship between commensurability and incommensurability matters for economic analysis because it suggests that sometimes we cannot–and indeed should not–fall back on a particular justification based on objectively weighing and ranking alternative courses of actions. The standard rational calculation clearly has limits. Much like Knight's view discussed above, this line of thinking has recently led Robb (2019) to conclude that some decisions and actions must be based on a leap of faith rather than rational calculation. But such conclusions still stem from the exclusive focus on choice as the main issue. They don't recognize the centrality of the need for ongoing aspirational work and resolving the conflict among incommensurable values. And they don't answer the question of whether there is a sense in which the conditions of heterarchy or incommensurability can be said to be preferable to conditions dominated by a single commensurating principle.

When looked at from the perspective of constrained maximization as an example of such single principle, heterarchy indeed looks like a breeding ground for logical contradictions and inefficiencies. However, Dekker and Kuchař (2017b) argue that tensions between the competing values can also be productive because a heterarchical system can exhibit more resiliency and more innovativeness than a

hierarchically organized one. Why this may be so is empirically an underexplored subject. But the discussion in this chapter suggests that one of the reasons may be that the ongoing need for conflict resolution implied by heterarchy requires an ongoing learning of new evaluating frameworks. In some respects, this may be less efficient than a constant application of a single evaluating principle. This probably explains why firms focus on profit maximization calculus rather than on the value conflict. It involves far less frictions because the evaluative principle is clear. But as David Stark has demonstrated in his work on heterarchy in organizations, "friction is not always a bad thing" (Stark, 2017, p. 387) and innovative action often stems out of the situations characterized by conflict of values. As he puts it, "it is when things do not fit together comfortably that novel recombinations become thinkable" (Stark, 2017, p. 388). Thus, heterarchy and incommensurability may indeed make organizations both efficient and creative by making the individuals interacting within them more adaptable and by stimulating learning.

4. Aspiration and the behavioral foundations of welfare economics

4.1 The dispute over the appropriate welfare criterion

As pointed out above, the standard view of the normative implications of the findings from behavioral economics research is based on the idea that acting in one's best interest is often inhibited by decision-making errors. These errors are induced by cognitive and informational limitations that prevent one from maximizing welfare by acting according to one's 'true' underlying preferences. The appropriate welfare criterion is thus the satisfaction of these latent, context-independent, well-considered, or consistent preferences. The challenge for policy makers is then how to help the 'planning' self to overcome the temptations of the 'doing' self (Thaler & Sunstein, 2008).

For more than a decade by now, Robert Sugden has been criticizing this standard framework of behavioral welfare economics, which he has labeled 'the new consensus' (Sugden, 2018). Specifically, he has targeted the idea that context-dependency of preferences implies a hierarchical ordering of preferences where the latent preferences reign superior over context-dependent ones in terms of welfare

maximization. Sugden points out that the new consensus clearly takes rational choice theory with its consistency axioms as the normative standpoint on which to identify those preferences whose satisfaction maximizes welfare. Thus, Sugden notes, it relies on a psychologically untenable notion of an inner rational agent residing in a constraining psychological shell (Infante et al., 2016). His critique is based on the central premise that what counts, and is indeed psychologically plausible, is the preference ordering in the moment of choosing. Preference satisfaction is central for Sugden as well, but he argues that there is no way to say that preferences over the future states held by the present self are more important than the preferences of the future self that will be revealed in the moment of the actual choice. Consequently, the appropriate welfare criterion must be the opportunity of available alternatives from which to choose, regardless of which context-dependent preferences the agent wants to act on.

Despite Sugden's claims of the psychological plausibility of his conception of the choice process, his account is based on a rather curious assumption that people are responsible agents in the sense that they at any time endorse all aspects of their psychology as their own, as part of who they are. His argument for the opportunity as the proper criterion of welfare relies heavily on the claim that people, in evaluating their own welfare, only care about the freedom to act on preferences they happen to have in the moment of choice. There is no sense in which a person might feel that there are parts of her psychology she would rather be without, or that she is suffering from self-control problems that feel alien to her. To the contrary,

> "the responsible agent asks of government only that it ensures him as wide a range of opportunities as possible. How he uses those opportunities is up to him, and he accepts sole responsibility for the consequences. He has no need to explain the decisions he has made, because they were no one else's business. And because they were his decisions, he can have no complaint against anyone else about how they turn out" (Sugden, 2018, p. 106).

In Sugden's view, time inconsistent behavior by a person is simply an instance of her holding two different context-dependent preferences. According to him, "this is not a self-control problem; it is a change of mind" (Sugden, 2018, p. 82). In this sense, Sugden's agent seems to be overly impulsive and clearly lacks any aspirational element.

A number of scholars have recently critically engaged with Sugden on these points. Schubert (2015) criticizes him for simply replacing one extreme (planning, acting on 'true' preferences) with another (impulsiveness, embracing any preference that one holds in the moment of choosing). He argues that what matters "is not the chance to satisfy whatever preference one may end up having in future periods, but, first and foremost, the chance to manage and develop one's preferences" (Schubert, 2015, p. 292). He reconceptualizes Sugden's opportunity criterion as an *opportunity to learn*. The normative standard is here not defined as a broad range of opportunities to choose from but by opportunities to learn new preferences. Dold and Schubert (2018) further develop this approach by explicitly building on the work of James Buchanan on aspiration and self-constitution. Contrary to Sugden, they argue that the agent may legitimately want to constrain himself in the future if such self-constraint is part of a free volitional process of 'becoming the man he wants to become'. In line with Hargreaves-Heap (2013), who argues that "welfare economics should be concerned with the conditions under which people's preferences form and not simply with how best to satisfy them" (p. 998), they argue that the focus of welfare economics should not be on the assessment of possible outcomes in terms of how much they satisfy people's preferences but on the evaluation of the processes that bring about preference learning.

I regard this to be an important upgrade to Sugden's account. However, the notion of learning implied by these authors is hardly satisfactory. Learning is considered to be a function of "trying out new preferences, discarding some of them, and keeping others" (Dold & Schubert, 2018, p. 233). It is also not clear how the agent decides whether he would need to employ some self-constraint in the first place. The account thus suffers from the fact that it assumes that simple exposure to different options will result in learning the new preference, that trying out new things is what brings about the change in preferences. But how can trying out some

good or service bring about a preference change if the individual does not possess the value that would enable him to evaluate the experience as worthy of being preferred to something else? The taste for Schoenberg is rarely developed simply by being exposed to his music. Preferences come about through what we do, not through what we are exposed to.

4.2 Proleptic reasons as opportunities to learn

Let me take stock of the discussion so far. I agree with Sugden's critique of the notion of inner rational agent and embrace his rejection of the satisfaction of latent preferences as a proper welfare criterion. But I also share the concern of his critics that there is something unsatisfying, limiting, and rather unrealistically demanding in his view of human psychology. I accept that learning new preferences is the part that Sugden neglects. But as I demonstrated in this chapter, learning requires much more than mere experiments in living and exposure to new things. It requires an active engagement with the resources in the environment that enable proleptic reasoning toward a new value. Any discussion of welfare effects based on opportunities to learn should thus include the discussion of the learning processes and the available environmental resources that make learning possible.

This requires a more nuanced understanding of the enabling function of incentives, rules, or constraints. Let us consider the case of Odysseus tying himself to the mast. This (self-)constraint indeed brings about the desired outcome, which is to not be lured in by the Sirens. And economists often think of the enabling functions of institutions in this way. Institutions as incentive structures enable certain behavioral patterns and outcomes by imposing constraints. But this is a very thin notion of what 'enabling' means. For example, alcohol prohibition may bring about lower alcohol consumption as a result. But it does not enable restraint. As pointed out earlier, believing something is valuable is not the same as valuing something, and prohibition does not teach the value of alcohol restraint. Similarly, Odysseus remains essentially the same man after the ordeal, despite going through the experience. I argue that the enabling function of incentives, rules, and other institutional factors in the environment must be based on the change in underlying values. But for this, it is not enough to simply be exposed to new experiences. Values

need to be learned through practice and aspirational process. (Self-)constraint does not have an enabling function. Proleptic reasons do.

I argue that the debate about the appropriate welfare criterion has focused too much on choice and not enough on the conditions that enable people to deal with the value conflicts. Sugden is right that limiting the available options is misguided. But I think he is right for the wrong reason. It is not that having lots of options is good because it makes possible satisfaction of any kind of preferences. Rather, I argue, options are good because they leave open the field of value contestation. Resolving value conflicts may be beneficial on its own because of the discovery and innovation it brings about. But as this chapter has shown, the crucial part here is the availability of the sources of proleptic reasoning in the environment. If life is about acquiring new values, and if this value learning is itself a beneficial process of resolving the conflict of values in the world that is inevitably characterized by multiple overlapping modes of valuation, then welfare considerations should indeed be focused on the availability of proleptic reasons. Not mere exposure, but proleptic reasons are the relevant opportunities to learn new values and to contribute to the resolution of the conflict.

4.3 Incentives, proleptic reasons, and intrinsic motivation: come full circle?

Let us now revisit the issue of intrinsic motivation. I have argued in this chapter that the accounts based on psychological theories of intrinsic motivation are unsatisfactory because they address only the underlying psychological mechanism of experience without explaining why people do what they do in the first place. Intrinsic motivation is typically cast in the standard framework of choice where the main issue is whether the choice is autonomous, or constrained by some interfering external factor. Values are mostly considered static, and as such only prone to crowding out. I have demonstrated that a more promising direction is to recognize the role of proleptic reasoning. Proleptic reasons enable people to act towards acquiring new values, not only acting from them, and thus to cope more effectively with situations that are characterized by several competing, but internally well-

ordered, modes of valuation. Crucially, people rely on their environments as sources of such proleptic reasons and thus clues for why people do what they do will be found in examining these environmental affordances, rather than in the internal psychological mechanisms.

But several questions remain unanswered. Why do people follow any particular proleptic reason? And what can we say about how different environments might stimulate learning and development? Do incentives play any role here, or is proleptic reasoning different in that is not incentive based? Has the discussion in this chapter shed new light on the problems of incentives, motivation, and aspiration, or do we have to again retreat to positing two fundamentally different types of motivation, such as intrinsic and extrinsic?

Tackling these questions calls for an examination of the relationship between incentives and proleptic reasons. At first sight, they seem to be fundamentally different. Incentives have power because they promise satisfaction of preferences. In contrast, proleptic reasons are not directed at satisfying wants but at generating them. But in an important way, one can also rely on incentives to keep attentive, to strengthen the proleptic reasons, so to speak. For example, Agnes Callard tells a story of a music appreciation student who promises herself a chocolate treat if she makes it through an opera without falling asleep. A 'purist' might say that the student is acting on the wrong reason (extrinsic), but the problem is of course that the 'right' reason (intrinsic value of music) she does not possess yet. Learning how to value music in this way is precisely the whole point of taking a music appreciation class. Nevertheless, it seems hard to reconcile an aspirational act, which is about acting towards a value that you do not yet possess; and an incentivized act, which means that you are acting according to a value that you do in fact possess. In this sense, incentives and proleptic reasons may well represent different types of motivations.

But incentives and proleptic reasons share an important characteristic: they are both external elements. In this regard, from the perspective of intrinsic motivation theories, aspirational process seems to be based on a lot of 'oppressive' and 'inauthentic' elements. A beginner is always inauthentic with respect to rules, precisely because he follows them religiously. Psychologists differentiate between

fully internalized and integrated norms on the one hand, and introjection on the other. Behaviors based on the former are characterized by freedom and flexibility and are referred to as intrinsically motivated. On the contrary, people guided in their behavior by introjected norms are responding "compliantly to those forces within them" (Deci & Flaste, 1995, p. 5) and so lack autonomy and authenticity. Only fully internalized and integrated norms can be part of intrinsic motivation. The rest is psychologically damaging. But this is hardly good social science. As Sugden (2018) points out, "the literature of intrinsic motivation invites us to aspire to the profoundly unrealistic ideal of an economy in which everyone's actions and efforts are coordinated to realize gains from trade, but in which no one is actually motivated to seek those gains" (p. 213). In this extreme sense, this is a world where there is not much learning and surely not much innovation. Additionally, any deliberate change in valuation and any learning of new evaluative frameworks automatically requires one to rely on external motivating factors, because the evaluative apparatus has not formed yet. In insisting on the primacy of intrinsic motivation, the crowding out approaches thus also seem to aspire to a world where no one is motivated to change any value or preference. This clearly goes against Knight's insight about the point of life.

I argue that intrinsic motivation, as presented in this literature, is in fact a secondary phenomenon. It can only take place once a value has been learned. In a situation characterized by a conflict among modes of valuation, where the resolution of the conflict requires learning a new value, the only possibility is to rely on proleptic reasons as external sources of motivation. I don't want to deny the fact that there are different ways of appreciating or valuing a particular activity, one of them being valuing it for the sake of itself. However, in order to account for this kind of valuation, we need to focus on learning. Intrinsic motivation is thus the outcome of a learning process. And this learning process depends on the availability of resources in the environment. What kind of role models are available? What do people around you do? Is it valued to go against the stream? This is indeed a process of socialization, crucially dependent on the cultural context, just not in the sense of passive conditioning. This socialization is based on aspirational agency that engages with the available opportunity set of environmental resources through ongoing social

practices. The so-called intrinsically motivated activity is not significant because it represents a type of motivation that is not oriented towards rewards. It is significant because it represents a practice through which a specific value conflict gets resolved. My general proposal, thus, is to take a step back and investigate the conditions that bring about the learning of values. It means that we should shift our focus from the mechanisms of choice and autonomy to the accounts of learning and resolving value conflicts.

5. Conclusion

The argument in this chapter has been based on a general observation that some people feel motivated to engage in certain activities without any apparent rewards or punishments. Their participation in the activity cannot be explained by reference to any payoff scheme. Psychological literature gives us a mechanism to understand what is happening when someone feels intrinsically motivated. However, it does not explain why people perform one activity and not another. In this chapter I have demonstrated that understanding this question requires us to pay attention to the agent's social and institutional environments.

I have argued that value learning through proleptic reasoning can account for the why question of human behavior. In the absence of available proleptic reasons, one may feel truly indifferent towards certain new opportunity. Thus, it is not enough to only have available opportunities for action. One also needs available proleptic reasons for acting on any of those opportunities. Callard (2018) even argues that "the aspirational theory is well placed to explain why those who have suffered from unusually harsh conditions in their upbringing are less responsible for failing to create themselves as good people and as valuers of the good" (p. 256). The questions of welfare should include considerations about the availability of these environmental resources, since, in the absence of proleptic reasons, an opportunity set as such, no matter how big, will have no effect on the actual welfare.

This chapter has also demonstrated that the standard framework of choice and preferences is unsatisfactory. Acting on proleptic reasons cannot be explained in terms of preferences, because the evaluative apparatus has not formed yet. One does not yet possess the value that would enable him to make an evaluative judgment

about the relative preferences. You can rank order the options, but such ordering will inevitably involve 'mistakes' because the value has not been learned yet. Explaining such action in terms of preferences inevitably leads to what Gigerenzer (2018) identified as a *bias bias*: a bias for explanations based on cognitive and behavioral biases. Furthermore, standard accounts cannot distinguish between actions undertaken to satisfy existing preferences and actions undertaken to learn new ones. In the standard framework, all acts of learning are basically acting against one's current values or preferences, and are thus behavioral errors. This underlines the need for shifting the focus to the study of value learning and the nature and role of proleptic rationality.

Understanding the nature of proleptic reasons enables us to better understand the activities people engage in. When a value is not fully grasped, acting on proleptic reasons will have different motivational effects than if the value would be fully grasped. A defect in the grasp of the value can thus motivate an action that resembles a lack of control over behavior. Attributing the lack of control to a failure of rationality and cognition will lead us down the wrong path. Rather, we must understand what the agent is trying to achieve and what kind of proleptic reasons are at his disposal. Empirically, this means a move away from laboratory studies to much more ethnographic empirical work that aims to understand the situations and the different evaluative frameworks that people aspire to use, as well as people's ideas about what those evaluative frameworks entail (Dekker et al., 2020). The challenge is not the autonomy of choice. The challenge is the learning of values as a process that deals with the underlying value conflict.

This argument also has important implications for the debate about the behavioral policy interventions. If weakness of will, and other inconsistencies of intertemporal choice, are not consequences of rationality failure but motivational failure, then it does not make sense to try to nudge people with more incentives that target their rational mind. We need to study motivation. Incentives–or nudging, or welfare state, or self-constitution–may bring about short-term behavioral change. They may also have a role in strengthening the proleptic reasons. But they do not, as such, bring about the transformational change in terms of learning new values.

For that, proleptic reasons as external resources guiding the practice are crucial, because engagement in a practice, and value learning that may come with it, is what has a potential to bring about the transformational change.

Conclusion

This thesis consists of chapters written on various topics from modern behavioral economics. However, the reader has probably noticed that the tone has been a predominately critical one. In particular, one of the main arguments put forward in this thesis is that behavioral economics does not offer a satisfactory account of the role of social environment and interactions within it. I have demonstrated that behavioral economics–just like neoclassical economics, the main target of its revisionist criticism–continues to conceptualize the role of the social environment exclusively through the analytical construct of a fundamentally isolated individual actor. Both Type I and Type II varieties of behavioral economics, as I have called them, have failed to fully account for the embeddedness of economic actors in their social and institutional environments. One might ask, then, why focus on behavioral economics? Why bother pointing out yet again the various shortcomings of using psychological insights in the social sciences? And why, in fact, frame the proposed new direction in terms of *behavioral economics* of social interaction?

To many scholars that are critical of psychology in similar ways as I have been in this thesis, the failures of behavioral economics on the social front simply reflect fundamental disciplinary limitations of psychology to contribute to economics as a social science that studies outcomes of interactions of many individuals. The argument, then, is that, to explain social life, economists should bypass psychology altogether and rather seek a deeper unification with sociology (e.g., Ross, 2014; Wagner, 2010). While I fully embrace calls for more sociologically informed economics, and indeed many of the crucial arguments I have advanced in the preceding chapters rely on insights from sociology, my aim in this thesis was not to argue against combining economics with psychology. Rather, I chose as a starting point the observation that in recent years behavioral economics has become fully integrated into the economics mainstream. Behavioral economics has arrived and should be taken seriously.

Throughout the last decades, much has been written on whether psychology should be in or out of economics. Yet there is scarcely a consensus on the matter. I consider this debate to be historically relevant. But I also believe it is uninteresting at this point in time. It may well be that there was once a time when economists were seeking to get rid of any psychological considerations in their theories, and when behavioral economics represented an active revolt against the limiting assumptions that practicing economists were required to adopt when working with the mainstream economic theories. These times are long gone. Today, behavioral economics is anything but a fringe contrarian discipline established on the basis of its opposition to the mainstream. On the contrary, the tools it has developed are part of the core toolbox that economists have at their disposal. Its approach is taught at the best universities, its research is reported at the most prestigious economic journals, and its proponents enjoy a widespread recognition, both within the academic community as well as among the popular audiences. Therefore, my answer to the question of why bother with behavioral economics is that I simply wanted to look at the current practice of economics. And in this practice behavioral economics represents an important component. My approach has been to evaluate and criticize the behavioral economics research program, as well as to formulate alternatives to it, on its own terms. This shifts the perspective from 'in-or-out' questions to exploration of what it is that economists actually do when they allegedly incorporate psychological insights. As this thesis has demonstrated, the key question is not whether psychology is used or not, but what kind of psychology is used.

What has emerged in the preceding chapters are the contours of a new type of behavioral economics. While Type I is primarily concerned with issues of rationality, and Type II with psychological well-being, the new type–I will call it Type III–centers on social interaction. Its main building blocks are intersubjective construction of meaning, the notion of the socially extended mind, and processes of aspiration and value learning.

This new type has several important implications. We can broadly divide them into three sets. The first set of implications is theoretical. Type III behavioral economics acknowledges the centrality of incentives, but it does not treat them as objective forces. It recognizes the importance of the motivation to pursue an activity

as an end in itself, but it does not situate such motivation in opposition to external incentives. It does not proclaim people to be either rational or irrational but shows how interactions with the environment enable the emergence of a rational conduct. It posits the processes of aspiration and value acquisition as important components of economic analysis, which goes against the usual conception by economists of a value as a (given) stock variable. It thus implies a shift in understanding the values from stocks to flows. In general, Type III represents a theoretical way forward for behavioral economics that corrects for the shortcomings of both Type I and Type II, and offers a more genuinely socialized and interactive account of economic action. Much work remains ahead to provide a more systematically unified account, though.

The second set of implications concerns the issues of welfare economics. Without a doubt, one of the central and most hotly debated contributions of behavioral economics has been in the form of its implications for welfare policy. It is thus only fitting that Type III offers its own perspective on the questions of welfare criteria as well. As we have seen, learning through interaction with the environment has implications for how we think about the policy interventions inspired by behavioral economics. If rationality is not something that can be posited in a clear-cut way, and if much of the economic activity is based on aspirational proleptic reasoning, this means that interventions will be limited as compared to nudging, for example, which has largely emerged as the gold standard of behavioral policy interventions. According to Type III behavioral economics, welfare policy cannot be a matter of trying to paternalistically nudge someone into a particular decision, since the guiding value cannot be fully specified in advance. In terms of a desired policy, this implies a general support for increased opportunities to acquire new values within a cognitively rich environment. Future work has to flesh out more concrete proposals based on specific problems in the policy space.

The third set of implications of the Type III behavioral economics are implications for thinking about economic change and innovation. When the locus of creative and innovative activity is neither the individual mind nor the institutions as incentive structures but coupled and entangled units of individual actors and their environments, economic change has to be studied on the level of these units. This implies that interactive creative communities, scenes, and institutionalized practices

will be much more important for the creative and innovative dynamics than the individuals and firms following their incentives and satisfying preferences. Throughout this thesis I have reported on and discussed some empirical work in this direction, but there is much left to be further developed in the future.

Bibliography

Abolafia, M. (2001). *Making markets: Opportunism and restraint on Wall Street*. Harvard University Press.

Alchian, A. A. (1950). Uncertainty, Evolution, and Economic Theory. *Journal of Political Economy*, *58*(3), 211–221.

Anderson, M. L. (2003). Embodied Cognition: A field guide. *Artificial Intelligence*, *149*(1), 91–130.

Angner, E. (2019). We're all behavioral economists now. *Journal of Economic Methodology*, *26*(3), 195–207.

Angner, E., & Loewenstein, G. (2012). Behavioral Economics. In U. Mäki (Ed.), *Philosophy of economics* (pp. 641–689). Elsevier.

Aoki, M. (2011). Institutions as cognitive media between strategic interactions and individual beliefs. *Journal of Economic Behavior and Organization*, *79*(1–2), 20–34.

Arnau, E., Ayala, S., & Sturm, T. (2014). Cognitive externalism meets bounded rationality. *Philosophical Psychology*, *27*(1), 50–64.

Ashraf, N., Bandiera, O., & Jack, B. K. (2014). No margin, no mission? A field experiment on incentives for public service delivery. *Journal of Public Economics*, *120*, 1–17.

Augier, M. (1999). Some Notes on Alfred Schutz and the Austrian School of Economics. *Review of Austrian Economics*, *11*, 145–162.

Bardone, E. (2011). *Seeking chances: From biased rationality to distributed cognition*. Springer Verlag.

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645.

Bates, R. H., Greif, A., Levi, M., & Rosenthal, J. L. (1998). *Analytic narratives*. Princeton University Press.

Bates, R. H., Greif, A., Levi, M., Rosenthal, J. L., & Weingast, B. R. (2000). The analytic narrative project. *American Political Science Review*, *94*(3), 696–702.

Becker, G. S. (1976). *The economic approach to human behavior*. University of Chicago Press.

Becker, H. S. (1953). Becoming a Marihuana User. *American Journal of Sociology*, *59*(3), 235–242.

Beckert, J. (2016). *Imagined futures*. Harvard University Press.

Beckert, J., & Aspers, P. (2011). *The Worth of Goods: Valuation and Pricing in the Economy*. Oxford University Press.

Bem, D. J. (1972). Self-perception theory. *Advances in Experimental Social Psychology*, *6*(C), 1–62.

Bénabou, R., Falk, A., & Tirole, J. (2018). Narratives, imperatives, and moral reasoning. *NBER Working Paper No. 24798*.

Bénabou, R., & Tirole, J. (2003). Intrinsic and Extrinsic Motivation. *The Review of Economic Studies*, *70*(3), 489–520.

Bénabou, R., & Tirole, J. (2006). Incentives and Prosocial Behavior. *American Economic Review*, *96*(5), 1652–1678.

Berg, N., & Gigerenzer, G. (2010). As-if behavioral economics: Neoclassical economics in disguise? *History of Economic Ideas*, *18*(1), 133–165.

Besley, T., & Ghatak, M. (2018). Prosocial Motivation and Incentives. *Annual Review of Economics*, *10*, 411–438.

Bohnet, I., & Oberholzer-Gee, F. (2002). Pay for performance: Motivation and selection effects. In B. S. Frey & et al. (Eds.), *Successful management by motivation*. Springer Verlag.

Boltanski, L., & Chiapello, E. (2005). *The New Spirit of Capitalism*. Verso.

Boltanski, L., & Thévenot, L. (2006). *On Justification: Economies of Worth*. Princeton University Press.

Boulding, K. E. (1969). Economics as a moral science. *The American Economic Review*, *59*(1), 1–12.

Bowles, S. (1998). Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions. *Journal of Economic Literature*, *36*(1), 75–111.

Bowles, S., & Polanía-Reyes, S. (2012). Economic Incentives and Social Preferences: Substitutes or Complements? *Journal of Economic Literature*, *50*(2), 368–425.

Brennan, G., & Buchanan, J. M. (1985). *The reason of rules: Constitutional political economy*. Cambridge University Press.

Bruni, L., & Sugden, R. (2007). The road not taken: How psychology was removed from economics, and how it might be brought back. *Economic Journal*, *117*(516), 146–173.

Brunswik, E. (2001). *The essential Brunswik: Beginnings, explications, applications* (K. R. Hammond & T. R. Stewart, Eds.). Oxford University Press.

Buchanan, J. M. (1979). Natural and artifactual man. In *The collected works of James M. Buchanan, Volume 1: The logical foundations of constitutional liberty (1999)* (pp. 246–259). Liberty Fund.

Buchanan, J. M. (1982). Order defined in the process of its emergence. In *The collected works of James M. Buchanan, Volume 1: The logical foundations of constitutional liberty (1999)* (pp. 244–245). Liberty Fund.

Callard, A. (2018). *Aspiration: The agency of becoming*. Oxford University Press.

Camerer, C. F., Issacharoff, S., Loewenstein, G., O'Donoghue, T., & Rabin, M. (2003). Regulation for conservatives: Behavioral economics and the case for "asymmetric paternalism." *University of Pennsylvania Law Review*, *151*(3), 1211–1254.

Camerer, C. F., Loewenstein, G., & Rabin, M. (2004). *Advances in behavioral economics*. Princeton University Press.

Cameron, J., & Pierce, W. D. (1994). Reinforcement , Reward , and Intrinsic Motivation : A Meta-Analysis. *Review of Educational Research*, *64*(3), 363–423.

Carpendale, J. I. M., Frayn, M., & Kucharczyk, P. (2016). The social formation of human minds. In J. Kieverstein (Ed.), *The Routledge handbook of philosophy of the social mind*. Routledge.

Cazzolla Gatti, R., Koppl, R., Fath, B. D., Kauffman, S., Hordijk, W., & Ulanowicz, R. E. (2020). On the emergence of ecological and economic niches. *Journal of Bioeconomics*, *22*(2), 99–127.

Chamlee-Wright, E. (1997). *The cultural foundations of economic development: Urban female entrepreneurship in Ghana*. Routledge.

Chater, N., Felin, T., Funder, D. C., Gigerenzer, G., Koenderink, J., Krueger, J. I., Noble, D., Nordli, S. A., Oaksford, M., Schwartz, B., Stanovich, K. E., & Todd, P. M. (2018). Mind, rationality, and cognition: An interdisciplinary debate. *Psychonomic Bulletin and Review*, *25*(2), 793–826.

Chetty, R. (2015). Behavioral economics and public policy: A pragmatic perspective. *The American Economic Review*, *105*(5), 1–33.

Clark, A. (1997). *Being There: Putting Brain, Mind, and World Together Again*. MIT Press.

Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, *58*(1), 7–19.

Cnossen, B., Loots, E., & van Witteloostuijn, A. (2019). Individual motivation among entrepreneurs in the creative and cultural industries: A self-determination perspective. *Creativity and Innovation Management*, *28*(3), 389–402.

Collins, R. (2004). *Interaction Ritual Chains*. Princeton University Press.

Cosmides, L., & Tooby, J. (1994). Better than Rational: Evolutionary Psychology and the Invisible Hand. *The American Economic Review*, *84*(2).

Cowen, T. (2007). *Discover your inner economist: use incentives to fall in love, survive your next meeting, and motivate your dentist*. Dutton.

Cowen, T. (2018). *Stubborn attachments: A vision for a society of free, prosperous, and responsible individuals*. Stripe Press.

Davis, J. B. (2010). Neuroeconomics: Constructing identity. *Journal of Economic Behavior & Organization*, *76*(3), 574–583.

Davis, J. B. (2011). Individuals and identity in economics. In *Individuals and Identity in Economics*. Cambridge University Press.

Davis, J. B. (2016). Economics, neuroeconomics, and the problem of identity. *Schmollers Jahrbuch*, *136*(1), 15–31.

DeCharms, R. (1968). *Personal causation*. Academic Press.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, *18*(1), 105–115.

Deci, E. L. (1975). *Intrinsic Motivation*. Plenum Press.

Deci, E. L., & Flaste, R. (1995). *Why we do what we do: the dynamics of personal autonomy*. A Grosset/Putnam Book.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum Press.

Dekker, E. (2016). Exemplary Goods: Exemplars as Judgment Devices. *Valuation Studies*, *4*(2), 103–124.

Dekker, E., & Kuchař, P. (2016). Exemplary Goods: The Product as Economic Variable. *Schmollers Jahrbuch*, *136*(4), 237–255.

Dekker, E., & Kuchař, P. (2017a). Emergent Orders of Worth: Must We Agree on More than a Price? *Cosmos + Taxis*, *4*(1), 23–34.

Dekker, E., & Kuchař, P. (2017b). Heterarchy. In A. Marciano & G. B. Ramello (Eds.), *Encyclopedia of Law and Economics*. Springer.

Dekker, E., & Kuchař, P. (2020). Lachmann and Shackle: On the Joint Production of Interpretation Instruments. *Research in the History of Economic Thought and Methodology*, *37B*, 25–42.

Dekker, E., & Remic, B. (2019). Two types of ecological rationality: or how to best combine psychology and economics. *Journal of Economic Methodology*, *26*(4), 291–306.

Dekker, E., Remic, B., & Dalla Chiesa, C. (2020). Incentives Matter, But What Do They Mean? Understanding the Meaning of Market Coordination. *Review of Political Economy*, *32*(2), 163–179.

Dellavigna, S., & Pope, D. (2018). What motivates effort? Evidence and expert forecasts. *Review of Economic Studies*, *85*(2), 1029–1069.

Denzau, A. T., & North, D. C. (1994). Shared Mental Models: Ideologies and Institutions. *Kyklos*, *47*(1), 3–31.

DeVaro, J., Maxwell, N., & Morita, H. (2017). Training and intrinsic motivation in nonprofit and for-profit organizations. *Journal of Economic Behavior and Organization*, *139*, 196–213.

Dewey, J. (1938). *Experience and education*. Macmillan.

Dickinson, A. M. (1989). The Detrimental Effects of Extrinsic Reinforcement on "Intrinsic Motivation." *The Behavior Analyst*.

Dold, M. F., & Schubert, C. (2018). Toward A Behavioral Foundation of Normative Economics. *Review of Behavioral Economics*, *5*(3–4), 221–241.

Dreyfus, H. L. (2014). *Skillful coping: Essays on the phenomenon of everyday perception and action*. Oxford University Press.

Dreyfus, H. L., & Dreyfus, S. E. (2005). Peripheral Vision: Expertise in Real World Contexts. *Organization Studies*, *26*(5), 779–792.

Emmett, R. B. (2006). De gustibus est disputandum: Frank H. Knight's reply to George Stigler and Gary Becker's 'De gustibus non est disputandum' with an introductory essay. *Journal of Economic Methodology*, *13*(1), 97–111.

Fehr, E., Klein, A., & Schmidt, K. M. (2001). *Fairness, incentives and contractual incompleteness* (No. 72; Institute for Empirical Research in Economics Working Paper).

Felin, T., Kauffman, S., Mastrogiorgio, A., & Mastrogiorgio, M. (2016). Factor markets, actors, and affordances. *Industrial and Corporate Change*, *25*(1), 133–147.

Felin, T., Koenderink, J., & Krueger, J. I. (2017). Rationality, perception, and the all-seeing eye. *Psychonomic Bulletin and Review*, *24*(4), 1040–1059.

Festré, A., & Garrouste, P. (2015). Theory and evidence in psychology and economics about motivation crowding out: A possible convergence? *Journal of Economic Surveys*, *29*(2), 339–356.

Forster, M., & Sober, E. (1994). How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *British Journal for the Philosophy of Science*, *45*(1), 1–35.

Foss, N. J., & Garzarelli, G. (2007). Institutions as knowledge capital: Ludwig M. Lachmann's interpretative institutionalism. *Cambridge Journal of Economics*, *31*(5), 789–804.

Foss, N. J., & Lindenberg, S. (2011). Managing joint production motivation: the role of goal framing and governance mechanisms. *Academy of Management Review*, *36*(3), 500–525.

Frey, B. S. (1997). *Not Just for the Money: an Economic Theory of Personal Motivation*. Edward Elgar.

Frey, B. S. (2018). *Economics of happiness*. Springer.

Frey, B. S., & Gallus, J. (2012). Happiness policy and economic development. *International Journal of Happiness and Development*, *1*(1), 102.

Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. *Journal of Economic Surveys*, *15*(5), 589–611.

Frey, B. S., & Oberholzer-Gee, F. (1997). The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out. *The American Economic Review*, *87*(4), 746–755.

Frey, B. S., & Stutzer, A. (2002). *Happiness and economics: How the economy and institutions affect human well-being*. Princeton University Press.

Friedman, M. (1974). Explanation and Scientific Understanding. *The Journal of Philosophy*, *71*(1), 5.

Gal, D., & Rucker, D. D. (2018). The Loss of Loss Aversion: Will It Loom Larger Than Its Gain? *Journal of Consumer Psychology*, *28*(3), 497–516.

Gallagher, S. (2009). Philosophical antecedents of situated cognition. In P. Robbins & M. Aydede (Eds.), *The Cambridge Handbook of Situated Cognition*. Cambridge University Press.

Gallagher, S. (2013). The socially extended mind. *Cognitive Systems Research*, *25–26*, 4–12.

Gallagher, S. (2017). *Enactivist interventions: Rethinking the mind*. Oxford University Press.

Gallagher, S. (2020). *Action and interaction*. Oxford University Press.

Gallagher, S., Mastrogiorgio, A., & Petracca, E. (2019). Economic reasoning and interaction in socially extended market institutions. *Frontiers in Psychology*, *10*(AUG), 1856.

Gibson, J. J. (1950). *The Perception of the Visual World*. The Riverside Press.

Gibson, J. J. (2015). The Ecological Approach to Visual Perception: Classic Edition. In *The Ecological Approach to Visual Perception*. Psychology Press.

Gigerenzer, G. (2000). *Adaptive Thinking: Rationality in the Real World*. Oxford University Press.

Gigerenzer, G. (2007). *Gut feelings: The intelligence of the unconscious*. Viking.

Gigerenzer, G. (2008). *Rationality for Mortals: How People Cope with Uncertainty*. Oxford University Press.

Gigerenzer, G. (2015). *Simply rational: Decision making in the real world*. Oxford University Press.

Gigerenzer, G. (2018). The Bias Bias in Behavioral Economics. *Review of Behavioral Economics*, *5*(3–4), 303–336.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482.

Gigerenzer, G., & Goldstein, D. G. (2011). The recognition heuristic: A decade of research. *Judgment and Decision Making*, *6*(1), 100–121.

Gigerenzer, G., & Selten, R. (2001). *Bounded rationality: The adaptive toolbox*. MIT Press.

Gigerenzer, G., & Todd, P. M. (2012). Ecological rationality: The normative study of heuristics. In P. M. Todd & G. Gigerenzer (Eds.), *Ecological Rationality: Intelligence in the World*. Oxford University Press.

Gigerenzer, Gerd., Todd, P. M., & ABC Research Group. (1999). *Simple heuristics that make us smart*. Oxford University Press.

Gneezy, U., & Rustichini, A. (2000a). A fine is a price. *Journal of Legal Studies*, *29*, 1–17.

Gneezy, U., & Rustichini, A. (2000b). Pay enough or don't pay at all. *Quarterly Journal of Economics*, *115*(3), 791–810.

Gode, D. K., & Sunder, S. (1993). Allocative Efficiency of Markets with Zero-Intelligence Traders: Market as a Partial Substitute for Individual Rationality. *Journal of Political Economy*, *101*(1), 119–137.

Gold, N. (2019). The limits of commodification arguments: Framing, motivation crowding, and shared valuations. *Politics, Philosophy and Economics*, *18*(2), 165–192.

Goldschmidt, N., Grimmer-Solem, E., & Zweynert, J. (2016). On the purpose and aims of the Journal of Contextual Economics. *Schmollers Jahrbuch*, *136*, 1–14.

Greif, A. (1994). Cultural beliefs and the organization of society: a historical and theoretical reflection on collectivist and individualist societies. *Journal of Political Economy*, *102*(5), 912–950.

Greif, A. (2006). *Institutions and the path to the modern economy: Lessons from medieval trade*. Cambridge University Press.

Greif, A., & Kingston, C. (2011). Political Economy of Institutions, Democracy and Voting. In N. Schofield & G. Caballero (Eds.), *Political Economy of Institutions, Democracy and Voting*. Springer.

Grüne-Yanoff, T., Marchionni, C., & Moscati, I. (2014). Introduction: methodologies of bounded rationality. *Journal of Economic Methodology*, *21*(4), 325–342.

Gul, F., & Pesendorfer, W. (2010). The case for mindless economics. In A. Caplin & A. Schotter (Eds.), *The foundations of positive and normative economics: A handbook* (pp. 3–42). Oxford University Press.

Hamermesh, D. S. (2013). Six decades of top economics publishing: Who and how? *Journal of Economic Literature*, *51*(1), 162–172.

Hands, D. W. (2010). Economics, psychology and the history of consumer choice theory. *Cambridge Journal of Economics*, *34*(4), 633–648.

Hands, D. W. (2014). Normative ecological rationality: normative rationality in the fast-and-frugal-heuristics research program. *Journal of Economic Methodology*, *21*(4), 396–410.

Harackiewicz, J. M., & Sansone, C. (2000). Rewarding competence: The importance of goals in the study of intrinsic motivation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 82–103). Academic Press.

Hargreaves Heap, S. P. (2013). What is the meaning of behavioural economics? *Cambridge Journal of Economics*, *37*(5), 985–1000.

Hausman, D. M. (2012). *Preference, value, choice, and welfare*. Cambridge University Press.

Hayek, F. A. (1937). Economics and Knowledge. *Economica*, *4*(13), 33–54.

Hayek, F. A. (1945). The use of knowledge in society. *The American Economic Review*, *35*(4), 519–530.

Hayek, F. A. (1952). *Sensory order: An inquiry into the foundations of theoretical psychology*. Chicago University Press.

Hayek, F. A. (2002). Competition as a Discovery Procedure. *The Quarterly Journal of Austrian Economics*, *5*(3), 9–23.

Hayes, A. S. (2020). The Behavioral Economics of Pierre Bourdieu. *Sociological Theory*, *38*(1), 16–35.

Hertwig, R., & Gigerenzer, G. (1999). The "conjunction fallacy" revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, *12*(4), 275–305.

Heukelom, F. (2014). *Behavioral economics: A history*. Cambridge University Press.

Heyes, A. (2005). The economics of vocation or "why is a badly paid nurse a good nurse"? *Journal of Health Economics*, *24*(3), 561–569.

Ho, K. (2009). *Liquidated: An ethnography of Wall Street*. Duke University Press.

Hoff, K., & Stiglitz, J. E. (2016). Striving for balance in economics: Towards a theory of the social determination of behavior. *Journal of Economic Behavior and Organization*, *126*, 25–57.

Hogarth, R. M. (2005). The challenge of representative design in psychology and economics. *Journal of Economic Methodology*, *12*(2), 253–263.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research. *ACM Transactions on Computer-Human Interaction*, *7*(2), 174–196.

Holmström, B., & Milgrom, P. (1991). Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, and Organization*, *7*, 24–52.

Hutchins, E. (1995). *Cognition in the Wild*. MIT Press.

Hutchins, E. (2014). The cultural ecosystem of human cognition. *Philosophical Psychology*, *27*(1), 34–49.

Infante, G., Lecouteux, G., & Sugden, R. (2016). Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics. *Journal of Economic Methodology*, *23*(1), 1–25.

James, W. (1890). *The principles of psychology, Vol. 2*. Henry Holt and Company.

Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *American Economic Review*, *93*(5), 1449–1475.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge University Press.

Kahneman, D., & Tversky, A. (1979). Prospect Theory : An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–292.

Karpik, L. (2010). *Valuing the Unique: The Economics of Singularities*. Princeton University Press.

Kelley, H. (1967). Attribution theory in social psychology. *Nebraska Symposium on Motivation*, *15*, 192–238.

Kirsh, D. (2010). Thinking with external representations. *AI and Society*, *25*(4), 441–454.

Kirzner, I. M. (1992). *The meaning of market process: Essays in the development of modern Austrian economics*. Routledge.

Kitcher, P. (1981). Explanatory Unification. *Philosophy of Science*, *48*(4), 507–531.

Kiverstein, J. (2018). Extended cognition. In A. Newen, S. Gallagher, & L. de Bruin (Eds.), *The Oxford Handbook of 4E Cognition*. Oxford University Press.

Klamer, A. (2016). *Doing the right thing: A value based economy*. Ubiquity Press.

Knight, F. H. (1935). *The limitations of scientific method in economics*. Harper and Brothers.

Knorr Cetina, K., & Bruegger, U. (2002). Global microstructures: The virtual societies of financial markets. *American Journal of Sociology*, *107*(4), 905-950+i.

Koenderink, J. (2014). The all seeing eye? *Perception*, *43*(1), 1–6.

Koppl, R. (2002). Big Players and the Economic Theory of Expectations. In *Big Players and the Economic Theory of Expectations*.

Koppl, R., Kauffman, S., Felin, T., & Longo, G. (2015). Economics for a creative world. *Journal of Institutional Economics*, *11*(1), 1–31.

Kreps, D. M. (1997). Intrinsic Motivation and Extrinsic Incentives. *The American Economic Review*, *87*(2), 359–364.

Kruglanski, A. W. (1978). Issues in cognitive social psychology. In M. R. Lepper & D. Greene (Eds.), *The hidden cost of reward: New perspectives on the psychology of human motivation* (pp. 19–29). Lawrence Erlbaum.

Kruglanski, A. W., Fishbach, A., Woolley, K., Bélanger, J. J., Chernikova, M., Molinario, E., & Pierro, A. (2018). A structural model of intrinsic motivation: On the psychology of means-ends fusion. *Psychological Review*, *125*(2), 165–182.

Kruglanski, A. W., Friedman, I., & Zeevi, G. (1971). The effects of extrinsic incentive on some qualitative aspects of task performance. *Journal of Personality*, *39*(4), 606–617.

Kuchař, P. (2016). Entrepreneurship and institutional change: The case of surrogate motherhood. *Journal of Evolutionary Economics*, *26*(2), 349–379.

Lachmann, L. M. (1971). *The Legacy of Max Weber*. Glendessary Press.

Lamont, M. (2012). Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology*, *38*, 201–221.

Lavoie, D. (2011). The interpretive dimension of economics: Science, hermeneutics, and praxeology. *Review of Austrian Economics*, *24*(2), 91–128.

Lavoie, D. (2015). The discovery and interpretation of profit opportunities: Culture and the Kirznerian entrepreneur. In L. E. Grube & V. H. Storr (Eds.), *Culture and Economic Action*. Edward Elgar.

Le Grand, J. (2010). Knights and knaves return: Public service motivation and the delivery of public services. *International Public Management Journal*, *13*(1), 56–71.

Leary, D. E. (1987). From Act Psychology to Problematic Functionalism: The Place of Egon Brunswik in the History of Psychology. In Mitchell G. Ash & William

Ray Woodward (Eds.), *Psychology in twentieth-century thought and society* (pp. 115–142). Cambridge University Press.

Lecouteux, G. (2016). From Homo Economicus to Homo Psychologicus. *OEconomia*, *6*(2), 175–200.

Leeson, P. T. (2020). Logic is a harsh mistress: welfare economics for economists. *Journal of Institutional Economics*, *16*(2), 145–150.

Lepper, M. R., & Greene, D. (1978). *The hidden costs of reward: New perspectives on the psychology of human motivation*. Lawrence Erlbaum.

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology*, *28*(1), 129–137.

Levin, J. (2016). Functionalism. In *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/win2016/entries/functionalism/

Levitt, S. D., & Dubner, S. J. (2005). *Freakonomics: A rogue economist explores the hidden side of everything*. Harper Collins.

Lewis, P., & Dold, M. (2020). James Buchanan on the nature of choice: Ontology, artifactual man and the constitutional moment in political economy. *Cambridge Journal of Economics*, *44*(5), 1159–1179.

Linson, A., Clark, A., Ramamoorthy, S., & Friston, K. (2018). The active inference approach to ecological perception: General information dynamics for natural and artificial embodied cognition. *Frontiers Robotics AI*.

Loewenstein, G. (1999). Because it is there: The challenge of mountaineering ... for utility theory. *Kyklos*, *52*(3), 315–343.

Loyal, S. (2012). Agency-Structure. *The Wiley-Blackwell Encyclopedia of Globalization*.

Luttmer, E. F. P., & Singhal, M. (2014). Tax morale. *Journal of Economic Perspectives*, *28*(4), 149–168.

Machlup, F. (1936). Why Bother with Methodology? *Economica*, *3*(9), 39.

Mäki, U. (2001). Explanatory Unification: Double and Doubtful. *Philosophy of the Social Sciences*, *31*(4), 488–506.

McCain, R. A. (1979). Reflections on the cultivation of taste. *Journal of Cultural Economics*, *3*(1), 30–52.

McCain, R. A. (1981). *Cultivation of Taste, Catastrophe Theory, and the Demand for Works of Art*. The American Economic Review.

McCullers, J. C. (1978). Issues in learning and motivation. In M. R. Lepper & D. Greene (Eds.), *The hidden costs of reward: New perspectives on the psychology of human motivation* (pp. 19–29). Lawrence Erlbaum.

Mears, A. (2011). Pricing Looks: Circuits of Value in Fashion Modeling Markets. In J. Beckert & P. Aspers (Eds.), *The Worth of Goods: Valuation and Pricing in the Economy* (pp. 155–177). Oxford University Press.

Mehta, J., Starmer, C., & Sugden, R. (1994). Focal points in pure coordination games: An experimental investigation. *Theory and Decision*, *36*(2), 163–185.

Mellström, C., & Johannesson, M. (2008). Crowding Out in Blood Donation: Was Titmuss Right? *Journal of the European Economic Association*, *6*(4), 845–863.

Menary, R. (2006). Attacking the Bounds of Cognition. *Philosophical Psychology*, *19*(3), 329–344.

Mijovic-Prelec, D., & Prelec, D. (2010). Self-deception as self-signalling: A model and experimental evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1538), 227–240.

Miyazaki, H. (2006). Economy of Dreams: Hope in Global Capitalism and Its Critiques. *Cultural Anthropology*, *21*(2), 147–172.

Morgenstern, O., & Schwödiauer, G. (1976). Competition and collusion in bilateral markets. *Zeitschrift Für Nationalökonomie Journal of Economics*, *36*(3–4), 217–245.

Mullainathan, S., & Thaler, R. H. (2015). Behavioral Economics. *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*, 437–442.

Nelson, J. A. (2006). *Economics for Humans*. The University of Chicago Press.

Nelson, J. A., & Folbre, N. (2006). Why a well-paid nurse is a better nurse. *Nursing Economics*, *24*(3), 127–130.

Newen, A., Gallagher, S., & de Bruin, L. (2018). 4E cognition: Historical roots, key concepts, and central issues. In A. Newen, S. Gallagher, & D. de Bruin (Eds.), *The Oxford Handbook of 4E Cognition*. Oxford University Press.

Nikolova, M., & Cnossen, F. (2020). What makes work meaningful and why economists should care about it. *Labour Economics*, *65*, 101847.

Norman, J. (2002). Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches. *Behavioral and Brain Sciences*, *25*(1), 73–96.

North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge University Press.

Núñez, R., Allen, M., Gao, R., Miller Rigoli, C., Relaford-Doyle, J., & Semenuks, A. (2019). What happened to cognitive science? *Nature Human Behaviour*, *3*(8), 782–791.

Osbeck, L. M., & Nersessian, N. J. (2014). Situating distributed cognition. *Philosophical Psychology*, *27*(1), 82–97.

Ostrom, E. (1990). *Governing the commons*. Cambridge University Press.

Ostrom, E. (2010). Beyond Markets and States. *The American Economic Review*, *100*, 641–672.

Petracca, E. (2017). A cognition paradigm clash: Simon, situated cognition and the interpretation of bounded rationality. *Journal of Economic Methodology*, *24*(1), 20–40.

Petracca, E., & Gallagher, S. (2020). Economic cognitive institutions. *Journal of Institutional Economics*, *16*(6), 747–765.

Pezzulo, G., & Cisek, P. (2016). Navigating the Affordance Landscape: Feedback Control as a Process Model of Behavior and Cognition. *Trends in Cognitive Sciences*, *20*(6), 414–424.

Plott, C., & Smith, V. L. (2008). *Handbook of Experimental Economics Results Volume 1: Vol. Volume 1*. North Holland.

Pommerehne, W. W., & Frey, B. S. (1992). The effects of tax administration on tax morale. In *Serie II-Nr.191* (Diskussionsbeiträge - Serie II No. 191, Vol. 2).

Preda, A. (2012). Interactions and decisions in trading. In K. Knorr Cetina & A. Preda (Eds.), *The Sociology of Finance*. Oxford University Press.

Prendergast, C. (1986). Alfred Schutz and the Austrian School of Economics. *American Journal of Sociology*, *92*(1), 1–26.

Prendergast, C. (2008). Intrinsic motivation and incentives. *American Economic Review*, *98*(2), 201–205.

Radin, M. (1996). *Contested commodities*. Cambridge University Press.

Rheinberg, F., & Engeser, S. (2018). Intrinsic motivation and flow. In J. Heckhausen & H. Heckhausen (Eds.), *Motivation and Action* (3rd ed., pp. 579–622). Springer.

Rich, P. (2016). Axiomatic and ecological rationality: Choosing costs and benefits. *Erasmus Journal for Philosophy and Economics*, *9*(2), 90–122.

Rizzo, M. J., & Whitman, G. (2020). *Escaping paternalism: Rationality, behavioral economics, and public policy*. Cambridge University Press.

Robb, R. (2019). *Willful: How we choose what we do*. Cambridge University Press.

Robbins, P., & Aydede, M. (2009). *The Cambridge Handbook of Situated Cognition*. Cambridge University Press.

Rogers, Y. (2006). Introduction to distributed cognition and communication. In K. Brown (Ed.), *The Encyclopedia of Language and Linguistic* (2nd ed., pp. 181–202). Elsevier.

Rogers, Y., & Ellis, J. (1994). Distributed cognition: An alternative framework for analysing and explaining collaborative working. *Journal of Information Technology*, *9*(2), 119–128.

Romaniuc, R. (2017). Intrinsic motivation in economics: A history. *Journal of Behavioral and Experimental Economics* , *67*, 56–64.

Ross, D. (2014a). *Philosophy of economics*. Palgrave Macmillan.

Ross, D. (2014b). Psychological versus economic models of bounded rationality. *Journal of Economic Methodology*, *21*(4), 411–427.

Roth, A. E. (2007). Repugnance as a constraint on markets. *Journal of Economic Perspectives*, *21*(3), 37–58.

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. The Guilford Press.

Schelling, T. C. (1960). *The Strategy of Conflict*. Harvard University Press.

Schubert, C. (2015). Opportunity and preference learning. *Economics and Philosophy*, *31*(2), 275–295.

Schutz, A. (1962). *Collected papers I: The problem of social reality*. Martinus Nijhoff.

Schutz, A. (1967). *The Phenomenology of the Social World*. Northwestern University Press.

Scitovsky, Tibor. (1992). *The joyless economy: the psychology of human satisfaction. Revised edition: Vol. Rev. ed*. Oxford University Press.

Selten, R. (2002). What is bounded rationality? In G. Gigerenzer & R. Selten (Eds.), *Bounded Rationality: The Adaptive Toolbox* (pp. 13–36). MIT Press.

Sent, E. M. (2004). Behavioral economics: How psychology made its (limited) way back into economics. *History of Political Economy*, *36*(4), 735–760.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*(1), 99–118.

Simon, H. A. (1956). A comparison of game theory and learning theory. *Psychometrika 1956 21:3*, *21*(3), 267–272.

Simon, H. A. (1997). *Administrative behavior: A study of decision-making processes in administrative organizations* (4th ed.). The Free Press.

Skinner, B. F. (1953). Some contributions of an experimental analysis of behavior to psychology as a whole. *American Psychologist*, *8*(2), 69–78.

Slaby, J., & Gallagher, S. (2015). Critical Neuroscience and Socially Extended Minds. *Theory, Culture & Society*, *32*(1), 33–59.

Smith, C. (2012). Auctions and finance. In K. Knorr Cetina & A. Preda (Eds.), *The Sociology of Finance*. Oxford University Press.

Smith, V. L. (1962). An Experimental Study of Competitive Market Behavior. *Journal of Political Economy*, *70*(2), 111–137.

Smith, V. L. (1967). Experimental Studies of Discrimination Versus Competition in Sealed-Bid Auction Markets. *The Journal of Business*, *40*(1), 56.

Smith, V. L. (1991). Rational Choice : The Contrast between Economics and Psychology. *Journal of Political Economy*, *99*(4), 877–897.

Smith, V. L. (1994). Economics in the laboratory. *Journal of Economic Perspectives*, *8*(1), 113–131.

Smith, V. L. (2003). Constructivist and ecological rationality in economics. *The American Economic Review*, *93*(3), 465–508.

Smith, V. L. (2008). *Rationality in Economics: Constructivist and ecological forms*. Cambridge University Press.

Smith, V. L. (2015). Conduct, rules and the origins of institutions. *Journal of Institutional Economics*, *11*(03), 481–483.

Smith, V. L., & Wilson, B. J. (2019). *Humanomics: Moral sentiments and the wealth of nations in the twenty-first century*. Cambridge University Press.

Stark, D. (2009). *The Sense of Dissonance: Accounts of Worth in Economic Life*. Princeton University Press.

Stark, D. (2017). For what it's worth. *Research in the Sociology of Organizations*, *52*, 383–397.

Steele, G. R. (2002). Hayek's Sensory Order. *Theory & Psychology*, *12*(3), 387–409.

Steiner, P. (2010). Gift-giving or market? economists and the performation of organ commerce. *Journal of Cultural Economy*, *3*(2), 243–259.

Stigler, G., & Becker, G. S. (1977). De gustibus non est disputandum. *The American Economic Review*, *67*(2), 76.

Storr, V. H. (2004). *Enterprising slaves and master pirates*. Peter Lang.

Storr, V. H., & Chamlee-Wright, E. (2010). *The political economy of hurricane Katrina and community rebound*. Edward Elgar.

Storr, V. H., Haeffele-Balch, S., & Grube, L. E. (2015). *Community revival in the wake of disaster: Lessons in local entrepreneurship*. Palgrave Macmillan.

Stroebe, W., & Frey, B. S. (1980). In Defense of Economic Man: Towards an Integration of Economics and Psychology. *Swiss Journal of Economics and Statistics (SJES)*, *116*(II), 119–148.

Stroebe, W., & Frey, B. S. (1982). Self-interest and collective action: The economics and psychology of public goods. *British Journal of Social Psychology*, *21*(2), 121–137.

Sturn, R. (2016). Scarce Means, Competing Ends: Lord Robbins and the Foundations of Contextual Economics. *Schmollers Jahrbuch*, *136*(1), 59–85.

Sugden, R. (1989). Spontaneous Order. *The Journal of Economic Perspectives*, *3*(4), 85–97.

Sugden, R. (2018). The community of advantage: A behavioural economist's defence of the market. In *The Community of Advantage: A Behavioural Economist's Defence of the Market*. Oxford University Press.

Sunstein, C. R. (1994). Conflicting Values in Law. *Fordham Law Review*, *62*(6), 1661–1673.

Svorenčík, A. (2016). The Sidney siegel tradition: The divergence of behavioral and experimental economics at the end of the 1980s. *History of Political Economy*, *48*, 270–294.

Swedberg, R. (2003). *Principles of economic sociology*. Princeton University Press.

Thaler, R. H. (1987). Anomalies: The January Effect. *Journal of Economic Perspectives*, *1*(1), 197–201.

Thaler, R. H., & Sunstein, C. R. (2003). Libertarian paternalism. *American Economic Review*, *93*(2), 175–179.

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

Thévenot, L. (2001). Pragmatic regimes governing the engagement with the world. In T. R. Schatzki, K. Knorr Cetina, & E. von Savigny (Eds.), *The practice turn in contemporary theory* (pp. 57–73). Routledge.

Thornton, P. A., Ocasio, W., & Lounsbury, M. (2012). *The institutional logics perspective: A new approach to culture, structure, and process*. Oxford University Press.

Throsby, D. (2000). *Economics and Culture*. Cambridge University Press.

Titmuss, R. (1970). *The gift relationship: From human blood to social policy*. George Allen and Unwin.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293–315.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business*, *59*(4), 251–278.

Vanberg, V. J. (1994). *Rules and choice in economics*. Routledge.

Velthuis, O. (2004). An interpretive approach to meanings of prices. *Review of Austrian Economics*, *17*(4), 371–386.

Velthuis, O. (2005). *Talking Prices: Symbolic Meanings of Prices on the Market for Contemporary Art*. Princeton University Press.

Wagner, H. (1983). *Alfred Schutz: An intellectual biography*. University of Chicago Press.

Wagner, R. E. (2010). *Mind, society, and human action: Time and knowledge in a theory of social economy*. Routledge.

Williamson, O. E. (2000). The new institutional economics: Taking stock, looking ahead. *Journal of Economic Literature*, *38*(3), 595–613.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, *9*(4), 625–636.

Woolley, K., & Fishbach, A. (2018). Underestimating the importance of expressing intrinsic motivation in job interviews. *Organizational Behavior and Human Decision Processes*, *148*(June), 1–11.

Zaloom, C. (2006). *Out of the pits: Traders and technology from Chicago to London*. University of Chicago Press.

Zawidzki, T. (2013). *Mindshaping: A new framework for understanding human social cognition*. MIT Press.

Zelizer, V. A. (1978). Human values and the market: The case of life insurance and death in 19-th century America. *American Journal of Sociology*, *84*(3), 591–610.

Zelizer, V. A. (2004). Circuits of Commerce. In J. C. Alexander, G. T. Marx, & C. L. Williams (Eds.), *Self, Social Structure, and Beliefs* (pp. 122–144). University of California Press.

Zelizer, V. A. (2005). *The Purchase of Intimacy*. Princeton University Press.

Zhang, J. (1991). The interaction of internal and external representations in a problem solving task. *Proceedings of the Thirteenth Annual Conference of Cognitive Science Society*, 954–958.

Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. *Cognition*, *98*(3), 287–308.

Nederlandse samenvatting

Wat motiveert mensen om iets te doen? Wat zijn de implicaties voor economische analyse, gebaseerd op prikkels, van intrinsieke motivatie? Economen accepteren inmiddels dat een antwoord op deze vraag moet worden gezocht op het snijvlak van economie en psychologie, maar de precieze relatie tussen economie en psychologie is nog onduidelijk. Deze dissertatie begint vanuit de vaststelling dat de opkomst van de gedragseconomie betekent dat het niet langer de vraag is of psychologie nu wel of niet in de economie hoort. De relevante kwestie is wat voor soort psychologie zijn weg vond naar de economie, en of dit de juiste en meest geschikte vorm van psychologie is voor economische analyse.

Aan het begin van het proefschrift betoog ik dat economen twee fundamenteel verschillende soorten gedragseconomie ontwikkelden, elk gebaseerd op een ander soort psychologie. Wat ik Type I gedragseconomie noem is een benadering die anomalieën probeert te identificeren in de toepassingen van de standaard rationele keuzetheorie. De meest voor de hand liggende voorstanders zijn Herbert Simon, Daniel Kahneman en Richard Thaler. In tegenstelling daarmee gaat Type II niet uit van de premisse dat de theorie van de rationele keuze beschrijvend onjuist is en we dus een correctie nodig hebben, maar van de premisse dat rationeel instrumenteel handelen niet leidt tot voldoening en persoonlijke groei. Voorbeelden hiervan zijn te vinden in het werk van Tibor Scitovsky en Bruno Frey. Afgezien van verschillen in theorie en analyse, verschillen de twee typen aanzienlijk in hun normatieve implicaties. Gedragseconomie Type I gaat over de bevrediging van 'echte' voorkeuren, vrij van cognitieve vooroordelen en andere psychologische beperkingen. Gedragseconomie Type II gaat over de behoefte om de innerlijke autonomie en identiteit tot uitdrukking te brengen, waarbij economische kosten en baten op een tweede plaats komen.

Het overkoepelende thema van de volgende hoofdstukken is dat een ander soort gedragseconomie zowel mogelijk als nodig is. De gedragseconomie die ik voorstel is niet uitsluitend gericht op de cognitieve capaciteiten en behoeften van het

individu, maar gebaseerd op het feit dat mensen zijn ingebed in een sociale wereld. Door deze inbedding serieus te nemen, moeten we rekening houden met de verstrengeling van individuen en hun sociale en institutionele omgeving. Dit is een perspectief dat zich bevindt op het gebied van sociale interacties, tussen de twee uitersten van Type I's instrumentaal perspectief en Type II's expressief perspectief. Het resoneert met Kenneth Boulding's gedenkwaardige citaat dat "economic man is a clod, heroic man is a fool, but somewhere between the clod and the fool, human man, if the expression may be pardoned, steers his tottering way" (Boulding, 1969, p. 10). Mensen die tussen deze twee uitersten leven, handelen en interacteren door en met de hulp van hun institutionele en sociale omgeving. Bij het uitzoeken wat het juiste is om te doen, worden ze niet alleen geleid door hun innerlijke impulsen, of door te reageren op een externe standaard van rationeel handelen, maar vooral door een constant proces van ontdekken en leren–door interactie met elkaar en met hun omgeving–over hoe de situatie op de juiste manier kan worden geëvalueerd. Om dit te bestuderen, hebben we weinig aan een gedragseconomie die puur redeneert vanuit individuele cognitie of van individueel psychologisch welzijn. We hebben de gedragseconomie van sociale interactie nodig. Een benadering die zich richt op de studie van intersubjectieve betekenis en voortbouwt op een inzicht uit de recente cognitieve wetenschap dat individuele geesten en hun omgeving epistemisch en ontologisch verstrengeld zijn.

In het eerste hoofdstuk heb ik de grenzen van de individualistische psychologie verkend via het probleem van de effectiviteit van prikkels aan te pakken. Het idee dat prikkels ertoe doen, wordt terecht beschouwd als een van de hoekstenen van het economisch denken. Maar verschillende soorten economisch onderzoek, waaronder de psychologisch onderbouwde die hier worden genoemd, hebben aangetoond dat prikkels niet kunnen worden behandeld als een objectieve natuurlijke kracht die het menselijk gedrag stuurt. Het hoofdstuk onderzoekt het verschil tussen situaties waarin bepaalde prikkels worden gezien als krachtige redenen om acties te veranderen, terwijl andere prikkels weinig of zelfs contraproductieve effecten hebben. Ik laat zien dat de sociale wereld een cruciale rol speelt bij het interpreteren van prikkels, omdat prikkels betekenis krijgen in relatie tot de sociale omgeving, sociale rollen en institutionele praktijken die mensen aangaan als onderdeel van hun

dagelijkse leven. De relatieve kosten en baten van het menselijk handelen zijn geen objectieve feiten los van hun sociale context, maar worden intersubjectief verkregen door processen van sociale interactie in een specifieke institutionele context. In dit opzicht kunnen prikkels averechts werken wanneer hun betekenis onduidelijk is of wordt betwist door een andere concurrerende betekenis.

Wanneer economen de ineffectiviteit van prikkels proberen uit te leggen, beroepen ze zich meestal op de psychologisch geïnformeerde literatuur over intrinsieke motivatie naar voren. In hoofdstuk 2 analyseer ik deze literatuur en beargumenteer ik dat het concept van intrinsieke motivatie door economen op inconsistente manieren is gebruikt, omdat de onderliggende theorieën van intrinsieke motivatie, geïmporteerd in de economie vanuit de psychologie, concurreren en tegenstrijdigheden vertonen ondanks het gebruik van dezelfde terminologie. Ik laat zien dat deze verschillen belangrijke implicaties hebben voor empirische studies intrinsieke motivatie en de op dit onderzoek gebaseerde beleidsinterventies. De relatie tussen psychologie en economie wordt gecompliceerd als we bedenken dat verschillende psychologische theorieën fundamenteel verschillende visies op de menselijke psychologie kunnen impliceren. Dit betekent dat het standaardargument voor het importeren van psychologische inzichten in de economie, namelijk dat psychologie de accuraatheid van de economische theorie vergroot, niet zo eenvoudig is als vaak wordt aangenomen door empirische gedragseconomen.

Hoofdstuk 3 ontwikkelt dit inzicht verder door de analyse van het concept van ecologische rationaliteit, dat recentelijk naar voren is gebracht als een alternatief voor de conceptie van begrensde rationaliteit. In dit hoofdstuk pleit ik voor het belang van het begrijpen van de onderliggende verschillen in zowel psychologische als economische benaderingen die worden gecombineerd als onderdeel van de inspanningen van gedragseconomen om marktresultaten te verklaren. Dit hoofdstuk zet ook een eerste stap op weg naar het nieuwe type gedragseconomie. Ik laat zien hoe een nieuwe manier van denken over psychologie en cognitie ons in staat stelt om af te wijken van het strikt individualistische en internalistische perspectief van de theorieën die we tot nu toe tegenkwamen. Dit type psychologie combineert bijzonder goed met de economische benaderingen die de rol van instituties in het economische

leven benadrukken. De instituties vervullen een cognitieve rol en ondersteunen zo de cognitieve capaciteiten van het individu.

In hoofdstuk 4 ontwikkel ik dit idee tot een nieuwe benadering voor het conceptualiseren van de omgeving en instituties in de economie. Ik beargumenteer dat de traditionele opvatting van de omgeving als een beperking voor individuele actie onjuist is, omdat het de belangrijke onderlinge afhankelijkheden negeert die ontstaan wanneer mensen zich verhouden tot die omgeving. Het hoofdstuk maakt gebruik van inzichten uit de recente cognitieve wetenschap over de 'extended mind' om aan te tonen hoe de studie van economie kan profiteren van het herconceptualiseren van de omgeving, niet als een beperking, maar als een hulpmiddel voor het maken van goede keuzes onder meer door feedback. Door mensen in staat te stellen gebruik te maken van de kennis die belichaamd is in institutionele praktijken, spelen deze middelen een sleutelrol bij het mogelijk maken van bepaalde economische acties, soorten redeneringen en de creatieve ontdekking van nieuwe kenmerken van de omgeving, of nieuwe potentiële acties binnen die omgeving, die het leerproces over de mogelijkheden tot actie en interactie verder voeden.

De thema's prikkels, motivatie, interacties tussen de individuen en hun omgeving, en leren door het ontdekken van nieuwe potentiële acties, worden in hoofdstuk 5 met elkaar verweven met behulp van Agnes Callards werk over aspiratie en het verwerven van nieuwe waarden en voorkeuren. Het hoofdstuk begint met de observatie dat psychologische literatuur over intrinsieke motivatie alleen een psychologisch mechanisme biedt om te verklaren waarom prikkels misschien niet effectief zijn, maar het schiet tekort om uit te leggen waarom mensen überhaupt aan activiteiten deelnemen, waarom ze de ene activiteit belangrijker vinden dan de andere, en hoe die waarden ontstaan. Ik beargumenteer dat het verwerven van nieuwe waarden moet worden gezien als het centrale proces. Dit overstijgt zowel de nadruk van Type I op rationele normatieve voorkeuren als de nadruk van Type II op de intrinsieke motivatie van het individu. Het hoofdstuk laat ook zien dat de exclusieve aandacht op beslissingen, of het nu de beleidsmaker of het individu is, misleidend is, omdat het veronderstelt dat degene die de beslissing maakt over het evaluatieve apparaat beschikt dat hem in staat stelt een relevant oordeel te vellen.

In plaats daarvan zouden we onze analytische focus moeten richten op de studie van processen en activiteiten tussen de uitersten van economische en heroïsche ethiek; en tussen de uitersten van alwetende beleidsmakers en autonome expressieve individuen. Dit andere perspectief laat zien dat mensen er voortdurend naar streven om waardeconflicten op te lossen en nieuwe waarden te leren door te handelen en interactie te hebben in en met hun sociale en institutionele omgeving.

The overarching theme of this thesis is that an alternative to what has in recent decades emerged as the mainstream approach to behavioral economics is both possible and needed: a behavioral economics that does not focus exclusively on the workings of individual minds but takes seriously the fact that people are embedded in a social world. Taking this embeddedness seriously means that we need to consider the deeply entangled and interactive nature of the relationship between individuals and their social and institutional environments. In figuring out what the right thing to do is, people are guided not by their inner impulses, or responding to some external standard of rational action, but primarily by a constant process of discovery and learning – through interacting with each other and with their environment – about how to appropriately interpret and evaluate the situation. To study this process, we need behavioral economics of neither individual cognition nor individual psychological wellbeing. Instead, we need the behavioral economics of social interaction: an approach that centers on the study of intersubjective meaning and builds on an insight from the recent cognitive science that individual minds and their environments are epistemically and ontologically entangled.