Worse than Complex
PETTER TÖRNBERG

Department of Energy and Environment
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2017
THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Worse than Complex

PETTER TÖRNBERG
Department of Energy and Environment
Division of Physical Resource Theory, Complex Systems Group
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2017
Thesis for the degree of Doctor of Philosophy

**Worse than Complex**

© Petter Törnberg, 2017
ISBN: 978-91-7597-534-4
Doktorsavhandlingar vid Chalmers tekniska högskola.
Ny serie nr 4215.
ISSN 0346-718X
Printed by Reproservice
Göteborg, Sweden 2017

CHALMERS UNIVERSITY OF TECHNOLOGY
Department of Energy and Environment
Division of Physical Resource Theory
Complex Systems Group
SE-412 96 Göteborg
Sweden
www.chalmers.se
Tel. +46-(0)31 772 1000

Author email: pettert@chalmers.se
Worse than Complex
Petter Törnberg
Department of Energy and Environment, Chalmers University of Technology

Abstract
This thesis engages with questions on the boundary between what has traditionally been understood as social and natural. The introductory essay contextualizes the specific contributions of the included papers, by noting and exploring a reinvigoration of “naturalism” (the notion of a continuity between the human realm and the rest of natural phenomena) under the banner of Complexity Science. This notion is put under explicit light, by revisiting the age-old question of naturalism and connecting ideas in complexity science with the work of e.g. Roy Bhaskar, Mario Bunge, William Wimsatt, and David Lane. A philosophical foundation for a complexity science of societal systems is thereby sketched, taking the form of an integrative and methodologically pluralist “complex realism”.

The first two papers provide a theoretical perspective on the distinction between social and natural: Paper I notes that societal systems combine two qualities that are commonly referred to as complexity and complicatedness into an emergent quality that we refer to as “wickedness”, and that is fundamentally and irreducibly different from either quality in isolation. This explains the recalcitrance of societal systems to the powerful approaches that exist for dealing with both of these qualities in isolation, and implies that they indeed ought to be treated as a distinct class of systems. Paper II uses the plane spanned by complexity and complicatedness to categorize seven different system classes, providing a systematic perspective on the study of societal systems.

The suggested approach to societal systems following from these conclusions is exemplified by three studies in different fields and empirical contexts. Paper III combines a number of theories that can be seen as responses to wickedness, in the form of evolutionary developmental theories and theories of societal change, to develop a synthetic theory for cultural evolution. Paper IV exemplifies how simulation can be integrated with social theory for the study of emergent effects in societal systems, contributing a network model to investigate how the structural properties of free social spaces impact the diffusion of collective mobilization. Paper V exemplifies how digital trace data analysis can be integrated with qualitative social science, by using topic modeling as a form of corpus map to aid critical discourse analysis, implying a view of formal methods as aids for qualitative exploration, rather than as part of a reductionist approach.

Keywords: Complexity, Naturalism, Critical Realism, Wicked Systems, Transitions, Evolutionary Developmental Theory, Digital Trace Data, Social Movements, Innovation Society
List of publications

CA, PT, AT developed the idea, and CA wrote the paper with contributions from PT and AT.

CA, PT developed the idea, and CA wrote the paper with contributions from PT.

CA, PT, AT developed the idea, and CA wrote the paper with contributions from PT and AT.

PT, AT developed the idea, PT collected the data, PT developed and ran model, AT conducted qualitative analysis, AT wrote the paper with contributions from PT.

PT, AT developed the idea, PT developed the model, PT conducted analysis, PT, AT wrote paper.
Publications not in this thesis


Acknowledgements

I wish to acknowledge the support of the EU-FET grants MD (no. 284625), INSITE (no. 271574) and ODYCEUS (no. 732942).

I would like to thank my supervisor, Claes Andersson, for showing me that you can survive in academia as an intellectual rather than merely an academic. My brother Anton Törnberg, David Avra Lane, Kristian Lindgren, Bill Wimsatt, Maurie Cohen, and Justus Uitermark, for making this thesis possible.

I also want to thank all my wonderful colleagues and friends at Physical Resource Theory at Chalmers University of Technology, at Gothenburg University, at ECLT Ca’Foscari University of Venice, at New Jersey Institute of Technology, and at the department of sociology at University of Amsterdam.

The synthesis was discussed and contributed greatly to by the participants\(^1\) at the workshops organized in Venice.

Of course, nobody, apart from myself, should be held accountable for the shortcomings of the arguments presented in this thesis.

Gothenburg, February 2017
Petter Törnberg

---

\(^1\)Including Manfred Laubichler, Sander van der Leeuw, John Odling-Smee, and Kevin Laland. For full list of participants, see the workshop series at www.insiteproject.org/slides/new-data/.
Contents

List of publications ........................................ ii
Acknowledgements ........................................... iv

I Introductory Essay: On the Limits of Complex Naturalism 1

Introduction ................................................. 3

1 Epistemological & Ontological Limits ................. 10
   1.1 Beyond the Open/Closed Dichotomy ............... 10
   1.2 Uncertainty in Wicked Systems ................. 25
   1.3 Boundaries of Wicked Systems ................. 28

2 Relational Limits ........................................ 31
   2.1 Narratives before Society ....................... 32
   2.2 Narratives in Society ......................... 34
   2.3 Narratives and Social Structures .......... 37
   2.4 Innovation in Society ....................... 39
   2.5 Innovation meets Uncertainty ............ 43

3 From Naturalism to Realism ......................... 45
   3.1 Society in Mainstream Complexity Science .... 47
   3.2 Fact and Action in Wicked Society .......... 50
   3.3 Explanation in Wicked Systems .......... 53
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4 Data Mining and Categorization in Wicked Systems</td>
<td>56</td>
</tr>
<tr>
<td>3.5 Facing Emergence in Wicked Systems</td>
<td>59</td>
</tr>
<tr>
<td>3.6 Integrated Pluralism - Tying Together Traces</td>
<td>63</td>
</tr>
<tr>
<td>3.7 To Change it</td>
<td>66</td>
</tr>
<tr>
<td>4 Conclusion</td>
<td>70</td>
</tr>
<tr>
<td>Paper I: Societal Systems - Complex or Worse?</td>
<td>103</td>
</tr>
<tr>
<td>Paper II: Innovation and the Anatomy of Wickedness</td>
<td>119</td>
</tr>
<tr>
<td>Paper III: An Evolutionary Developmental Approach to Cultural Evolution</td>
<td>157</td>
</tr>
<tr>
<td>Paper IV: Modelling Free Social Spaces and the Diffusion of Social Mobilization</td>
<td>181</td>
</tr>
<tr>
<td>Paper V: Combining CDA and Topic Modeling: Analyzing Discursive Connections Between Islamophobia and Anti-Feminism on an Online Forum</td>
<td>205</td>
</tr>
</tbody>
</table>
Part I

Introductory Essay: On the Limits of Complex Naturalism
Introduction

The traditional boundaries between natural and social systems seem increasingly fragile and contested in a world that is now widely seen as having entered into the geological époque of the Anthropocene (Steffen et al., 2007; Waters et al., 2016), characterized by human society as the globally dominating shaping force on Earth’s geology and ecosystems. The idea of the natural world existing independently of the social world is increasingly hard to sustain as profound transformations driven by human activity ravage global ecosystems (Moore, 2014). This changing relationship between the natural and social has highlighted a confounding asymmetry, between our capacity to transform the world through technological innovation, and our profound lack of understanding of the very world which our own strength has established. Humanity’s effects on nature have been characterized as a “perfect storm” of interrelated crises of increasing scale and magnitude (Bai et al., 2015; Crutzen, 2002; Pievani, 2014).

Simultaneously, a host of parallel developments are unfolding in scientific theory. Old theory and assumptions, in a range of disciplines dominated by natural scientific methods, are being undermined by an explosion of new data and analytic methods, in particular centered around notions such as “Big Data” and “complexity” (e.g. Anderson, 2008; Ball, 2012; Laland et al., 2015). This is both opening up new opportunities and unveiling deep-rooted problems in traditional natural scientific approaches, challenging Humean² notions of causal law, pre-
dictability and hierarchy. The onslaught of new data and methods is undermining basic simplifying assumptions, upon which much of the apparent predictive and explanatory power of time-tested theory rested, and as a response, new theory and new models are developing to deal with the revealed complexity of the natural world. A related pattern of development can be seen in a range of disciplines traditionally dominated by quantitative and formalist approaches, e.g. in the collapse of the neoclassical paradigm and the pluralist explosion of economic theory (e.g. Fontana, 2010; Hodgson, 2014), in the Ancient DNA revolution in archeology and paleoanthropology (e.g. Larson et al., 2007; Meyer et al., 2014), in the questioning of the assumptions of time-scale separation between ecosystem and evolutionary dynamics in ecology (e.g. Fussmann et al., 2007; Odling-Smee et al., 2003; Post and Palkovacs, 2009), and in the growing recognition of the limits of the Modern Synthesis within evolutionary theory (e.g. Oyama et al., 2003; Pigliucci et al., 2010).

New methods developed to respond to the discovered messiness of these systems enable an extension of formalistic approaches to fields previously considered outside of their scope. This can be seen within the humanities and social sciences, where traditional approaches in various realms increasingly have to compete with computational modeling and large-scale surface studies of digital trace data. These shift the focus to motifs rather than meaning, but are capable of finding intricate patterns through brute-forcing of immense data quantities (e.g. Anderson, 2008; Conte et al., 2012; Macy and Willer, 2002). While these developments are sprawling and multifaceted, “Complexity Science”\(^3\) can be identi-

---

3This term is contested and describes a discipline with many – often contradictory – directions. As used in this essay, the term will specifically denote an important direction within the larger discipline centered around, and developing from, the Santa Fe Institute (see Galison, 1997), representing a mainly formalist and simulation-based approach to complexity, with its roots in the natural sciences, that has proven highly capable of analyzing many types of complex systems that have otherwise been impenetrable to formal approaches. This direction is what Morin (2008) and Byrne and Callaghan (2013) refer to as “restricted complexity”.

---
fied as one of its most important standard-bearers. Complexity Science focuses on abstract systemic dynamics that have proven to be powerful sources of analogies across various empirical domains, exemplified by how important factors in as diverse fields as urban innovation, human travel, animal metabolism, and basic physics have been found to conform to the same universal scaling laws (Betten-court et al., 2007; Brockmann et al., 2006; Freedman and Index, 1971). This has been taken to imply a diminished methodological relevance of the boundaries between the natural and social world, as both are increasingly understood to be “complex” (e.g. Gilbert, 2010; Helbing, 2012; Lazer et al., 2009; Mitchell, 2009).

These parallel developments – (i) increasing entanglement between social, natural and technological systems, (ii) increasing understanding of the complex nature of many natural systems, and (iii) increased application of natural scientific methods in social systems with access to new data and methods – together lead to a situation in which the expanding impact of societal systems on nature – paradoxically – result in further coaching of society in terms of technological and natural scientific problems and methods (see also Malm and Hornborg, 2014). The increasingly fuzzy boundary between natural and societal systems is leading to a renewed expansion of methodology from the natural to the social sciences.

While there is nothing wrong per se with interdisciplinary transfer, these new methods come with hard-to-detect stowaways: implicit ontological assumptions – which may have been debated in their natural scientific contexts long ago but are now taken for granted – sneak into the study of social systems; meta-theoretical underpinnings about the nature and organization of systems that are rarely made explicit. These contain tacit and unexamined answers to questions like: What are the real entities of the social world (Byrne, 2002, p.136)? Are higher-level organizations (like firms, tribes, and states) fully explainable in terms of the preferences of their members, or are higher-level organizations also social individuals with their own properties and powers? Can individual action, meaning, and values be disregarded in the study of causality in human societies? Questions like these are not only matters of philosophical curiosity, but have profound implications for how we can and should research, manage, and think about social systems.

This essay revolves around revisiting, with new complexity-informed eyes, an age-old notion that is at the heart of all of these questions: naturalism, the idea of a continuity between the human sciences and the sciences of the rest of natural phenomena (see also e.g. Bhaskar, 1978; Danto, 1967). Naturalism signifies

Furthermore, the word “science” deserves a note of its own: it is, in this essay, used in the broader sense, in line with the German “Wissenschaft”, Swedish “vetenskap”, or slavic “nauk”, which do not include the positivist/formalist connotations of their English counterpart.
the completion of the Copernican dethroning of man from any meta-natural position, advocating that there is no need to appeal to extra-natural qualities – such as conscience, intentionality or meaning – to account for or understand human society. One of the more influential naturalist tendencies, which Khalil (1995) calls “crude naturalism”, adds to this that the methodology of the natural sciences can and should be employed also within social science. This flavor of naturalism proposes a unification of the sciences in concordance with positivist principles. This has furthermore tended to imply a Cartesian-Newtonian paradigm, associated to reductionist and scientist notions, that ontologically incorporates a shallow realism, and, methodologically, the type of formalist, linear and equilibrium-based notions epitomized by neoclassical economics (Blaikie, 2007, p.178).

The defense from social science, to what from their perspective is seen as disciplinary imperialism (Vinck et al., 2010), has importantly been in the form of social constructionism, i.e. asserting the role of social action in the production of a variant social world, and emphasizing that all beliefs are constructions (Byrne, 2002). Influential here has especially been the program that Pawson and Tilley (1997, p.21) refers to as “hermeneutics II”, which “adds the twist that we cannot, therefore, get beyond constructions”. This anti-naturalist tradition, traceable back to Weber and Dilthey (see e.g. Dilthey et al., 1989; Weber, 2009), posits a methodological cleavage between the natural and social sciences, through a view of the subject-matter of the social sciences as consisting of meaningful objects, and hence suggesting that their aim is the elucidation of the meaning of these objects (Bhaskar, 1978).

The recent reincarnation of naturalism under the banner of Complexity Science has, however, come to be seen as a third direction in relation to naturalism, as it ostensibly does not bear the same “crudeness” as its predecessors. It subscribes to naturalism but brings a fundamental criticism against the linear, stasis-focused approaches of the Cartesian-Newtonian paradigm, and thereby manages to speak to both the postmodernist and the positivist traditions (Cilliers, 1998). Complexity Science has been described as going beyond the limits of what is often called “reductionism” by noting that “the ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct the universe” (Anderson et al., 1972, p.393). Instead emphasis is placed

---

4This is closely related to the notion of a “clockwork universe” – the scientific paradigm based on the idea that exact and empirically testable predictions of the future state of a system can be had, given knowledge of initial conditions and of the universal laws governing the system. This furthermore includes the implicit assumption that external forces can be either controlled or excluded. The Cartesian-Newtonian paradigm is also associated with, and the foundation of, the linear modeling framework, which adds notions of equilibrium, additivity, and homogeneity.
on the study of how parts interact to form complex systems with global, novel, emergent qualities and patterns. As Ball (2012, p.IX) puts it, in the Complexity Science perspective, the traditional conceptualization of naturalism

... remains valid but it often drew on the wrong analogies. Society does not run along the same predictable, ‘clockwork’ lines as the Newtonian universe. It is closer to the kind of complex systems that typically preoccupy statistical physicists today: avalanches and granular flows, flocks of birds and fish, networks of interaction in neurology, cell biology and technology. (Ball, 2012, p.IX)

The focus is no longer stasis and equilibria, but non-linearity, far-from-equilibria, and self-organization (Waldrop, 1993). This has provided a naturalist account that to many seems more plausible, as social phenomena indeed seem to have more in common with emergent physical processes – i.e. “collective effect[s] [...] that cannot be deduced from the microscopic equations of motion in a rigorous way and that disappears completely when the system is taken apart” (Laughlin, 1999, p.863), exemplified by superfluidity or the fractional quantum Hall effect – than with classical Newtonian notions of particles and laws. What remains of traditional crude naturalism is a clear focus on formalist methods, and an underlying, implicit notion of complexity science as a step toward “making social science scientific”: bringing it out from the crude infancy of discursivity and into the scientific adulthood of mathematical and algorithmic rationalism.

This essay aims to put the naturalist tendency of Complexity Science under the light of explicit study, and to explore to what extent this “middle way” naturalism can in fact be said to be less crude than its predecessors, as well as how it relates to the third direction to naturalism proposed by Bhaskar (1978, p.2), in the form of “a qualified anti-positivist naturalism, based on an essentially realist view of science”. By connecting ideas in Complexity Science to the work of e.g. Roy Bhaskar, Mario Bunge, William Wimsatt, and David Lane, an alternative philosophical foundation for a complexity science of societal systems will be sketched, taking the form of an integrative and methodologically pluralist “complex realism” (Reed and Harvey, 1992).

This essay takes the position that the question of naturalism is fundamentally a question of ontology5, and to explore it, we will need uncover and examine the ontological assumptions implicit in scientific practice. Bunge (1979a, p.15)

5“Ontology” is here understood as a specification of “the particular entities and processes postulated by some substantive scientific theory” (Bhaskar, 2013, p.30): an ontology describes what are the entities in the world, their attributes and powers.
argues that “Every theoretical view of society [...] has two components: an ontological and a methodological. The former concerns the nature of society, the latter the way to study it.” Complexity Science brings both of these to the table – meta-theory that speaks usefully to a social ontology and a toolkit of powerful methods – but the latter goes in important ways in the opposite direction of the former by being singularly focused on formalisms. Therefore, I join Elder-Vass (2007c, p.228) in arguing that what is needed in the Complexity Science study of society is a well-articulated social ontology that is open to revision, which would also be conducive to the improvement of the methodology (which is not the same as “the set of methods”).

The aim is thereby to cast light on the borderlands between the natural and social sciences with the purpose of developing a complexity-informed understanding of the ontological distinctiveness of social systems, by attempting “to analyze and to systematize the ontological categories” (Bunge, 1977b, p.12). This categorization is performed on a plane spanned by dynamical and structural complexity (Érdi, 2007), relating to Khalil’s (1995) separation between “artificiality” and “individuality”. Based on this, the essay argues for “a turn to ontology” (Perona, 2007) in Complexity Science’s study of society, by sketching a complexity naturalism that does not invoke the reductionism of ontological individualism. This argument entails a degree of integration between the domains of the social and natural sciences, in the sense that we may be able to apply similar ontological frameworks to both – but not in the traditional reductionist or scientistic sense (as Khalil 1995 notes, the ontological implications of using ecosystems as a metaphor for society depends completely on how one views ecosystems). This may in fact mean that the epistemological conclusions of this metaphor may not be an ecology-inspired sociology, but just as well a sociology-inspired ecology.

The structure of this essay is as follows:

First, the exploration of the limits of complex naturalism begins with what Bhaskar (1979) refers to as the *ontological* and *epistemological conditions* of the social sciences, and specifically the question of openness. Following Khalil (1995), I will use “ontological individualism” rather than the more common “methodological individualism”, due to the ambiguity of “methodology” in this context. "Open”/"Closed” are here used in the critical realist sense, which, it should be noted, is quite different from the understanding in physics. Physics understands open systems as systems whose borders are permeable to energy, and emphasizes that closed systems must eventually attain a time-independent equilibrium state according to the second law of thermodynamics. This means that all living systems are – by definition – open, as they maintain themselves by increasing the entropy of their environment. The critical realist conception is instead of an analytical distinction, seeing a gradient between open and closed, and focuses on the isolation between structural levels of the systems that are the prerequisite for regularities, rather than on the direct exchange of energy/matter...
seen as linked to complexity, but the linkage has been rather unclear: the social scientific understanding of complexity tends to see openness as a categorical subset to complexity, while the natural scientific approach tends to approach systems as closed. This essay uses what in practice amounts to a comparative case study between system categories to create a taxonomy of system ontologies, in which the open/closed dichotomy and the simple/complex is split into a higher resolution plane, spanned by *complicatedness* (or “structural complexity”) on one side, and *complexity* (or “dynamic complexity”) on the other. Following this, we look closer at the primary category of systems relevant for the study of society. This section leans on Paper I and Paper II.

Secondly, the essay focuses on how the complexity of societal systems plays out in societies, in particular relating to what Bhaskar (1979) called the relational condition of the social sciences. We focus on the dialectical relation between agency and structure by looking at the historical evolution of human society and the specifics of human cognition, bringing in notions of meaning and narrative understanding. This leans on Paper III (as well as on work that is still to be published, e.g. Törnberg and Andersson 2016).

Finally, this metatheoretical exposition is used as basis for a “turn to ontology” (Perona, 2007) in the complexity approach to societal systems, suggesting that while the complexity approach is certainly needed to study mass-dynamics in societal systems, they need to be based in epistemology and methodology compatible with the real ontology of such systems. Urry (2003, 2005, 2012) has famously argued that social science has experienced a “complexity-turn” – and now it may be due time for Complexity Science to experience a corresponding “society-turn”. This section leans on Paper IV and V, as well as on contributions not included in this thesis (Andersson and Törnberg, 2016; Törnberg, 2016a,b).

(e.g. Danemark et al., 2001). This tends to be understood as to also include systems whose interaction with their environment is constant, stationary or orderly enough to allow fixed ontologies to be assumed. This conception can perhaps, as in Reed and Harvey (1992, p.359), be referred to as “ontological” – as opposed to “thermodynamical” – openness. See chapter 1 for more in-depth definitions.
Chapter 1

Epistemological & Ontological Limits

1.1 Beyond the Open/Closed Dichotomy

A central separation between social and natural systems emphasized by philosophers of science is that social systems tend to be more open than natural systems. Systems are considered closed if they are “cut off” and isolated from external influences, allowing them to operate under fixed conditions (e.g. Archer et al., 2013; Bhaskar, 2013; Collier, 1994; Von Bertalanffy et al., 1950). Bhaskar (2013) specifies two conditions for such closure: 1) The inner condition: no qualitative change in the object under study; its internal mechanisms will stay the same (Psillos, 2008); 2) The outer condition: the relationship between the causal mechanisms and the mechanisms in their environment in which they act, have to be constant (or at least stationary). This essentially means that the generative mechanisms of phenomena operate independently from intervening mechanisms, which leads to regularities in system dynamics.

Fleetwood (2016) attempts a definition of closure focusing on event-level regularity, based on how the concept is used in e.g. Bhaskar (2013), Lawson (1989, 2014) and Mearman (2006):

Parts of the social world characterised by (stochastic and/or probabilistically specified) regularities between events or states of affairs of the form ‘whenever event or state of affairs x then event or state of affairs y’, are closed systems, and parts of this world not characterised by such regularities are open systems. (Fleetwood, 2016, p.1)
Simon (1991) also looks at the open/closed dichotomy, but on the basis of system structure, showing that closed systems can be described as “near-decomposable” – the reason that they can be seen as “isolated” from external influences is that they are separated into distinct organizational levels, and on those levels into distinct entities interacting through defined interfaces (see figure 1.1). Such levels bring separations of timescales, ensuring that the ontology of the system level will be relatively fixed during a relevant “short run” – a time scale that is long enough for interesting dynamics to occur, but short enough for the assumptions about the interfaces to remain valid. This is what allows one to study them as if they were cut off from external influences; rigorous quantitative analysis becomes viable because qualitative change happens on substantially longer time scales than those of the analyses. Due to this difference in time scales, we can assume and formulate a fixed ontology – a specification of the entities and the “rules of the game” – and see how it plays out. The greater the separation of scales between the internal and the external environment, the greater will the difference in size and speed of the dynamics on these two levels be, and the more generous will the short run be; i.e. the more interesting things will have time to happen. For example, a suitable “short run” for the study of traffic would be between minutes and hours. Over time scales shorter than minutes not much would happen – other than the movement of the pistons and rods in the engines of the cars – and if we move to several days, the dynamics would more or less repeat itself. Moving to even longer time scales, roads, types of vehicles, regulations and so on would begin to change.

In other words, what allows us to study these systems is that they have a number of distinct levels. On closer inspection, we see that such systems can be separated into two types, depending on the way that their levels emerge from the underlying components. This is basically the fundamental insight of Complexity Science: some systems are more like cars, others like flocks of birds (see Mitchell (2009) for an overview of complexity; see Bajec and Heppner (2009) for an overview on bird flock dynamics). In Paper I and II, we refer to the former as complicated and the latter as complex, a separation that corresponds to Érdi’s (2007) separation between “structural” and “dynamical” complexity.

As Paper I and II show, these system categories have distinctly different proximal causes: complicated systems develop through orderly pre-determined specifications of assembly or morphogenesis (Slack, 2009), and are characterized by, in general, having their distinct levels not through emergence\(^1\), but its opposite,

\(^1\)This difficult notion will be dealt with in more detail in chapter 3, but its frequent ambiguity warrants also more directly subscribing to a definition: as will be seen, I agree with Kaidesoja
aggregativity (Wimsatt, 2007, p.274-276), of their underlying elements: a smaller number of heterogeneous, functionally differentiated, adapted elements interacting in a relatively ordered way. Complex systems tend to develop through self-organization (Kauffman, 1993), and their structural levels form through emergence, i.e. the mass-interaction of a large number of simple and homogeneous entities (e.g. Goldstein, 1999). We will now look closer at these two categories of near-decomposable systems.

Complicated systems are exemplified by technology and organisms: systems organized in level hierarchies that may pack a very large numbers of components into delineable compartments. This enables strongly simplified assumptions, as it limits the permitted patterns of component interaction, and hence very little knowledge about the surrounding system is needed to operate locally on its components. Components are slaved by the larger system, meaning that they are fully aligned, permitting adaptation into fine-tuned and non-redundant machineries: this is what allows us to build spacecrafts with the capacity to land with high precision on planets millions of kilometers away. These properties, in particular non-overlapping component functionality, also make them quite easy to study and predict, since the mechanisms are naturally isolated to specific components, and component interaction is simple. This means that we can reduce the system over its interactions and break it down to its components. Its lack of dynamical complexity is what allows us to make ceteris paribus assumptions, and isolate the function of each component. In other words, we may usefully test our theories through the use of controlled experiments, which works exceedingly well on such systems. Since each component is structured to perform certain functions, reality can be captured in analogous abstractions biting over the entire system; the use of what Hayles (1999), irreverently, calls the “Platonic backhand” is hence quite reasonable for approaching such systems. Closed systems include the realm that Khalil (1995) refers to as the “artificial”, and, as Byrne (2002) concludes, (2009) that both Bhaskar’s and Bunge’s use of the term suffer from certain ambiguity and imprecision, and instead use Wimsatt’s (2007) understanding (quite compatible with both Bhaskar and Bunge’s larger frameworks). Wimsatt (2007) defines emergence in negative terms, holding that a given system property is emergent if it breaks one or more of the conditions of aggregativity, implying a dependence on their mode of organization.

2E.g. any programmer will have intimate experience with attempting to construct complicated systems: they are basically the goal of object-oriented design (Calero et al., 2006; Riel, 1996).

3This understanding of experiments also shows clearly why they should not be considered “natural”, but that they are models, with built-in assumptions about the target systems. They are virtual in the same sense as anything that is created for scientific use.
1.1. BEYOND THE OPEN/CLOSED DICHOTOMY

this includes also mechanistic scientific models; in other words, our models can in this case match the ontology of the system under study.

Complicated systems tend to display structural hierarchy, which links them to perspectives that have experienced a revival in the analysis of biology (e.g. Hull, 1980; Koestler and Smythies, 1972; Weiss, 1971) and social theory (e.g. Miller, 1978; Williamson, 1975). What we mean by hierarchy here is not the idea of an outside commander dictating the activity of the organization, but rather a nested system, where components are embedded within other components in a level-upon-level organization.

Complex systems are exemplified by herds, traffic, and social networks, and are well-described by Morin’s (2008) “restricted complexity”: phenomena that are the dynamically emergent product of interaction among a large number of relatively simple agents from a few component classes (Johnson, 2002). The emergence at play in such systems is specifically micro-emergence, in which the interaction occurs between elements on the same ontological level, resulting in macroscopic qualitative novelty (e.g. Bedau, 1997; Corning, 2002; Holland and Wolf, 1998). Such emergence, and dynamical emergence in general, appears “surprising” due to our inability to intuitively follow complex dynamics – long chains of causation undermine our ability to predict outcomes.

Since mechanisms are not located in specific components, complex systems are not characterized by precision, but rather by parallelism, adaptivity and feedback: this permits resilience over components – if one component breaks, others can dynamically step into its place (Scheffer, 2009). This clearly constrains the usefulness of experiments, and the Platonic backhand in general, since we cannot isolate mechanisms to specific components. Dynamical complexity means that, in a sense, mechanisms are complex and distributed within the interactions between the components rather than within single components: one does not find the intelligence of the anthill in any specific ant (e.g. Dorigo and Stützle, 2009; Wahde, 2008). In a sense, agency is all there is in complex systems, but it is a highly limited agency, which only plays out locally. Due to this, complex systems are much simpler than they may appear, at least given the tools to successfully deal with emergence in systems with very large numbers of interacting entities.

Using simulation – Hayles’ (1999) “Platonic forehand” – we can test our theories about system mechanisms by “growing” systems from models of the underlying entities (Epstein, 1996). Since studying individual components tends to give little clues, we are often forced to resolve to ad hoc assumptions regarding the behavior of these components. While this is a risky game, since such systems
are characterized by both equifinality (a phenomenon may rise from radically different conditions) and multifinality (similar conditions may result in very different outcomes), it is made possible by the simplicity of the components. In other words, in the same way that we can reduce complicated systems over their dynamical simplicity, we can reduce complex systems over their structural simplicity. Even so, since complex systems are generally characterized by chaos – the flip-side of the resilience coin – and an inherent lack of precision, we cannot predict their future states, but only their general dynamics (e.g. Cvitanovic et al., 2005, p.146–149).

**Figure 1.1:** A near-decomposable system, conceptually illustrated in two ways. Because of time scale separation, the outer environment can be regarded as static, and the inner can be similarly disregarded.

**Figure 1.2:** An attempt to illustrate a poorly decomposable system. Because of lack of clear system demarcations and time scale separation, it can be unclear what outer and inner environment would even mean.

**Non-Decomposable Systems**

As we have seen, our study of complicated systems relies on the assumption that they are dynamically simple, and our study of complex systems on that they are structurally simple. However, Paper I shows, there is no reason to assume that a system cannot display both complexity and complicatedness, which hence would mean that we cannot reduce it over either axis. This perspective, of viewing complexity and complicatedness as two separate system properties, allows us span
1.1. BEYOND THE OPEN/CLOSED DICHOTOMY

a plane of system types, permitting a systematic categorization of ontologies in which both closed and open systems are represented, and in which open systems combine complexity and complicatedness (see figure 1.3.)

While complex and complicated systems differ in some important ways, they also share some similarities. A crucial similarity is that they are both structured into nested hierarchies, with each level forming the building blocks for the next, which is what allows the systems to be reduced downward into distinguishable subsystems (Simon, 1991). One important reason that systems can become decomposable is when they are what Khalil (1995) refers to as “artificial”: they lack individuality of their own as they are fabricated, either through natural selection or by the hands of purposive agents, to fulfill a purpose. William Wimsatt’s (e.g. 1986) concept of “generative entrenchment” shows how such selection will result in structured systems, where each level functions as an alphabet for generating the next, as this is the most adaptive and flexible system structure. This is also described by what in assemblage theory is called “coding” (DeLanda, 2006).

The components of open systems, on the other hand, are not completely artificial, but have at least some level of individuality. This means that, since there is no higher functionality of the system, and mechanisms therefore cannot be linked to a function, the generative mechanisms will not be located exclusively in neither relations or in components. This in turn implies that the messy real cannot be reduced to an abstract form, since such a form would require a functional ideal: hence, the Platonic backhand will not work, since their mechanisms are not located into specific components; the Platonic forehand will not work since their components are too complicated to be ad hoc deduced. They hence become non-decomposable, and there is, in the general case, no separation between time scales: anything may interact with anything, and a modification may therefore impact any part of the system. This means that invariant empirical regularities do not obtain, and open systems are therefore denied decisive test situations for their theories; ceteris paribus can never be assumed (Bhaskar, 1979).

Just as with closed systems, we can distinguish a number of different types of open systems based on their level of dynamical and structural complexity, again stemming from different developmental histories: trans-complicated, trans-complex, sub-wicked and wicked systems (see Figure 1.3 and Paper II).

**Trans-complicated systems** are complicated organizations of components with separate agendas, exemplified by organizations with human components, or biological individuals (e.g. of different species) with separate channels of reproduction. Complexity enters as an increased density, and lower regularity, of interac-
Figure 1.3: A map over the ontological categories, as spanned by dynamical and structural complexity. See Paper II.

tions: for example, while an exhaust manifold is precisely an exhaust manifold, a human component will connect a system to just about all sectors of society and in a wide variety of ways (a seamless web; Hughes 1986.) These systems do have some collective functionality, but also some level of internal competition between components, which tends to break down the level hierarchy. System alignment must be actively maintained by dedicated systems, which is costly and carries the risk of failure. Failure, i.e. components adapting to their own aims and goals at the expense of the whole (i.e. “defecting” in game theoretic terms), can go from having detrimental effects on system structure, to being exceedingly dangerous for the system. For example, when cells begin to compete with other cells within an organism, racing toward becoming the fittest unicellular phenotype within this selection environment – i.e. the cancer cell – this typically spells the end of both the organism and the cell germ-line, as their only long-term means of reproduction remains the holistic system (although there are fascinating examples of cancer evolving transmissibility, hence leaving the selection-pressure of
1.1. BEYOND THE OPEN/CLOSED DICHOTOMY

the holistic system altogether, e.g. Pearse and Swift 2006, as well as immune responses evolved to combat this transmissibility, see Alderton 2016 – an example which highlights the quasi-porous boundaries between inside/outside systems, as well as between system levels; see below.)

The trans-complicated nature of organizations can be clearly seen in the tension between organizational and occupational perspectives in the study of behavior in organizations (see Orr, 1996; Van Maanen and Barley, 1982): while the former tends to describe a functional machinery with limited patterns of interaction between parts, the latter emphasizes informal story-telling and interactions cutting through organizational compartments. There is clearly also a broad spectrum of organizational forms, between the strict hierarchical military organization – close to Weber’s (2009) *ideal bureaucracy* – to loosely organized open source groups, characterized by sub-wickedness (see below) rather than trans-complicatedness. These organizational forms afford different strengths and weaknesses: the former a high level of precision, predictability and stability, while the latter permits high levels of innovation and adaptability. Some process frameworks, such as *Scrum*, explicitly aim for self-organization, in some ways mimicking hunter-gatherer groups (Dingsøyr et al., 2012), by institutionalizing sub-wickedness through e.g. story exchange (Martin, 2003).

**Trans-complex systems** represent the harnessing of affordances of complex systems by adding elements of persistent complicated organization to complex systems, exemplified by various systems in the “sharing economy” (e.g. AirBnB, Uber, see e.g. Hamari et al. 2015), smart grids (see e.g. Clastres, 2011), social media movements (e.g. Anonymous, see e.g. Beraldo 2016), and terrorist networks (e.g. Bohorquez et al., 2009). If the epitomizing example for complex systems is flocks, the corresponding for trans-complex systems is an organized herd: the self-organization of the system is put under a simple scaffolding structure to align and direct it toward some central goal (Wimsatt and Griesemer, 2007). In other words, these are often loosely organized groups based on disseminated designs, shared views, and norms for alignment, rather than direct top-down control. By optimizing either structural features of the interaction between components (as in social media) or of the components themselves (as in selection-induced morphological change of animal behavior in literal herding, see e.g. Marshall and Weissbrod 2011), a level of structural complexity is induced. Non-decomposability develops specifically in the interaction between structural change and dynamical emergence in these systems (see also Lane, 2016).
Sub-wicked systems are wicked systems that are small enough to fit into the range of human cognition, and that have not outgrown our capacity to design or govern them. In other words, what delimits sub-wicked from wicked is the limits of human comprehension, which is, however, highly relevant in the methodological context. Sub-wicked systems can be exemplified by local social contexts such as families or workplaces, and early human societies. If societal systems – the realm of human politics – are wicked, then social systems – the realm of everyday human, and certain other mammal, interaction – are sub-wicked.\(^4\)

Sub-wicked systems exhibit wicked problems: they are recalcitrant to formal methods, but they are small enough for us to handle cognitively. It is no coincidence that we possess the capacity to do so: we are adapted specifically for dealing with sub-wicked systems, as human intelligence developed in, and perhaps even in response to, exactly such a context (e.g., Read, 2012; Tomasello et al., 2012; Van der Waal, 1982). The nature of wickedness is captured by the feedback interaction between emergence and the patterns resulting from that emergence, as individuals are capable of not only adapting to other individuals, but also to patterns emerging from precisely this interaction – what Goldspink and Kay (2007) call “reflexive emergence”. Hence, acting demands the ability to deal with constant social innovation: intrigues, new constellations, secrets, lies, and the relations between others and between others and oneself (Read, 2012).

In human groups, the primary mean of dealing with this is the narrative, which is simultaneously a way of theorizing behavior, exchanging meaning, and structuring community, i.e. just like the interaction modalities of social platforms shape online communities, narrative is a naturally evolved interaction modality that constrains and shapes social communities (c.f. e.g. Brown and Duguid, 1991; Orr, 1996). Narrative thinking provides “the genetically transmitted possibilities for interaction, resulting in relations and patterns, provide the framework in which the single individuals can realize their specific behavior and thereby jointly create the group-specific social structure.” (Hendrichs, 1983, p.739). Narratives are not only evolved to allow negotiation between the interests of individual actors and the collective, but are also a form of reasoning native to sub-wicked systems, as they embody an understanding of interaction of adapting systems:

\(^4\)This separation between social and societal/political existed in Greek understanding, and resonates with Hannah Arendt’s thinking: “This special relationship between action and being together seems fully to justify the early translation of Aristotle’s \(\zoon politikon\) by \(\textit{animal socialis}\), already found in Seneca, which then became the standard translation through Thomas Aquinas: \(\textit{homo est naturaliter politicus, id est, socialis}\) ("man is by nature political, that is, social"). More than any elaborate theory, this unconscious substitution of the social for the political betrays the extent to which the original Greek understanding of politics had been lost.” (Arendt, 1958, p.39)
1.1. BEYOND THE OPEN/CLOSED DICHOTOMY

narratives are capable of handling a number of key characteristics of sub-wicked systems such as heterogeneity, contingency and a multilevel nature (Richardson, 1990).

The resulting structure that forms from the interactions between such individuals are, like complex systems, robust in that, within limits, individuals and behavior are exchangeable without altering the social order (Hendrich, 1983). Hence, the social order has a reality of its own which acts on individuals and their behavior. The structures preexist individuals, and form them to a greater extent than to which the structure is changed by their specific individuality (Hendrichs, 1983). As Khalil (1995) puts it, the organization of groups of humans or of her close relatives is not an artificial entity or a vehicle used by organisms for preconstituted strategies. Rather, the organization is an individual with its own distinctive traits which are passed from one generation to the next through learning. This sort of individual influences the behavior of its members concerning rank, attachments, friendly relations, role divisions, and profiles. (Khalil, 1995, p.410)

An important glue of these social entities is what Tuomela (2007, p.338) calls “we-mode thinking and acting”: the members think and act for the group’s use and benefit. This makes it difficult to deny them status as “social agents in a genuine sense” (Niiniluoto, 2007, p.419), even with rather restrictive definitions of agenthood (e.g. Sibeon, 2004). These entities cannot only be regarded as “social”, “organizational” or “collective” actors (Mouzelis, 1991; Sibeon, 2004), but should also be seen as being endowed with agential causal powers (Pettit, 2009). That emergent structures in these systems should hence be considered real, is what results in the particular relationship between part and whole that is characteristic of open systems in general, and wicked systems in particular (see Goldspink and Kay, 2007). This relationship is in practice a dialectical one, as Levins and Lewontin (1985) put it:

“Part” and “whole” have a special relationship to each other, in that one cannot exist without the other, any more than “up” can exist without “down.” What constitutes the parts is defined by the whole that is being considered. Moreover, parts acquire properties by virtue of being parts of a particular whole, properties they do not have in

---

5Khalil instead refers to the broader group of “mammals” here, however, this is no longer believed to be the case – see Read (2012).
CHAPTER 1. EPISTEMOLOGICAL & ONTOLOGICAL LIMITS

isolation or as parts of another whole. It is not that the whole is more than the sum of its parts, but that the parts acquire new properties. But as the parts acquire properties by being together, they impart to the whole new properties, which are reflected in changes in the parts, and so on. Parts and whole evolve in consequence of their relationship, and the relationship itself evolves. These are the properties of things that we call dialectical: that one thing cannot exist without the other, that one acquires its properties from its relation to the other, that the properties of both evolve as a consequence of their interpenetration. (Levins and Lewontin, 1985, p.3)

As Fuchs (2007, p.18) argues, this interpretation of emergence amounts to “a reformulation of dialectical philosophy,” despite the lack of explicit acknowledgment of the “dialectical tradition and heritage of the philosophy of nature in the line of Friedrich Engels’ and Hegel (Fuchs, 2003, p.74) (one exception to this is Carneiro’s (2000) wonderful study of the transformation of quantity into quality, providing a complexity-perspective on the law stipulated by Hegel (Bukharin, 1925) and studied by Marx and Engels in Anti-Dühring and Dialektik der Natur.)

The capacity for the component to adapt to the whole means that events and acts play out on multiple levels, and that the boundaries dividing different levels are quasi-porous: there is, for example, no biological phenomenon whose causes and consequences play out in only one context. Hendrichs (1983) provides an example of this multi-levelness:

When performing their defecation ceremony at a specific place on the boundary of its territory, dik diks do at least five things at once: they excrete urine and dung; they mark their territorial boundary with optical and olfactory signals; they claim their territorial ownership up to that point; they strengthen their attachment to that place; they strengthen the integration of their group. (Hendrichs, 1983, p.741)

This interconnection between levels is also expressed in that sub-wicked systems have two types of emergence, while complex systems display only one. Complex systems display what Gilbert (2002) calls “first order emergence”; in which interactions among individual components result in a whole. Subwicked systems also display what Gilbert (2002) calls “second order emergence” and what Goldspink and Kay (2007) call “reflexive emergence”, where the components are able to recognize and adapt to the emergent products of their own interaction, resulting in “the amazing variety and mutability of social [and societal]
1.1. BEYOND THE OPEN/CLOSED DICHOTOMY

systems” (Bunge, 1998, p.8). This implies that we may also consider some emergent structures as having capacity for agency, again illustrating the reality and capacity for agency of structures in such systems.

“Wicked systems” is a reference to Rittel and Webber’s (1973) term “wicked problems”, which describes a class of problems that are characterized by a set of epistemological griefs; e.g. that they lack definitive formulations, that it is unclear when and if we are finished solving them, that they are caused by – and are the causes of – many other similar problems, and that they require uniquely tailored solutions. The reference implies that this is in fact not only a description of isolated problems as such, but that it captures something fundamental about the nature of the systems that generate such problems.

Wicked systems are arenas of and for innovation, with their constituents constantly trying to outsmart one-another, reaping their own benefits, reacting to threats from other constituents, as exemplified by large human societies and ecosystems over evolutionary time. This produces a situation where complicated organization and complex dynamics are in a constant state of re-negotiation, constantly challenging any settlement of the system into a level hierarchy, constantly facing the system with qualitative novelty that other components have to react to. This has two immediate results. First, the dynamics of wicked systems cannot be understood in terms of functions, but rather as emergent externalities from underlying component interaction. Secondly, qualitative change, i.e. change in kind,

6The question may here arise whether wicked systems are in fact “systems” at all, if they fulfill no functionalities and have unclear boundaries. Cambridge Dictionary defines “system” as “a set of connected things or devices that operate together”. Oxford English Dictionary defines it as: “1. An organized or connected group of things. 2. The whole scheme of created things, the universe. 3a. A group or set of related or associated things perceived or thought of as a unity or complex whole. 3b. A set of persons working together as parts of an interconnecting network.” etc. These definitions clearly cover wicked systems as well, as they are often “perceived or thought of as a unity of complex whole” and they do consist of “connected things or devices that operate together” (at least in the broader sense of “together”). This seems to imply that wicked systems indeed fulfill the dictionary definition of the term (one might object that the often ill-defined boundaries of their components challenge – or at least make problematically recursive – definition 1, but in any case it is clear that they are commonly “perceived or thought of as a unity or complex whole”). So far so good. However, there is a second line of criticism against the term, basically arguing that the concept “system” itself is associated to what we call “complicatedness”; as Jenkins (2010, p.142) argues, “notions of system may encourage us to tell the wrong story about humans”. Bunge (1999, p.5) responds to this with that: “trying to avoid the word ‘system’ just because of its association with Parsons or Luhmann is like boycotting the word ‘nation’ only it is abused by nationalists.” While I remain unconvinced about the ultimate usefulness of nations, I do find that using “system” is both legitimate and indispensable. That being said, it should also be noted that what I mean by “system” in this context also resonates strongly with Delanda’s (2006) notion of “assemblage”.
related to what Archer (e.g. 2013a,b) calls *morphogenesis*, is the *modus operandi* of wicked systems, and to disregard it in their study “signifies nothing less than the wilful obliteration of [the] very subject matter” (Arendt, 1958, p.57).

It is the innovation of underlying actors that upsets any level hierarchical organization, thereby ruining prospects for near-decomposability, by constantly rewriting the “rules of the game”. As Wicked systems are non-decomposable, short runs (see section 1.1) are not just hard to find, there is no guarantee that there even exists a meaningful short run – wicked systems may in fact be seen as systems that largely lack relevant short runs and thereby also opportunities for powerful formal modeling. Levels of organization have been described as “stable foci of regularity and predictability”, and as such, the existence of levels of organization in itself must be expected to act as attractors to adaptive processes: they should self-reinforce and self-stabilize over time (Wimsatt, 1994) since adapting systems evolve in such a way as to minimize uncertainty in their environment (Levins, 1968). However, as Wimsatt (1994) points out, this is only half the story. In a competitive situation, i.e. a situation with what Khalil (1995) calls individuality, entities under competition (be they organisms, organizations or humans) will themselves seek to be as unpredictable as possible to their competitors, which would make it adaptive to also break up level hierarchies.

Wimsatt (1975, p.181–185) furthermore argues that Simon’s principles take only ease of design and assembly into account, not optimality of function. Optimality of function, of course, may be under strong selection pressure, and when it is we should expect this to cause breakdowns in level-hierarchical organization. The reason is that there is no convincing argument for why a style of organization that simplifies assembly and design would also make for optimal function. Intuitively this expectation seems to be carried out in reality. Technological artifacts that are mass-produced (strong pressure for adaptability, cheap assembly and easy maintenance) contain more standard components, and are simpler in their architecture, than ones that are highly specialized and produced only in very few numbers.

The wickedness of human communities developed from an initial sub-wicked organization of great ape foraging groups and communities (Grove et al., 2012). Through the development of increasingly tight and multifaceted cooperation (Tomasello et al., 2012), early hominin (after the concestor\(^7\)) communities may have begun to accumulate the physiological and cultural affordances necessary for cooperative functioning.

\(^7\)The concestor refers to the last common ancestor of humans and chimpanzees (Dawkins and Wong, 2005). The nature of the concestor can be inferred from strong similarity in lifestyle and morphology between fossil great apes (Wrangham and Pilbeam, 2002) from the relevant time space 5-7 million years ago (Kumar et al., 2005).
to transition from social/sub-wicked to societal/wicked, through the “seamless” (Hughes, 1986) integration between the social and technical, transforming, eventually, the emerging human community into a larger interconnected sociotechnical system (Geels, 2004). Technology here functioned simultaneously as powerful modes of interaction between actors, while also themselves being interactors (Hull, 1988). Due to the complicated and near-decomposable structure of technology, technological artifacts can maintain vast and heterogeneous arrays of interaction and thereby integrate cultural systems by tying its various domains together. Hence, through such socio-technical structures, human culture became capable of maintaining interaction systems where every node is densely connected to just about all domains of the web. The seamless web – which is barely but importantly discernible in chimpanzees technology-assisted extractive foraging (e.g. Biro et al., 2003) and “politics” (Van der Waal, 1982) – was thereby simultaneously integrated through weakly constrained interaction, and separated, through specialization. This is a procedural recipe for full-scale wickedness.

Interactions in these seamless webs have a strong enveloping competitive component but display also the whole spectrum of ecological interactions (competition, symbiosis, neutralism, parasitism, commensalism and amensalism; see Sandén and Hillman 2011, p.407). Symbiotic interactions may give rise to self-organized systems toward the trans-complicated and trans-complex regimes; e.g. bundles of value chains as described by Sandén and Hillman (2011, p.404-406). Parts and levels may over time co-adapt to become increasingly co-dependent; compare with examples of symbiotic origins of complicated systems (see Leigh, 2010; Roze and Michod, 2001). The boundary between wickedness and trans-qualities is thereby porous.

Components act and react within neighborhoods in the seamless web, and, since each is part of many neighborhoods, change is liable to propagate across the system. Dynamically and macroscopically, this leads to two dialectical dynamical regimes: transition and lock-in. Transitions are self-propagating waves of qualitative “reconfigurations” of and by components, traveling across neighborhoods in the seamless web (Geels, 2002; Lane and Maxfield, 1997). These may form potentially system-wide cascades of change (Geels, 2011; Lane, 2011a; Lane et al., 2009b; Schiffer, 2005). However, if locally beneficial reconfigurations cannot be made, change will be resisted, and if such criteria, posed by large numbers of strongly interconnected components, are combined, the range of actually viable innovations will be strongly constrained and channeled. The result is a lock-in, such as by a dominant design (Utterback and Abernathy, 1975) or a
sociotechnical regime (Geels, 2002; Rip and Kemp, 1998). The combined effects of cascades and entrenchment of effects is a potentially unlimited horizon in time and scope for consequences of actions.

Due to the strong and heterogeneous connections that crisscross wicked systems, it is impossible to divide such systems them into realistic pictures: only the full system is enough to represent the system, and that will never be theoretically achievable (see also Cilliers, 1998, 2002). Any “picture” captured will necessarily be from a perspective, and rarely subject to universal agreement. Even if we could obtain a “realistic picture”, this would frequently not help much since the system changes unpredictably over time – including as a direct result of us interacting with it.

Since there is no axis of reduction, there is no native method fit to fully cover wicked systems, and hence, as Paper I discusses, the attempts to formally deal with wicked systems have generally focused on treating them either as complex or complicated systems – “the reductionist and the functionalist approaches extend the tool found successful in one domain to decipher the other” (Khalil, 1995, p.414-415) – neither of which matches their ontological nature and neither of which can be said to have been very successful. They are, in Archer’s (1996) terminology, conflating the systems either downward or upward.

In the context of social theory, the former – ontologically individualist reductionism – relates to what Gilje and Grimen (1992) call the action paradigm, in which actors are viewed as free agents whose interaction leads to varying types of phenomena. In this view, structures and systems are merely relatively stable patterns, either emerging from or simply constituted by aggregated individual action. In relation to social meta-theories, this implies treating society as ontologically flat, as – to a certain extent – in assemblage theory and analytical sociology (DeLanda, 2006; Hedström, 2005) (see chapter 3).

<table>
<thead>
<tr>
<th>Primacy:</th>
<th>Action paradigm</th>
<th>Fact paradigm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bottom-up</td>
<td>Top-down</td>
<td></td>
</tr>
<tr>
<td>Complexity:</td>
<td>Dynamical (complex)</td>
<td>Structured (complicated)</td>
</tr>
<tr>
<td>Ontology:</td>
<td>Individualism</td>
<td>Holism</td>
</tr>
</tbody>
</table>

The latter – ontologically holistic reductionism – relates to functionalism, and more broadly to what Gilje and Grimen (1992) call the fact paradigm, i.e. in which social facts are the primary object of study: institutions, social structures, organizations etc. The term “social structure” is used in a strong way, while “agency” is used weakly, i.e. as simply ways in which structures are reproduced. This encapsulates methodologically holistic (e.g. social fabric matrix, see
1.2. UNCERTAINTY IN WICKED SYSTEMS

Hayden 1982, and institutional dynamics, see Radzicki 1988), functionalist and system-focused approaches, as well as Luhmann’s (1986; 1995) system theory (as Wan (2011) convincingly argues), and is traditionally understood (e.g. by Danermark et al. 1997) to have Émile Durkheim\(^8\) as its early proponent. (It can also be seen rather explicitly represented in Mumford’s (1966) concept “mega-machine”, in which society is seen as an “archetypal machine composed of human parts”.)

Hence, from a social theory standpoint, the combination between complexity and complicatedness is linked to the meeting between the fact paradigm and the action paradigm, between agency and structure, a meeting whose resulting ontology is either “dualist” (as Giddens, 1984) or a “dualism” (as Archer, 2000). (The resulting implications of wicked systems on the view of social structures will be discussed in Chapter 3.) The intuitive, but non-formalized, alternative to these two directions, is to cast wicked systems into the realm of subwickedness by using narratives as the methodology. This is an attractive option since it matches the ontology of wicked systems (as we will see in Chapter 2). That narratives are compatible with the combination of complicatedness and complexity does, however, not mean that they are very capable of dealing with either – indeed, this is why such systems are perceived as overwhelming to begin with.

### 1.2 Uncertainty in Wicked Systems

As we have seen, complicated systems permit exact prediction, while complex systems are more uncertain: due to nonlinearity, one may only predict their general dynamics, rather than specific states. Wicked systems are defined by an even deeper level of uncertainty, which lies at the core of what dealing with wickedness is about: both prediction of future states, and of future dynamics rely on the possibility of decoupling “the game” from “the rules of the game”. But in wicked systems, the game and its rules frequently change dynamically on similar time scales: qualitative – and, from the perspective of decomposition, ontological – change is occurring on the same time-scales as quantitative change.

This means that wicked systems are less like a game in which defined entities play according to set rules, and more like a boiling pot of change where each discernible shape and structure, resulting from the interplay of a variety of counteracting forces, may very well dissolve as quickly as they evolve. Yet, entity

\(^{8}\)Boudon (1981, p.155) however refers to this view as a result of a “superficial reading of Durkheim”, that has given rise to “the myth of Durkheimian holism” (Cherkaoui et al., 2008, p.18). Durkheim’s empirical analyses are in fact “much richer, subtler and promising” than this would suggest (Cherkaoui et al., 2008, p.39).
interaction, game-play and gradual quantitative change are ubiquitous, as many
structures display surprising persistence, but this may just as soon be replaced by
rapid fundamental transitions and rapid qualitative change, as quantitative change
becomes qualitative (Carneiro, 2000).

This perspective on level of near-decomposability of different system do-

mains gives a new perspective on the separation between static ontologies, char-

acterized by the belief that “change is only a momentary departure from equilib-

rium or harmony, which would be the ideal state of affairs” (Bunge, 2011, p.20),
and dynamic ontology, in which the central thesis is that “statis is a particular and
ephemeral case of process: that every state of a thing is either the initial, interme-
diary or final phase of a process” (Bunge, 2011, p.20). The latter is exemplified
by Emirbayer’s (1997) relational sociology:

Sociologists today are faced with a fundamental dilemma: whether
to conceive of the social world as consisting primarily in substances
or in processes, in static ‘things’ or in dynamic, unfolding relations.
Large segments of the sociological community continue implicitly
or explicitly to prefer the former point of view. Rational-actor and
norm-based models, diverse holisms and structuralisms, and statisti-
cal ‘variable’ analyses – all of the beholden to the idea that it is
entities that come first and relations among them only subsequently
– hold sway throughout much of the discipline. But increasingly,
researchers are searching for viable analytic alternatives, approaches
that reverse these basic assumptions and depict social reality instead
in dynamic, continuous, and processual terms. (Emirbayer, 1997,
p.281)

This separation is not as much a binary dichotomy as a gradual scale; Bunge
(2001, p.32-33) distinguishes (in a somewhat biased terminology) between “rad-
ical” and “moderate” dynamism, where the former invokes the full phanta rhei
of Heraclitus, and sees the world as constituted by processes or events rather
than entities, while the latter admits that “some traits remain invariant throughout
certain changes” (Bunge, 1977c, p.279) and that change is sometimes enabled
exactly by this permanence.

The position on this scale is not only defined by the system category, but can
also vary over time. For example, the ontological stability of society is rooted
primarily in the stability of material culture, and ontological change is driven
primarily by innovation and changes in the material base (see e.g Archer, 2014;
Bauman, 2013; Danermark et al., 1997; Elder-Vass, 2017). Hence, as the power
1.2. UNCERTAINTY IN WICKED SYSTEMS

relation between these two factors change, as material culture becomes more liq-
uid, not least due to digitalization, human society becomes more prone to qual-
itative change – in other words moving toward a more dynamic ontology (see
Törnberg, 2016a).

Dynamics of qualitative change is exceptionally difficult to study, even com-
pared to complex and chaotic systems. For example, the stock-market, while
infamous for its unpredictability, is only quantitatively chaotic: on the relevant
time-scale, it is ontologically fixed, and hence, it is only chaotic in relation to its
state, not its kind (Byrne and Callaghan, 2013). The unpredictability of the stock-
market is an example of what we can call “second order chaos”, in reference
to Gilbert’s (2002) “second order emergence”: if predictable patterns emerge,
they will be discovered and undermined by constituent agents. Conversely, in
wicked systems, the capacity of agents to detect emergent structures allows such
second-order chaos to take a qualitative form (what could be rightly described as
“ontological chaos”).

Because of this, consequences of action in wicked systems is shrouded in
deep uncertainty, described by Lane and Maxfield (2005) as an “ontological un-
certainty”: not about the truth of well-defined propositions (“truth uncertainty”),
nor about the meaning of a given statement (“semantic uncertainty”), but about
what entities that inhabit the world, how they may interact, and how interactions
and entities change through interaction (Bonifati, 2010). Uncertainty keeps us
from aligning action to respond to future ill effects (game theory; e.g. Gintis 2000; Ostrom 1990), but it also prevents us from designing effective interven-
tions without high likelihoods of causing unexpected troubles in other domains.

Clearly, however, there are many important cases where we can surely make
assumptions of near-decomposability also in wicked systems, and where we thus
are able to bring powerful scientific approaches to bear. For the purposes of
complexity science, it would seem reasonable that certain subsystems - such as
crowd behavior, protein-folding, or the ceteris paribus fate of a new trait in a
population - can be argued to fit this description. The dynamics of cars and
people play themselves out over much shorter time scales than that on which
urban systems, roads, traffic regulation and so on, change. Such phenomena are
also often ephemeral, which bounds the problem even further. For example, at
night the traffic jam dissipates and leaves no traces that affect tomorrow’s traffic.

But what about evolutionary societal and ecological phenomena more in gen-
eral? For example, what about sociotechnical transitions, evolutionary radiation
events, or other wicked problems? Wicked systems in general are open systems,
in which many and far-flung types of processes co-exist, co-evolve and have an
impact on each other on overlapping timescales and levels of organization. They involve discontinuous, qualitative change as well as cascade effects (Lane, 2011a) whereby change strongly and rapidly feeds back into the conditions for further change. Such systems are, to say the least, hard to contain in a Simonean compartment with a “short run” over which, for example, transitions can be studied against the background of an unchanging external environment.

Put in another way, the type of hierarchy that micro-emergence assumes does matter in wicked systems, as Bickhard (2000, p.326) argues, “emergence presupposes a notion of levels”, but neither causality nor structure is restricted to the hierarchies: we have interpenetration and overlaps, as well as multi-directional causality. It is not only that “an \( n \):th level system is composed of things on level \( n-1 \)” (Bunge, 2004b, p.133), but it may also include elements from any other levels. Simon (1991) was of course aware of the existence of such interpenetrations, but nevertheless, assumed that there was enough hierarchical structure to make modeling possible, and thus emphasized that which falls within the hierarchies, proposing that hierarchical models provide an adequate approximation (in line with the spirit of the time). In complex and complicated systems, such interpenetrations are indeed exceptions, but in wicked systems they are instead the norm. More figuratively speaking, if complicated and complex systems are like onions, where you can neatly peel each layer from the next with only some thin slimy strings connecting the two, wicked systems are more like a mango: any attempt at separating them into levels, or in fact even trying to peel them, will most likely result in a gooey – but delicious – mess.

1.3 **Boundaries of Wicked Systems**

Due to the high level of interconnection and dynamic complexity in wicked systems, it tends to be difficult to define system boundaries. Decomposable systems generally have clear defined boundaries, and in turn consist of distinguishable entities. For example, the birds in a flock might interact in complex ways leading to unpredictable emergent dynamics, but at least the birds themselves can be distinguished as clearly defined, separate entities that are unlikely to evolve during the time-scale of the flock. This is often not the case in wicked systems. As Cilliers (2001) puts it, what we call wicked systems

have structure, embodied in the patterns of interactions between the components. Some of these structures can be stable and long lived [...], whilst others can be volatile and ephemeral. These structures
1.3. BOUNDARIES OF WICKED SYSTEMS

are also intertwined in a complex way. We find structure on all scales. [...] [N]on-contiguous sub-systems could be part of many different systems simultaneously. This would mean that different systems interpenetrate each other, that they share internal organs. How does one talk of the boundary of the system under these conditions? (Cilliers, 2001, p.4-6)

Furthermore, since ideas of spatial continuity do not apply to these systems, one of the foundations on which we traditionally base the notion of boundaries is turned on its head. As Cilliers (2001) puts it:

We often fall into the trap of thinking of a boundary as something that separates one thing from another. We should rather think of a boundary as something that constitutes that which is bounded. This shift will help us to see the boundary as something enabling, rather than as confining. [...] [An] implication of letting go of a spatial understanding of boundaries would be that in a critically organised system we are never far away from the boundary. If the components of the system are richly interconnected, there will always be a short route from any component to the “outside” of the system. There is thus no safe “inside” of the system, the boundary is folded in, or perhaps, the system consists of boundaries only. Everything is always interacting and interfacing with others and with the environment; the notions of “inside” and “outside” are never simple or uncontested.9 (Cilliers, 2001, p.5)

So not only are wicked systems under constant ongoing ontological change, but their boundaries are far from as clear as positivist science tends to imagine them. In wicked systems, entity interaction – on and between all levels – is ubiquitous and central to the dynamics of the system. But since there are also relations with the surrounding environment, it is generally not obvious where the boundaries are to be drawn. It is more of a question of framing: we frame the system by describing it, but reality constrains where the frame can be drawn (Cilliers, 1998, 2001). The boundary is neither only a construction nor only a natural thing – it is a mix and an ongoing interaction between these (Richardson and Lissack, 2001).

---

9 It is interesting to note that the dominance of boundaries can in fact be said to be a general fact of high-dimensional systems. To see this, consider the ratio between surface area $S_n$ and volume $V_n$ for a n-dimensional hypercube: $\frac{S_n}{V_n} = \frac{2n\pi^{n-1}}{n!} = \frac{2n}{\pi}$, now clearly as $\lim_{n\to\infty} \frac{S_n}{V_n} = \infty$. 
This clearly has implications for how to approach wicked systems scientifically. While complex systems require radically new scientific methodologies to deal with the intricacies of relational reduction, the poor decomposability of wicked systems calls for something far more radical still (Castellani and Hafferty, 2009). The constant ontological transformation clearly implies a weaker type of knowledge claims, and Cilliers (1998; 2001) accordingly suggests a significantly less universal conception of scientific knowledge: as contextual, local and specific in time and space. This may sound postmodernist in a negative sense, but there is a significant difference between this and full-blown relativism. That the possibilities for prediction and description are limited does not mean that anything goes, as the most radical postmodernist theorists could lead us to believe; the world can be known, even if that knowledge is contextual and time limited. While this has implications for positivism, it in no way downplays the importance of scientific work. Quite the opposite: that our knowledge of a system is only local and temporary emphasizes the importance of knowing how to learn about a system (Byrne and Callaghan, 2013). It however affects how to scientifically relate to the systems under study, as Actor-Network theorist Law (2004, p.7) puts it: “... in this way of thinking the world is not a structure, something that we can map with our social science charts. We might think of it, instead, as a maelstrom or a tide rip. Imagine that it is filled with currents, eddies, flows, vortices, unpredictable changes, storms, and with moments of lull and calm”.

But our scientific knowledge will not only be contextual and time-limited: because of the constant ontological transformation we have no stable ground to stand on required for a reduction of the system; if we are to be sure that the dynamics of the system is the same, we cannot represent a complex system with anything less complicated than itself (Cilliers, 2001). Since simplifications are of course necessary for any meaningful scientific work, this basically means that any representation will necessarily be flawed, and we cannot even know in which way it is flawed.

In wicked systems involving humans, the question of framing and system boundaries is decidedly at its hardest. In such systems, the choice of framing will not only be a scientific question, but an ethical question, as the choice will necessarily affect the reality of the system (Byrne and Callaghan, 2013; Cilliers, 2000): constructing a boundary for a system can mean that this boundary becomes more present in the system; temporary structures may gain longevity by being described. This is part of what Bhaskar (1979) calls the relational condition of the social sciences, which is one of the ways that wickedness specifically plays out in societal systems, which will be the focus of the next chapter.
So far, the separation has dealt with different factors connecting to openness and closure, and the nature of social structures, and we have seen that the separation does not cut cleanly between social and natural systems, but that it is rather, as Khalil (1995, 1999) argues, a question of system organization. The contribution thus far can be seen as a continuation of Khalil’s work. However, Khalil focuses only on what Bhaskar (1978) refers to as the ontological (i.e. the activity-, concept-, and space-time-dependence of social structures) and epistemological (i.e. the openness of social systems) limits of naturalism, and neglects what Bhaskar calls the relational conditions of social science. Our description of society cannot completely suffice with these naturalistic notions about openness/closure and complexity, as these miss “that the self-organization of society is not something that happens only blindly and unconsciously but depends on conscious, knowledgeable agents and creative social relationships” (Fuchs, 2007, p.27). Human agents – the components of the social realm – are entities with unique causal powers – a fact that we cannot overlook when formulating our ontology of these systems. In Bunge’s (1998, p.122) words: “Human beings are the creators, reformers, and destroyers of all human social systems, and social laws and rules are nothing but the patterns of being and becoming of such systems”.

The next section approaches this condition, but from a rather different viewpoint: we will look at how the relationship between structure and emergence evolves and plays out in social systems, and from that review the implications of the relational condition (relating to Paper III and Törnberg and Andersson 2016). We begin our exploration of human social systems just across the boundary into the realm of natural systems.
2.1 Narratives before Society

Our chimpanzee cousins (*Pan troglodytes, Pan paniscus*) illustrate a picture that most probably represents the natural origins of wickedness and human societies. Hominins (*Homo* and extinct relatives) diverged from Pan 5-7 million years ago, and fossil remains suggest that Pan has remained similar ever since. Apart from their only very slightly complicated tools, they have no higher level of organization than the community, which is organized through friendships and maintained by intimate daily contact. Like human hunter-gatherers still, and many other species (e.g. wolves), they have a fission-fusion organization where smaller foraging groups continually form and dissolve within the community. This permits the community to grow relatively large and cohesive while not putting too much pressure on the land and its resources. Large fraction of their time is spent in the upkeep of their relations, through touching, grooming and exchanging favors, building a mutual trust between each member of the community that allows them not only to (locally) hunt together, but also to lend assistance in fights between communities and in internal conflicts and so on. This way of maintaining a group imposes clear limits on the size of the group: two chimpanzees who have not previously met cannot know whether to trust one another, and because of this, when chimpanzee groups grow too large, they tend to destabilize and rupture into two separate groups (Moffett, 2013, p240-241). Furthermore, since intimate trust is the only form of collaboration, communities never cooperate, but compete fiercely for territory and food (de Waal 2005; bonobo communities mostly mingle amicably but remain separate and compete for both territories and females; see Kano and Ono-Vineberg 1992).

Human groups attained the capacity to expand this limitation by letting communication – gossiping and story-telling – supplement direct intimate interaction (Dunbar, 2004). Gossiping let humans know about the trustworthiness of other humans in the group, without requiring direct interaction, thereby allowing human troops to build trust through reputation-driven indirect reciprocity (Nowak, 2006): more tightly and across significantly larger groups than other apes. This however put great pressure on cognitive capacity, as navigating such groups demands the ability to deal with constant social innovation: intrigues, new constellations, secrets, lies, and the relations between others and between others and oneself (Read, 2012). According to the “Machiavellian intelligence” (e.g. Byrne and Whiten, 1988) and “cultural intelligence” (van Schaik and Burkart, 2011) hypotheses, these requirements were key drivers of the evolution of human intelligence and large costly brains (Aiello and Wheeler, 1995).

Regardless of the validity of these hypotheses, which are currently under
2.1. NARRATIVES BEFORE SOCIETY

strong development, it is clear that gossiping and narratives are deeply ingrained in human cognition: Michotte’s (1963) experiments in attribution of causality provides clues on just how deep. In these experiments, observers were shown two or more small, colored rectangles in motion on a screen, and when asked to describe what they saw, they intuitively imposed elaborate cause-and-effect stories in which the moving rectangles were assigned intentionality and meaning through intricate plots, exemplified by descriptions such as “the red ball hit the blue ball” or “the red ball is chasing the blue ball” (see also the Thematic Apperception Test, Murray 1938). Narrative structuring assigns intentions and cause-and-effects to sequence of experienced events by drawing them into unified plots, through which they take on significance and meaning (Ricoeur, 1980). There is certainly some flexibility in this description, but the narrative has to represent events in a way that is perceived as coherent, i.e. that fulfills implicit assumptions about human behavior and cause-and-effects. The structuring process that configures events into such plots is interactive or dialectical, moving between a temporal meaning that might explain or show a connection among the events and the events’ resistance to fitting the construction (Polkinghorne, 1991).

While narrative competence emerges at an early age and is culturally universal (Mancuso, 1986), its specific coherence relies on a culturally specific understanding of human characteristics, which are at the same time constructed by the narrative; “the narrative constructs the identity of the character, what can be called his or her narrative identity, in constructing that of the story told. It is the identity of the story that makes the identity of the character” (Ricoeur, 1992, p.147-148). Narrative structuring shares some similarities to the visual configuration described by Gestalt psychology, in which, for example, three dots can be seen as the angle points in a triangular figure (Gurwitsch, 1964). In the same way, narrative structuring has a part-whole or Gestalt organization: just as spatial organization (e.g. of a kitchen) consists of topological relations (up, down, left, right, next to, inside, etc.), temporal organization (e.g. a trip to the store) consists of causal and enabling relations. In this way, narrative structure is used to make meaningful the actions of actors, public individuals and groups, and governments and institutions, based on implicit behavioral assumptions. This capacity, according to Lakoff’s (1987) theory, moves us to recognize the patterned bodily experience of going from an initial state, through a sequence of events, to a final state. This source-path-goal schematic pattern serves as the metaphoric origin for the type of temporal organization that makes the elements of episodes and stories understandable as parts of a temporal whole (Polkinghorne, 1991).

Narratives are not only used for the understanding of external actors; they
are also widely believed to be the way we understand ourselves (Polkinghorne, 1988). When people we meet tell us the story of their lives, we are not surprised that they have one – in fact, we would be surprised (or even worried) if, in trying to get to know someone, all he or she had to share were a series of unconnected events (see Lacan’s notion of a psychotic structuring Lacan 1960. Stories are important for having a vision of the self and to understanding and connect to others. The reconstruction of a coherent self-narrative has been held as a therapeutic goal since Freud’s inauguration of psychoanalysis, which Spence (1982) describes as the art of turning disordered pieces of information from patients into coherent stories. This view conceptualizes anxiety as an evolved imperative to narratively relate ourselves to the larger community, and shows clearly how narratives may have evolved to align strongly individualist members to form a functional group in early communities.

2.2 Narratives in Society

However, while gossiping and story-telling allowed Homo sapiens to form larger and more stable bands, even gossip has its limits: the maximum “natural” size of a group bonded by gossip is about 150 individuals (Dunbar, 1992, 1995; Dunbar and Shultz, 2007). As we pass this threshold, informal organization starts breaking down, and there starts to be the need for things like formal ranks, titles and rules to keep the group together. Clearly, while gossiping allowed a powerful extension of group size for early humans, it was not a foundation which could hold together cities and empires with thousands, and eventually hundreds of millions of inhabitants.

The way that this transition occurred holds central significance for the relationship between individuals and structures in society, and for how we can understand societal systems. Passing this threshold occurred through an exaptation (Bonifati, 2010) of narratives, in which the use of narratives and story-telling no longer only included individuals as actors, but also collective actors and social structures. In other words, the cognitive tools that we developed to deal with sub-wicked systems, story-telling and gossiping, were applied to deal also with emergent social structures: the groups and social structures became actors in narratives (Henshilwood and d’Errico, 2011).

This most likely first played out in the form of sizable – in recent egalitarian hunter-gatherers sometimes 2,000 strong – communities composed of several bands whose members were distinguished on the basis of multiple society-specific, and socially acquired, labels (Moffett, 2013). These features made it
2.2. NARRATIVES IN SOCIETY

easy for humans, unlike wolves and chimps, to recognize members of other
groups at a glance (Diamond, 1992, p.220). Through such labeling, they accom-
plished what Moffett (2013) calls “Anonymous Society”, where one no longer
need to personally know another member to know how to relate.

This allows large groups to cooperate, as strangers can cooperate success-
fully by identifying one another through symbols and belief in common myths.
Through this common ground, social entities such as organizations and commu-
nities, attain “the causal power to influence the behaviour of human individu-
als” (Elder-Vass, 2007d, p.465). Instead of organizing through the interaction
between every single individual, they organized toward a mythical idea of a col-
lective, rooted in common stories and labels (e.g. Johnson and Krüger, 2015).
Large-scale human cooperation – whether a modern state, a medieval church, an
ancient city or an archaic tribe – is ultimately rooted in common narratives which
enable mutual trust: two tribesmen who have never previously met can cooperate
on the basis of their common membership in a thought in-group, defined by a
complex web of shared belief in stories about the world. Similarly, today, two
persons who have never met can cooperate and exchange goods through their
common belief in ideas such as trade, capitalism, and the value of money.

In small, traditional societies, common stories of ghosts, gods and spirits may
have functioned to cement the social order, and today, our modern institutions
function on the same basis (Harari, 2014). Just as rumors and stories function
in small groups to assign characteristics to the actors involved, we today assign
human characteristics, agency, and even emotional states to imagined emergent
actors. This can be observed in brands being seen as trustworthy or manly, or in
the depiction of “the market” as worried or stressed. Such characteristics hardly
make sense as descriptions of imagined collectives, but to an extent, the sto-
ries become self-fulfilling prophecies (if you own stocks, and you understand the
market as a “worried” actor, this is likely to affect your trading decisions so as to
induce volatility.)

In this way, narratives are used not only to understand the inter-human world,
but also the roles and behavior of institutions and social structures: humans not
only adapt to emergent structures, but they do so by assigning meaning and roles
to them. They shape our behavior and organize our society, yet they – strictly
speaking – do not exist outside of our collective stories. Over the years, there has
developed an incredibly complex network of such stories. Within this network,
ideas such as “the market” can accumulate immense power. We navigate this
symbolic network, using stories as building blocks to construct new stories.

This symbolic network has increasingly come to dominate how we under-
stand the world and ourselves, in what Baudrillard (1994) calls the “precession of simulacra”. As this operates in the same symbolic landscape in which we understand ourselves, this is also the lens through which we see ourselves: we relate ourselves to the categories and groups to which we see ourselves as being part. The dual reality of humanity is hence mirrored in our individual selves: Lacan (1960) described this fundamental split as the *mirror stage* — in which the Symbolic, and Imaginary is formed, separated from the Real — the relation between which determines our very structuring: our narratives can never fully capture reality, which is reflected in cracks and tears in our selves. In other words, just as the defecation ceremony of *dik diks* is multi-dimensional, so does human symbolism and narratives cut across levels — but in an immensely more flexible and powerful way.

Let us, briefly, connect this historical exposition of the nature on narrative and the evolution of societal systems to contemporary social theory on the agency-structure relationship. As we have seen, wickedness is defined by the feedback interaction between complexity and complicatedness — the dialectics between agency and structure: just as in complex systems, patterns emerge from the interaction between constituents, in wicked systems, however, this emergence is reflexive (Goldspink and Kay, 2007): the patterns are observed, and met by adaptation from the individuals in the system. The narrative nature of human wicked systems means that in societal systems, this interaction between social structure and emergence takes a particular form: narrative, meaning, intentionality, and so on, are the building blocks of social and societal systems. This does not mean to argue that individuals should be treated as only products of narratives. As Kaidesoja (2007, p.82) points out: “it is surely one thing to say that the conversations, in which biological individuals engage in their lives, in many ways shape and modify their powers, and another to claim that people are nothing but conversational constructs.”

As we have furthermore seen, social actors, unlike the components of systems in the natural world, are capable of collective intentionality (Searle, 2010, 1995, 2006), sophisticated communication and creative collaboration, including what Sawyer (2003) calls “improvised dialogues” that make “distributed creativity” possible. This capacity to evaluate the social world is not limited to the external, but just as important is the capacity of self-awareness and reflexivity: human beings, as Sayer (2005) points out, are “evaluative beings”, continually monitoring and assessing their own behavior and that of others, in a narrative form. Evaluative beings tend to engage in what Archer (2003) calls “internal conversation”, that is, processes of continuous “internal deliberation”, which is “self-reflexive
2.3. NARRATIVES AND SOCIAL STRUCTURES

because it is a self-critical exercise” (Archer, 2003, p.105-106). The importance of these dialogue has been stressed in different contexts by a number of leading social theorists, such as Giddens (e.g. 1992), Archer (e.g. 2000), Sayer (2005), Elder-Vass (2007c, 2010), and Mouzelis (2007). Fleetwood (2008) has even gone as far as arguing that that reflexive deliberation via internal conversation is the “process that links social structure and agency” (Fleetwood, 2008, p.260). This, however, seems to underestimate other factors, such as more deep-lying social instincts, including emotions (Elster, 1999; Emirbayer, 1996), as well as what Bourdieu (2005) terms “habitus”.

This idea of self-evaluation also resonates with Taylor’s (1989) discussion of “strong evaluation”, that is, human beings’ capacity to evaluate their own preferences and beliefs, and thus to form “second-order desires” (see e.g. Callinicos, 2004). This is essentially what allows humans to change the games that they are playing, as they are playing them, since the notion of internal conversations emphasize that “our relationship to the world is not simply one of accommodation or becoming skilled in its games, but, at least in some ways, one of wanting to be different and wanting the world and its games to be different” (Sayer, 2005, p.35). This brings us into the importance of meaning and value in the context of human social structures in general, and in their qualitative change in particular.

2.3 Narratives and Social Structures

The exposition about narratives in human communities in this chapter has served to show how “agents are constrained and enabled by ideas, rules, norms and discourses” (Kurki, 2008, p.228), and that “‘meanings’ or ‘ways of conceiving’ that are dominant come to inform the intentions and the actions of agents” (ibid, p.224). As we have seen, both ideational and material aspects have to be brought in to adequately provide a view on the causal roles of social structures (Elder-Vass, 2017). Hence, when causation in social systems occurs through symbols and meaning, analysis becomes interpretation, or Weber’s Verstehen. This is part of what Bhaskar (1979) calls the relational condition of social science, in that it implies that social researchers are equipped with unique tools to understand the dynamic in such systems, since they are themselves part of their own research subject. This gives them access to the actors’ points of view, through the understanding of meaning. Such a research method implies entering into the shoes of the other, and treating the actor as a subject, rather than an object of our observations (Morehouse, 1994; Spradley, 2016). It also implies that unlike objects in the natural world, human actors are not simply the product of the pulls and
 pushes of external forces: individuals are seen to create the world by organizing their own understanding of it and giving it meaning.

Mechanisms in natural systems exist regardless of their meaning, while social structures are what they are through their meaning for the underlying individuals (Geertz, 1994). For example, if money had no meaning for individuals, it could not function as an explanatory mechanism: it would not be able to affect behavior. In other words, it is the interpretation of social structures by the underlying actors that grants them causative abilities; value and meaning are the stuff of the social realm, and so research is inherently about meaning. Because social structures exist only in virtue of the activity they govern, they do not exist independently of the conception that the agents possess of what they are doing in their activity; i.e. some theory of these activities. Since social structures are themselves social products, social activity must be given a social explanation, and cannot be explained solely by reference to non-social parameters (Bhaskar, 1979, 2010).

Because of this, hermeneutics is often seen as replacements to causal explanations of events within social science (Ricoeur and Thompson, 1983), but, as Archer et al. (2013) argue, this is taking it too far. Without hermeneutics, we cannot understand the meaning of an action, but identifying the meaning is not enough to explain what brought about the act: there are also beliefs, intentions, motivations, etc., at play (Bhaskar, 2010). There is hence a double hermeneutics in play in social systems, as “meaning has to be understood, it cannot be measured or counted, and hence there is always an interpretive or hermeneutic element in social science” (Sayer, 2000, p.17). In natural systems, the scientist attempts to interpret and create meaning in the object, but in social systems, the object has already been interpreted by the objects of study, an interpretation which is even part of its causal capacity. The formation of concepts is not only a part of the scientific work, as in the natural sciences, but also part of the scientific object. This is the “hermeneutic premises” (Collier, 1994) of the social sciences. Because of this, our view of societal systems is necessarily historical, value-laden and “situated”: there is no view from nowhere, so while our perspectives and knowledge is necessarily partial and relative, it is the best we can hope for: without somewhere to stand, no knowledge is possible (Nagel, 1989).

In short, since value and meaning are the stuff of the social realm, and what grants social facts their explanatory power, there will necessarily be an interpretive or hermeneutic element in the social sciences. In fact, the “mechanism” metaphor should be used with care, as it implies a problematic Humean notion of causality (e.g. Harré, 1985): social mechanisms are not mechanical, and the adoption of such a conception of causality “makes it very difficult ... even to
suggest a plausible theory of human agency” (Ellis, 2002, p.197). We can talk about generative mechanisms of social structures because they do make something happen, but while doing so, we must not forget that the effects of structures are mediated by agency: in social life, nothing happens without the activation of the causal powers of people (Carter and New, 2005). Hence, values and meaning are deeply connected to qualitative change, since humans are capable of navigating and transforming them through social action. For example, when studying how changing pay structures affect employee behavior in companies (e.g. Lazear, 2000), it is easily forgotten that the desires and values underlying these behaviors had to be taught, as workers would otherwise work no longer than necessary to meet their traditional needs. Or put in another way: sociotechnical change is to a large extent about change in values and meaning – and hence contains hermeneutic elements – showing how qualitative change is inextricably entangled with meaning (Geels, 2005; Geels and Schot, 2007). The nature and effects of such change in wicked systems will be the topic of the next section.

2.4 Innovation in Society

We have so far looked at how the relationship between structural and dynamic complexity plays out in social systems, and we will now turn toward a second central feature of wicked systems: the way that innovation drives constant qualitative change and the increase of these two complexities.

As has already been noted, wicked systems are deeply connected to innovation: on the micro-level, they are arenas of and for innovation, in which their competing constituents try to outsmart one-another; on the macro-level, they are characterized by a combination of self-propelling cascades of transformation, and periods of stasis and locks-ins. These cascades are self-propagating waves of qualitative “reconfigurations” of and by components, unfolding distributedly and locally in “the adjacent possible” (Kauffman, 1996, 2000) and propagating across neighborhoods in the seamless web, potentially with system-wide implications (Geels, 2002, 2011; Lane and Maxfield, 1997; Lane, 2011b, 2016; Schiffer, 2005). In other words, qualitative change in wicked systems is driven by positive feedback, change driving change, making both constant innovation and deep uncertainty hallmark properties of wicked systems.

Wicked systems both enable open-ended innovation, and are themselves produced by it. They enable innovation since the complicatedness (afforded by the technical) and the complexity (afforded by the social) are both necessary components of open-ended innovation. Without the structural complexity, it is impossi-
ble to construct sophisticated and specialized systems; unstructured system interactions would make for unmanageably design spaces, in practice impossible to explore through creative processes (Stankiewicz, 2000). Without the dynamical complexity, it is hard to imagine any efficacious adaptation process, as these rely on exploration of design spaces through e.g. parallelisms, feedback, and mass-interaction. They are produced by innovation, since the innovations themselves are structurally complex, and become part of a dynamically complex, seamless web. This positive-feedback dynamic is not a functionality of the system: their lack of what Khalil (1995) calls artificiality means that their macro-level is not aimed at fulfilling any functions, but that it is merely the emergent and aggregated sum of externalities of the underlying innovation processes.

These features of wicked systems also play out in a particular way in human society. This relates to what e.g. Lane (2011a); Lane et al. (2009b); Lane (2016); Lane et al. (2011) call the “Innovation Society”: a society where innovation is no longer just a means of solving problems, but where innovation is ideologically sublimated and has become entrenched at the very heart of how society functions – where innovation is important in itself quite regardless of what gets innovated. The Innovation Society is a society organized around the dynamics of its own wicked nature.

The development of such a society was crucially enabled by the capacity of social structures to affect their own constituents, which has allowed the innovation feedback cycle of wicked system to become reified into a goal in its own right, elevating it from emergent dynamics to a social goal. This has led to a situation where constant innovation has become entrenched in society (Wimsatt, 1986), and thereby necessary for its functioning and stability. This dependency suggests a rather paradoxical stability: we have become locked into a state of constant explosive change (plus ça change, plus c’est la même chose, indeed!) Through this process an inherent property of wicked systems has become a core value in society, expressed culturally as an ideology that permeates it (Lane et al., 2011)

That innovation has become a project for innovation itself is, as a natural phenomenon, an entirely new thing – unique to human culture. But it is not an essential feature of human culture, and it has not always been that way. The idea that we can improve society and our own quality of life by innovation is characteristic of the Enlightenment and signifies a drastic shift in ideology: from the view that we ought to preserve a God-given social order to the view that that we ought to use science (in a broad sense) to understand the world and master it so as to increase our wellbeing. There are of course several sub-ideologies that
propose different ways of organizing innovation to achieve such improvements in well-being – most importantly based on either bottom-up self-organization or top-down management – neither of which has proven itself to be potent as solutions to the sustainability problems that we face today.

Lane et al. (2011) summarize the Innovation Society and its ideology as follows:

Our society’s dependence on innovation cascades is expressed in, and sustained by, an increasingly widespread way of thinking, which we will term the Innovation Society ideology. This ideology underlies almost all current discourse about business strategy and governmental policy. The following four propositions form its central core: (1) the principal aim of policy is sustained economic growth, interpreted as a steady increase in GDP; (2) the engine of this growth is innovation, interpreted as the creation of new kinds of artifacts; (3) Which new kinds of artifacts have value is decided by the market; (4) the price to pay for not innovating, or for subordinating innovation to other values, like cultural enrichment or social justice is prohibitively high: competition, at the level of firms and of national economies, dooms dawdlers to failure, which translates into economic decline and social chaos.

The capacity of societal systems to relate to their own emergent phenomena was also key to permitting the expansion of the human sociotechnical system: it allowed it to be subjected to itself – its process and conditions – permitted it to improve its own function, and through this, increase the speed and magnitude its cycle of innovation. The “Great Acceleration” (Moore, 2014; Steffen et al., 2015a; Waters et al., 2016) may be understood as the quantitative signature of this qualitative explosion in diversity. But the development can also be seen directly in a veritable explosion in types of artifacts: if 3 million years ago, our ancestors had essentially one kind of artifact, and 50 000 years ago, maybe a few hundred, today’s inhabitant of New York City can choose among over $10^{10}$ different bar-coded items (Lane, 2016).

The introduction of new artifacts necessarily involves changes in new patterns of interaction among people, not only through the use of the artifacts, but also through, for example, their production, marketing and maintenance. There is an inextricable linkage between the introduction of new artifacts into a society and transformations in the social relations and organization of that society. As people’s living conditions and social relations transform around the presence of
the artifact, they may become incompatible with the institutional and organiza-
tional structures of the old; structures that used to facilitate now become dead-
weight. This is a point of conflict, where new organization may replace the old, 
where agency can play an important role, as the developing structures have not 
yet become entrenched. A society’s institutions emerge in interaction with the ar-
tifacts and the technology of that society, through highly unpredictable feedback 
processes: on the small scale, a new office computer system may result in new 
company work processes; on a large scale, the hand-mill co-evolved into a society 
with feudal lords, and the steam mill into a society with industrial capitalists. The 
institutions that form around technologies need to be in a sense compatible with 
the technology with which they interact: when institutions become misaligned 
with the artifacts underlying them, instabilities occur, creating the opportunity for 
social and technological change. As the rapid progress of artifact innovation con-
tinues, such societal instabilities are continually sparked on all levels of society. 
Societal structures effective in harnessing the possibilities of available artifacts 
gradually turn into shackles as the artifacts continue to evolve. The breakdown of 
structure leads to an “era of ferment”, where a set of alternatives are competing 
openly for the development of new structures co-evolving with the development 
of new technology (Geels, 2006; Grübler, 2003).

Such feedback process of co-evolution results in unpredictable social and 
technological transformation, making it highly difficult to achieve intended so-
cial effects through technological development. But such technological change 
is in any case the exception: artifacts are generally not evaluated on the basis of 
the transformative effects they will have on societal structure, but only evaluated 
locally. It is this type of atomistic evaluation processes that decides the value 
of new artifacts: it is the what we refer to when we say “the market”. Atom-
istic rationality has taken over a larger and larger part of what used to be part of 
the realm of political decisions, while politics is increasingly left to priming the 
pump of innovation. Progress is seen as inevitable, whatever it may entail (see 
also Ellul, 1967).

The stability of this system is based on competition on all levels, which, 
through the interlinkage of markets associated to neoliberalism, has now become 
global (e.g. Harvey, 2007). This competition out-crowds everything but more in-
novation. Any attempts to subordinate innovation to other values, like cultural 
enrichment or social justice, are made impossible by competition at the level of 
individuals, firms and national economies. Competition dooms any potential 
Samaritans to failure, which at the national level – which ostensibly has some 
level of political play – would translate into economic decline and social chaos.
This in practice undermines any attempt at going against the stream, except on the global level – where necessary structures to scaffold agency are largely lacking\(^1\) (Dunford, 2000).

### 2.5 Innovation meets Uncertainty

The production and use of innovation will naturally result in some environmental and social externalities: changes that cannot be predicted (Sveiby et al., 2009). As society becomes more entangled and wicked, it is becoming increasingly difficult to respond to these externalities, as the dynamical consequences of action in such a web is shrouded in ontological uncertainty (Lane and Maxfield, 2005). This uncertainty not only prevents us from designing effective interventions, without high likelihoods of causing unexpected troubles in other domains, but it also keeps us from aligning and organizing action in the first place (e.g. Gardner et al., 1990; Gintis, 2000). This is indeed illustrated by the enormous difficulties in organizing an adequate response to climate change. That it, indeed, is even harder can be seen by comparison with similar global challenges just a few decades ago (Gareau, 2013; Laube et al., 2014).

If these externalities of innovations are left without response, they accumulate until they pass certain thresholds and develop into full-scale crises (e.g. Rockström et al., 2009). In the case of biological innovation, such crises seem to become increasingly rare as the structural complexity of the system increases. With more species, ecosystems become more stable, and the size of the available design space increases, both resulting in increased capacity to respond to externalities (Allesina and Tang, 2012; Stankiewicz, 2000). In society, however, innovation seems to have the opposite effect. As we have seen, more innovation makes our society more and more wicked and uncertain, and hence makes its problems harder and harder to anticipate and respond to. Hence, society will produce more crises, they will be harder to solve, and the system will develop toward increasing instability (see also Ponting, 2007; Tainter, 1990).

Uncertainty forces us to be shortsighted by preventing us from building sufficient certainty for large-scale alignment and action. A shorter and shorter foresight horizon, combined with a virtually un-bounded horizon for consequences of actions, makes wicked systems susceptible to self-undermining: what we typically refer to as unsustainability. Societal evolution is thereby prone to spontaneously and collectively embark on pathways leading to new dynamical regimes.

\(^1\)There have certainly been an increase of multi-national political organizations, in part in an attempt to meet this development, but the balance of power has clearly shifted.
CHAPTER 2. RELATIONAL LIMITS

that may be arbitrarily disadvantageous (e.g. the Anthropocene; Steffen et al. 2015a). Control demands a global overview, but growth and change is local and demands no such overview, so wicked systems may outgrow any capacity for governing them.

As organized efforts to respond strategically become more difficult, as the anticipation of the effects of actions becomes clouded in ontological uncertainty, they are met instead by responses whose consequences are never even taken into account: innovation responds only to the myopic opportunities that crises provide. In other words, like the alcoholic curing a hangover, society tries to solve its problems by applying more of what caused them. This is a condition that Beck (1992) calls “reflexive modernity”, in which modernity “becomes its own theme”, as the focus of innovation increasingly becomes to alleviate the negative effects of previous innovation. But every such ‘solution’ is not only unlikely to solve the highly entangled problem (in part as it would imply undermining their own raison d’être) but also has the potential to produce new problems.

As Lane (2016) observes, the result is a cycle of problem-solution-problem with potentially disastrous effects. Lane exemplifies this with the obesity epidemic, initiated by a large surplus in cheap available calories, brought about by innovation cascades in agriculture. This resulted in innovations in food processing to provide higher returns to producers and distributors from the cheap calorie surplus. This, in turn, resulted in changing patterns of consumption, followed by a rapid increase in obesity rates. The market responded to this problem with waves of innovation in the diet and pharmaceutical industries, with huge market successes, but no discernible effect in decreasing the obesity epidemic. The social results of these dynamics have been catastrophic: the obesity epidemic is today seen by many in the public health community as the principal public health challenge of the twenty-first century.

Through the lens of the innovation ideology, the wicked problems we are facing tend to start to look like engineering problems, i.e. solvable through more innovation. As Kingsnorth (2011, p.x) points out, even the green movement have fallen into this perspective, seeing unsustainability “as an engineering challenge which must be overcome with technological solutions guided by the neutral gaze of Science, [which] has forced it into a ghetto from which it may never escape”. Through this lens, the reinvigoration of naturalism, following from society’s increasing impact on environmental systems, seems less a result of an increased understanding of the complexity of societal systems, and more a sign of a pervasive innovation ideology.
As we have seen, while it is increasingly uncontroversial to claim that society is complex, complexity seems to represent a rather large number of system properties. In approaching society, the assumption of mainstream complexity science has been that societal complexity is essentially similar to the complexity of the kind of systems with which it has shown great success, such as bird flocks or fish schools (Mitchell, 2009). The undeniable structural complexity of society has been seen as merely a complicating factor: the only thing missing in the attempts to understand society is more time, effort and funding. This has not only been the foundation for the way that society has been approached methodologically, but also scientifically, as it has been seen as a green light for the same type of positivist application of formal methods that have proven successful in their application to complex and complicated system.

The above systemic examination of ontological categories implies that this assumption has been erroneous. Most importantly, it seems that the complexity of most natural systems is in fact rather different from that of society: the former is “complex” and the latter is “wicked”. This separation can indeed be seen in how the “mainstream” understanding of complexity, associated e.g. to the Santa Fe Institute, differs from the understanding of many social scientists of what has been understood to be the same system category. This division is reflected in e.g. Morin’s (2008) separation between “restricted complexity” and “general complexity”, as well as in Byrne’s (2005) “simple” and “complex” complexity. In the terminology of this essay, the former describes complexity, and the latter wickedness.

Wicked systems are certainly complex, in the sense that they display dynam-
ical complexity, but they are not only complex. They are heterogeneous, interconnected, nonlinear, far-from-equilibria, emergent and adaptive, as suggested by Complexity Science, but also open and contingent, with any patterns only local in time and space, and subject to ubiquitous qualitative change. This is furthermore related to the way that emergence plays out, with restricted complexity approaches focusing on the emergence of the whole from the parts, thereby neglecting, or even rejecting, the role of top-down causation, while generalized complexity approaches instead emphasize feedback between structure and emergence (Elder-Vass, 2010).

This calls for a science of wickedness that, in the words of Reed and Harvey (1992, p.359), “treats nature and society as if they were ontologically open and historically constituted; hierarchically structured, yet interactively complex; non-reductive and indeterminate, yet amenable to rational explanation”. Since wicked systems are not closed, we cannot approach them only through formal models, and hence, as Cilliers (2002, p.X) puts it, the study of what we call wickedness “is not going to introduce us to a brave new world in which we will be able to control our destiny; it confronts us with the limits of human understanding.” Our understanding of complex systems has implied a way to untangle their intricate webs of causation, but a deeper understanding of wicked systems instead seems to imply learning about the boundaries of knowledge and just how “how little we can know about the world” (Koppl, 2010).

This does not have postmodernist implications for whether we should do science, but it does carry significant implications for how we should do science. While different framings of the same system are possible, reality does have a say in how they are made. Wickedness suggests a perspective which accepts neither positivism nor relativism: it recognizes that our scientific descriptions of reality are social constructs, but also that “they are constructs made by reality and therefore shaped by reality” (Byrne and Callaghan, 2013, p.33). In other words, this complexity speaks to realist (rather than postmodernist) social theories (Walby, 2007).

This connection to realist theory can also be seen in the way emergence needs to be understood in wicked systems. This term has generally, implicitly or explicitly, been understood as epistemological concept: “‘emergent’ was construed as ‘unexplained’ by means of contemporary theories” (Bunge, 2003, p.13). This view was first represented by British emergentists in the late-19th early 20th centuries, but can now be identified in a range of work on complexity, including
3.1. SOCIETY IN MAINSTREAM COMPLEXITY SCIENCE

that of Ernst Mayr\(^1\), Tony Lawson\(^2\), Keith Sawyer\(^3\), Peter Hedström\(^4\), and Niklas Luhmann (see Wan, 2011, p.67). A central part of wicked systems, however, is what Kaidesoja (2009) calls an “ontological concept of emergence”, which is real and unaffected by our knowledge of its processes: “Emergence is often intriguing but not mysterious: explained emergence is still emergence” (Bunge, 2003, p.21). Emergence is an aspect of dynamical complexity, implying relationally distributed mechanisms, not a function of our knowledge. The wicked systems perspective hence follows the ontological understanding of emergence of e.g. Archer, Bhaskar, Gell-Mann, Searle and Elder-Vass: “Emergence is the idea that a whole can have properties (or powers) that are not possessed by its parts-or, to put it more rigorously, properties that would not be possessed by its parts if they were not organised as a group into the form of this particular kind of whole” (Elder-Vass, 2007a, p.28).

The view, and the wicked systems perspective in general, has much in common with Mario Bunge’s “emergentist-systemist” philosophy (e.g. Bunge, 1979a, 2000a,b), and with the related (see Danermark et al., 1997, p.4) critical realism of e.g. Archer et al. (2013), and Bhaskar (2013), in particular Reed and Harvey’s (1996; 1992) “complex realism”, further developed by e.g. Harvey (2009), Byrne (2002, 2004, 2005, 1998), Byrne and Ragin (2009), and Byrne and Callaghan (2013). This section will serve to see how these ideas connect and relate to wickedness, to see how Complexity Science can learn from social theory regarding how to approach wicked systems in general, and society in particular. We begin by looking at the mainstream complex systems ontology and its development.

3.1 Society in Mainstream Complexity Science

Complexity Science developed through and around new computational methods, and at the heart of this methodology lies computer simulation, which crucially

\(^1\)E.g. “the characteristics of the whole cannot (not even in theory) be deduced from the most complete knowledge of the components” (Mayr, 1982, p.63)

\(^2\)E.g. something is “emergent if there is a sense in which it has arisen out of some ‘lower’ level, being conditioned by and dependent upon, but not but not predictable from, the properties found at the lower level.” (Lawson, 2006a, p.176)

\(^3\)E.g. “at the global system level are patterns, structures, or properties that are difficult to explain in terms of system’s components and their interactions” (Sawyer, 2005, p.4)

\(^4\)E.g. “social emergence refers to social properties that cannot, in practice, be predicted by knowing everything there is to know about the pre-emergent properties of the parts” (Hedström, 2005, p.74)
brings the capability to describe the entities and interaction rules of dynamical systems so as to put it all “into motion” (Fontana, 2006). The typical model in this tradition has a microlevel of interacting nodes existing in a pre-defined environment. Having set up the rules and the environment, the system is allowed to play out, and the results and patterns that emerge from the often long causal chains of interaction are studied. This is a highly flexible methodology that made it possible to study and visualize dynamics that are inaccessible both to analytical mathematics and to unaided human cognition. This can be viewed as an extension of the study of the micro aggregation from additive cases, as in classical linear and mathematical methods, to situations where “the whole is more than the sum of its parts”.

As so far described, complexity science could be seen just as a methodological toolkit allowing the study of a broader range of phenomena than previous tools – which would be all well and good. But the “social life” of these methods (Law et al., 2011) has increasingly led to the development of a corresponding social ontology: the labeling of society as a “complex system” clearly goes beyond a methodological claim, and into the realm of ontology (Fontana, 2010). As Perona (2007) argues, Complexity Science is guilty of a “fallacy of misplaced concreteness”: in Lawson’s (2005) terminology, the “ontic” (description of reality) is not separated from the “theoretic” (descriptions of the models), meaning that reality becomes seen as artificial and closed.

This is playing out much in the same way as the development of neoclassical economics, which is widely understood to have grown its ontological perspectives on the economy on the basis of its methods, as Debreu (1986, p. 1265) puts it: “as a formal model of an economy acquires a mathematical life of its own, it becomes the object of an inexorable process in which rigor, generality, and simplicity are relentlessly pursued”. Through this process, an equilibrium-based ontology developed within economics, as the assumptions required for the application of the mathematical methods increasingly turned into theory about the nature of the world, and factors that could not readily be brought into the mathematical machinery were simply disregarded. Models were no longer separate from theory, as methodology transformed into epistemology, in turn – over time – transforming into ontology (Fontana, 2010) – or, put more simply: economists found a hammer, and everything started to look like nails.

In Complexity Science, that hammer is primarily simulation. This has led to an understanding of society that emphasizes micro-level interaction, which can be seen in definitions of complexity. Johnson (2009, p.1) defines complexity as “the study of the phenomena which emerge from a collection of interacting ob-
3.1. SOCIETY IN MAINSTREAM COMPLEXITY SCIENCE

jects”. Similarly, Mitchell (2009, p.13) describes a complex system as “a system in which large networks of components with no central control and simple rules of operation give rise to complex collective behavior, sophisticated information processing, and adaptation via learning and evolution”. Holland (2006, p.1) agrees, but is almost even more restrictive by stating that complex systems “are systems that have a large numbers of components, often called agents, that interact and adapt or learn.” This is very much in line with Joshua Epstein’s understanding of society as a complex system growing from the bottom up (Epstein, 1996).

All of these fit in well as descriptions of the ontology of most simulation models, but as descriptions of society they seem overly focused on the individual agent level. As Lane (1993, p.194) puts it, these models “offer only very limited scope to the emergence of new structures—and, so far, none at all to the emergence of higher-level entities.” Social structures are seen as merely patterns emerging from the behavior of underlying agents, or as Jarvie puts it (cited in Bhaskar 2010, p.55), in these perspectives “army is just the plural of soldier”. Despite Complexity Science being associated to a criticism of reductionism, it seems to fit neatly into the realm of ontological individualism. As Epstein (2006, p.37) readily admits, “classical emergentism holds that the parts (the microspecification) cannot explain the whole (the macrostructure), while to the agent-based modeler, it is precisely the generative sufficiency of the parts (the microspecification) that constitutes the whole’s explanation! In this particular sense, agent-based modeling is reductionist.”

While such assumptions may be appropriate in complex systems, where only first order emergence applies (Gilbert, 2002), the “reflexive” (Goldspink and Kay, 2007) second order emergence of wicked systems is completely at odds with such an understanding, as it emphasizes the reality and causal capacities of emergent structures. These are essentially what enables the structural complexity of such systems. Hence, while the methods of mainstream complexity may provide helpful aid in the first half of our science of wickedness, i.e. casting light on the heterogeneous, interconnected, nonlinear, far-from-equilibria, emergent, and adaptive nature of society, its methodology is based on an ontological perspective that does not allows us to approach other aspects of wickedness – i.e. that it is also open and contingent, with patterns only local in time and space, and ubiquitous qualitative change – let alone the value-laden and meaningful nature of society in particular. Let us therefore look at alternative ways of looking at the complexity of society, which takes into account both structure and emergence, by leaning on ideas developed in critical realism and in Mario Bunge’s emergentist-systemism.
3.2 Fact and Action in Wicked Society

As we have seen, the wickedness perspective argues for the causal capacities and reality not only of the components, but of the structures that emerge from their interaction. As we have also seen, this meeting between structure and dynamics mirrors a metatheoretical meeting between the “fact” and “action” paradigms in sociology (Gilje and Grimen, 1992), and indeed between structure and agency – one of the most central questions of sociology.

In this tension between what basically amounts to whether to treat society as a complicated or complex system – both of which are subsets of reductionism – the wicked system perspective clearly points toward a third option, where we instead attempt to take both agency and structure into account, and where reality is seen as a dialectic between them (e.g. Wight, 2006). Such a theoretical perspective has been developed by theorists such as Archer, Layder, Pawson and Sayer – an “analytical dualism” (e.g. Archer, 1982, 1996; Bhaskar, 2008; Layder, 1985) that starts from the ontological claim, following from the “ontological concept of emergence”, that structures and agents each possess distinct properties and powers in their own right, referred to as *sui generis*, and that they are very different type of entities, rather than as two parts of the same process (as in Anthony Giddens’s (1984) “structuration theory”). This has multiple implications, such as that the goal of social analysis is to keep structure and agency apart, to study the link and interaction between them, not to reduce one to the other (Danermark et al.,
While social structures are not “powerful particulars” like individuals, in the sense that they can produce “observable effect in certain conditions and in a relatively autonomous way” (Kaidesoja, 2007, p.81), they are endowed with \textit{sui generis} properties and powers. As we have seen, they are characterized by anteriority: they precede agency, not in the sense that they could exist without human action, but in the sense that humans only reproduce or transform social structures, they do not create them. Property relations, linguistic systems and legal systems are existing features of the world into which we are born, they are not things that we create at birth. Their anteriority points to another important property: they are relatively enduring. They are granted this longevity through the material expressions that are always part of social practices, in a mutual dependence between material practices and the formation of meaning (Danermark et al., 1997; Törnberg, 2016a). Social structures also have powers: they are capable of “motivating or discouraging, constraining and enabling certain sorts of human action” (Carter and New, 2005, p.10), for example: unequal wealth distributions will tend to constrain the poor and enable the rich. Hodgson’s (2009) concept of “reconstitutive downward causation” is also useful here, capturing the elusive notion that “the whole, to some extent, reconstitutes the parts” (Hodgson, 2009, p.168).

Individuals are similarly equipped with \textit{sui generis} properties and powers: they are self-conscious, reflexive, emotional, intentional, cognitive, and so on. Their reflexivity and symbolic abilities endow them with the powers to formulate plans, organize project, pursue interests, etc.: it is people who make history; they inhabit the social world, and are able to reflect upon, seek to change, or even overthrow the social structures, according to their own interests and views.

The understanding of emergence of structure from agents emphasizes that there is no singular “humanity”, but only plural and heterogeneous mortals, giving their politics its organic, interconnected and contingent nature (Arendt, 1958). Hence, social structures are not planned, but often unexpected: people do not marry to reproduce the nuclear family, nor do they work to reproduce the capitalist economy, but these are nevertheless the unintended consequences of their activity (Bhaskar, 2010). Social structures are both the cause for and caused by action: we tend to behave according to the wage-labor relations, which in turn reproduces the structure of wage labor, in turn generating new action, and so on. This interplay between social structure and agency occurs over time, meaning that their emergence takes the form of a continual process. “Causality, in virtue of its transitivity, gives aid and comfort neither to the holist nor to the individualist. The causal chain just keeps rolling along,” (Sober, 1980, p.95).
CHAPTER 3. FROM NATURALISM TO REALISM

Individuals, and the emergent structures following from their relations, contingently combine to produce second and third-order emergent structures. This means that the world is not only differentiated, but stratified: it has different levels characterized by different properties. This is similar to the idea of emergent multi-level systems in complexity science (e.g. Beurier et al., 2002): but with the difference that it is not excluded to micro to macro emergence, where macro-patterns emerge from a set of interacting micro-objects. Instead, the capacity of humans to relate and act upon emergent structures results in that emergence tends to go in more than one direction: interaction is not limited to a single stratum; emergence can occur from interaction between social structures and the actors that underlie them. This idea of emergence as going in all directions rather than only “upward”, is a central difference between natural and social systems. One way to describe this is as going from a Darwinian “population thinking” to an “organization thinking”, in which no relevant population can be discerned, and variation/selection are inadequate to describe change, which is rather based on a modality of “organizational self-transformation” (Lane et al., 2009a). While the natural world is often seen as hierarchical, the social world is better described as consisting of sets of nested structures, and its effect on actors as “a plurality of interpenetrating constraints deriving from many recognisable ‘levels’ looping back and around each other” (Dyke, 1988, p.64). In fact, as we have seen, this hierarchical perspective likely does not even apply to many of the systems in the natural world, but is rather a feature of the models that have been applied in this context (see e.g. Hendrichs, 1983; Khalil, 1995).

This takes us away from an understanding of “explanation” that is in line with crude naturalism’s tendency of reducing phenomena to underlying levels, exemplified by analytic sociology’s idea of “carefully dissecting” social processes into their underlying component parts and their actions (Hedström, 2005, p.73). Instead of, as suggested by analytical sociologists and complexity theorists alike, asserting the individual level as a *conditio sine qua non* for social scientific explanations, we should aim for a “multiscaled social reality” (DeLanda, 2006, p.34-40) – because regardless of how “complex” the individualism of these approaches, it is still individualism (Wan, 2011). The multiscaled perspective suggests a social explanation that makes structure part of the process, having a methodology and ontology that allows for nested but interpenetrating systems with causal powers running in all directions.

This emphasizes tracing the interaction within and between *strata*, as emergence is not seen as a macro appearing from the dynamics of the micro, but as a continual process between levels. Emergence can be explained – at least
in principle – in terms of elements and interactions, as suggested by Bunge’s (1979b) “rational emergentism”. This implies a rejection both of individualism and holism in favor of an approach that, while explaining phenomena in terms of generative mechanisms in deeper strata of existence, does not see the higher properties and powers as “explained away” by such an explanation (Danermark et al., 1997; Elder-Vass, 2010), since “explained novelty is no less novel than unexplained novelty” (Mahner and Bunge, 1997, p.29). We can, at least in certain cases, trace relational emergence of phenomena to underlying constituents and their relations, which “allows higher level properties to be explained scientifically,” (Elder-Vass, 2007b, p.415) but due to equifinality and multifinality, we cannot directly link a macro to a micro, meaning that such explanation “does not allow them to be replaced with properties of the parts in causal explanations because it is only when the parts are organized into this particular type of higher level system that the causal power exists” (Elder-Vass, 2007b, p.415).

While reductionism is deeply problematic as a research strategy, amounting to “the methodological principle according to which (micro)reduction is in all cases necessary and sufficient to account for wholes and their properties” (Bunge, 2012, p.178), reduction is often desirable and fruitful, such as when part of a productive research strategy in wicked systems, “reduction does not imply leveling: it relates levels instead of denying that they exist” (Bunge, 1977a, p.79). Reduction is important, as emergence should be explained rather than dodged, while ontological novelty at every level should at the same time be acknowledged. Both the phenomenon and its underlying mechanisms remain as real and with their separate powers and properties: social explanation always involves a meeting between structure and agency, and is played out as a co-acting between people and the cultural and social structures they encounter, use and embody; “structures which position them, motivate them, circumscribe their options and their capacity to respond” (Carter and New, 2005).

3.3 Explanation in Wicked Systems

This also has implications for the type of explanation employed. Abbott (2001b, p.164) separates between two such approaches in sociology: the variable-based approach, which focuses on stochastic realizations and uses correlations between variables to try to find causal links, and the narrative-based approach, which instead focuses on identifying mechanisms and categories by looking at patterns in data (Abbott, 2001b; Abell, 2004; Calhoun, 1998; Griffin, 1993).

The variable-based explanation, related with a regularity or succession the-
ory of causality associated with empiricism, implies treating wicked systems as complicated, through the formulation of thought-up entities – “variables” – that are purported to interact through causes and effects. Causality is here understood as the relationship between these entities, which also requires that they are fixed over time, and that they exhibit fixed set of attributes. This is hardly the general case in wicked systems, although, of course, such assumptions may be more legit and useful in some cases than in other. Variable-based analysis usually also assumes that causes are independent, in the sense that there is no interaction between causal factors. This is also hardly the case in wicked systems (Marini and Singer, 1988). However, there is also a more fundamental problem to the approach: variables do not exist in the real world (Byrne, 2002). Bunge (1985, p.138) puts this succinctly: “in science we handle changing things, not changeless ones, let alone thingless changes”. Variable-based approaches have a tendency of losing sight of these rather important facts, and that it is “individuals whose lives provide the data for the models. Although variables rather than individual people may become the subjects of the statistician’s narrative, it is individuals rather than variables who have the capacity to act and reflect on society” (Elliott, 1999, p.101-102). This implies a Hempelian banner of causality, in which “causes” tends to essentially mean “causal variables” (Abbott, 2001b; Skocpol, 1984), hence describing causality between entities existing in the model, not in reality. Even a model that successfully accounts, in a statistical sense, for the variation in some phenomenon, can still tell “us rather little about just what is going on at the level of social processes and action that underlie [...] the interplay of the variables that have been distinguished” (Goldthorpe, 1997, p.9). Hence, this seems to be yet another example of the social life of methods, as the variable-based approach implies an understanding of the system as artificial and complicated.

Narrative-based explanation essentially relies on treating wicked systems as subwicked, on being compatible with how actors of these systems understand their environment, and on fundamental human cognitive capacities in relation to such systems. Such an approach has the benefit of matching the ontology of the system5, suggesting the possibility of “narratives as a fundamental foundation of

---

5It should be noted that our language itself brings with it pieces of implicit ontology, that are brought into our narratives, mathematical models, and everything in between. Such a conception of the role of language is of course part and parcel of e.g. discourse analysis, but has also been the subject of some interesting formal study, in particular in the context of probability theory. An example of this is Goodman’s new riddle of induction, which shows how Bayesian induction is based on the assumption of our language’s concepts matching natural kinds (Godfrey-Smith, 2009). This illustrates how ontological assumptions are part of language, how formal models are
complexity research” (Byrne and Callaghan, 2013, p.202). This match can be seen in the affordances in narratives when it comes to complex causality, time, and multi-level explanation. As Abbott (2001a, p.101) puts it: “reality occurs not as time-bounded snapshots within which “causes” affect one another ... but as stories, cascades of events. And events, in this sense, are not single properties, or simple things, but complex conjunctures in which complex actors encounter complex structures.” Temporal sequence is crucial in these stories, and a different order of events may produce different outcome (Griffin, 1993), as well as for allowing for both multiple temporal levels and the real complexities of causality (Ricoeur, 1980).

In narrative theory, the focus is on events rather than variables, and the entities involved are not ontologically fixed as the story plays out – entities participate in events and change over time. This takes us from a “push-type causality” to one which requires tracing of events, and one which allows patterns on different time-scales and structural level to interact and play out a common story (Poole et al., 2000). An event and process focused approach does however not imply advocating a process ontology, in the sense of seeing entities not as “the fundamental categories of being” but as “derivative of or based in process” (Sawyer, 2005, p.134); as Sawyer continues, “an empirical focus on practice does not require a process ontology. One could accept the traditional ‘entity’ view that individuals and groups both exist and nonetheless argue that it is methodologically necessary to study situated practices” (Sawyer, 2005, p.134).

Narrative-based explanation focuses on identifying mechanisms through pattern, which signifies a move from aggregation and variance, to categorization and pattern-finding, and using comparison between cases to find relevant mechanisms (Byrne, 2005; Byrne and Ragin, 2009). The two research strategies proposed by critical realists are highly relevant here, as they allow connecting events and phenomena to mechanism-based explanations: retroduction and retrodiction. Retroduction is to identify the causal powers that act upon events, the entities that possess these properties and powers, and the underlying mechanisms (Sayer, 2010), while retrodiction aims to explain how these can combine to produce the events in question (Danermark et al., 1997). This view and approach to explanation speaks especially to a current development within Complexity Science, with a move from simulation models based on ad hoc assumptions, towards research on digital trace data, for example in the emerging discipline of computational social science (see e.g. Conte et al., 2012; Jungherr, 2015). This development has

---

often built on less rigid foundations than one may conclude, and constitutes a more formal way to shed light on the depth of the problem of dealing with change of kind.
clear potential for supporting a more ontologically plausible approach to wicked systems.

### 3.4 Data Mining and Categorization in Wicked Systems

As we have seen, dynamical complexity implies that mechanisms can be distributed within a system, rather than located in any specific component, implying that “cause is a property of complex and contingent mechanisms in reality” (Byrne and Ragin 2009, p.515, see also Coverdill et al. 1994). The contextual nature of wicked systems implies that cause is not universal or permanent, but rather local and temporary. Such a perspective has traditionally been understood through qualitative rather than quantitative reasoning, in part since “quantitative” tends to mean “variable-based”. In other words, a narrative and complex causation perspective has tended to imply that we rarely use formal tools when approaching data. This is problematic, as complexity is a real and important part of these systems and, as we have seen, high structural and dynamic complexity limits how far we can go without the aid of formal tools. In other words, neither the traditional quantitative or qualitative approaches are adequate in approaching wicked systems.

The increasing proliferation of digital trace data and sophisticated analytic methods, implying a convergence of qualitative and quantitative approaches, has important potential here. These seem to be capable of supporting a narrative-based reasoning while at the same time being capable of increasing the capacity to deal with high structural and dynamic complexity. This potential is however not without possible issues: many argue that the digital humanities and computational social science are fostering weak, surface analysis, rather than deep, penetrating insights, much due to the exclusive focus on computational models (e.g. Kitchin and Lauriault, 2014). The fields are argued to be reductionist and crude in their application and interpretation of the techniques – sacrificing context, clarity, and critique for the automatic identification of large-scale patterns, predicated in the notion that breadth could replace depth and context as basis for interpretation. This seems to in turn be based in the confusion of the mere identification of pattern with the explanation of human behavior; as Kitchin (2014) succinctly points out: “It is one thing to identify patterns; it is another to explain them”. Even if, as Anderson (2008) argues, such methods are commercially valuable due to their predictive capacities, this does not necessarily make them useful for scientific inquiry: “It is possible to predict well without explaining anything about what is going on” (Hedström, 2005, p.107). As Kurki (2008, p.166) puts
3.4. DATA MINING AND CATEGORIZATION IN WICKED SYSTEMS

it, “what is important in tracking causal connection is not identification of law-like regularities or empirical observerables, but, rather, the description of the real properties, structures and generative mechanisms that underlie the actualization of events and their empirical observations”.

This tendency of pattern-finding in digital trace data research may furthermore be linked to a tendency of thinking that interpretation can be altogether removed from the equation, and that content analysis should be analytic, quantitative, and what is understood as “objective” (Anderson, 2008; boyd and Crawford, 2012). This ambition is deeply flawed. Whether one admits it or not, the construction of the corpus, tool, and statistical results are all types of interpretation – it does not become less, but more, interpretative, with more steps between corpus and conclusion (Byrne, 2002; Kritzer, 1996). This only leads to the reification of what basically amounts to a form of glorified variables, whose subjective nature is merely hidden under a thin veneer. Through formal models, one may conceal subjectivity behind computational or mathematical forms of representation, hiding the normative decisions of framing by enciphering the assumptions and normative choices in technical code (Feenberg, 1991).

There are however directions of digital analysis that are built on more theoretically sound foundations. These methods tend to build on using comparison and categorization, which in turn build on similarity/difference rather than variance analysis, with the aim of finding patterns that give leads on the underlying mechanisms which shape phenomena in the observed processes (Skocpol, 1984). The idea of using comparison and categorization is a hopeful one, as it steps away from the complicated system idea of seeing causes as single factors whose presence inevitably generates an effect and whose absence means that the effect does not occur, toward a conception of causality more compatible with wicked systems. The typical example of such a research approach is Darwin’s *Origin of Species*, where a large number of case studies were combined to find a powerful emergent generative mechanism (Byrne and Callaghan, 2013).

Two examples of such systematic comparison techniques are Qualitative Comparative Analysis and cluster analysis (Cooper et al., 2012). These, in particular QCA, come out of the realist tradition’s focus on case study based research, in which case studies has been regarded as the primary research design (Easton, 1998), due to their potential to, in the case of single-case studies, reveal the generative mechanisms in specific contexts, and in the case of multi-case comparisons, identifying situations or contexts in which similar mechanisms operate (Ackroyd, 2009). This way of combining narrative explanation within cases, and systematic comparison across, suggests a middle path between quantitative and qualitative
social research (Cooper et al., 2012), since, in contrast to variable-based methods or controlled experiment which deals in single causes, such methods are capable of including more than one cause. Causation is in these methods understood in terms of a combination of factors in interaction, and there is explicit recognition of “causal complexity” (Coverdill et al., 1994, p.57). Furthermore, these methods are capable of dealing with contingency and contextual dependency of social causality, as they are capable of investigating situations where equifinality and multifinality apply.

These techniques are also useful as a starting-point to think of other developing analytic techniques, that go outside the quantitative-qualitative divide in similar ways. For example, Paper IV shows how it is possible to approach certain text analytic techniques through a similar theoretical lens, by using Topic Modeling (Blei, 2012) in combination with close-reading to explore large quantities of textual data. Topic models cluster textual data, and is commonly used to give an overview over large data quantities. However, while informative, such analysis runs the risk of reifying resulting “topics”, in the same way as variables are reified within variance analysis: we end up essentially telling a story of the rise and fall of something that does not actually exist, and to add insult to injury, the complexity of the algorithms turns it into even more of a black-box than most traditional variance studies.

We instead suggest an alternative approach, using topics not as reified entities, but as a basis of a type of explorative sampling, and a way to provide a map through which the material may be explored in various levels of detail. This exploration was performed using a custom-developed text analysis platform, designed for this type of computationally-assisted exploration, which treats computational methods not as a replacement, but as aids to human intuition and interpretation, and an extension of human cognitive capacities. In this integrated interpretation environment, automatic analysis provides a discursive landscape that the researcher can “zoom” into, and combine various types of analysis to find traces of the underlying meaning. The framework does permit the use of some variance-analysis as well, but in a way in which these are explicitly designed to function as variance traces, rather than as reifications.

The notion underlying this platform is essentially that methods will always have a social life (Law et al., 2011), meaning that they will always tend to drive ontology: human cognition has an inherent tendency to reify and create stories around whatever measures we devise\(^6\). Hence, it becomes necessary to design

\(^6\)This is part of a wider argument regarding the relationship between the technological and the social, or base and superstructures, if you will, that was hinted at in chapter 2.
methods in such a way that they afford a methodology that is in line with a reasonable ontological understanding of the system under study.

The approach to data that is supported by this platform is not a matter of testing pre-established hypotheses, but rather of a quantitative (Martindale and West, 2002) or computational hermeneutics (Mohr et al., 2015): continuing re-engagement with the data that mixes levels of interpretation and analysis, predicated on the idea that “[e]xploration is the real and serious game” (Byrne, 2002).

3.5 Facing Emergence in Wicked Systems

Categorization and comparison can indeed be highly useful for finding macro-patterns in data and to use as basis for the identification of generative mechanisms. However, it will not necessarily take us all the way to identifying the underlying mechanisms at play. Dynamical complexity may often lead to phenomena resulting from unintuitive combinations of factors, which we can do no better than black-boxing under the label of “emergence”. Just as we cannot approach digital trace data without algorithmic aid, we cannot unpack such black-boxes, or, in Wan’s (2011) terminology, solve the “problem of transformation”. We may ponder potential generative mechanisms, but without aid, we have no way of determining whether those in fact even have the potential of generating the identified causal patterns.

This is where simulation models come in. Simulation permits quantitative studies of mass-interaction and a de-mystification of emergence (an idea that has in a more mysterious and metaphysical form been around since Aristotle’s Metaphysics: “the totality is not, as it were, a mere heap, but the whole is something besides the parts”, cited in Cohen 2010):

Simulations are partly responsible for the restoration of the legitimacy of the concept of emergence because they can stage interactions between virtual entities from which properties, tendencies, and capacities actually emerge. Since this emergence is reproducible in many computers it can be probed and studied by different scientists as if it were a laboratory phenomenon. (DeLanda, 2011, p.6)

By translating narrative descriptions into a simulation model, we can see whether they produce the hypothesized emergent result (McGlade, 2014). However, due to equifinality and multifinality, doing so will not tell us as much as one might think: if they do produce the emergent phenomenon, we cannot know for sure that they are indeed the generative mechanisms at play in the real system,
and if they do not, we cannot even know that they are not the generating mechanisms. This does however not mean that such models are useless: what we do get is a link between a micro-level mechanism and a macro-level phenomenon in a complex system, which may or may not “shine through” in the actual system, as there is not a single mechanism at play but rather multitudes of interacting mechanisms. This is the reason why we should not think of open systems in terms of laws, and why it is instead more useful to talk about tendencies, “which may be possessed unexercised and exercised unrealized, just as they may of course be realized unperceived (or undetected) by anyone” (Bhaskar, 2010, p.13); a system may have a tendency for certain phenomena, but it might not be realized in the empirical system if it is inhibited by other mechanisms. In wicked systems, causes are not as single factors whose presence inevitably generates an effect and whose absence means that the effect does not occur, rather, cause is a property of a combination of a range of local, temporary and contingent mechanisms.

This means that one should approach empirical regularities with some caution, as they will be local in space and time. Phenomena are complex entangled effects from the influence of multiple interacting mechanisms, reinforcing or cancelling each other out, and so objects should be seen as having forces whether these forces are displayed or not. Conversely, the question of the most important mechanism for a specific phenomenon can only be decided on a case-to-case basis based on specific empirical investigations.

Due to this complex causation, many wicked problems are so unique and contingent that modeling makes no sense. In some cases, however, certain mechanisms do tend to “crowd out” others, and shine through into the macro-level system despite of the multitude of other interacting factors, possibly generating regularities: “Over restricted regions of time-space certain mechanisms may ... be reproduced continuously and come to be (occasionally) apparent in their effects at the level of actual phenomena, giving rise to rough and ready generalities or partial regularities, holding to such a degree that prima facie an explanation is called for” (Lawson, 1995, p.26). This is related to the concept of “control parameters” (e.g. Scheffer and Carpenter, 2003) in complexity science: at given points of time, certain mechanisms tend to dominate the system dynamics, and tuning certain parameters may disproportional affect the system. Using simulations, we can go some way in determining the relative importance of a given mechanism for the system dynamics: simulation conclusions are usually only tested for stability in the parameter space, but it is also possible – although only rather arduously – to explore their stability in the algorithmic space, i.e. in the space of interaction with other mechanisms. This can be informative with regard to whether an emer-
3.5. FACING EMERGENCE IN WICKED SYSTEMS

Gent effect will be likely dominate a system despite embedment in the richness of real open systems. For example, congestion emerges quite robustly in systems of interacting cars in traffic (e.g. Bando et al., 1995): we have found that entire classrooms of students independently developing models – according to whatever assumptions and methods of implementation they find reasonable – will almost in every case produce models exhibiting the same phenomenon, implying an algorithmically highly stable phenomenon.

This furthermore illustrates an alternative way of thinking of models than what is represented by e.g. Hedström (2005), who departs from an analytical sociology perspective. Hedström refers to models that do not engage with data as “fictions”, arguing that they have no connection to reality from which their “isomorphism” and “validity” can be assessed – they are “no mode of calibration” (Hedström and Åberg, 2005). This idea of validation and isomorphism has exactly the platonic flavor to it which Hayles (1999) picks up on in her critique: it presupposes an idea that models are useful to the extent that they match an imagined ideal form of system – a form that is just perpetually out of our reach. The result of trying to reach for that platonic world is models that become increasingly complicated and specific, in the attempt to reach that ever-evasive “realism”. Since they cannot intermix with the viewpoints, knowledge and experiences of the participants, this causes them to lose their real source of usefulness (e.g. Klosterman, 2012). They will not only fail the goal of being “realistic” representations of reality, but also their goal of functioning as useful metaphors and pedagogical tools helping us building a better intuition for complex dynamics. As Macy and Willer (2002, p.147) put it: “making these models more realistic inevitably adds complexity that undermines their usefulness as tools for theoretical research if we can no longer figure out how the model produces a given result.”

Good simulations can function as computationally-assisted thought-experiments, contributing ideas about how emergence can play out in closed systems, and we can use them as crutches for our flawed intuition for complex dynamics when describing and thinking about wicked systems.

At their best, models allow us to develop concepts about emergence like “threshold”, “tipping-point”, “feedback”, or “cascade”, that have been immensely useful in elucidating emergent dynamics in wicked systems. Such concepts do not in themselves constitute claims: “it is undisputed today that concepts can neither be true nor correct, but that they are only instruments, which prove to be more or less suitable when it comes to the correct ascertainment of truths and/or untruths”\(^7\) (Luhmann 1992, p.390 cited in and translated by Wän 2011, p.36).

\(^7\) In the German original: “Unbestritten ist heute, dass Begriffe weder wahr noch richtig sein
CHAPTER 3. FROM NATURALISM TO REALISM

Bunge’s words, concepts cannot be “tested, because they neither assert nor deny anything. Hence there are no true or false concepts: concepts can only be exact or fuzzy, applicable or inapplicable, fruitful or barren” (Bunge, 1996, p.49). These concepts, if proved the former rather than the latter, can then be used in statements and theories the world, and it is the world that determines the truthfulness of these statements and theories (Christis, 2001). This view of simulation, as a way to develop powerful metaphors (Thrift, 1999), as well as concepts for complex dynamics that may then be used to formulate claims about wicked systems, represents a radically different view on the use and understanding of modeling compared both to analytical sociology and mainstream Complexity Science. The reductionist perspective often uses ‘merely metaphorical’ as a form of critique, but in a wicked world, metaphors are what we live by (Byrne and Callaghan, 2013, p.6).

The main limitations of the use of simulation models in wicked systems is hence not, as Hedström and Åberg (2005) might suggest, the necessity to make assumptions about underlying mechanisms, but rather that they are limited to first order emergence. This is not a problem for the results of simulation as such, since first-order emergence is important in a large range of systems. It does, however, constrain what can be studied with their aid. Remember that wickedness means that the rules of the game change on the same timescale as the game is played, and in interaction with the gameplay. Simulation models in comparison, operate by fixed rules, which have pre-assigned meaning: “[e]very interaction is a product of rules because the agents are only autonomous in terms of the implementation of their rules” (Byrne, 2002, p.136). Taking real social action – the way individuals actively interact with social structures – into account requires an understanding of how humans understand emergence. This, as we have seen, implies theories of meaning, value, and narratives. Because of this, it is hard to imagine modeling that would support the general study of second-order emergence (Gilbert, 2002), as they would need to share the wickedness of the system under study.

There is, however, an underused potential in including institutions and social structures as real and active actors in the simulations. For example, I have supervised a student project looking at the conflict between opinion dynamics on the micro- (e.g. relative agreement) and macro-level (e.g. spatial voting). Con-

können, sondern dass sie nur Instrumente sind, die sich als mehr oder weniger geeignet erweisen, wenn es um die richtige Feststellung von Wahrheiten bzw. Unwahrheiten geht.” (Luhmann, 1992, p.390)

Genetic programming is an exception to this that proves the point: the results of such systems tend to be as flexible as they are hard interpret.
3.6 INTEGRATED PLURALISM - TYING TOGETHER TRACES

Tradictory results are often found on these two levels (e.g. Deffuant et al., 2002; Downs, 1957), and the aim was to see which effect that would dominate a system in which both played out. Such simulations, which grants causal status to social entities, may allow us to better understand emergence “that runs in every direction” (Byrne and Callaghan, 2013), and not only micro-to-macro, even if they cannot fully take social action into account and do not bring us into the realm of second order emergence.

Again, the almost singular focus on micro-emergence does not need to be a deal-breaker if we are operating from the perspective of the complexity-toolkit-as-computational-hermeneutics. The purpose is not to completely explain or capture the system, but rather to contribute yet another heuristic tool to a toolkit, bringing a host of qualitative and quantitative techniques, to be brought into an integrative and iterative engagement for the understanding of social systems. Such an approach is exemplified by Paper V, which suggests that part of the observed importance of free social spaces for radical social mobilization is emergent and based in network structural effects, in the form of interaction between network clusters and political deviance. Rather than singularly focusing on the model in itself, as a reified entity, the paper attempts to integrate these ideas into a larger narrative grounded in the literature. This is not an easy task, considering the tension between “the gang [who] can count but don’t know what they are counting” (Byrne, 2002, p.15) and the gang who “can’t count, won’t count, and assert that counting is a vile and perverse activity which ought not to be allowed” (Byrne, 2002, p.15) running high, but it is a necessary one if simulation is to become a constructive part of social scientific work.

3.6 Integrated Pluralism - Tying Together Traces

Simulation models are not unique in their assumptions of closure, as “all theorising in science […] involve some partial or temporary closure” (Hodgson, 2006, p.3). As we have seen, openness means that there are no universal methods that can be used to study all aspects of reality. The bad news is that this, in practice, lays waste to any project aiming at “realistic” rather than “unrealistic” models and assumptions: all theories are abstractions, partial, and unrealistic (Mäki, 1992).

The good news is that this does not mean that modeling is meaningless, but rather that it speaks to the increasingly influential philosophy of pragmatism (Baert, 2005); “All models are wrong, but some are useful” (Box et al., 1987), or put inversely: “Everything is lawful, but not everything is beneficial” (1 Corinthians 10:23). This proposes seeing models as tools, rather than descriptions of re-
ality: “a theory is not really to be believed to give us a truthful picture of what the world is like, it is rather to be used as a useful tool for whatever purposes there may be” (Mäki, 2001, p.10). But it does not necessarily lead to a complete instrumentalism, as there is, at least to some extent, a “difference between descriptively false and descriptively incomplete statements” (Hedström, 2005, p.20), and the usefulness of a model is “a product of the fact that [it] contains a good deal of truth” (Wimsatt, 2007, p.392).

While wicked systems cannot be reduced to any single model, different models are capable of casting light at different aspects of them – giving us what (Byrne, 1998) calls “traces of reality”. This can be illustrated with economic theory, in which, undeniably, neoclassical, complexity and heterodox economics all have contributed many useful models and studies. This is despite their fundamentally contradictory – and, I would argue in the case of neoclassical economics, blatantly erroneous – understanding of the ontological nature of economic reality (for a complex realist perspective on economic theory, see see Törnberg 2016b). As this example shows, even mathematics need not be employed as a map of reality or for prediction, but can also be approached as explanatory heuristics (Hodgson, 2013; Sugden, 2000). The key point is to not confuse the “ontic” and the “theoretic” (Lawson, 2005) in either direction, i.e. not to reject methods due to issues with their ontology, nor to reject the reality due to issues with one’s method.

This integrative approach of letting an “ontology that proclaims both the diversity and the unity of the world” (Bunge, 1973, p.162) guide the methodology, which Danemark et al. (1997) call “critical method pluralism”, Bunge (1973, p.162) calls “integrated pluralism” and Mitchell (2002) calls “integrative pluralism”, has been systematized in various ways (e.g. method triangulation, combined operations and mixed strategies; e.g. Thurmond 2001) and is gaining significant traction in sociology (Danemark et al., 1997). This view suggests that social science should be question-driven, allowing the nature of their inquiry to guide their methods, rather than the other way around, while avoiding “the confusion between a factual item such as a mechanism, and any of its models” (Bunge, 2004a, p.375). This in turn requires the formulation of an ontology of the system under study; a fixed-point from which to pose the questions, or, as Lawson (2006b, p.47) puts it, “to focus competently on specific aspects requires an understanding of the totality”.

Binding together the various traces of reality, acquired through a multitude of methods, into a coherent whole requires narratives: the narrative has strong aligning and integrating functions and can form the “glue” in iterative cycles
3.6. INTEGRATED PLURALISM - TYING TOGETHER TRACES

of sub-wicked approaches. The ontological flexibility of the narrative allows it to bind together and interpret results from multiple sources for traces of reality, while sketching out their limitations in scope, hence functioning as descriptions of the trajectories of wicked systems. There will clearly come multiple stories from such scientific work – all potentially compatible with the available data – and it should be noted that the integrative approach aims at bringing together not only different models, but also stories: there is usually more value viewing them as perspectives that cast light from different direction on the same issue, rather than as competing truths; we should hence “try to integrate all the fields of knowledge that study the same objects” (Wan, 2011, p.53).

Such an integrative approach also implies a clear step away from the notion of science as ideally objective and external, admitting instead for the relational condition of social science: that the scientist is an inseparable part of the subject of study, hence inevitably affecting reality through its study (Bhaskar, 1979). This can be illustrated by the development of the postmodernist theory itself, as the postmodernist crisis in knowledge was not caused by theory alone, but is a response to real changes in society. Again, reality has a voice in the formation of theory; it is the complexity of postmodern society that has generated the crisis of knowledge – but these changes in theory have also had important implications for that same reality: again, the two interact and co-evolve.

Since reality cannot be captured in its entirety, but only from a frame of reference, and we will impact the system by deciding on a perspective, theorizing becomes a type of practice. Furthermore, reality speaks to us through how our theories affect it: if theory is practice and prediction is impossible, the only way to evaluate our work is the change that our practice makes. Hence, while in closed systems, objective study may be possible, and the point can be to predict, for societal systems, the point is to change it. This is not to say that theorists should become activist – it is to claim that they already are.

The real question is whether we should be activists for the hegemony and the status quo, or for societal change. Formal models are no less activist because they conceal normative decisions of framing by enciphering them in technical code (Feenberg, 1991). But, as Byrne and Callaghan (2013) note, there is a clear tendency for such ostensible neutrality of scientific work to serve what Gramsci called the hegemony. The reason for this tendency is the social life of the necessary assumption of fixed ontologies that makes the world seem like it is playing out from a set of fixed rules, with the resulting emergent effects seeming natural and inevitable. Mainstream Complexity Science has often been guilty of this, as exemplified by the description of Pareto wealth distributions as a “natural” result.
of interaction dynamics, implying a certain inevitability by neglecting the possibility for qualitative change of underlying rules (perhaps to a system in which poverty is not “natural”). Again, such limitations are fine when they are merely methodological – it is with their transformation into a tacit and unexamined ontology/ideology that they become problematic.

An alternative to being affected by the social life of one’s methods in this way, is to take a more explicit and active engagement with the system under study. Indeed, Cilliers (1998) focused strongly on the ethical implications of Complexity Science, and based on this he argued that it was even unethical to attempt to engage with societal systems from the outside. He is not alone in these considerations. Gerrits (2012) argues in the context of policy interventions that it is necessary for “the complexity researcher to be engaged with her subjects, both for deeper scientific understanding and better policy information” (Gerrits, 2012, p.181); we need to engage with the actors who are part of the situation, and to take account of their respective narratives, if we are to fully understand their meaning. As Byrne and Callaghan (2013, p.208) argue: “good social knowledge of complex social systems is based on co-production between social scientists and the human agents in the field of investigation”. This notion of co-production takes us beyond dialogue and into the realm of action.

3.7 To Change it

As we have seen, social and environmental unsustainability is strongly linked to the emergent properties of wicked systems: these systems are arenas for competition fought out through innovation, and this constant innovation has in human society been reified into a goal in its own right. This innovation is resulting in an increasingly global society, leveraged by increasingly diverse and versatile technologies and strategies (e.g. Beinhocker, 2006, p.6), and fueled by accelerating resource use (Steffen et al., 2015a), which has transformed human society into a major driver of planetary systems (e.g. Biermann, 2014; Steffen et al., 2015b; Zalasiewicz et al., 2011). At the same time, increasing sociotechnical interconnectedness, fueled by strongly innovation-oriented economic and ideological systems, has made this society more prone to uncontrollable cascade effects, with lower resilience to stress and an amplification of problems from local to more global scales (e.g. Folke et al., 2010; Helbing, 2013; Lane, 2011a). This development has brought about increasingly severe and interconnected crises; e.g. the

---

9Thereby also connecting with the rather large strand of literature associated to action research, e.g. Argyris et al. 1985; Byrne 2011; Reason and Bradbury 2001; Whyte 1991
climate crisis, biodiversity crisis, refugee crisis, financial crisis, and so on. The political response to this “crisis of crises” (e.g. Beddoe et al., 2009; Lane et al., 2011) – which tend to emphasize prediction, planning and control – has been inadequate both in its ability to align around a common direction and in leveling action (e.g. Castree et al., 2014; Haasnoot et al., 2013; Leach et al., 2010; Loorbach, 2010). Instead, the political has largely been reduced to priming the pump of innovation, in an attempt to innovate out of a hole that was created by innovation – disregarding overview and direction for the benefit of locally driven change. This is resulting in cycles of problem-solution-problem, in the end producing merely profit and new problems (Lane, 2016).

This points to a confounding asymmetry in the human condition – between our ability to transform the world and our profound lack of understanding of the very world which our own strength has established; our power over the world seems to be increasing at the same rate as our ability to control this power is diminishing. This asymmetry is rooted in a fundamental aspect of the seamless webs of wicked systems, in which understanding and controlling demands a global overview, but growth and change requires only local and myopic action. The former entails untangling a socio-eco-technical web of feedback and threshold effects, intertwined drivers and deep uncertainty; the latter, in comparison, merely entails the innovation of artifacts that thrives locally in the societal system over the short term. Hence, our world is pushed forward in its track not by rational decision-making based on in-depth understanding of the problems at hand, but rather through the invention and marketing of products; cascades of myopic innovation launching us in directions that we cannot predict or control (see Lane, 2011a). More succinctly put, promethean powers are being exercised in a society of beings too absorbed in consumption to take any responsibility for the world (Arendt, 1958, p.XV).

This fundamental unsustainability harks back to the development of a sociotechnical system – the interweaving of social life with threads of materiality – that enabled the dynamics of self-driving innovation. This afforded an emergent feedback process of constant acceleration, putting mankind at the mercy of the thoughtless invention of every artifact which is technically possible, no matter how destructive. The unsustainability of these artifacts is only a competitive disadvantage in so far as it becomes a problem before the civilization has had time to eliminate its competition, which in practice means that there is little selection pressure for long-term sustainability. In other words, while the current crisis of crises is neither an historical coincidence, nor a biological exigence, it is a strong historical attractor following from the development of a sociotechnical innovation
system.

Our unsustainability is deeply linked to a defining aspect of wicked systems – the fact that humanity as a whole does not fulfill any functionality; our collective behavior is merely the accumulation of externalities of underlying processes that are never evaluated on the systemic level. Humans are all “the helpless victims of a mechanism which is nothing but the cumulative effect of their own strategies, engendered and amplified by the logic of competition of everyone against everyone.” (Bourdieu, 1998, p.27). In other words, there is no collective “us” that has taken us to where we are – “at no moment did the species vote for it either with feet or ballots” (Malm and Hornborg, 2014, p.64). But neither is there a “humanity” that could take charge of events, and consciously make its own future: human beings are plural and mortal, and it is these features of the human condition that give politics both its openness and its desperate contingency: no snowflake in an avalanche ever feels responsible.

This does not mean that this condition cannot be changed, or that we cannot construct governance structures that can lever our collectivity into new “societal actors” (Knight, 2015) capable of instilling our processes of innovation with a necessary level of collective agency. Negotiation and narratives have the power to form local alignment in seamless webs, allowing us to steer the direction of innovation, and thereby of wicked systems. In Hannah Arendt’s words: “when plural persons come together to bind themselves for the future, the covenants they create among themselves can throw ‘islands of predictability’ into the ‘ocean of uncertainty’” (Arendt, 1958, p.xix). Technological change has enabled new forms of collaboration and new ways to come together, which is simultaneously enabling alternative economic organizations, and clashing with important tenets of the old systems (Castells, 2002, 2011; Mason, 2016).

Innovation is by definition needed to solve the meta-crisis, but at the same time, it is also what caused the crisis – innovation as constituted by our current processes of innovation. Since the necessary social transformations is unlikely to be generated by the same structures and processes that caused it, what is needed is meta-innovation: an alternative sociotechnical regime for innovation. But as we have seen, there appears to be no off-the-shelf approaches for organizing innovation in ways that lead to a sustainable path into the future. The meta-perspective of viewing the structures and processes of innovation as regimes (Geels, 2005; Grübler, 2003) – constructing niches and themselves becoming entrenched – was useful to understanding the problems we face, and can be similarly useful in finding a solution: new forms of innovation and production, unattached to the structures of the current regime of innovation and instead belonging to a new cluster
of societal organization, can contribute to the formation of scaffolding structures facilitating the development of new forms of organization for harnessing innovation.

Harnessing innovation implies developing scaffolding structures capable of directing and supporting iterative processes of innovation. Through such structures, collective agency can be instilled into the economic sector, as exemplified by social movements experiments in organizing democratic alternatives to the atomistic market, for evaluating innovations (see e.g. Fenton-Smith, 2015). This literally brings in collective agency into the innovation process, and suggests a way in which scaffolding structures for supporting social innovation can emerge within civil society (Lane et al., 2011). Underlying this is an alternative view of socio-eco-technological systems, characterized by multidimensionality, path-dependency and unpredictability (e.g. Bai et al., 2015; Beddoe et al., 2009; Berkhout, 2002; Folke et al., 2002, 2010; Rip and Kemp, 1998). These qualities – all related to ideas about complexity – are seen as irreducible root causes of the crises that we face and of our inability to predict, prevent and deal with them. The main lesson is that we must harness and embrace these troublesome qualities rather than vainly attempt to plan and control them away.
Chapter 4

Conclusion

This essay has aimed to cast light on the borderlands between the natural and social sciences with the purpose of developing a complexity-informed understanding of the ontological distinctiveness of social systems. It was noted that, while complexity has been associated to an influx of naturalism, the ontological conception of emergence in fact provides a solid foundation for a critique against crude naturalism, as it suggests the existence of exclusively social entities, requiring their own methodological approaches. This was investigated by attempting “to analyze and to systematize the ontological categories” (Bunge, 1977b, p.12) by looking at the interaction between binaries like individuality/artificiality, emergence/hierarchy, agency/structure, and complexity/complicatedness (Khalil, 1995). This categorization has served as a foundation of a naturalistic perspective on naturalism, with the conclusion that society, as well as some natural systems, are fundamentally different in type from the systems among which they are typically being placed, and that they need to be approached in another way.

We have furthermore noted two ironies in that the reinvigoration of crude naturalism is driven by, on the one hand, increasing impact of social systems on the natural world, and, on the other, the development of digital data and computational methods:

For the former, it seems obvious, as Malm and Hornborg (2014) point out, that the increasing acknowledgement of society’s impact on nature should logically imply a more profound engagement with social and cultural theory, not in coaching society in terms of natural science. That humans alter geological layers does make social science more relevant for geologists, but it does not make geologists more equipped to study societal processes: as is becoming painfully clear,
society mostly drives climate change, not the other way around.

The latter is exemplified by Manovich (2016): “Digital is what gave culture the scale of physics, chemistry or neuroscience. Now we have enough data and fast enough computers to actually study the ‘physics’ of culture”. The irony here is that digitalization is part of a development toward technology becoming increasingly liquid (e.g. Archer, 2014; Bauman, 2013), thereby undermining the ontological stability of culture (e.g. Elder-Vass, 2017). This means that formal methods are in fact becoming less capable of dealing with society, as their necessary assumption of fixed ontology are becoming increasingly problematic, and their “short runs” increasingly short (Simon, 1991). In other words, approaching society as a closed system – as physics, if you will – has in fact never been less appropriate.

The underlying naturalistic idea that drives both of these developments – that formal methods constitute a development toward increased predictability, and toward making social science “scientific” – is fundamentally misconceived. As we have seen, the lack of precision is not a problem of the methods, but part of the ontological nature of the system: if your prediction is exact in an uncertain system, it is certainly exactly wrong. Car repair doesn’t become a social science just because you start interviewing carburetors – it just makes you a terrible mechanic.

The category of systems to which society belongs – whose components display both individuality and artificiality (Khalil, 1995) – are non-decomposable (Simon, 1991), which in turn means that their mode of change is, to important extent, of kind, not of degree. To understand such change, methods based on fixed ontologies – including quantification and measurement on a ratio scale – will not be adequate: change of kind implies a change of the very meaning of what is measured. Instead, cause-and-effects in such systems can be accessed through the comparison of categories of effects, and from there deducing the causes – according to the idea that “where there is difference in the effect, there is a difference in the cause” (MacIver 1942, p.26, cited in Byrne and Callaghan 2013). However, nota bene that this in no way contradict a computational approach, but simply implies that the nomothetic program may not be an appropriate foundation for such a pursuit. Computational tools may be highly beneficial for aiding a program of systematic comparison, as illustrated by the computationally relatively simple tools associated to Qualitative Comparative Analysis (e.g. Rihoux and Ragin, 2009).

As we have seen, the claims that emergentism and Complexity Science are anti-reductionist seem to be in practice misleading: complex individualism is still individualism. The difference between the Platonic backhand and forehead
(Hayles, 1999) is not between reductionism and emergence, but between component and relation reductionism. Since these abstractions are rendered as real, the corresponding ontologies follow from a lack of distinction between the real and the artificial. What we need is not to go from “factors to actors” (Macy and Willer, 2002), but from artificiality to reality (Khalil, 1995).

Instead of approaching naturalism by continuing to “extend the tool found successful in one domain to decipher the other” (Khalil, 1995, p.414-415), this essay has suggested taking a turn to ontology in the complexity scientific approach to societal systems. A set of compatible ideas were found primarily in critical realism of e.g. Archer et al. (2013), and Bhaskar (2013), and in particular Reed and Harvey’s (1996; 1992) “complex realism”, further developed by e.g. Harvey (2009), Byrne (2002, 2004, 2005, 1998), Byrne and Ragin (2009), and Byrne and Callaghan (2013), as well as in Mario Bunge’s related “emergentist-systemism” (e.g. Bunge, 1979a, 2000a,b).

Based on a review of these ideas connected to the ontological categorization of Wicked Systems, this resulted in an approach to societal complexity that:

(i) steps away from the ontological individualism and reductionist explanation of mainstream Complexity Science, to one that views emergence not as something that necessarily develops bottom-up, but that can run between levels and in any and all directions. This furthermore suggests a type of explanation that does not merely describe underlying elements, but traces processes through the structures and elements involved.

(ii) focuses on qualitative change – changes in kind rather than only in degree – suggesting a step away from approaching data through variable-based methods, to one that instead emphasizes categorization and pattern-finding, accepts that wicked systems cannot be fully represented in closed models, and therefore calls for an approach that sees the scientific practice as interpretation rather than as analysis, and aims for exploration rather than complete explanation or “realistic” representation.

(iii) sees to complex rather than simple causation, and tendencies rather than laws. This reframes models from attempts to match or access a platonic ideal form of the system – inevitably resulting in their reification – to indicators giving us traces of reality (Byrne, 1998). This changes our understanding of the toolkit of complexity from realistic representations of reality, to parts of a set of “computational hermeneutics”; crutches that can help our intuition navigate cognitively difficult terrain, but that should never be expected to walk on their own.
(iv) takes an integrative and methodologically pluralist approach, allowing ontology to guide methodology and focusing on the specific with an understanding of the totality, tying together a range of traces to a coherent narrative.

(v) steps away from notions of science as objective and external, admitting that the social scientist is an inseparable part of the subject of study, meaning that theory becomes a form of practice: we change the world by understanding it, and we understand it by changing it (Byrne, 2002). Since description becomes a form of criticism in an unequal and unsustainable world, this in turn implies active engagement.

To conclude, it should be noted that this does not only bear relevance within the confines of scientific research; we are ravaging our world with a “perfect storm” (Pievani, 2014) of interrelated crises of increasing scale and magnitude, driven, ultimately, by our confounding lack of understanding of the very world which our own strength has established. It is increasingly understood that this “crisis of crises” (Beddoe et al., 2009) calls for a fundamental transformation in how we organize our society – what some call a Great Transition (Raskin et al., 2002). But at the same time as the capacity to understand large-scale change is starting to seem increasingly essential for our survival as a species, scientific research risks withdrawing – driven by scientism and empiricism – to a focus on small, empirically verifiable study, with any theory left implicit and tacit. But to see the human impact on the Earth system, and allow us to gaze toward a more just, equal, and sustainable future, a bigger picture is needed, capable of evoking the experience of seeing the first photographs of Earth from space: showing our precious sphere floating in infinite black emptiness; reminding us that we live on a precarious razor-thin interface between rock and emptiness, between a vast past and an unknown future, in a moment that is not only remarkable in beauty but perhaps also in brevity. We need a science capable of seeing not only the frailty of that reality, but also that another world is possible.


Archer, M. S., 2014. Late Modernity. Springer.


Fenton-Smith, R., 2015. Can soup change the world? [Online; accessed 2016-12-20].

URL https://goo.gl/h301Is


URL http://science.sciencemag.org/content/323/5915/721


Manovich, L., 2016. Online post. [Online; accessed 2016-12-20].

URL https://goo.gl/GdcJhm


Moore, J. W., 2014. The capitalocene, part i: On the nature and origins of our eco-
logical crisis. unpublished paper, Fernand Braudel Center, Binghamton Uni-
versity.

Morehouse, R. E., 1994. Beginning qualitative research: A philosophic and prac-


Mouzelis, N., 1991. Back to sociological theory: The construction of social or-
ders, new york: St.

Mouzelis, N., 2007. Habitus and reflexivity: Restructuring bourdieu’s theory of


(Eds.), Approaching Truth: Essays in Honour of Ilkka Niiniluoto. London:

Nowak, M. A., 2006. Five rules for the evolution of cooperation. science
314 (5805), 1560–1563.


Orr, J. E., 1996. Talking about machines: An ethnography of a modern job. Cor-
nell University Press.


mental systems and evolution. Mit Press.


Pearse, A.-M., Swift, K., 2006. Allograft theory: transmission of devil facial-


Paper I
Societal systems – Complex or worse?

Claes Andersson a,*, Anton Törnberg b, Petter Törnberg a

a Complex Systems Group, Division for Physical Resource Theory, Department of Energy and Environment, Chalmers University of Technology, 412 96 Göteborg, Sweden
b Department of Sociology and Work Science, University of Gothenburg, Box 720, 405 30 Göteborg, Sweden

ARTICLE INFO

Article history:
Available online 22 July 2014

Keywords:
Complexity
Complicatedness
Wicked problems
Wicked systems
Models
Theory

ABSTRACT

The basic observation that we explore in this paper is simple but, we argue, rich in consequences: societal systems combine two qualities that are commonly referred to as complexity and complicatedness. We address the problem that societal systems remain recalcitrant despite the development of powerful approaches for dealing with both of these qualities. The root of this problem we identify to be that the combination between complexity and complicatedness is emergent; i.e. fundamentally and irreducibly different from either quality in isolation. This means that neither class of such approaches can be expected to work well on their own. It also means that the obvious strategy of combining theory for complexity and complicatedness may be much more challenging than envisioned. In short, systems where complexity and complicatedness is mixed ought to be treated as a distinct class of systems. Noting a connection to what has long been called “wicked problems” we hereby outline such a class of systems that we call “wicked systems”. We introduce a simple model and heuristic and discuss some implications for theorizing and modeling.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

A mounting scale and frequency of societal and environmental crises has increasingly brought about a realization that we must broaden the range of factors that affect the direction of societal evolution to include societal and environmental values. This means not least that we must break out of the hegemony of the present and near future. But to do so has turned out to be more easily said than done. Societal change and innovation has always been a process that basically unfolds spontaneously, without much overall control and monitoring, and the dynamics of societal systems is also more and more being identified as inherently hard to control and predict. In short, society, and the larger global systems that it is part of, are being identified as complex systems.

All of this clearly changes the landscape for policy (see e.g. Byrne, 2005; Leach, Scoones, & Stirling, 2010; Scoones, Leach, Smith, & Stagl, 2007). The question of how to predict and optimize the future is changing into an acceptance of the futility of such aims and the aiming for other goals such as resilience and sustainability. On the one hand, this raises serious questions about the efficacy of many of the standard policy tools, most of which are designed under entirely different assumptions about how societal systems work. Indeed, it challenges even our basic intuitions about how societies evolve. But, on the other hand, it has also opened up the promise of entirely new types of analytical and policy tools, based on ideas about how we can

* Corresponding author. Tel.: +46 31 7723127.
E-mail address: caand210@hotmail.com (C. Andersson).

http://dx.doi.org/10.1016/j.futures.2014.07.003
0016-3287/© 2014 Elsevier Ltd. All rights reserved.
dynamically steer and scaffold society by engaging more directly with its causal mechanics; for example more bottom-up approaches like the management and design of social networks of actors (e.g. Lane & van der Leeuw, 2011) or the management of innovation dynamics (e.g. Loorbach, 2010).

The application of complexity science is today increasingly suggested as a solution to the new set of challenges that arises from this change in outlook. As we shall see, it is not a trivial matter to delineate complexity science, and it probably cannot be done to universal satisfaction. What we do here is to identify a mainstream of complexity science, with a dominant position from which is has come to set the research agenda, and in effect also thereby define what the concept of complexity means. This mainstream represents a mainly formalist and simulation-based approach to complexity, with its roots in the natural sciences, that has proven highly capable of analyzing many types of complex systems that have otherwise been impenetrable to formal approaches. In the social sciences, the promise is basically the ability to internalize complexity into models of large-scale societal phenomena. The hope is that, like holistic narrative-based approaches, complexity science could deal with historical path-dependent change, but in a more formal and analytically powerful way, and with the crucial addition of providing a theoretical grasp of non-linear dynamics and emergence. It would thereby be able to address a widely recognized need for what we might call a higher theoretical level of resolution (see e.g. Geels, 2010; Holtz, 2011; Malerba, 2005), and be key to a move “from promising to delivering” with regard to this new complex outlook on society.

But despite what looks like an obvious match, and despite a good deal of effort in this direction, the success of complexity science has been mainly limited to rather simple social systems, such as crowds, traffic or evacuation, where the complexity of human behavior is reduced to simple rule-following. Success in the study of societal systems in their full complexity has been much more limited. Although concepts like path-dependency, attractors, tipping points and chaos have provided basic lessons that have transformed deeply seated ideas about causality in society, these are highly general lessons that have proven hard to operationalize and they are not strongly represented in policy work. In many – if not most – branches of social science, complexity science remains rarely or at least superficially used. Complexity science appears to remain in a state of being perpetually promising in relation to social science.

We think complexity science is crucial for understanding of how societies work, but that to really make a difference it must be more carefully placed into the context of other basic approaches to understanding systems. We will argue that it is really suited for dealing with only a subset of those systems where we are hampered by the presence of non-linearity and emergence.

Our aim in this paper is to make this problem more explicit and workable, and to contribute an analysis of why complexity science, and indeed formal approaches in general, has proven to be so hard to apply to the social sciences. The paper is structured as follows. First, we introduce the distinction between complexity and complicatedness that is central to our argument, declaring in the process what we mean by “complexity science”. Second, we go on to characterize systems that mix these two qualities as “wicked systems” and introduce a model and a heuristic that allows us to chart the relations between problems, systems and methods. Our third point is the question of what wickedness is and why “wickedness” is so wicked” – i.e. what is it that makes these systems so hard to address using formal approaches. Fourth, we argue that combining approaches for dealing with complexity and complicatedness separately, which is otherwise a methodological strategy that suggests itself, has worked poorly so far. Using the introduced tools, we offer an analysis of why this would be so problematic. Sixth, and finally, we offer some thoughts about further implications and future directions.

2. Complex, complicated and wicked

We begin from a distinction between complexity and complicatedness (also referred to as dynamical and structural complexity; see e.g. Erdi, 2008). These two system qualities are often juxtaposed and contrasted for the purpose of explaining what complexity science is really about. A Google search for “complex vs complicated” yields a wealth of examples, and the nature of these demonstrates that this is a distinction that is perceived of as consequential and of practical relevance. When opposed in this way, complexity is associated with bottom-up self-organization1 – like the behavior of a school of fish or a crowd – while complicatedness is associated with top-down organization, such as in engineering. The distinction is really used for explaining what complexity is, by specifying something that it is not, but that it might otherwise be mistaken for, namely complicatedness. So complexity, then, would roughly be what we intuitively think of as complexity, but minus complicatedness. That may not go a long way as a formal definition of complexity, or even as an intelligent discussion about what complexity can be argued to be, but we think it does say something about the practice of complexity science.

A quick glance at the history of complexity science may tell us something about how this practice and (largely tacit) meaning of the term complexity has emerged. From very early on, the Santa Fe Institute (SFI) came to act as a powerful uniting and aligning force in what can today be referred to as complexity science. The SFI was the first dedicated research

---

1 A minimal characterization of self-organization, which is a central concept in complexity science, typically includes the emergence of order without any centralized or external description of the order that emerges. In particular, it is microscopic order that, through a dynamics, gets extended to macroscopic scales. Although it has been developed mostly in physics, biology and for social systems, as part of a larger complexity science movement, it originally arose in neither of those fields but in psychology (Ashby, 1947), for more on the history of the concept of self-organization, see Shatz (2008).

2 An earlier systems-theory era notion of this sort of complexity is that of interactional complexity (Wimsatt, 1975) which is defined as the degree of cross-coupling in a system.
center for complexity science, founded in Santa Fe, New Mexico in 1984, by a group of highly influential scientists, many of which were active at the nearby Los Alamos National Laboratory, with roots in the Manhattan Project, and thereby also in the origins of scientific computing and dynamical systems theory in general; see e.g. Galison (1997). Many important ideas about complexity are of course older than, and unrelated to, the SFI and its precedents. Qualitative social science, for instance, often recounts a very different heritage (in systems theory, see e.g. Sawyer, 2005; Vasileiadou & Safarzyska, 2010). But it remains the case that, from the 1980s and onward, the SFI came to define a mainstream of complexity science, and thereby also, in practice, the concept of complexity as understood by scientists, policymakers and the public.

The SFI was created as an explicitly multidisciplinary center, but although it was, and remains, highly multidisciplinary, it is not as methodologically diversified. The methodology that was (and still is) mainly pursued at the SFI is formal, quantitative and closer to natural science and quantitative social science. As the heart of this methodology, as a critical enabler, lies computer simulation and the capability to describe the entities and interaction rules of a dynamical system and put it all into motion. This is an extremely flexible methodology that made it possible to study and visualize dynamics that used to be utterly inaccessible to the human mind – aided or unaided. Above all, it makes possible a systematic inquiry into emergence in dynamical systems.

Fascination for this new capability and its potentials was a powerful impetus in the formation of complexity science worldwide with the SFI as a hub. Furthermore, although it predates, and is not necessarily a proper part of it, complexity science is underpinned by a set of fundamental concepts for dealing with non-linear dynamical systems, such as bifurcation, path-dependency and chaos etc, most of which originate from chaos theory (see e.g. Cvitanovic, Artuso, Mainieri, Tanner, & Vattay, 2005; Ott, 2002) but also from other traditions such as dissipative systems theory (Prigogine & Nicolis, 1977) and synergetics (Haken, 1977).

The typical model in this tradition has a microlevel of agents/nodes existing in a pre-defined environment. What is studied (and subject to empirical testing) are the patterns that arise on an emergent macrolevel from the dynamical interaction between these agents/nodes. Individual traditions and scientists may be more or less strongly aligned with this mainstream, but anyone claiming to work with “complex systems” must relate to it in one way or the other.

So our “definition” of complexity is (i) that complexity as a concept is constructed by the main of the community that works with complexity; (ii) that the concept of complexity thereby reveals the limits of applicability of a particular theoretical toolbox. So we think that it points to (iii) a specific class of systems that happens to be amenable to analysis using that particular toolbox. What is key to our differentiation between complexity and complicatedness is this: although complexity and complicatedness are clearly linked in numerous ways, they still present us with two radically different sets of methodological and theoretical challenges.

So what about societal systems in this mix? Are they complex or complicated? On the one hand, societies are undeniably complicated with their multi-level organization and bewildering array of qualitatively different and interacting entities. Systems theories for example seize upon what is seen as an irreducible complicatedness of societal systems. Yet society is also often, and quite convincingly, argued to be a complex system in the bottom-up self-organization sense (e.g. Ball, 2012; Castellani & Hafferty, 2009; Sawyer, 2005), and it can certainly be argued that much of its complicated structure arises from bottom-up rather than top-down processes. We clearly see no reason why systems could not be both complicated and complex at the same time, and societies would appear to be an excellent example of such a type of system. Systems that combine these two qualities will here be referred to as “wicked systems”. We will soon go into the details of what that means, but for now we will just use the term as a label.

It is precisely for those “wicked systems” that we claim that scientific practice and expectations become confused, and where we think that the compass for future scientific development has been lost. Let us illustrate this by considering some different viewpoints on society and complexity, all representing insights, but insights that we think are hampered by the lingering confusion that we here aim to address.

For example, when Ball (2012) speaks of society as a complex system, he appears to mean complexity in a wide sense that encompasses both complexity and complicatedness. But the models that are covered really deal with complexity only in the more narrow sense used here. We may only presume that he thinks that complexity of the structural sort (i.e. complicatedness) should be explained by models working on the basis of dynamical complexity (i.e. complexity in the here-used dichotomy). This is a micro-macro view, quite typical of positivist science, and of mainstream complexity science, where micro is seen as prior to macro, and thus the universally proper locus of explanation.

Walby (2007) provides a lucid analysis of what complexity science (or, theory, in her parlance) could and should contribute to sociology, and her picture of societal systems is clearly one of what we refer to as “wicked systems”; i.e. combined complexity and complicatedness. But in her analysis of how this contribution should occur, where she applies the concepts of complexity science, she ends up where others have frequently ended up: in a tantalizing but unfilled promise of something more and deeper; in a perpetual feeling that we are close to a breakthrough and a paradigm shift. In their native

---

1 In the mission statement of the Santa Fe Institute one can see that the commitment to quantitative approaches is quite explicit: “SFI combines expertise in quantitative theory and model building with a community and infrastructure able to support cutting-edge, distributed and team-based science” (our emphasis).

2 See also Byrne (2005) concept of “simple complexity”, and Morin (2007) concept of “restricted complexity”, which are in line with how we conceive of this complexity science mainstream; also Byrne and Callaghan (2014) who, from the standpoint of Morin’s terminology, discuss the dominance of this mainstream.
domain, the complexity concepts that she applies are crisp, well defined and methodologically powerful. When applied to systems that are both complex and complicated they still make a lot of sense, but they also somehow seem to lose their theoretical bite.

The distinction that we propose between complex and wicked systems can also be interpreted in the context of the sociological discussion about the meaning of complexity in societal systems by e.g. Byrne (2011), Byrne, 2013, Byrne and Callaghan (2014). Their view of societal complexity is referred to as “complex realism”, which can be described in terms of sociological traditions as critical realism informed by complexity science (Harvey & Reed, 1996; Reed & Harvey, 1992). Using the terminology of Morin (2007), these authors differentiate between two perspectives on complexity as “restricted complexity” and “general complexity”, where the latter is similar to our description of wicked systems, and the former to what we here refer to as mainstream complexity. However, the separation between general and restricted complexity is treated as simply two separate perspectives on the same thing, namely “complexity”. The point that we wish to make is that these two perspectives are better understood as descriptions of different things: general complexity does not actually describe complexity, but rather wickedness. We understand complexity as part of wickedness, but we do not thereby see wickedness as a type of complexity. This can perhaps be likened to how the fact that the color red is a component color of yellow does not mean that we can describe yellow as a type of red. Our point therefore fundamentally transforms the meaning of this discussion: general complexity is in our view not a generalization of restricted complexity, it is a description of something of which restricted complexity is a component, but that is qualitatively different than restricted complexity.

Surprisingly perhaps, the possibility of systematically exploring the consequences of mixing complexity and complicatedness appears not to have been explicitly pursued. As things stand, the mainstream of complexity science may be aware that complexity and complicatedness are distinct qualities, but complicatedness in complex systems is not seen as a fundamental problem. Complicated or not, they are complex, and that is what is seen as fundamentally important: extending mainstream theory to deal with them is seen as challenging and hard, but, essentially, gradual and cumulative work. Sociological complexity thinkers disagree with mainstream complexity science about how complexity ought to be understood. Notably, however, they do not really disagree that, in the end, complexity ought to be one single thing. Our analysis will lead us to a different conclusion. We do not see societal systems as a type of complex systems, but as a type of system where complexity is mixed with complicatedness, yielding an emergent quality – wickedness – to which neither complexity science, systems approaches, mathematical models or combinations between them lend themselves very well.

3. Understanding the co-evolution of methods, problems and systems

We will now introduce and explore a simple model and heuristic aimed at helping us to better understand relations between methods, problems and systems. We focus on the question of why mainstream complexity science would be so hard to extend to “wicked systems”, but we think this question is more general than that. We think it boils down to why formal science in general would be incapable of dealing with these systems.

In Figs. 1–4 we chart out system types, problems and theoretical approaches on the basis of these two system qualities. The corners of the plane that is described give us four ideal system types. Systems that are neither complex nor complicated (bottom-left corner) we call “simple systems”, systems that are complex but not very complicated we call “complex systems”, systems that are complicated but not very complex are labeled “complicated systems” and, finally, systems that are both complicated and complex we call “wicked systems”. We may view these figures as either four-field graphs or graphs with continuous axes, corresponding to whether we prefer to see complexity and complicatedness primarily as binary qualities or graded quantities.

We choose the term “wicked systems” in recognition of a potentially deep connection (whose exact nature remains to be worked out) between this class of systems and what has been called “wicked problems”. The term “wicked problems” was first coined in management research by Horst Rittel (briefly introduced by West Churchman, 1967) to characterize a class of problems that failed to fit into the molds of the formal systems theoretical models that were being applied across the board at the time with considerable confidence. Just about any large-scale societal problem can in fact be confidently put into the category of wicked problems: starvation, climate change, geopolitical conflicts, social disenfranchisement, and so on. All these are problems that escape definition and where there is a constant feeling that the efficacy of proposed solutions is called into question not only with regard to feasibility and adequacy but also with regard to the risk of creating cascades of other problems that are impossible to foresee and that may be even worse than the initial problem (see also Leach, Scoones & Stirling, 2007; Scoones et al., 2007). Explicating the concept, Rittel and Webber (1973) conclude that the domain of wicked problems in social systems is vast – it includes just about any problem short of trivialities. In West Churchman (1967) words, what we do with wicked problems is to either tame them by creating “an aura of good feeling and consensus” or by “carving off a piece of the problem and finding a rational and feasible solution to this piece”; this would appear to well describe also our generalization of “wickedness”. By considering “wickedness” as a system quality, we generalize to also be able to speak of things like wicked dynamics, wicked phenomena and wicked systems.

---

1 At least not recently and with the benefit of a more developed science of complexity at hand. But see Wimsatt (1975) for an early and interesting analysis that clearly goes a long way in Morin’s direction; also re-printed in Wimsatt (2007).
In Fig. 1 we map three groups of formal theoretical approaches into the complexity–complicatedness plane: “mathematical theory”, “systems based theories” and “complexity science”. Under the rubric of “mathematical theory” we place theory and models mainly based on closed-form equations (whose scope have later been expanded with numerical methods), most importantly in this context neoclassical economic theory, such as mainstream macroeconomics approaches, which are characterized by strong assumptions about agent rationality and equilibrium that serve to make models mathematically tractable. When we refer to systems based theories we mean approaches that rely on holistic ontologies;
systems theory is a central example of such approaches although the category is wider than that. What can be more generally termed “systems thinking” (e.g. Weinberg, 2001), for example, permeates science exceptionally widely today. Systems based theories break with the reductionist tradition of mathematical theory by making the structure of societal systems explicit on a larger scale, deeming them to be irreducible: this has expanded the reach of formal methods from the simple corner towards the complicated corner, generally following the example of engineering. Finally, complexity science has revolutionized science on a fundamental level by covering an important flank that used to be so hard to deal with that it was nearly entirely unexplored before the 1980s. It mainly expands from mathematical theory, but also to an important extent from systems based theories with which its shares a strong attention to feedback processes. As we illustrate in Fig. 4, agent-based simulation, or more generally “descriptive simulation” (see Edmonds & Moss, 2005), expands from the simple-and-complex towards the complicated-and-complex side, and in doing so it comes to share features with systems based approaches; more on this later in the paper.

Societal systems would have their center of gravity near the wicked corner of the graph (see Fig. 2), but since formal approaches are unable to access many of the problems that societal systems present us with, they will selectively address sub-problems that happen to fall into their domains or transplant problems from near the wicked corner to the corners of their preference (see Fig. 3). In the former case we obtain important but limited “snapshots” of the system in question and we are, as we will discuss later, faced with the problem of how to combine such snapshots; see also Wimsatt (1975). In the latter case we may get spurious results where strong assumptions mean that the benefit of accessing formal methods of analysis does not warrant the price in realism.

Most complexity scientists will readily admit that their influence is concentrated to the region near the complex corner. But since complexity science defines itself more generally as dealing with systems of high complexity (falling toward the top of the complexity axis) there is no systematic recognition that something about complex systems may change qualitatively along the complicatedness axis. In summary, it is easy to identify powerful scientific approaches for dealing with all parts of the plane except for the wicked corner.

To begin understanding what a mix between complexity and complicatedness entails, we should first note that we are mixing a primarily structural quality – complicatedness – with a primarily dynamical quality – complexity – and that both of these on their own are theoretically challenging. To make matters even worse, the way in which they intermix cause these qualities to fuse into something quite unlike either quality in isolation. This is what we mean when we say that the combined quality of “wickedness” is emergent. Complexity and complicatedness can be seen as mutually reinforcing in societies and ecosystems – our two principal examples of wicked systems. Self-organization here generates, changes and maintains macro structure, and macro structure, in turn, scaffolds and creates a multitude of arenas for self-organization. We will now delve at least somewhat deeper into to this observation in order to understand better how and why wicked systems come to mix

---

Organisms may seem strange bedfellows with machines near the complicated corner. But although organisms are generally closer to the Wicked corner than machines they are in fact quite strictly organized. Just like machines they generally do not change during their lifetime: they have an assembly/ontogenetic phase followed by another phase during which they have a nominal ecological function (in some instances this pattern will recur, such as in metamorphosing species like dragonflies, but this happens in a highly pre-ordained way.)
complexity and complicatedness: Where do they come from? How are they maintained? What sets them apart from complex and complicated systems?

4. Wicked systems as poorly decomposable systems

The constituents of wicked systems constantly try to outsmart one-another, reaping their own benefits, reacting to threats from other constituents. They constantly enter into new constellations, dissolve old constellations, and react to the immediate situation around them. What we get is a situation where complicated organization and complex dynamics is in a constant state of re-negotiation, constantly challenging any settlement of the system into a level hierarchy, constantly facing the system with qualitative novelty that other components have to react to. They are more like arenas for interaction between entities (typically complicated systems like organisms and artifacts) than they are like systems with overall top-level functions and agendas of their own. When we imagine a desirable society, we do tend to imagine it as a good arena for interaction, yielding equitable and sustainable patterns of interactions for its denizens.

Such systems, as we know, can be tremendously persistent despite all this upheaval. In fact, since they lack an overall function (at least in a straightforward sense, see e.g. Schlosser, 1998), the very concept of them breaking down is not straightforward. A breakdown of societal or ecological systems typically does not mean that they cease to operate (like a car that broke down and sits at the side of the road), but that the conditions for (and makeup of) the components change in a dramatic way; species going extinct, humans starving, being oppressed and so on. We may expect them to be persistent in the sense of having continuity, but less persistent in the sense of providing an arena for interaction that we see as desirable.

It is not hard to imagine why systems of this description would be challenging to understand formally – but can it be understood formally why they cannot be understood formally? The benefit of achieving such an understanding could be substantial as it could offer a more detailed map of what, more precisely, the methodological problem with these systems consists in. Such an understanding, if it can inform us about what approaches are likely to work in which contexts, can also serve to make it easier for formalists and non-formalists to come out of their trenches and collaborate. We will now propose one way of understanding why wicked systems are so recalcitrant in a way that is both formal and intuitive and that should be accessible to formalists and non-formalists alike: the organization and dynamics of wicked systems make them poorly accessible to approaches that rely on what Simon (1996) called near-decomposability, which is to say just about any conceivable formal theorizing.3

Like any formal scientific approach, complexity science and systems theories strive to isolate systems for independent study such that model properties can be fixed and formally defined. Ceteris paribus, formal models are desirable since they promote clarity and may be amenable to powerful analytical tools based on mathematics and computation. As Byrne (2005) states, the problem with formalist science is by no means that it does not work, it is "... that it works where it works and it

---

3 Using Simon’s model also has the added benefit that it has been highly influential in shaping both practice and conceptual work in systems and complex systems theory.
does not work everywhere". But what really makes a system suitable for such an approach? Under what circumstances, more precisely, will formal methods run into trouble?

Simon (1996) introduced the concept of “near-decomposability” to explain in a clear and systematic way what conditions that need apply for a system to be possible to study in a formal and controlled manner. The central observation is that if a system is to be possible to study in isolation, its dynamics cannot be importantly disturbed by outside influences. In Simon’s parlance we should be able to identify an internal environment where the dynamics that we study takes place, and an external environment that can be assumed to be static, or at least to be variable only in highly regular ways. The boundary between the internal and external environment is the delimitation of the model, and it is referred to as the interface. What we study with a model is then an internal environment. Hierarchical system organization is important here: our internal environment constitutes the external environment of the objects that populate it, and that we deal with only via their interfaces. We deal, therefore, with objects only in the form of interfaces, for example their interactions with other objects in the studied system. The beauty of all this is that it makes the world manageable: we declare our system as autonomous from external disturbance and we hide any complexity and complicatedness residing on lower levels of the hierarchical organization.

We may study this internal environment during what Simon refers to as the short run: a time scale that (i) is long enough that our objects’ interfaces are meaningful and for important dynamics to have time to happen and (ii) short enough that our assumptions about the interfaces remain valid. The greater the separation of scales between the internal and the external environment, the greater will the difference in size and speed of the dynamics on these two levels be, and the more generous will the short run be; i.e. the more interesting things will have time to happen. For example, models of particle physics can gainfully be formulated in this way because those systems exhibit a clear and clean scale separation. Engineered systems, as Simon (1996) points out, are designed to fit into above description.

Importantly, this scheme can, potentially ad infinitum, be nested hierarchically. At each new level we can reduce both complexity and complicatedness back down to manageable levels again, and this is what allows us to construct systems that, taken as wholes, are mindbogglingly complicated; most notably in engineering, but they same principle applies to biological organisms. The parts of such a system can be improved independently, with respect to identifiable functions, as long as those functions in the system are retained. For example, it is straightforward to replace the engine of a car with another engine with the same function, but, say, an improved fuel economy. In fact you can do anything to a component as long as you do not alter its interface. Significantly, the components of technological artifacts and organisms also have no separate agendas; it makes no sense for the engine of a car to benefit on the expense of the car as a whole since engines are meaningless objects except as part of a functional whole. Multicellular life, with cell types, tissues and organs, is no different in this respect.

Selection for top-level functionality often yields hierarchical systems with near-decomposable levels (such as engineered artifacts or organisms) since such an organization increases evolutionary adaptability (Simon, 1996; Wimsatt, 1975).

In many important cases we can surely make assumptions of near-decomposability for societal systems, and when we can we are able to bring powerful scientific approaches to bear. For the purposes of complexity science it would seem reasonable that certain subsystems, such as traffic or crowd behavior, can be argued to fit this description. The dynamics of cars and people play themselves out over much shorter time scales than that on which urban systems, roads, traffic regulation and so on, change. Such phenomena are also often ephemeral, which bounds the problem even further. For example, at night the traffic jam dissipates and leaves no traces that affect tomorrow’s traffic. Similarly may be argued for certain highly abstractly conceived phenomena that depend on persistent features such as network dynamics, geography, basic resource constraints, strategic dilemmas and so on.

But what about societal phenomena more in general? For example, what about sociotechnical transitions or other wicked problems? Sociotechnical systems in general are open systems, in which many and far-flung social, technical and natural processes co-exist, co-evolve and have an impact on each other on overlapping timescales and levels of organization. In many cases these problems unfold across time scales of decades or more. They involve discontinuous, qualitative change as well as cascade effects (e.g. Lane, 2011) whereby change strongly and rapidly feeds back into the conditions for further change. Such systems are, to say the least, hard to contain in a Simonean compartment with a “short run” over which, for example, transitions can be studied against the background of an unchanging external environment. The fundamental problem for complexity science in this context, and really any approach that relies on these ontological assumptions, is that on the time scales of sociotechnical change almost everything in society really is “changing with everything else” (Malerba, 2005): there is no relevant “short run” for a model to operate in; there is no way of cutting the system into distinct and persistent levels of organization.

5. The obvious way forward and why we keep bogging down

We may cruelly summarize the formal theoretical situation as follows. Macroeconomic models drag society towards the simple corner of the plane illustrated in Figs. 1–4, which brings it under the sway of analytical or numerical mathematical methods. This makes for supremely formal and powerful models, but it often comes at the cost of moving societal problems far from their “empirical home”. Systems theoretical approaches emphasize the complicatedness of society and attempt to understand, design and steer it as a complicated but not complex system, approximately in analogy with machines and

---

* A human can for instance make decisions (a typical interface feature) over a time scale of minutes, but hardly on a time scale of milliseconds.
organisms. By doing this, society is moved from near its native wicked corner to near the complicated corner. Although systems theories catch more of the structure and dynamics of society than macroeconomic theory does, they again bring society out of its right element in order to subject it to certain tools of analysis and design.

The call of complexity science is that society is neither simple nor complicated – it is a complex system – so we should use these new complexity methods for understanding it (see e.g. Ball, 2012; Epstein, 2007; Epstein & Axtell, 1996; Squazzoni, 2008, among many). In evidence that this is a promising track, successes in understanding certain simple and well delimitable social systems, that happen to fall closer to the complex corner, are brought forth. These are often viewed as “steps along the way”; a way that is envisioned as perilous but negotiable using the established methodology, either by linking the formal models or by scaling them up. But by emphasizing complexity in this way, we think that society is again frequently moved from its wicked corner. This time toward the complex corner and, although in a different direction, again out of its right element.

If society is both complicated and complex, it would appear particularly reasonable to use a combination between approaches that successfully deal with complicated and complex systems to provide a more integrated view. For example to combine systems theoretical methods with a complexity approach. This is also in practice what many are attempting to do. Multi-agent simulation (e.g. Gilbert & Troitzsch, 2005), which is basically an extension of simple complexity modeling (such as in physics and chemistry), could for instance be described as such an approach: agents, interaction modalities and environments are designed a priori in a fashion that recalls a systems approach, but they are subsequently let loose in a dynamics where emergent patterns arise, so they are clearly also complexity models. One can also imagine multi-agent and other models being combined in the form of modules in a more explicit hierarchical systems ontology along the lines proposed by Ostrom (2007). This is also an approach typical to Integrated Assessment Modeling (Rotmans & Asselt, 2003) where systems of models are constructed to deal with problems involving large-scale, often global, dynamics; societal sub-systems are usually represented simplistically using general equilibrium economics models. Agent-based models have frequently been used as modules in such models for some time (see e.g. Pahl-Wostl, 2002).

But are such simple combinations, where models organize complexity into a complicated structure, necessarily suitable for providing a fuller picture of these wicked systems? We would say that they are not. In wicked systems, the rules and entities are not only hard to uncover, they change as a result of the dynamics itself. There are no neat scale separations in wicked systems that allow us to separate ontology from state in the general case. This is what Lane and Maxfield (2005) refer to as “ontological uncertainty” and it is also the reason why wicked systems are “worse than complex”.

Another reason why combination works poorly is that the formal methods that we thereby combine derive their power not from the presence, but from the lack of complexity or complicatedness in systems. They rely on simplifications of either the structure or the dynamics of systems (or both), and for wicked systems we can frequently do neither while maintaining acceptable levels of realism. For example, engineering does not work because technological systems are complicated. It works because technology can be constructed such that it is not complex. For complexity science, even a cursory review of the literature clearly reveals that it excels specifically in dealing with systems toward the top-left corner of this plot; i.e. systems that are complex but not very complicated. Combined approaches thereby combine the weaknesses rather than the strengths of the constituent approaches.

6. Discussion

Can this theoretical lacuna near the wicked corner be filled – just like the area around the complex corner is being filled by complexity science? We think that a great challenge for the future lies precisely in understanding wicked systems. In this paper we have focused on the preliminary task of establishing this category of wicked systems, and as a result there has been little space to pursue all the implications that we see as important. Some of these, such as what the implications of our argument would be on the structure-agency debate, would demand an entire article on its own for a satisfying treatment. However, the concept of wicked systems clearly has implications for this discussion concerning the analytic connection or co-relation between individual and social processes, since the conditions under which we see scale separation, and the effect that scale separation has on the efficacy of different approaches, are central to our argument. On the basis of the foundation that we lay here, the exploration of these issues will have to wait for coming papers, as will indeed a more thorough interpretation of the implications for constructing and imagining futures.

In Section 5 we concluded that the obvious strategy that (explicitly or implicitly) mixes approaches for dealing complexity and complicatedness in formal models faces serious problems. The reason for this is that, as we argued in Section 4, formality fundamentally rests upon an assumption of non-wickedness. We think this contributes to an explanation of why narrative\footnote{Narrative is of course the oldest form of theorizing about social systems. There is quite a deal of literature on the concept of “narrative” and how it is important for not only “wrapping our minds around” society, but also to deal with its constant state of change from innovation, re-interpretations and the potential for taking different perspectives on the same thing. There is no room here for a review of this literature, so we point the interested reader to e.g. Geels and Schot (2010) and Lane and Maxfield (2005) and further references therein.} qualitative approaches have been so hard to move away from despite vigorous efforts to
do so. Arguments to this effect have so far typically been put forward by qualitative scientists in the language of, and for a receptive audience of, qualitative scientists. What we have attempted here could be described as a formal, yet intuitive, critique of formal approaches to wicked systems. Our choice to argue our point in the framework of Simon (1996) should be understood on this background. It was not only suitable for the purpose. Being based, as it is, on robust and time-tested systems thinking, we also think it has a chance of making sense to both quantitative and qualitative scientists.

In Fig. 5 we add also narrative approaches to the mix, reflecting our observation that the use of narrative in science (and otherwise) is strongly correlated precisely with wickedness. The advantage of narrative is that, unlike formal approaches, it can at least begin to operate in the wicked zone. It can handle a number of key characteristics of wicked systems such as their heterogeneity, their contingency and their multilevel nature. We generally concur with the analysis of Byrne and Callaghan (2014) and Regin (2009) that narrative, and in particular historical case study, is important for understanding what we mean by wicked systems. The future, of course, is always attached to a future history, and an imagined history has all the characteristics of a remembered history. Narrative in the role of constructing what we might call historical case studies of the future has also been discussed in the context of future studies for similar reasons (see e.g. Miller, 2007). In our conceptualization, another way of specifying what the strength of narrative is with respect to wicked systems, is that it is the only mode of theorizing that is not obviously married to assumptions of a lack of either complicatedness or complexity; which, as we will soon get back to, does not mean that it is particularly good at dealing with their presence.

Beyond its use in theorizing, which goes back all the way to the roots of philosophy, the suitability of narrative for understanding matters concerning social life is hardly a happy coincidence. It is linked to our cognitive apparatus and as such a result of evolutionary cognitive adaptation to a multi-million year history of living in communities that are simultaneously complex and complicated; see e.g. the so-called “Machiavellian intelligence hypothesis” (e.g. Byrne & Whiten, 1989; de Waal, 1982; Whiten & Byrne, 1997; Whiten & van Schaik, 2007).

But this also points to the limitations of narrative and the necessity of formal approaches. In particular in the context of modern societal systems, our abilities (while impressive on a zoological comparison) may be quite insufficient for our purposes. Our ability to handle complicatedness is quite limited, not least by a limited short-term working memory (see e.g. Coolidge & Wynn, 2005, 2006; Read, 2008), and our ability to make strict inferences and abstractions without the use of mathematics and formal logic is poor. When it comes to our ability to handle complexity, it is worse than poor: it is treacherous.

Consider for instance Schelling (1978) tipping model as an example. The arguably most remarkable contribution of this model is that it formulates both a surprising question and a surprising answer at the same time. The model basically shows that even tiny preferences about what neighbors that we would like to live among can quickly lead to complete segregation. The whole dynamics thereby turned out to work in a way that is qualitatively different than what our unaided intuition would tell us. Non-linearity and emergence are treacherous because they often throw us off without us even noticing it. We think that we understand the problem. We may even think that the solution to the problem is so obvious that we need not even bother to consider it very closely.

To tackle complexity head-on we absolutely need formal models, and in particular simulation models since they provide a flexibility that we cannot do without. The cognitive engine that runs the narrative must be complemented with formal

---

![Diagram](image.png)

**Fig. 5.** We here map narrative/qualitative approaches together with the formal approaches that we mapped in Fig. 1.
theorizing to get past impasses where reliance on intuition will get us no further, and complexity is perhaps the steepest of those impasses.

The reader may here interject that narrative theorizing is used in conjunction with formal models already. This is certainly true. The problem is that a clear and formulated understanding of (i) why it would be a good idea to do so, and (ii), how it ought to be done, is notably lacking. In particular we are lacking an understanding in a format that would be acceptable to at least a notable fraction of those close to the formalist mainstream of complexity science. Under the formalist and reductionist\(^\text{10}\) ideals that govern the mainstream of complexity science, narrative based strategies clearly remain on the fringe. They are maximally tolerated as something that we are constrained to doing “in the meanwhile”, before we succeed in coming up with formal models. Many indeed see narrative approaches as precisely what complexity science should aid us in “moving past”. We hope to have contributed to a better foundation for questioning, in a systematic way, that sentiment of a \emph{a priori} rejection of qualitative approaches.

As indicated in Section 2, the system quality that we here call wickedness has been described under other labels in the literature, and typically as a type of complexity. For example, Byrne (2005) speaks of “complex complexity” as opposed to “simple complexity”, and Stewart (2001) talks about “incondensable complexity” as opposed to “simplicity”. In both these cases the former concept clearly concern what we call wickedness. Byrne and Gallagher (2014) and Morin (2007) use the term “general complexity” in opposition to “restricted complexity” where the former, however, contains, rather than is complementary to, the latter. It would thereby correspond to the entire top region of our complexity–complicatedness plane (see Figs. 1–4). The set of problems that we associate with wickedness, and the confusion that surrounds them, also come to the fore clearly in the literature about hierarchies and organizational levels. See for example the analysis by Lane (2006) of, on the one hand, the partly overlapping and partly contradictory views on the relation between hierarchy and complexity by Holland (1998), Anderson (1972) and Simon (1996), and, on the other hand, views about whether the concept of hierarchical levels is useful at all in social science by Latour (1993) and Ginzburg (1989).

To summarize our contribution in this context, we have identified wickedness as a distinct quality that is related to both complexity and complicatedness, but as an emergent combination between these qualities, rather than as a type of either quality. We illustrated this in Section 2 metaphorically by stating that red is a component color of yellow, but yellow is not thereby a type of red. Using the introduced diagram we have been able to place systems, approaches and problems in to the context of each other, facilitating a better understanding of their interrelations and, not least, of why systems near the wicked corner of the diagram are so hard to understand. We think that one major benefit of our conceptualization here is that it removes the otherwise so strong prior notion that societal systems should be understandable by extending the toolbox of mainstream complexity science. Having identified that complexity science has mainly evolved to deal with a related, yet distinctly different, type of systems, we confirm the intuition that complexity science has an important part to play, but we have also redefined its role and invited other approaches in alongside with it. We have not offered much on the issue of how to combine such approaches, but considering the vast scope of that question, we think that setting the stage even slightly better for addressing it is an important step forward.

By shifting the picture of society and complexity in this way we hope to have opened up new avenues of thinking on this important issue, which is central not only to future studies, but in any context where the evolution of wicked systems is considered. The theoretical challenges that we discuss here are by no means unique to the social sciences, and that we may be well served by looking to developments in other fields for inspiration, as a guide and as a source of models. Most importantly, evolutionary biology, has over the past few decades undergone a dramatic transformation. The new theoretical landscape that has emerged (sometimes broadly referred to as an “Extended Synthesis”, referring back to the “Modern Synthesis” and the formation of neo-Darwinism; see e.g. Pigliucci & Müller, 2010) has invigorated a class of biological theories\(^\text{11}\) that maintain a grounding in quantitative theorizing but that also has deep similarities with \emph{qualitative} theories in the social sciences. In the context of our thesis, what is particularly interesting is that they have clearly begun tackling what we call wickedness, and not least that it addresses the question of how such systems evolve and how they are generated and maintained. This opens up an interesting prospect for exchange between the biological and social sciences on a different level than what has formerly been the case; see for example the application of such theories to cultural evolution by Andersson, Törnberg, and Törnberg (2014).

\textbf{Acknowledgements}

\(^{10}\) Practitioners of mainstream complexity science would take strong issue with being thus labeled, and to a large extent it is all a matter of what we compare with. If we compare with natural science and quantitative social science, they do take issue with fundamental reductionist assumptions by the very fact that they accept and systematically study emergence. On comparison with more qualitative approaches, however, their focus on studying emergence from the perspective of basic organizational levels means that they tend to disregard multi-level organization and still come across as reductionist.

\(^{11}\) E.g. Developmental Systems Theory (e.g. Oyama, Griffiths, & Gray, 2001), Generative Entrenchment (e.g. Winsatt & Griesemer, 2007) and Niche Construction (e.g. Odling-Smee, Laland, & Feldman, 2003)
We wish to acknowledge the support of the EU-FET grants MD (no. 284625) and INSITE (no. 271574). The paper has benefited from discussions with David Lane, Björn Sandén, Duncan Kushnir and several others at our department.

References


Paper II
Paper III
An Evolutionary Developmental Approach to Cultural Evolution

by Claes Andersson, Anton Törnberg, and Petter Törnberg

Evolutionary developmental theories in biology see the processes and organization of organisms as crucial for understanding the dynamic behavior of organic evolution. Darwinian forces are seen as necessary but not sufficient for explaining observed evolutionary patterns. We here propose that the same arguments apply with even greater force to culture vis-à-vis cultural evolution. In order not to argue entirely in the abstract, we demonstrate the proposed approach by combining a set of different models into a provisional synthetic theory and by applying this theory to a number of short case studies. What emerges is a set of concepts and models that allow us to consider entirely new types of explanations for the evolution of cultures. For example, we see how feedback relations—both within societies and between societies and their ecological environment—have the power to shape evolutionary history in profound ways. The ambition here is not to produce a definitive statement on what such a theory should look like but rather to propose a starting point along with an argumentation and demonstration of its potential.

Introduction

Highly schematically one can speak of two dominant approaches to Paleolithic culture: (i) an ecological/economical approach (EA) that is largely based on behavioral ecology and economical constraints like time consumption, portability, and so on; (ii) a physiological approach (PA) that emphasizes physiology, in particular cognition, as an enabler and constraint on culture (see also Foley and Lahr 2003 for a similar factorization). Neither approach makes sense on its own, and controversy typically concerns the relative importance of the forces that they represent. Both moreover share a common neo-Darwinian model of adaptation where natural selection is seen as the sole provider of evolutionary direction; to the extent that other factors enter into consideration, they sit uneasily on the margins of this framework.1

The PA is pervasive but rarely championed explicitly; Richard Klein typically serves as its embodiment (and lightning rod) in the literature (Klein 1995, 2000; Klein and Edgar 2002). Its persistence stems largely from that it is hard to see how a cognitive technological potential could have persisted untapped over extended periods of time. This logic dictates that technological stasis and transitions must be tied to cognitive capabilities that, due to strong selection for the tools and strategies that they afford, would rapidly come to realize their highest expressions. These capabilities are typically envisioned as providing distinct powers—such as the ability to conceive of higher dimensional structure, symbolism, multicomponent tools, and so on—that will leave distinct traces in the archaeological record. Transitions would then be true watershed events, triggered by the advent of new cognitive capabilities and marking the first appearance of new behavior and the last appearance of old obsolete behavior, with the best known example no doubt being the concept of "behavioral modernity" (see, e.g., Henshilwood and Marean 2003).

The EA is often contrasted to the more paleontological (in terms of both artifact and hominid taxonomy) approach of the PA. It emphasizes geographically and temporally varying environmental selection pressures, and cognition is seen more as a capability for variability than as a set of fixed capacities. The EA therefore maintains a much more generous comfort level to archaeology and palaeontology being "out of sync," such as when one hominid form exhibits considerable variability. Weiss (1992) noted that Paleolithic time is often marked by "waves of cultural invention"—a concept that certainly has appeal.2

We here propose that the same arguments apply with even greater force to culture vis-à-vis cultural evolution. In order not to argue entirely in the abstract, we demonstrate the proposed approach by combining a set of different models into a provisional synthetic theory and by applying this theory to a number of short case studies. What emerges is a set of concepts and models that allow us to consider entirely new types of explanations for the evolution of cultures. For example, we see how feedback relations—both within societies and between societies and their ecological environment—have the power to shape evolutionary history in profound ways. The ambition here is not to produce a definitive statement on what such a theory should look like but rather to propose a starting point along with an argumentation and demonstration of its potential.

---

1. Indeed, elements of developmental reasoning can be found throughout the paleoanthropology literature, and we will employ several of these elements; examples include the invention cascades model by Schiffer (2005), the broad spectrum hypothesis by Flannery (1969), the role and significance of modified lithic technology outlined by Davidson and McGrew (2005), and the evolution of home bases by Rolvand (2004).
ation in cultural style and complexity (e.g., within *H. sapiens*), or when several hominin forms share style and complexity (e.g., between *H. sapiens* and *H. neanderthalensis*). So since there is indeed a considerable “lack of sync” between hominin stones and bones, the PA has come to be seen by many as a Procrustean bed for archaeology, picking and choosing what appears to be of use for reconstructing an elegant hominin phylogeny.

To mention only a small selection of reviews and finds that contribute to the feeling that the PA is under siege empirically, McBrearty and Brooks (2000) and McBrearty (2007) consider a wide range of Middle Stone Age (MSA) reports of artifact types often taken to be diagnostic of the Upper Paleolithic (UP). In a recent review, Shea (2011) finds that the idea of Mode 4 technology being unique to some behaviorally modern UP/Later Stone Age (LSA) variant of *H. sapiens* has no archaeological support. McCall (2006) and Soriano et al. (2007) review the MSA Howiesons-Poort evidence concluding that it was a separate and recursive tradition not ancestral to UP technology. Hovers and Kuhn (2006) collect studies of intermittent “modernity” in the European Lower Paleolithic (LP) and Middle Paleolithic (MP). Schlebusch et al. (2012) revealed, based on genetic evidence, that the Khoe-San of southern Africa diverged as a group circa 100 kya (kya = 1,000 years ago), which is at least 50 kya before the UP/LSA and a presence of *H. sapiens* outside of Africa and the Levant. Some LP sites, moreover, give evidence of notable complexity in nonlithic culture, for example, Gesher Benot Yaaqov (Goren-Inbar 2011), Schöningen (Thieme 1997, 2005), and Bilzingsleben (Mania and Mania 2005; Mania et al. 1994). Finally, Bednarik (2011) documents a large collection of possible paleoart finds from the LP and onward.

But the EA has still not been able to muster a satisfying alternative that really absorbs and defuses the PA account. First of all, the expected covariation between technology and biotopes does not become clearly manifest until well into the UP (Bocquet-Appel and Tuffreau 2009; Stiner and Kuhn 2006). Second, its explanations hover dangerously close to what in evolutionary biology is known as the just-so-story format (Gould and Lewontin 1979); in arguing how localized phenomena can be seen as environmentally adaptive, highly specific hypotheses may be produced, but these may be blamed for offering little beyond a restatement of their own basic assumption: “because it was adaptive.” Cultural evolution appears as underdetermined as ever: sometimes changing dramatically for no apparent reason and sometimes sapping through substantial external stress without much change at all. But we suspect an even more fundamental reason behind this chronic lack of a resolution. Neo-Darwinian models are based on a priori given selection pressures, which basically reduces adaptation to optimization. This general model of how evolutionary change happens is firmly entrenched and shapes questions and answers alike. It leads us into seeing processes as uninteresting transients, into assuming that change must be due to exogenous triggers, and into thinking that any and all features of an adapted system should match some feature of an environment that is external and autonomous. With no disagreement on this level between the EA and PA perspectives, there is full tacit agreement that there really should be a perfect match between potential and expressed cultural capability. When there is not, the resolution will be that we are wrong either about the capabilities or about the environment. Although culture frequently takes center stage, it is itself notably absent from the stage as a factor in its own evolution; there is no room for explanations having to do with how culture is organized, how it develops, and how it interacts with other processes in the neo-Darwinian framework (see also Sterelny 2011). The evolutionary constraints that the PA and EA represent are clearly supremely relevant, but could the whole underlying model of adaptation be misguided?

The present work represents the view that the evolution of culture is channeled by endogenous developmental mechanisms and an intermingling between cultural and noncultural processes. By this, culture does not only become a crucially important factor in its own evolution, it also becomes coupled with processes unfolding over both longer and shorter scales of time and space. We argue that we need to develop what we call an evolutionary developmental approach if we are to understand the origins and history of culture, and we view our effort as a direct parallel of the corresponding approach to evolutionary biology (see, e.g., Arthur 2011; Odling-Smee et al. 2003; Oyama, Griffiths, and Gray 2001).

The ultimate purpose of this paper is to advocate an evolutionary developmental approach to studying socio-technical evolution. To do this, we will develop a provisional evolutionary developmental theory by combining a set of complementary models that all illuminate innovation processes in different ways and on different levels of organization. We will apply this theory to a set of brief case studies. The aim is not to provide a final statement but rather to propose a starting point and a way of being more concrete than what an argument entirely in the abstract would have been.

We begin by motivating the need for such a theory in general. We then introduce the three models that we will adapt and combine, developing a synthesis as we go along. The concepts introduced are next elucidated using two short case studies: the first focusing on organization and the second on process. This is followed by a discussion about transitions, stasis, and adaptation, and why the record of lithic technology can be expected to poorly represent hominin capabilities and adaptability.

2. Equilibrium models based on such an optimization logic can certainly be highly useful. What we take issue with is the (often tacit) implication that it usefully characterizes the overall evolutionary process.
The Evolutionary Development of Sociotechnical Organization: A Synthetic Model

Why We Need Evolutionary Developmental Models

Early *Homo* was the first animal as far as we know that had in its possession a piece of knowledge that could not be biologically inherited and that it simply could not afford to lose. But summoning culture to our service was a Faustian bargain: its tremendous powers came at the price of eternal servitude. It also set the stage for an extensive coevolution between hominid culture, ecological niche, and physiology. Socially and materially transmitted culture truly became a part of us and we of it; it is shaped around us, and we around it, just like any organ that happens to be a material part of our bodies.

The relation between ourselves, culture, and our environment mirrors that between heritable material, organism, and environment, and the latter set of relations has been explored by evolutionary developmental theories in biology. Evolutionary developmental biology (see, e.g., Arthur 2011) tells us that variation, and thereby evolution, cannot be understood at the genetic level—in particular, not if we are interested in evolution in the long run. Put simply, variation in phenotype must happen via variation in developmental trajectories, so development strongly determines what can and cannot arise, and while genes play a central role in development, their role is highly complex and cannot be disentangled from the rest of the developmental system. Niche construction theory (see, e.g., Laland et al. 2007; Odling-Smee et al. 2003) goes even further and breaks apart the neo-Darwinian delimitation between what adapts and what gets adapted to. It sees organisms as not just adapted to their ecological environments but also as adapting their environments to themselves. In the end, it becomes impossible to decouple these processes from each other. This position and approach to studying evolutionary biology are also represented in a more generalized form in developmental systems theory (see, e.g., Griffiths and Gray 1994; Oyama et al. 2001).

These arguments apply even more strongly for culture, where social structure, institutions, artifacts, and environmental modifications constitute an adapted system that is also the engine of its own adaptation. As evolution unfolds, this apparatus is all the time present: it catalyzes, scaffolds, and gets modified by evolutionary change both internally and externally. Ideas, cultural representations, memes, routines, or whatever we want to call them here, play a role roughly analogous to genes (although not quite; see, e.g., Anderson 2008, 2011b), which is to say that they are necessary but only in a wider developmental context. So just like proponents of evolutionary developmental theories argue that we must understand ontogeny and the interplay between organisms and their environment in order to understand organic evolution, we argue that we must understand the processes and structures of culture as well as how they interact with noncultural processes in order to understand the evolution of culture.

These developmental theories challenge the neo-Darwinian view of evolutionary change in ways that we believe are at least as important for understanding how culture evolves. We think that many of the current paradoxes and enigmas in hominid evolution may be theoretical artifacts rather than true features of the system under study; a simple change of perspectives may turn some of these paradoxes into workable problems or even dissolve them (see also Sterelny 2011). For example, on a developmental view, the absence of notable external change no longer implies stasis, nor the other way around. Questions of the nature-or-nurture type are no longer posed on either a binary either-or form or a quantitative form; saying that something is 70% due to genetic factors makes as little sense as saying that the operation of an internal combustion engine is 70% due to the fuel and 30% to its mechanisms. Also, large effects do not necessarily call for large causes, and the original causes of large effects may well be mundane, idiosyncratic, and probably impossible to find out anyway. Besides, they are no longer even the most interesting part of the story. What is really interesting are the processes and organization that often stabilize the system but that sometimes amplify small events into landslides. These processes combine environmental, physiological, and cultural factors, and they are weaved together in bootstrapping processes that largely feed on themselves; exogenous events and processes of course remain important, but they are not the sole sources of explanations.

If there is a need to frame Darwinism differently in evolutionary biology, it would hardly be surprising if it needed to be framed differently in paleoanthropology too. But we cannot use biological theories as they stand to understand culture. In fact, biological theories of this sort will mostly be used as a source of overarching inspiration. We will mostly use elements from innovation research set in present times, where the processes and organization of societies are there studied together in a way that suits our purposes. What we do here is select three models that each cover an important part of the dynamics that we think should be better understood in cultural evolution. Together they become more than their sum; they have to be readapted from their original areas of application, and we need to stitch them together into a coherent framework. We will do this as we introduce them.

**Exaptive Bootstrapping**

First out is what we call the exaptive bootstrapping model (see Lane 2002, 2011; Lane and Maxfield 2005; Lane et al. 2009). Innovation processes are conceptualized as transformation processes in an agent-artifact space where any transformation potentially triggers a reconfiguration of the entire fabric. New configurations of agents may change how artifacts...
are used and trigger changes in the artifacts. These changes in turn have an impact on agents’ roles and interactions. Five steps in what we call the internal innovation cycle are identified: (i) new artifact types are designed to achieve some particular attribution of functionality; (ii) organizational transformations are constructed to proliferate the use of tokens of the new type; (iii) new patterns of human interaction emerge around these artifacts in use; (iv) new attributions of functionality are generated to describe what the participants are obtaining or might obtain from them; (v) new artifacts are conceived and designed to instantiate the new attributed functionality (Lane 2011, 69). Since points i and v coincide, we have a bootstrapping process that, since each cycle may lead to more than one new cycle, is prone to cascading.

Exaptation (Gould and Vrba 1982)—the principle by which an artifact can be put to uses other than those originally intended—enters between steps iii and iv and is identified as a major mode of innovation (see also, e.g., Davidson and McGrew 2005; Rolland 2004), where the expactive process is described but not named). The internal innovation cycle brings out the dichotomy of radical versus incremental innovation, where the former involves changes in function while the latter is of a faster-better-cheaper type. The latter is less prone to cascading, since it does not directly disrupt configurations in agent-artifact space.

To further illustrate how internal innovation cycles can play themselves out, let us consider the application of culturally transmitted strategies. Application of cultural knowledge entails a production of situational knowledge that adapts and combines strategies to local conditions. Most such knowledge is for immediate use and discard (Lorentz 1977). Some will, however, be candidates for generalization and incorporation into the culturally transmitted package. So by simply using culturally carried knowledge, possible improvements in the zone of proximal development (Vygotsky 1978) will reveal themselves to the practitioners (Schiffer 2005). This process is necessary for adaptation to changing circumstances to be possible: retaining what appears to work, removing what appears not to work, and solving problems. Each addition will furthermore reveal new sets of possible novelty, calling for even more change in a cascading mode of growth.

Davidson and McGrew (2005) explain how lithic technology may be crucial in this respect, since it lends itself to general usage of exceptionally wide applicability, such as cutting, scraping, and piercing. Apart from affording the potential to branch out into many types of more specialized tools, lithic tools may also be used for fashioning (and from the early MP/MSA become part of; see, e.g., Mazza et al. 2006; Rots and Van Peer 2006; Rots et al. 2011) secondary implements, giving a tremendous leverage on their range of uses and on their propensity for generating open-ended skill systems. Around these implements, new patterns of use may appear, paving the way for ever more addition and change. Just as important in this respect may be the rarity of suitable raw material, as evidenced by its conscious transportation in the Oldowan (e.g., Braus et al. 2008; Stout et al. 2010). Even without diversifying or specializing function, there was a consistent potential for sophistication through the entire production chain, from raw material procurement to tool use: higher quality raw material, caching of raw materials and tools, exchange, sharing, control over the shapes of produced flakes, retouch of worn tools to renew or change properties, and so on (for a review of how lithic technological organization is currently viewed, see Andréfsky 2009).

But there are also innovation processes mediated by environmental and physiological systems. If we look at the internal innovation cycle as analogous to evolutionary developmental biology processes, the external innovation cycle can be seen as the sociotechnical counterpart of niche construction.

Let us consider a hypothesis about the MP-UP transition along lines similar to the broad spectrum revolution hypothesis (Flannery 1969; Stiner et al. 2000) and mutual exclusion between cultural groups (O’Connell 2006). The LP and MP saw a development from a broad opportunistic subsistence base (e.g., Steele 2010) toward a more narrow subsistence focused on big game predation (e.g., Hoffecker 2009; Pérez-Pérez et al. 2003; Villa and Lenoir 2009). With the UP, this development is followed by a new broadening of the subsistence base, where big game hunting is complemented by a much wider array of low-level resources (see, e.g., Hoffecker 2009; Richards and Trinkaus 2009; Richards et al. 2000; Stiner et al. 2000). This development can be tied to the emergence of higher technological variability in UP technology (see, e.g., Hoffecker 2009; Kuhn and Stiner 2001; Stiner and Kuhn 2006), and it can be suspected to produce an ecological ratchet effect (see fig. 1). A broader subsistence base reduces risk and increases the ecological carrying capacity, which leads to higher population density. This would in turn inevitably lead to smaller areas per capita for subsistence and to the destruction of the previous niche of relying more heavily on big game. So there is no going back from such innovations, and for each step that is taken the pressure to take another step in the same direction will remain strong.

External innovation cycles may continually have operated on any scale and via any external system that can act as a ratchet, most importantly physiological adaptation such as changed cognitive capabilities. Its main preconditions may have been around for a long time, and its trigger would be some kind of novelty behind which an unknown amount of generativity happens to hide. Such a portal innovation may be entirely inconspicuous since its generativity is not intrinsic to it: each step creates the conditions for new steps and acts in various ways to block backsliding. It is thought, for example, that modified lithics arose in relation to large mammal

4. It might however also cause cascades under certain circumstances, e.g., via threshold effects. If the cost of some practice drops, it may suddenly become useful in applications that were formerly uneconomical; this would lead to exaptation and functional change.
scavenging (e.g., Heinzelin et al. 1999; Semaw et al. 2003; Domínguez-Rodrigo et al. 2005); the future generativity of this technology millions of years down the line obviously played no role in its original adoption and development.

### Generative Entrenchment

The generative entrenchment model (Wimsatt 1999; Wimsatt and Griesemer 2007) was originally devised in evolutionary biology as a model of how new features always assume the presence of, and adapt to, earlier features. If exaptive bootstrapping describes the explosive side of innovation, generative entrenchment describes how basically the same processes cause the dynamics to jam up to generate powerful dynamical lock-in effects. The central insight is that apart from the fitness of the novelty itself, we must also consider the fitness of the cascade of reconfigurations that must take place for novelty to fit into an existing and fine-tuned system. The more fundamental the locus of change, the more other things will be affected by the change, so the larger will the cascade effects be and the higher the likelihood that it will bring about maladaptive effects that cancel whatever benefit that the original novelty brought. This tends to lead—as it has in the organization of genetic networks (Shubin et al. 1997), biological body plans (Arthur 2011), and modern society (Wimsatt 2013)—to a hierarchy whose low-level parts have general and abstract functions that rarely change and whose high-level parts are progressively more adaptively flexible.

Generative entrenchment brings out a second spectrum of innovation types: deep versus superficial innovation. Deep innovation affects lower parts of the hierarchy, tending to trigger larger cascades. These will be rare but have dramatic and sweeping effects. Superficial innovation affects higher and more flexible parts and tends to generate smaller cascades. They will be more frequent and less disruptive. Radical innovation is therefore progressively less likely to enter into the system the more fundamental the affected component is. Substantive refinement of function through incremental innovation will however be present in fundamental components since these are likely to remain for a long time in a stable context. In effect, this means that the bottom levels will be selected for exaptability through generality: the wider the range of functionalities that they afford in dependent parts further up in the hierarchy, the more adaptable will the system as a whole be.

### The Multilevel Perspective

Also addressing the proneness of cultures to lock-in is the multilevel perspective (see Geels 2010; Geels and Kemp 2012; Geels and Schot 2007), which focuses on what conditions that need to obtain for fundamental transitions to take place. Different areas of activity—say, storage, food preparation, domestic fire handling, construction of dwelling structures, foraging, hunting, raw material procurement, and so on—would here be referred to as sociotechnical regimes organized into a...
sociotechnical system. Regimes evolve to become specialized and separated to reflect differences in function. But at the same time they will become entangled and coordinated: horizontally because they are all part of a larger project of maintaining the well-being of the society and vertically because they will rely on and share more entrenched and fundamental technology and practices. Regimes will come to serve certain functions in relation to one another, and by fitting well into the sociotechnical system, important and well-connected regimes have an “internal fitness” that can trump even substantial “external fitness” of challenging regimes. Regimes will also be resilient to radical changes in their function for the exact same reason. Most innovation will be incremental and channeled into the regimes, essentially making them better, faster, and cheaper while preserving their function. Moreover, the hierarchical arrangement described by generative entrenchment will apply both within and between regimes: some parts of them will be more free to change while others will be more entrenched, and in the same way, some regimes will be more dynamical and others more static. The sociotechnical system, along with all other relevant aspects of the environment (other groups, ecology, climate, etc.), collectively form what is referred to in the multilevel perspective as a landscape. Changes and shifts in this landscape are generated both within and outside of the sociotechnical system, often gradually and over time. Since the landscape contains the regimes, it will pull these, as it were, and generate stress in their configurations. As stress builds, rifts may form in the sociotechnical system as regimes come to fit in poorly, reducing their internal fitness, and offering windows of opportunity for fundamental change. Whether or not a transition really will take place depends on the availability of challenging minor regimes.” Minor regimes fill important but more specialized roles, and they serve an important role in the innovation dynamics because they sometimes turn out to be exaptable into more major roles. Modern examples include things like specialized types of propulsion, materials, energy production, and so on in niche applications (e.g., military, space) where ad hoc inventiveness is called for, or for that matter major regimes unique to certain countries, cultures, and so on. The reason why regime-level radical innovation needs a long period of relative isolation as a minor regime is that whole new systems of practices do not appear overnight and a minor regime must have developed quite a bit before it can hope to take on even a crippled major regime.

Brief Demonstrations
We will here employ the concepts that we have introduced to analyze two brief case studies. The primary purpose is to demonstrate the applicability of the framework to historical processes more in general, to test whether the framework promises to have any bite as a method for integrating and analyzing a heterogeneous collection of stylized facts.

The Synthetic Model in Context
Let us consider (mostly) the organization of Acheulean communities from the Gesher Benot Yaaqov (GBY) site (~800 kya) as described by Goren-Inbar (2011). Rich assemblages are found in 15 horizons covering a time period of approximately 100 ky, and they provide evidence on a level of resolution that is rare for sites of such antiquity.

Two major regimes can be particularly well established: (i) raw material procurement and (ii) lithic production. GBY hominids primarily used basalt as raw material in their giant core Acheulean lithic technology. The basalt was quarried from nearby locations whose selection demonstrates intimate knowledge of rock properties relevant for tool production. Slabs were reduced to smaller pieces using methods out of which at least two have left persistent traces: large rock percussion and levers, the latter being inferred from characteristic notches present on slabs. Many parts of the quarrying practices cannot be reconstructed, but it is clear that quarrying must have been a collective activity. What emerges is a coherent sociotechnical regime consisting of a system of multiple types of artifacts, individuals, knowledge, and sites, and whose continuity can be confirmed over a time span of 50 ky (Sharon et al. 2011). Since it can be inferred, as Goren-Inbar (2011) does by considering an ethnographic case (Pétrequin and Pétrequin 1993), that considerable detailed knowledge and non-trivial strategies were needed, this continuity indicates that the regime was passed on across generations with great fidelity.

We can safely infer that the quarrying regime was particularly strongly entangled with the tool production regime. This regime can also be confirmed over a span of 50 ky and was no less based on transmitted expert knowledge. Both soft and hard percussors were used, and a high level of knowledge and control is evidenced by the fact that despite great variability in blank properties, the resulting tools were remarkably uniform. Indicative of transmission is also a substantial similarity with African Acheulean tool types and morphology of the same era. Moreover, Goren-Inbar (2011) notes that basalt is hard to process, and basalt knapping is considered to represent the highest level of expertise among ethnographically studied biface knappers in Irian Jaya (Pétrequin and Pétrequin 1993).

We also see the base of a hierarchy: although the higher
level regimes leave few persistent traces, the low-level lithic tools were obviously used in several roles. Some direct uses of lithic tools, all of which can be interpreted as representing sociotechnical regimes, have however been documented, such as nut cracking (Goren-Inbar et al. 2002) and butchering (e.g., Rabino-
vich et al. 2008). There is also evidence of wood-
working (Goren-Inbar et al. 1992), which strongly suggests a general material processing use of basic lithic tools; this has also been verified for other Acheulean sites, such as in Schön-
ingen (~400 kya; Thieme 1997) and Penin (~1.5 mya [million years ago]; Domínguez-Rodrigo et al. 2001). At Schöningen, wooden javelins were found along with cutting tools and retouching tools, indicating both the existence, connection, and something about the internal structures of three regimes: (i) a tool production regime; (ii) a higher level woodworking regime supported by the tool regime, and (iii) a big game hunting regime supported by the woodworking regime. Im-
plied are, for example, raw material provision regimes and a carcass-processing regime.

The GBY site also provides early evidence of persistent domestication of fire (Goren-Inbar et al. 2004), predating widespread use by several hundred millennia (Roebroeks and Villa 2011; Rolland 2004). The same goes for the systematic use of fish as a source of nutrition, which is seen at the GBY site long before this becomes commonplace in the Upper Paleolithic. Fire use appears likely to have been a local major regime that was strongly linked to many other practices, while the exploitation of fish more likely represents a locally minor regime complementing more major subsistence strategies. Ex-
ceptions like these are the rule in the archaeological record. For instance, despite the strong emphasis on big game hunting during the MP, numerous examples exist of exploitation of small and hard-to-catch resources (see Gaudzinski-Windheu-
der and Niven 2009), which is not to speak of the MSA, which is replete with examples of local styles, ornaments, and tech-
nologies (McBrearty 2007; McBrearty and Brooks 2000). Al-
though there are no particular reasons to believe that the GBY instances were ancestral to much later universalized regimes, they demonstrate the existence of developed practices that constitute minor regimes that could act as fodder for radical innovation.

Early Human Sociotechnical Organization: Flexibility and Conservatism

This case study will follow hominin organic and sociotech-
nical evolution through the emergence of a rudimentary hunter-gatherer lifestyle and the spread of erectoids across the Old World. The path that we have chosen through the thicket of archaeological and paleontological evidence (and their inter-
pretations) is one that emphasizes the interplay between diet, physiology, ecology, and the generativity of modified lithic technology.

The emergence of the Oldowan around 2.6 mya (Semaw et al. 2003) and its spread across east and south Africa would be the first step toward establishing what we might call a human sociotechnical organization. But it is likely that the organization of the overall Oldowan sociotechnical system would still essentially be that of the great apes, with a set of quite separate regime systems rather than a hierarchy of re-
gimes. The reason why this can be suspected is that butchering is the only verified use of Oldowan tools before the end of that period. This suggests that the full generativity of lithics (see the “Exaptive Bootstrapping” sec.) does not appear to have come into play yet at this point. While there exist in-
dications of uses of lithics besides butchering in the early Oldowan (Beyries 1993), the evidence is tenuous (Domí-

nuez-Rodrigo et al. 2005). So it appears likely that if such uses existed, they would have constituted minor regimes—
paving the way for, but not yet constituting, a more complex organization that lay ahead.

We propose that the human sociotechnical organization originated in the realization that lithics afforded manipulation of a wide range of materials besides animal carcasses. This general affordance provided exceptional leverage in terms of the range of implements that could be made (Davidson and McGrew 2005). This is an example of what we mean by a portal innovation: its generative effects on the sociotechnical system cannot have been part of the reason why it was initially adopted into a general role. Modified lithics would sneak in under and transform virtually all activity that relied on pro-
cessed materials, and in the process it would also over time spawn new such activities that had previously been unthink-
able; e.g., spears,hafted points, big game hunting, hide work-
ing. The scene was set for a major reorganization of the sociotechnical system: from something resembling the or-
ganization of great ape technology to a hierarchy with a strongly entrenched base and more and more flexible regimes upward.

Signs of such a shift appear toward the end of the Oldowan and the beginning of the Acheulean. The earliest direct sign is evidence found in late east African Oldowan tools (Beyries 1993; Keeley and Toth 1981) of the application of lithics to a wider variety of materials, and in Acheulean tools of similar age (Domínguez-Rodrigo et al. 2001). Another sign suggestive of an external innovation cycle is concurrent transformations involving physiology, cultural organization, and ecological in-
teractions. The shift between H. habilis and H. ergaster brings many physiological changes, such as an increase in body mass, encephalization, a reduction in masticatory musculature and postcanine tooth size, and an adaptation to a more modern mobility pattern with an essentially modern postcranial anat-
omy and thermoregulation (Plummer 2004; Wood and Strait 2004). It also brings evidence of behavioral changes including food sharing, change in land use patterns, and in general the emergence of a rudimentary hunter-gatherer lifestyle (see, e.g., Snodgrass et al. 2009 for a review). The first hominids found outside of Africa (at Dmanisi, Georgia, ∼1.75 mya; Vekua et al. 2002) are also dated approximately to this period, and H. erectus is subsequently found all over the Old World in a wide
range of biotopes. The thesis that a new sociotechnical organization can have emerged that provided key abilities from early on is buttressed (i) by the fact that Dmanisi tools are Oldowan in character rather than Acheulean, (ii) by the primitive physiology of the Dmanisi hominids (they were not fully developed erectoids; see Yekua et al. 2002), and (iii) by the low frequency of bifaces in east Asia and the generally low affinity to western Acheulean bifaces in those that have been found (Norton et al. 2006; Pettaglia and Shipton 2008). This pattern would then be consistent with a migration initiated early during the external innovation cycle. The characteristic Acheulean bifaces would be an effect as much as a cause of this transition.

The early human sociotechnical organization certainly witnessed substantial flexibility relative to ecology and climate, but notably no trend toward diversification in resource exploitation. To the contrary, as mentioned in the Exaptive Bootstrapping section, the trend was toward narrowing and climbing higher in the trophic system, and the Neanderthals finally occupied a niche at the very top, typically specializing heavily on one or a few species (see, e.g., Gaudzinski 2006; Gaudzinski-Windheuser and Niven 2009; Kuhn and Stiner 2006; Stiner 2002). A logical development would be to follow the path of refining exploitation strategies for the best yielding resources that were already targeted and for which considerable sociotechnical adaptation had been achieved: that is, to move to larger game and toward being able to target them as efficiently as possible, that is, fat-rich prime adults rather than only young and infirm animals. The early modern sociotechnical organization would be hierarchical and exhibit a system of linked regimes, but they would primarily be subservient to an ever-narrowing set of top-level regimes by which the groups would live and die. Few of these, perhaps (at least locally) only one, would be directed at obtaining food. The sociotechnical system developed to maximize output and minimize risk, but it did so path-dependently and constrained by its own organization. The sociotechnical organization of the UP/LSA, which we might have called the early human sociotechnical organization if we were to deal with it here, would be tempting to characterize as a break with the pattern of homing in more and more closely on bigger and bigger prey to that of being able to home in on several targets at the same time and to be able to swiftly change targets as needed. This is clearly not possible to achieve by the incremental improvement of a big game hunting regime, and reliance on such a regime must have been exceptionally hard to break out of. Everything from technology to hunting practices, dwelling places, group structure and size, social interaction strategies (within and between groups), and conceptions of value and identity would be coadapted to the occupation of this niche and contribute to a powerful lock-in.

**Entrenchment, Entanglement, Hierarchy, and Lock-In**

**Archaeological Homogeneity**

Taphonomic processes have left us almost exclusively with two inert parts of ancient hominids to pore over: their skeletons and their small lithic technology. But if lithic technology mainly, and in particular early on, belonged to basic and strongly entrenched regimes, such a bias could play tremendous tricks on us today. Adaptability and inventiveness would largely be archaeologically invisible, and rare signs of behavioral complexity beyond what is evidenced by lithics would seem either paradoxical or too fantastic to even consider seriously.7 It would only be at the rare points in time when innovation affected the lithic core technology that clearly detectable traces of sociotechnical change would be left behind. During the prolonged periods of deep lock-in, lithics would remain relatively static, and the archaeological record will be uneventful and uniform despite considerable climatic and ecological variations across time and space (see, e.g., Bocquet-Appel and Tuffreau 2009; Foley and Lahr 2003; Kuhn and Stiner 2001; Stout et al. 2010). Most lithic technology may simply not have been in direct contact with geographically and temporally variable parts of the environment but have interacted on the basic level of materials such as meat, wood, straw, hides, bone, antler, and so on. Inferring cultural stasis and homogeneity on the basis of stasis and homogeneity in lithics could in other words be akin to arguing that mice and elephants must be similar because of a high similarity on the cellular level—we would compare them on a level that is strongly generically entrenched and that simply is not the main locus of variability. Indeed, with respect to raw material availability, a point at which lithic skills were strongly and directly exposed to environmental variability, considerable variation does occur over time and space (see Kuhn and Stiner 2001).

**Regime Change: Adaptation and Revolution**

Due to functional entanglement and specialization, regimes cannot easily undergo radical innovation even when put under substantial pressure. This pressure, however, may open the door for the replacement of one regime by another regime (see “The Multilevel Perspective” sec.). At the rare points when this successfully took place in the lower parts of the sociotechnical system, cascading (see the “Exaptive Bootstrapping” sec.) reconfiguration could lead to revolutionary change; when it occurred in the upper parts, the overall system would be less disturbed. The multilevel perspective cites two main preconditions for regime replacement: (i) a window of opportunity whereby one or more regimes begin to fit in poorly and (ii) readily deployable alternatives to such regimes,

7 For example, sealing H. erectus (Bednarik 2003, 2011).
that is, minor regimes. If either of these two preconditions is lacking, status quo will remain (e.g., Geels and Schot 2007).

The basic mechanics of window-of-opportunity formation can be summarized as follows. Stress builds up as sociotechnical regimes strive to configure themselves to match the external environment, horizontally between the regimes, and vertically to the locked-in core parts of the system. Over time, stress may reach a point where internal adaptation between parts of the system begins to lag considerably. Strong selection pressures and new affordances will come to persist without being effectively pursued, and regimes will find it increasingly hard to accommodate needed changes. The internal fitness that a major regime enjoys as a central and well-integrated part of the sociocultural system will be diminished, and we may speak of a window of opportunity opening up since external selection pressures now have stronger relative importance.

The most common type of window was probably due to changes in the relative availability of resources. We expect that the sociotechnical system was adapted to deal expeditiously precisely with that sort of change through superficial reconfigurations. But over longer time scales, drawn-out directional climatic change occurred, not least due to glaciation cycles, and substantial organic evolution unfolded. This could slowly open up deep rifts in the sociotechnical system, beyond what could be dealt with by superficial reorganization. As stress increased, windows would open up more and more frequently and widely, with deeper and more frequent reorganizations leaving detectable but still isolated and recursive archaeological traces behind. Then, eventually, a deep radical innovation would turn out to be a portal innovation and come to sweep the board. This would be what we now detect as a shift between the major technological modes.

As suggested in the Archaeological Homogeneity section, radical innovation seems not to have been the weapon of choice for adaptation. The entrenched core regimes would instead have been under pressure to undergo incremental innovation to become as configurable as possible so as to make the top tiers more flexible and thereby reduce the need to change the core. What were the evolutionary dynamics of core regimes? Delagnes and Meignen (2006), for example, trace the fate of different debitage methods (Levallois, discoidal, and Quina) across the MP in France, and although the Quina method appears late, the pattern conforms well to what Kuhn (2006) finds in a similar study of the Italian MP, namely that “flexibility seems to have been manifest mainly through redeploying and recombining a limited range of technological options, not as the development of entirely new technological solutions.” Kuhn (2006, 116) furthermore notes that whatever temporal trajectories can be discerned are in any event not in the direction of the UP.

The top of the hierarchy was the organ of external adaptation in the sociotechnical system, and most regime shifts would thereby be superficial and quite uneventful. Regime replacement, we hold, is realistic only if a suitable regime that could be redirected without too much deep modification already existed. Developing a functional regime from scratch under pressure would probably have had a very low likelihood of success: such a regime would be unrefined, as it would not represent accumulated time-tested knowledge, and those forced to begin practicing it would not be highly skilled. The presence of minor regimes was therefore likely to be highly important for adaptability, and mobility was probably the main strategy before sociotechnical flexibility became highly developed. As an entirely conjectural example (along with those in the sec. “The Synthetic Model in Context”), imagine that the bow and arrow emerged as a toy for children, or perhaps a display weapon, and underwent a good deal of functional refinement in that role, that is, that it was turned into a functioning but minor regime with practices, materials, and production routines worked out. Then, if one day conditions called for a broadening or alteration of the subsistence base, this technology would be available as an option and would not need to be conceived ex nihilo. With respect to minor niches, complexity breeds complexity: the larger and more diversified the culture, the more nooks and crannies will there be in which such alternative ways of doing things can thrive without being dominant.

Summary

In evolutionary biology, developmental models scaffold natural selection into a richer system of processes that arise from the organization that evolution itself brings about. Natural selection is seen as necessary but not sufficient for understanding the patterns of evolution. We have argued that the corresponding move should be made in the study of cultural evolution. To this end, we introduced a provisional synthetic theory where three components from innovation research—which is an area where sociotechnical organization and process have lately been studied extensively—were brought together under a more overarching logic borrowed from evolutionary developmental models from biology. The intention was to demonstrate that a theory tailor-made for the purposes can be developed. In some cases, concepts were immediately useful, but in other cases they must be customized, abstracted, and reapplied, and several “bridging concepts” had to be introduced.

The result should not be viewed as a ready and delivered framework, and the case studies should not be seen as its definite explanations and predictions. It should be viewed as a demonstration of a possibility and a starting point that is sufficiently developed to provide leads for further exploration.

We think that a suggestive picture of the evolutionary development of culture emerges as a result of this exercise, if so only in outline. Unless earlier hominid culture operated on radically different principles than that of more recent societies—and there are no reasons to believe that they did—we have strong reasons to believe that such cultures came to be organized hierarchically, with interlinked but compar-
m entalized and specialized regimes, subject to strong conser- vatism at the base and successively more flexibility toward the top.

We furthermore think that this organization was far more than just an adaptation to an environment; it opened up a whole universe of new possibilities for the future. The emergence of such an organization, which we have here suggested can have occurred at the Oldowan-Acheulean transition, could even be seen as a cultural version of the Cambrian Explosion and the emergence of multicellularity in the biological world. It also suggests a candidate differentiation between human culture from animal culture on the basis of organization.

We hypothesized that modified lithic technology would have played a crucial role in this development by establishing itself as the enabler and precondition—directly or indirectly—for just about all hominid cultural activities. By serving general material processing roles, it made secondary implements and strategies possible, and these would form the upper tiers of this hierarchy. With lithics entrenched at the base of the sociotechnical system—analogous to cells in multicellular organisms—they would not need to change very much. Their main design criterion would be to make change possible in the upper tiers. In other words, they would be adapted to allow flexibility, and we should not be surprised if they themselves were quite inert in the face of ecological heterogeneity. The normal mode of change in response to ecological heterogeneity would instead have been superficial reconfigurations in the upper levels of the sociotechnical system, and the system as a whole would be adapted to allow such flexibility. The mode of sociotechnical change over time would be characterized by lock-ins and bootstrapping processes: phenomena that we cannot even begin to consider unless we apply an evolutionary developmental perspective.

But due to changes in systems external to culture—not least physiological evolution and changes in ecology and climate—change also in the more fundamental parts of the sociotechnical system would be inevitable sooner or later. When such change did happen, it probably occurred in geographical areas where the environmental conditions were such that the threshold to such change locally happened to be lower. Once the lock-in was broken, we identified what was referred to as an external innovation cycle as a particularly important type of change process. The external innovation cycle is related to a lock-in roughly the way that positive self-reinforcement is related to negative self-reinforcement: it is basically the same mechanism operating in a different dynamical regime, and both are due to the nonlinearity of evolutionary processes. This innovation process is inspired by the logic of niche construction theory whereby also systems external to culture are related to negative self-reinforcement: it is basically the same mechanism operating in a different dynamical regime, and both are due to the nonlinearity of evolutionary processes. This innovation process is inspired by the logic of niche construction theory whereby also systems external to culture are

8. The editors of this journal may care to note that Pope and Brumm’s important paper was rejected from Current Anthropology.
things that turned into landslides: my favorites are the plano-
convex cores from Olduvai DK (Leakey 1971) or Peninj (de
la Torre et al. 2003) that could be rejuvenated and become
Levallois cores, and the fact that the intensification of the
exploitation of gazelles in the Natufian was followed by the
domestication of goats, which had been a minor element of
the faunal exploitation previously (see data in, e.g., Munro
2004).

It is, therefore, paradoxical that the authors rely on GBY
and the claim that its archaeology results from a single culture
that lasted 50,000 years—longer than the whole time modern
humans have been the dominant species on the planet. But
the essence of culture is that it is transmitted from culture-
holders to the culturally naive and that in that process of
teaching and learning there is ample opportunity for what
Henrich (2004) described as failure in cultural copying. Here,
I think, the explanation lies in a little-emphasized aspect of the,
surely inappropriate, comparison with toolmaking in
modern Papua New Guinea: the raw material is described as
hard-to-process basalt. Here then is a context in which equi-
finality might arise from the limited ways in which the raw
material can be worked—hence the uniformity may not be
due to inexplicably perfect and long-lasting cultural trans-
mission but a convergence on a method that works, perhaps
on many discrete occasions, including ultimately recent Papua
New Guinea. The attempted analogy actually works the other
way around.

Erella Hovers and Anna Belfer-Cohen
Institute of Archaeology, The Hebrew University of Jerusalem,
Jerusalem 91905, Israel (hovers@mscc.huji.ac.il, belferac@mscc.huji .ac.il). 13 VI 13

Archaeology is notorious for being an empirical discipline
that borrows from many other, theory-driven branches of
research. We do not often see the archaeological record—a
proxy for cultural evolution (albeit a partial one, as stated
correctly by the authors)—being investigated by experts from
other disciplines, whose ambition is to offer “a starting point
. . . an argumentation and demonstration” for a new devel-
omental approach to sociotechnical evolution. By the au-
thors’ own admission, the formal approach advocated here
borrows elements from evolutionary, optimal foraging, and
general systems theories, as well as the recently developed
theory of niche construction, in order to make sense of Pa-
leolithic culture. It is precisely because archaeologists are com-
pulsive borrowers of theories (e.g., Binford 2001) that much
of this new offering had already been embraced, for better or
worse, by prehistorians.

The advocated approach attempts to account for gene-cul-
ture-environment interfaces, but it fails to consider how par-
ticular cultural histories influence group and individual de-
cision making when interacting with their environment
(physical and social) at a given time. The authors claim that
a powerful sociotechnical lock-in, the emphasis on meat eat-
ing, had been operating and growing stronger from the Ol-
dowan to the Middle Paleolithic Neanderthals—perceived as
committed to carnivory. This time-transgressive intensifica-
tion directed cultural evolution in all its aspects. Thus, lithics
“evolved” mainly to accommodate the drive of getting more
meat in a more efficient way. The authors consider the MP-
UP transition as a break with the constriction imposed by
homing in more and more closely on bigger and bigger prey,
and the appearance of behavioral flexibility.

This scenario is inconsistent with the archeological data
throughout the relevant time span. The Oldowan techno-
system is potentially the only aspect of the Oldowan that
differs from the great apes. Still, the invention of Oldowan
stone flaking is built on behaviors known among apes (Hovers
2012; Read and van der Leeuw 2008). It is for this reason
that we disagree with the statement that Oldowan stone tools
respond to meat-processing needs and are associated strictly
with butchering. Great ape tool use does not encompass
butchering (and in fact does not include intentional stone
flaking in the wild; McGrew 2010). Moreover, isotopic, tooth
microwear, and tooth calculus findings indicate that the diet
of hominins from the Oldowan through the Middle Paleo-
lithic (as well as later) was variegated, regionally diverse, and
far from portraying the strong commitment to meat eating
envisioned by the authors (Cerling et al. 2013; El Zaatari et
al. 2011; Henry, Brooks, and Piperno 2011; Hockett 2012;
Speth 2010; Ungar and Sponheimer 2011; Ungar 2012).

Finds from the Acheulean site at GBY indeed speak to
developed cultural practices of early hominins. We concur
that such finds need not necessarily be considered as ancestral
to much later universalized regimes. Indeed, such finds may
evidence the now widely accepted concept that even major
inventions might have disappeared and reappeared as part of
the multitude of processes that constitute the “evolution of
culture” (e.g., Henrich 2004; Hovers and Belfer-Cohen 2006;
Powell, Shennan, and Thomas 2009; Premo and Kuhn 2010;
Shennan 2001). How, then, could they serve as “fodder for
radical innovation” in much later times? Implicit in the au-
thors’ statement is an entrenched belief in linearity, contrary
to the main thrust of their approach. Their statement also
disregards the role of historical trajectories (e.g., context of
cultural transmission, demographies) that are proximal trig-
gers for the emergence and diversity of prehistoric cultural
innovations and institutions.

As archaeologists trained in lithic analysis, we are the first
to recognize the importance of lithic technology in under-
standing cultural evolution. Still, this is a far cry from the
reductive statement that lithic technology is “the enabler and
precondition . . . for just about all hominid cultural activities,”
a prime mover of all things. Perceiving of lithics as analogous
to cells in multicellular organisms defines culture and its prod-
ucts as just another biological organ. Biological evolution has
an important role in shaping the human condition, but only
up to a point. Culture, including its material manifestations, cannot be explained through biological models except on a very abstract level of discourse. This is probably because one has to take into consideration the fact that as complexity rises, it can change the rules of the game. "Culture", made humans "too clever by half" in the sense that our species acts contra its biological "good" (Dawkins 1989).

Andersson, Törnberg, and Törnberg made a commendable endeavor to test a theoretical model through archaeological case studies, yet it results in a "just so story." In spite of the emphasis on flexibility and connectivity of cultural processes and structures, the suggested model is rigid when it comes to explaining the processes inferred from the archaeological record. The authors came up with a typology of processes rather than explanatory insights. What is sorely needed is a record. The authors came up with a typology of processes to explaining the processes inferred from the archaeological and structures, the suggested model is rigid when it comes to explaining the processes inferred from the archaeological record. The authors came up with a typology of processes rather than explanatory insights. What is sorely needed is a middle-range theory that will bridge between valued theories and empirical archae-ological data on the other, and the historical circumstances that are responsible for cultural evolution.

Timothy A. Kohler
Department of Anthropology, Washington State University, Pull- man, Washington 99164-4910, U.S.A. (tako@wsu.edu). 19 VI 13

Andersson and colleagues offer a set of stimulating and valuable suggestions for archaeologists seeking to explain innovation in the archaeological record from an evolutionary perspective. I also appreciate their general argument that cultural evolution ordinarily interacts significantly with development (the "evo-devo" perspective) and should be analyzed in that way.

It seems to me, however, that their account misunderstands some aspects of current and recent directions by archaeolo- gists employing evolutionary processes in an explanatory fashion. These directions have developed largely in the Anglo-American literature, drawing inspiration primarily from biology and population genetics, whereas the approach taken by Andersson et al. seems to stem ultimately from a systems approach (among several) would be to consider semantic in- formation in brains or external storage devices (e.g., parietal art) to be "like" the genotype and the eventual expression of such acquired information (through behavior, language, art- tifacts, etc.) in some environment to be "like" the phenotype. Then "cultural evo-devo" would study how the interaction between semantic information and the natural and social environment shapes behaviors and artifacts and ultimately, at the population level, the cultural evolutionary trajectory. As in DIT approaches, the focus here is on the transmitted in- formation. This in fact seems to be the approach taken here as well: "Ideas, cultural representations, memes, routines, or whatever we want to call them here, play a role roughly anal- ogous to genes."
As Andersson et al. continue to develop these ideas, I would urge them to better integrate their perspectives with existing approaches to evolutionary archaeology and to be more aware of intellectual antecedents (e.g., does Flannery's [1968] attention to feedback processes not invalidate their claim to have discovered in feedback relations “entirely new types of explanations for the evolution of cultures”?); to be more explicit about the structures of the populations being discussed (e.g., are intergroup competition and cultural group selection the likely motor for the innovations they seek to understand?); and to consider the implications of their arguments for the study of causation in evolving sociotechnical systems (see, e.g., Laland et al. 2011 on the necessity of employing reciprocal causation to understand the behavior of such systems, providing a logical underpinning for employment of concepts such as niche construction theory).

Claes Andersson and his coauthors argue that traditional models of cultural evolution need to be expanded to include elements of “evolutionary developmental theories” in order to overcome the limitations of what they perceive as simplistic applications of external selective dynamics in our understanding of cultural and technological evolution. They take their inspiration from similar debates within evolutionary biology that have, over the last 3 decades, led to an expansion and transformation of evolutionary theory. As a result, parts of evolutionary theory—the newly emerging fields of developmental evolution and evolutionary developmental biology (evo-devo), two conceptually distinct, albeit related approaches—are now also focused on the origin of variation as a distinct explanatory problem, one that requires understanding the mechanisms of individual development as part of any explanation of evolution.

One of the central questions for developmental evolution has been the problem of innovation. In evolutionary biology, innovation mostly refers to novel phenotypes, often those connected to body plan evolution, that are not simply variations of already existing characters. The main challenge has been to explain such phenotypic innovations in light of highly conserved genes and especially gene numbers. Simply put, comparative genomics has falsified the deceptively elegant assumption that new phenotypic characters are simply the product of new genes (remnants of this view are still widespread in form of the “gene for” ideology). The answer to the problem of innovation is now found within the complex regulatory architecture of the genome and its role in mechanistically explaining development. In most basic terms, increasing evidence suggests that (often small) changes in underlying gene regulatory networks can explain the emergence of new characters (innovations), while redundancy in and conservation of specific parts of these networks at the same time explains the long-term stability of other phenotypic features, such as body plans.

Current developmental evolution has a well-developed research program for studying phenotypic innovations, or more precisely, inventions. Taking a lead from the economic theory of innovation, developmental evolutionists have also adopted Schumperter’s distinction between invention and innovation, where invention refers to the developmental origin of a novel phenotype, whereas innovations are those inventions that successfully spread though populations and leave a phylogenetic signal. The latter then connect considerations of internal developmental mechanisms related to the origin of (novel) variation with the external dynamics of natural selection.

But in the case of innovations or evolutionary novelties, these external dynamics are not simply those of natural selection. An innovation or a novel phenotypic character by definition opens up a new niche or changes the dimensionality of existing niches. Therefore, the new evolutionary theory of innovation combines developmental mechanisms that can explain the origin of inventions with those processes that describe the constructive interactions of organisms with their environment and with each other (niche construction, natural selection).

As this brief overview of trends in current evolutionary theory shows, Andersson et al.’s approach is grounded in a broader framework of theoretical approaches that aim to understand innovation from a complex adaptive systems perspective. Their paper thus points to a larger theoretical question: Is it possible to develop a unified theory of innovation across domains that is based on insights derived from perspectives of developmental evolution, niche construction, and complex adaptive systems?

In my view there is reason to be optimistic, but we have a long way to go. Studies like Andersson et al.’s that try to frame case studies about lithic technology within the context of developmental evolution models, such as the external innovation cycle of cultural evolution, reveal the usefulness of these novel perspectives beyond the narrow domain of re-wiring gene regulatory networks and phenotypic evolution. Research that analyzes patent data, institutional diversity, the development and evolution of cities, or the dynamics of knowledge systems all point in the same direction, namely that innovation can be defined as specific shifts in the state space of complex systems at the intersection of internal regulatory and external dynamics.

So where does this leave us? For once, this initial definition of innovation is very abstract. Therefore, we need more empirical case studies so that we can see to what degree innovation dynamics are similar and different within and across domains. Based on these descriptive cases, we also need to build predictive models that can be tested. It is one thing to frame cases of innovation within a specific theoretical point
Cross-disciplinary endeavors such as the one reported here are to be applauded, but there is of course inherent risk, especially if one is entering somewhat unfamiliar territory. This is where we see Anderson and colleagues, who, in an effort to promote their evolutionary developmental approach to cultural evolution, have either overlooked or ignored an enormous body of literature that already views culture in the way they suggest: as a system of descent with modification mediated by the pillars of inheritance, variation, and sorting. This oversight has led them to build a straw man and to unknowingly reinvent a first-generation cultural evolutionary wheel.

The authors, for example, are nowhere near the first behavioral scientists to examine culture from an "evo-devo" perspective (e.g., Mesoudi and O'Brien 2008; O'Brien et al. 2010), and although they mention in passing niche construction theory (NCT), they apparently are unfamiliar with the literally dozens of articles and journal issues devoted to NCT and culture (e.g., Kendal, Tehrani, and Odling-Smee 2011; Laland, Odling-Sme, and Myles 2010; O'Brien and Laland 2012). The myriad issues covered there are critical for the kind of argument Anderson and colleagues are trying to make: that, as Lewontin (1983:280) famously put it, "organisms do not adapt to their environments; they construct them out of the bits and pieces of the external world."

Unfamiliarity with the literature, both biological and cultural, leads Anderson and colleagues to state that in evolutionary biology, "Natural selection is seen as necessary but not sufficient for understanding the patterns of evolution. We have argued that the corresponding move should be made in the study of cultural evolution." We have two comments. First, in evolutionary biology, natural selection is not seen as necessary. Rather, what is necessary is a mechanism that sorts among variants. That mechanism could be, and often is, selection, but it could also be stochastic sorting, or drift, which is perfectly capable of "causing" evolution. Second, decades of work on cultural evolution stemming from the efforts of Cavalli-Sforza and Feldman (1981) and Boyd and Richerson (1985), among others, have taken account of not only the important role played by stochastic sorting but myriad other aspects of a Darwinian cultural evolution. Importantly, cultural evolution is not merely an analogy to, or metaphor for, biological descent with modification but rather an evolutionary process mediated by variation, inheritance, and sorting (Lycett 2011).

Of significance in the context of Andersson and colleagues' discussion, archaeological studies have been at the forefront of much of this work (e.g., Bettinger and Eerkens 1999; Eerkens and Lipo 2005; Lyman and O'Brien 1998; Mesoudi and O'Brien 2009; Neiman 1995; O'Brien and Lyman 2000, 2002; Shennan 2002), including those studies dealing specifically with issues of variation within entities such as "Acheulean" (e.g., Kempe, Lycett, and Mesoudi 2012; Lycett 2008, 2011), which is the subject of Anderson and colleagues' case studies. Curiously, the authors cite several papers by Steven Kuhn, and even thank him for "valuable discussions and feedback," but they ignore his insightful paper "Evolutionary Perspectives on Technology and Technological Change" (Kuhn 2004).

Returning to the authors' claim that selection is the only visible process in cultural evolutionary research, they note that the two dominant approaches to the study of Paleolithic culture—one ecological/economic and the other physiological—"share a common neo-Darwinian model of adaptation where natural selection is seen as the sole provider of evolutionary direction." This simply is untrue, whether the subject is the Old World Paleolithic or the New World Paleolithic. If it were, why do we find articles with titles such as "An Experimental Test of the Accumulated Copying Error Model of Cultural Mutation for Acheulean Handaxe Size" (Kempe, Lycett, and Mesoudi 2012) or "Acheulean Variation and Selection: Does Handaxe Symmetry Fit Neutral Expectations?" (Lycett 2008) or "The Accumulation of Stochastic Copying Errors Causes Drift in Culturally Transmitted Technologies: Quantifying Clovis Evolutionary Dynamics" (Hamilton and Buchanan 2009)?

The last title brings us to a critical component that is all but missing from Anderson and colleagues’ discussion, namely, cultural transmission, which is the vehicle of cultural inheritance. Ignoring the vast literature on transmission allows the authors to state that culture is "notably absent from the stage as a factor in its own evolution; there is no room for explanations having to do with how culture is organized, how it develops, and how it interacts with other processes in the neo-Darwinian framework." Again, this is an untrue and naive statement. The authors might look, for example, at work on the organization and evolution of cumulative culture (e.g., Enquist, Ghianda, and Eriksson 2011; Tomasello 1999). And, even a cursory glance at any of the many works on NCT and culture will completely undermine the notion that culture is "notably absent from the stage as a factor in its own evolution."
biological evolution by including nongenetic traits transmit-
correctly that cultural evolution defined as an extension of
jectories having the same beginning point? Some have argued
radically from the trajectory leading to
question: Why did the trajectory leading to
and trait selection alone, still unanswered is a fundamental
lutionary process than is obtained through focusing on traits
these provide a richer and more complete picture of the evo-
erative entrenchment, and a multilevel perspective. As the
authors note, none is specific to cultural evolution. Though
exaptive bootstrapping, the external innovation cycle, gen-
erative random mutation process to “innovation that allows
transformation changed innovation from an externally ex-
genous random mutation process to “innovation that allows
innovation evolution that fundamentally redefined what con-
ominin evolution that fundamentally redefined what con-
strained the evolutionary process with regard to the cultural
side of Homo sapiens” (Read 2012; Read et al. 2009). The
transformation changed innovation from an externally ex-
genous random mutation process to “innovation that allows
innovation innovation” (Read et al. 2009). This “innovation innovation” (Read et al.
reversed the previous pattern of functionality at the group level emerging from functionality at the individual
level, to functionality at the individual level being derived
from functionality introduced at the group level.

We can see the reversal in the development during the
Upper Paleolithic of “an external cognitive architecture by
which hominins achieved social extension within local groups
and a wider community” (Gamble 2010:32), thereby tran-
sceding individually framed, cognitive abilities through
group-level organization of individual cognitive abilities. The
“cognitive architecture” enabling this social extension derives
from the “culturally constructed systems of kinship” (that)
provide the basis for all the other culturally based forms of
social organization that arose with modern Homo sapiens” (Leaf and Read 2012:19). Culturally constructed kinship sys-
tems whose organization is expressed linguistically through a
kinship terminology enabled social relations to be extended
ted socially and having an impact on behavior makes the
difference.

The cultural side of Homo sapiens, though, is not deter-
mined through social transmission. As the authors comment,
in the neo-Darwinian framework there is “no room for ex-
planations having to do with how culture is organized, how
it develops, and how it interacts with other processes” (see
also Lane et al. 2009; Wimsatt and Griesemer 2007). Lacking
is a critical innovation transformation introduced during
ominin evolution that fundamentally redefined what con-
stitutes the evolutionary process with regard to the cultural
side of Homo sapiens” (Read 2012; Read et al. 2009). The
transformation changed innovation from an externally ex-
genous random mutation process to “innovation that allows
innovation innovation” (Read et al. 2009). This “innovation innovation” (Read et al.
reversed the previous pattern of functionality at the group level emerging from functionality at the individual
level, to functionality at the individual level being derived
from functionality introduced at the group level.

We can see the reversal in the development during the
Upper Paleolithic of “an external cognitive architecture by
which hominins achieved social extension within local groups
and a wider community” (Gamble 2010:32), thereby tran-
sceding individually framed, cognitive abilities through
group-level organization of individual cognitive abilities. The
“cognitive architecture” enabling this social extension derives
from the “culturally constructed systems of kinship” (that)
provide the basis for all the other culturally based forms of
social organization that arose with modern Homo sapiens” (Leaf and Read 2012:19). Culturally constructed kinship sys-
tems whose organization is expressed linguistically through a
kinship terminology enabled social relations to be extended

The authors identify several processes central to the evo-
lution story applied to culture: the internal innovation cycle, exaptive bootstrapping, the external innovation cycle, gen-
erative entrenchment, and a multilevel perspective. As the
authors note, none is specific to cultural evolution. Though
these provide a richer and more complete picture of the evo-
lutionary process than is obtained through focusing on traits
and trait selection alone, still unanswered is a fundamental
question: Why did the trajectory leading to Homo diverge rad-
cally from the trajectory leading to Pan, despite both tra-
jectories having the same beginning point? Some have argued
incorrectly that cultural evolution defined as an extension of
biological evolution by including nongenetic traits transmit-

Table 1. Expansion and extension of the Darwinian model

<table>
<thead>
<tr>
<th>Darwinian model</th>
<th>Modern synthesis</th>
<th>Expanded synthesis</th>
<th>Social/cultural extended synthesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Variation</td>
<td>Gene mutation</td>
<td>Evo-devo theory</td>
<td>Cultural idea systems</td>
</tr>
<tr>
<td>Inheritance</td>
<td>Mendelian inheritanc</td>
<td>Plasticity and accommodation</td>
<td>Language and other symbolic forms of communication</td>
</tr>
<tr>
<td>Natural selection</td>
<td>Population genetics</td>
<td>Niche construction</td>
<td>Generative grammars</td>
</tr>
<tr>
<td>Contingency</td>
<td>Epistemic inheritanc</td>
<td>Replicator theory</td>
<td>Enculturation</td>
</tr>
<tr>
<td>Speciation and trends</td>
<td>Genomic evolution</td>
<td>Evolvability</td>
<td>Cultural models</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multilevel selection</td>
<td>Intentionality and purposive behavior</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Genomic evolution</td>
<td>Recursive reasoning</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Imitation/phenotypic inheritance</td>
<td>Theory of mind</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Social cognition*</td>
<td>Technological systems</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Social behavior*</td>
<td>Role systems</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Organization systems</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Symbolic inheritance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Invention and innovation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Social boundary</td>
</tr>
</tbody>
</table>

Note. Cols. 1–3: derived from figure 1.1 in Pigliucci and Müller (2010a). Col. 4: proposed extension addressing cultural and social evolution.

* Additions to Pigliucci and Müller’s expanded synthesis.

Dwight W. Read
Department of Anthropology and Department of Statistics, University of California, Los Angeles, California 90095, U.S.A. (dread@anthro.ucla.edu). 6 VI 13

The authors credibly integrate some of the new ideas of the expanded synthesis of evolutionary theory (Pigliucci and Müller 2010b)—ideas that have expanded the modern synthesis (Mayr and Provine 1980: figure 1.1)—into a developmental approach for explicating cultural evolution. Relevant to their argument, the expanded synthesis can be extended to encompass ideas relating to the evolution of social/cultural systems (table 1, col. 4).

Whereas the modern synthesis focused primarily on gene evolution through mutation, inheritance, and natural selection, the expanded synthesis has focused on endogenous processes affecting the development and expression of traits, not just their selection as optimal solutions to externally imposed change. The authors suggest that accounts of cultural evolution should focus similarly on endogenous processes relating to the development and formation of cultural phenomena. The goal is laudable; the means proposed for doing so are incomplete.

The authors identify several processes central to the evolution story applied to culture: the internal innovation cycle, exaptive bootstrapping, the external innovation cycle, generative entrenchment, and a multilevel perspective. As the authors note, none is specific to cultural evolution. Though these provide a richer and more complete picture of the evolutionary process than is obtained through focusing on traits and trait selection alone, still unanswered is a fundamental question: Why did the trajectory leading to Homo diverge radically from the trajectory leading to Pan, despite both trajectories having the same beginning point? Some have argued incorrectly that cultural evolution defined as an extension of biological evolution by including nongenetic traits transmit-
in time and space beyond the local group, and the boundary of a community was thereby no longer limited by the scope of face-to-face interaction, as is the case with the nonhuman primates (Read 2012). Instead, the social system was transformed into a relation-based form of social organization expressed linguistically through a kinship terminology system (Read 2012). The culturally defined kinship terminology system provides the foundation for the social organization of hunter-gatherer societies from which more extensive forms of social organization have evolved.

The kinship terminology system neither emerges from patterned behavior of individuals (Leaf and Read 2012:16) nor provides functionality except through the group level: individually knowing a kinship terminology provides no functionality to that individual. Instead, functionality for the individual arises from a group collectively having and sharing a kinship terminology system, thus reversing the sequence for the expression of functionality implied by neo-Darwinian evolution. As a consequence, the social boundary for small-scale societies is determined by those who can mutually determine they are kin, using the kinship terminology as a symbolic computational system (Read 2001, 2007); hence the boundary became the consequence of an internal, rather than an external, process, in the manner discussed by the authors. The reversal in the expression of functionality implies that cultural evolution is not derived from evolution of individual traits, genetic or otherwise, but from evolutionary processes acting on the structure and organization of cultural idea systems (Leaf and Read 2012:14).

A Story Better Told Elsewhere

Andersson et al.’s critique of what they call the ecological/economic and physiological approaches to Paleolithic culture are well taken. In the case of humans, understanding our past certainly depends upon an evolutionary theory of culture, as they argue. However, they seem unaware of a large literature on cultural evolution and gene-culture coevolution making the same point. It goes back to papers in the 1950s and 1960s (Alchian 1950; Campbell 1960; Gerard, Rapoport, and Kluckhohn 1956). Mathematical modeling of the processes of cultural evolution and gene-culture coevolution began in the 1970s, and major synthetic work was published in the 1980s (Boyd and Richerson 1985; Cavalli-Sforza and Feldman 1981; Lumsden and Wilson 1981). For a recent book-length review of the current state of the field, see Mesoudi (2011). Archaeologists and paleoanthropologists have been significant contributors (e.g., Bettinger 1991; Eerkens and Lipo 2005; Powell, Shennan, and Thomas 2009; Shennan 2002; Tostevin 2013). The ability of geneticists to understand something of our evolutionary and demographic history has turned up many genes that are candidates to have been involved in episodes of culture-led gene-culture coevolution (Laland, Odling-Smee, and Myles 2010; Richerson and Boyd 2010).

Reply

We wish to thank the commenters for their valuable feedback and hope that in this brief response we will be able to clarify at least some of the issues that were raised. In retrospect, we regret that our decision to take out a section on dual-inheritance theory (DIT) appears to have been a source of some irritation and misunderstanding. O’Brien and Lycett, for example, come away with the impression that what we suggest is that culture ought to be viewed “as a system of descent with modification mediated by the pillars of inheritance, variation, and sorting.” No wonder that they think we are reinventing the wheel! Kohler summarizes this critique in a constructive way: “Andersson et al. might have considered developing their ideas within the context provided by DIT. At a minimum, they should give clear reasons for not doing so.” We will here try to do the latter, and by that to at least partly do the former by explaining how DIT is related to the proposed approach. Then we will move on to more specific comments.

The choice overall was to contextualize our story more with debates about transitions and large-scale patterns than with debates about the modeling of microlevel mechanisms. We see DIT as a valuable source of models, but our approach for understanding the evolutionary process as a whole is very different. As Laubichler argues, our work is to be understood precisely as part of a greater project of understanding innovation in complex adaptive systems—and by innovation we mean radical rather than just gradual innovation.

DIT focuses strongly on population thinking—whether it is selection, drift, or other sorting mechanisms that are in action—as the key to the whole story of evolution. Let us first stress that we have nothing against population thinking—we think it is essential for understanding evolution, and it interests us greatly (see, e.g., Andersson 2011). But population thinking is basically all that DIT can deal with, and we think that more is needed. Boyd and Richerson (2005), for example, describe DIT as a combination of population genetics and rational choice theory, and we believe that this does sum it up quite well. DIT never strays far from this set of basic models, which also embody what is seen as the essence of what it means to be scientific in the first place. These basic tools bring analytical power, but they also bring strong constraints and biases in terms of what types of questions that can be addressed. Radical innovation—that is, transformation...
of social structure through qualitative novelty—is something that we think is absolutely central for understanding cultural evolution, and it is not in the set of things that DIT does a good job of modeling. Here we need also organization thinking (see Lane et al. 2009; Read and Lane 2008), although we think Kohler is right that systemic perspectives have tended to deemphasize population thinking too much. We do not agree, however, with Kohler’s comment that “DIT explicitly deals with ‘how culture is organized, how it develops, and how it interacts with other processes.’”

We are certainly of the opinion that bottom-up modeling is necessary for understanding cultural evolution—how would we otherwise be able to understand emergence in complex and massively parallel dynamical systems? But we do not think that the microlevel can be seen as prior to higher levels of organization. Microdynamics generates high-level organization—but high-level organization also scaffolds microdynamics so the arrow does not just point from the bottom and up. Consequently, we think that it is important not to postpone the task of getting the macrosopic patterns right for an undefined future (that we think will never arrive) where they would be explained from the microlevel. Getting this right, we think, demands being more methodologically synthetic and broad than is the case with DIT.

As suggested by Richerson and Bettinger, let us consider Alex Mesoudi’s (2011) book *Cultural Evolution* as a recent and reasonably representative statement by a major contributor to the DIT tradition. We first note that Mesoudi covers evolutionary developmental theory in about one page and uses the term in a much narrower sense than we do—which includes a host of partly aligned traditions that pay attention to structure on meso and macro levels; see also table 1 in Read’s comment. We concede that using the term the way we do, even if we specify that we use it as an umbrella term. It is also instructive to see how Mesoudi (2011) differentiates between a general Darwinian and a neo-Darwinian account of cultural evolution. He does this in a way that is very different from ours and that illustrates why DIT researchers probably find our account confusing: they may not agree with us that DIT is neo-Darwinian. Mesoudi’s points include: whether evolution is Darwinian or Lamarckian, whether transmission is particulate or not, and whether variation is blind or not. These are classical issues that have been debated—in our opinion, with little obvious progress and consequence—in the context of universal Darwinism, evolutionary economics, and DIT for a long time; we have participated to some extent in that debate (e.g., Andersson 2008, 2011b).

Mesoudi concludes that cultural evolution is primarily Darwinian rather than neo-Darwinian. These points of differentiation, however, represent more incremental adjustments to the same formalist framework. The rise of developmental thinking, by contrast, signifies a break with the view that evolution can be understood on a single level of organisms or genes, that is, solely in terms of population thinking; see also Laubiichler’s comment. While DIT importantly does challenge a number of standard neo-Darwinian assumptions, such as by introducing frequency-dependent fitness and other selection biases, the picture that Mesoudi paints (and that is amply evident in the literature) is still fundamentally neo-Darwinian on points that we see as much more significant than those listed above.

We are well aware that—as O’Brien and Lycett point out—some of the elements that we consider to be part of evolutionary developmental theorizing—not least niche construction theory (NCT)—have been used previously in the context of cultural evolution. But it is certainly true that the bulk of the applications of NCT to culture has concerned how culture affects genes. This is not without importance to us, but what is more central to our argument is how culture affects culture. That is, pointing the reciprocal arrow of causation in NCT back from culture to culture. This is something that both Odling-Smee and Laland (personal communication, 2013) see as an interesting possible development of NCT that has not been much pursued so far.

The points raised by Hovers and Belfer-Cohen about our rendition of the MP-UP transition illustrate another effect of strong competition for space in this paper. This is important to respond to for two reasons. The first is that we maintain that although other resources besides meat were clearly used during the MP (e.g., El Zaatari et al. 2011; Henry, Brooks, and Piperno 2011), we find no evidence that the strong reliance on a big game meat diet during the MP (and indeed well past it) as such would be in question. MP strategies for utilizing plant resources do not appear to come close either to MP hunting strategies in complexity or to plant-processing strategies in the UP (Piperno et al. 2004; Revedin et al. 2010). We would be inclined to interpret these as minor regimes, while big game hunting would be the single major food-getting regime. So although a more thorough case study would go more into such details, we think that the main message still stands.

The second is the question of how to evaluate the case study examples. They are not unique predictions of the framework, and the framework could easily be used for constructing very different accounts—just like any Darwinian framework can. There is a lot of room for disagreeing on how history played itself out. What we advocate is a new way of conceptualizing the evolution of culture—not a single model of what happened, although of course we have our favorite ideas, subject to revision over time in the face of evidence.

This brings us to the issue of whether this is really a new way of looking at cultural evolution. We readily admit that we need to better integrate intellectual antecedents such as Flannery. This includes also some of the recent work that O’Brien and Lycett point us to, where O’Brien and others, 10. And partly to our hypothesized role of lithics, although we do not see how their conclusion about the role of lithics flows from their statement about its origins.
for example, apply Michael Schiffer’s models (e.g., O’Brien and Bentley 2011). Schiffer’s work has inspired us greatly, but we do need to integrate this further. We are thereby certainly not the first to find new ways of understanding cultural evolution through feedback processes, and older systems theories are treasure troves for our purposes. But there is an important difference between systems theories and complex-systems theories. Emergence—in the sense of macroscopic features being irreducible (in practice or in principle; e.g., Bedau 1997, 2003; Chalmers 2006)—is central to systems theories and complex-systems theories alike. But while systems theories posit emergence, complex-systems theories explicitly study emergence. Systems theories predefine systems as “connected boxes” and address the dynamics of these systems while complex-systems theories ask how these “boxes” and their connections appear and change over time. We think that combining these approaches to systems will be necessary, and that is the direction in which we would like to go.

Davidson raises the question of whether the GBY really represents a single culture that lasted across 50 ky. It is of course very hard to have an intuition for whether that is reasonable or not, but it does raise some interesting questions. We think three points are worth mentioning in this context. One is that parts of culture could be very persistent, while other parts may be highly ephemeral. More specifically, some of the “bottom parts” that we argue would be genetically entrenched we think could survive both 50 and 500 ky and much longer. We agree that there is ample room for transmission failure and would like to additionally point to our own work on precisely that issue (Andersson 2011a, 2013).

The second point is that the capacity of earlier hominins for maintaining culture appears to have been consistently underestimated. Recent research about the cultural capacity of great apes suggests rethinking the minimal cultural capabilities of early hominins. To take only one example, Mercader et al. (2007) archaeologically trace chimpanzee nut cracking back 4,300 years; 4,400 surely is not 500,000, but 4,300 is a lower bound, and H. erectus were not chimpanzees.

The third point is that the issue of culture and transmission is typically posed as whether features are maintained using transmission or whether they are externally stabilized (e.g., by properties of the raw material and hominin physiology) and thereby cannot be properly described as cultural; this pertains also to a comment by O’Brien and Lyckett. But we think one factor does not rule out the other. Say that limitations apply on the capacity for transmitting cultural knowledge (e.g., Andersson 2011a; Henrich 2004), would it not be reasonable that artifacts that are strongly externally scaffolded would be attractive since they would obtain a lot of stability “for free” in terms of transmission? In our view, culture is not just transmission—it is a whole system that stabilizes itself by whatever means and that we could just as well describe as autopoietic as being based on transmission.

We agree with Read that we did not address the question of why Homo and Pan diverged culturally in such a drastic way, and we view this as an example of a case that should be pursued. We also agree that the question of group cohesion and the notion that Homo reinvented a way of maintaining large groups from scratch during the course of the Paleolithic are highly interesting, and we think that it falls well into the type of explanations that our framework is intended to help producing. Concerning the expanded synthesis: finding a suitable umbrella term is a problem, and we thank Read for pointing us to this table. The expanded synthesis, we would like to note, is not as well synthesized as the modern synthesis, and it contains methodological friction that the Modern Synthesis did not contain.

Laubichler’s comment complements the article well by making some of the arguments about the parallel biological case much better than we were able to do it. We fully agree that this is part of a greater project of understanding innovation in complex adaptive systems, and we think this is also what is so exciting about it: the interest for coming up with such a theory is very wide, and we believe that it is the logical next step that needs to be taken—rather than to keep holding our breath for a single-level bottom-up understanding to explain it all.

—Claes Andersson, Anton Törnberg, and Petter Törnberg

References Cited


Paper IV
Modelling free social spaces and the diffusion of social mobilization

Anton Törnberg & Petter Törnberg

To cite this article: Anton Törnberg & Petter Törnberg (2016): Modelling free social spaces and the diffusion of social mobilization, Social Movement Studies, DOI: 10.1080/14742837.2016.1266243

To link to this article: http://dx.doi.org/10.1080/14742837.2016.1266243

Published online: 21 Dec 2016.
Modelling free social spaces and the diffusion of social mobilization

Anton Törnberg* and Petter Törnbergb

*Department of Sociology and Work Science, University of Gothenburg, Gothenburg, Sweden; bDivision for Physical Resource Theory, Department of Energy and Environment, Complex Systems Group, Chalmers University of Technology, Gothenburg, Sweden

ABSTRACT
Free social spaces have long been emphasized in the social movement literature. Under names such as safe spaces, social havens, and counterpublics, they have been characterized as protective shelters against prevailing hegemonic ideologies and as hubs for the diffusion of ideas and ideologies. However, the vast literature on these spaces has predominantly focused on internal dynamics and processes, thus neglecting how they relate to the diffusion of collective mobilization. Inspired by formal modeling in collective action research, we develop a network model to investigate how the structural properties of free social spaces impact the diffusion of collective mobilization. Our results show that the assumption of clustering is enough for structural effects to emerge, and that clustering furthermore interacts synergistically with political deviance. This indicates that it is not only internal dynamics that play a role in the relevance of free social spaces for collective action. Our approach also illustrates how formal modeling can deepen our understanding of diffusion processes in collective mobilizations through analysis of emergent structural effects.

Far more than a place to eat and drink, the nineteenth-century Parisian working-class café offered a diverse and changing working population a unique protected space, enabling the growth of a proletarian public sphere with its own order, structure, and rituals. In this sense, the cafes have been described as a potential bridge between the ordinary world and revolution and helped foster ‘a latent class consciousness’ that came to have large-scale political consequences (Haine, 1998). Similarly, the southern black churches, removed from white control and repression, were key in nurturing and sustaining southern civil rights protest. The church served as a meeting place where social protest could incubate and mature and where resistance against slavery was discussed, organized, and mobilized. In Islamic countries, Mosques and Bazaars have played a similar role, not only in the Arab spring, but also in earlier uprisings (Bennani-Chraibi & Fillieule, 2003). These social spaces were central for their communities, as they created possibilities to develop both a repertoire of tactics and strategies, but also to foster oppositional consciousness, protected from a climate of intense repression (Morris, 1986). Block clubs, tenant associations, theatres, union halls, student lounges, and separatist women groups, these are all examples of free spaces that in various ways seem to serve as clandestine embryos for mass mobilization and insurgency.

CONTACT Anton Törnberg anton.tornberg@gu.se

© 2016 Informa UK Limited, trading as Taylor & Francis Group
As these examples illustrate, there is hardly any doubt that preexisting social groups and counter-spaces have a large impact on mobilization for many reasons, not only by constituting a protective shelter against repression and the prevailing hegemonic ideologies in society, but also by comprising a hub for diffusion and circulation of ideas and ideologies. For this reason, the vital role and significance of such free social spaces have long been emphasized within the social movement literature, using concepts such as free social spaces (Evans, 1979), social movement communities (Buechler, 1990), and counterpublics (Fraser, 1990). The dynamics taking place within these spaces have been the matter of extensive analysis from a wide range of different perspectives. However, the literature shows that the role and relevance of these spaces have in general been emphasized primarily in relation to when and how they contribute to the development of collective identities, oppositional cultures, and collective action frames, and to what extent they contribute to generating a cultural challenge that precedes or accompanies political mobilization. In other words, while there has been extensive focus on the internal processes going on within such spaces, less is known about the role they play in relation to collective mobilization and the diffusion of social movements.

Using a diffusion perspective, this article takes a somewhat different approach by addressing the overarching question: What are the emergent network structural effects of free social spaces on the diffusion of mobilization? More specifically, we investigate the interesting, yet ambiguous, relation between ‘publicness’ and ‘privateness’ in free social spaces: How open, or respectively, how closed, should free social spaces be in order to most effectively spread their innovation to surrounding society? And how does this relate to how political or deviant the space is, compared to mainstream society? This represents a difficult balance between being open enough to make new allies and influence surrounding society, but at the same time, remaining sufficiently isolated from the hostile and powerful to maintain their function.

Using formal modeling and tools from social network analysis, we develop a model that provides tentative answers to these questions, for further empirical and theoretical exploration. With this approach, our hope is to contribute both to a better understanding of how free social spaces relate to open rebellion and the diffusion of social movements in society, but also to provide a contribution to the overall literature linking social movements and computational modeling. Elegant, stringent computer models may be published in mathematical journals, but in order to have any influence within the social sciences, a broader empirical and theoretical context is necessary. By explicitly relating to existing theory, we argue that models and formal methods can fulfill an important complementing purpose in theory development by enabling systematic investigation of dynamics that would be beyond the comprehension of our unaided intuition, thus constituting a way to deal with mass dynamics and emergence.

We begin by elaborating different concepts associated with free social spaces and provide a brief overview of this research field and some of the existing theoretical and methodological issues. Following this, we will turn to the issue of diffusion of social movements and begin to sketch out the role and function of free social spaces in processes of diffusion. Following from this, we schematically distinguish between two main functions of these spaces that are often emphasized in the literature. In this way, these sections serve as a literature overview and a way of rooting the model in the field, and also provide a theoretical rationale for the assumptions underlying the model. In the next part, we introduce a network model primarily based on two functions, and analyze the result. In the subsequent discussion, we zoom out and further discuss the results in relation to the literature, showing the relevance and implications for the specific field of inquiry. With this approach, we hope to illustrate the benefits of formal network analytical tools, and how these methods potentially can help contribute to further advancing the field.

Free social spaces

During the last decades, we have seen a general shift in emphasis in social theory from a focus on the dominant spaces of capitalist society toward the societal periphery to the marginal spaces used by
those located at the fringes of society (Hetherington, 2003). This tendency is also reflected in social movement theory, with an increased focus on how peripheral social groups and counter spaces relate to cultural resistance and the formation of alternative oppositional identities.

Evans (1979) was among the first to emphasize that limited public spaces are required for challengers to develop collective action frames. Evans coined the term free spaces, focusing on the role of autonomous spaces where collective identities, oppositional cultures, and collective action frames can develop. She thus laid the ground for an emerging research field focusing on these types of phenomena from various theoretical perspectives and using different terms, such as free social spaces (Fisher & Kling, 1987), social movement communities (Burchler, 1990), counterpublics (Fraser, 1990), protected spaces (Tétreault, 1993), safe spaces (Gamson, 1996), spatial preserves (Fantasia & Hirsch, 1995), havens (Hirsch, 1993), sequestered social sites (Scott, 1990), cultural laboratories, spheres of cultural autonomy (Taylor & Whittier, 1995), submerged networks (Melucci, 1989, 1996), and abeyance structures (Taylor, 1989).

As Polletta (1999) argues in a lucid review of the field, these are basically different words for the same thing, generally referring to:

- small-scale settings within a community or movement that are removed from the direct control of dominant groups, are voluntarily participated in, and generate the cultural challenge that precedes or accompanies political mobilization (Polletta, 1999, p. 3).

A more recent conceptualization can be found in Diani (2013), who inclusively refers to these as dense subcultural networks, distinguishing such spaces as a separate mode of coordination of collective action, aside from social movements, coalitions, and organizations. Without taking any position regarding potential additional connotations of these various terms, we will hereby use the term free social space to refer to all these kinds of spaces. These perspectives share a common view that the mobilization of protest is facilitated by a group’s ability to develop and maintain a set of beliefs and common cultural frames that contradicts those of dominant groups. Thus, these concepts share a focus on the ideas, beliefs, and the collectively shared grievance and frames of understanding that are viewed as central drivers for protests, and that autonomous spaces are central in these social processes.

However, while referring to basically the same phenomenon, scholars have focused on somewhat different aspects and functions of these spaces, partly because different empirical studies have analyzed their role in distinct social and political climates with varying degrees of repression. Theorists thus variously describe them as subcultures, communities, institutions, organizations, and associations, which indeed are different referents. Some scholars such as Boyle and Evans (1986, p. 18) represent them as physical spaces since, in their account, the term ‘suggests strongly an “objective”, physical dimension – the ways in which places are organized and connected, fragmented, and so forth’. Others, such as Scott (1990) argue that these kinds of spaces could just as well be constituted by linguistic codes and hidden symbols that evade the gazing eyes of the dominant and the powerful.

However, a common denominator in most of the literature is an often explicit and unilateral focus on the internal dynamics and the structures of these spaces. As a consequence of this, there is a vast amount of rich and comprehensive research focusing on what is going on inside these spaces in the form of, for example, processes of cultural framing (e.g. Boyle & Evans, 1986; Fantasia & Hirsch, 1995; Haine, 1998; Hirsch, 1990; Scott 1990). However, this poses the intriguing question of how these internal processes relate to open rebellion and the diffusion of social mobilization. Often, the concept of free social spaces is used as a link to converge structuralist models emphasizing networks, resources, and organizations with the focus of new social movement theorists on cultural processes. This means that the notion of free social spaces is often used to highlight the specifically cultural dimensions of prior networks, and thus tends to remain ‘a black box grafted onto a political opportunity model or a structural breakdown one’ (Polletta, 1999, p. 2).

Debates thus typically concern internal dynamics such as to what extent these spaces are actually independent from surrounding hegemonic ideologies and the influence of powerful groups and actors, or how different characteristics such as group size and intimacy affect the possibility of free expression.
and whether it is primarily the isolation or the specific cultural content that has the largest impact on the mobilizing potentials of free social spaces. Hence, while there is a wide consensus that free social spaces are relevant for mass mobilization, it is more ambiguous how or if cultural forms occurring within these spaces can turn into overt opposition, and the role they play in actually igniting insurrection.\footnote{There seems to be a tendency among scholars to argue that some form of external process, some ‘triggering event’, or certain political opportunities, is required to ignite overt mobilization or to create structural potential for mobilization. For example, Fantasia and Hirsch (1995) argue that acute social and political crises are required to set off mass mobilization, and thus assign free social spaces in themselves a relatively small role. Similarly, Scott (1990) stresses the necessity of a relaxation of surveillance/repression for infrapolitics to transform to overt defiance.}

In other words, we still seem to lack any clear conceptualization of the role such free social spaces play in relation to fast transitions and unexpected, sudden mobilization and the spread of ideas and behavior across and between groups and movements in society. As Polletta (1999, p. 8) argues, ‘the free space concept simply posits a "space" wherein those dynamics occur, without specifying how, why, and when certain patterns of relations produce full-scale mobilization rather than accommodation or unobtrusive resistance.’

We argue that one way of addressing these issues is to depart from the field of diffusion studies of social movements. This is a disciplinarily separate, but arguably theoretically adjacent, field of research studying the diffusion of social movements and mobilization in society that we argue could prove key to understanding the structural effects of the dense networks of these spaces in relation to open rebellion.

**Diffusion of social movements**

Students of social movements have long been interested in the way movements influence each other and spread from one site to another; in other words, how they diffuse. Actions in social movements affect other actions, and they should therefore not be understood as simply isolated independent responses of external conditions. Clearly, diffusion processes are involved and some scholars have even gone so far as arguing that diffusion is the central process of social movements (Myers & Oliver, 1998).

Diffusion has been highlighted from various perspectives and the diffusion literature has historically had a significant impact on the field of social movements, and increasingly so in recent years due to globalization and new communication technology (Chabot, 2010; Givan, Soule, & Roberts, 2010; Tarrow, 2005; Wood, 2012).

Diffusion is here understood in a broad, inclusive sense, as a ‘flow of social practices among actors within some larger system’ (Strang & Meyer, 1993, p. 488). A diffusion perspective has been developed and applied in various fields, and often in relative isolation from one another, including, for example, the spread of new technology, changes in political regimes, and within social movements.\footnote{The concept of diffusion can help to account for the wave-like character of protest cycles and the occurrence of what Tilly, Tilly, and Tilly (1975) called ‘rebellious centuries’, followed by long periods of tranquility.}

Diffusion processes are generally understood as complex, multidimensional processes, highly dependent on what is being diffused, and how the diffusion occurs (Givan et al., 2010; Wood, 2012). For instance, there are several different pathways of diffusion and objects can diffuse both through direct interpersonal networks and through indirect channels, such as mass media (for a discussion on this, see e.g. Chabot, 2010; Givan et al., 2010; Tarrow, 2005, 2010). Different pathways of diffusion have different consequences, and are also relevant for what kind of social practices are being diffused, for example: whether they require multiple contacts to spread, which is often referred to as complex contagion. This is generally the case, for example, when it comes to identities, behavior, or interpretive frames. Other diffusing objects, such as information and most diseases, might diffuse through single contact or through indirect ties, often referred to as simple contagion. Thus, since the process of diffusion depends on what is actually being diffused, it is necessary to distinguish between different kinds of diffusing objects. There are ways to formally separate between these types of diffusion processes using formal models, something we return to later.
To sum up, the application of diffusion theory and network approaches to social movements constitutes a well-established field that has lately received increased attention, much due to recent computing and software developments and new access to virtual networks and digital data (Crossley, 2007; Diani, 2011; Diani & McAdam, 2009; Krinsky & Crossley, 2014). However, while the theoretical understanding of the impact of social networks on mobilization has become more complex, most existing empirical studies in the field have focused particularly on how social networks and preexisting social ties affect individual behavior and whether individuals get involved in collective action or not (della Porta, 1988; Diani & Mische, 2015; Goodwin, 1997; Kitts, 2000; McAdam and Paulsen 1993; Macy, 2001; 2007). As Diani (2009, p. 200) has argued, network developments and new access to virtual networks and digital data (Crossley, 2007; Diani & McAdam, 2009; Krinsky & Crossley, 2014; Sandell, 1999; Tindall, 2015. For an exception: Diani 2015; Mische, 2008). As Diani (2003, p. 8) argues, ‘it is much rarer that the overall configuration of networks linking individual activists is assessed in order to evaluate the potential for collective action in a given collectivity’. Thus, how the structure of the network and concepts such as centralization, clusters, and density affects collective action is under-researched. Also, most existing network analytic studies of social movements tend to focus upon snapshots, capturing the network configuration at a specific moment in time, thus neglecting the dynamics of change (Stevenson & Crossley, 2014; Stohl, 2008).

This can probably partly be related to the difficulties of collecting detailed empirical data on whole populations of individual activists and their interactions, thus rendering most established empirical methods ineffective (Diani & Mische, 2015; Kitts, 2000).

One way of investigating these issues is to use formal network modeling. But while network approaches to social movements have indeed engaged with broader social movement theories and debates, there has been less penetration of formal network analytical tools and diffusion models into social movement research and theorizing (Diani, 2011; Myers & Oliver, 2008). When formal models have been applied, they have seldom been explicitly related to the overall literature on social movements, but have rather formed what Diani (2011) has called an internally well-connected, but isolated field. Many of the existing formal models depart from the large group problem in collective action (Olson, 1965), i.e. the question of why large group collective action is possible when free riding is the rational decision for each participant. Marwell and Oliver (1993) suggested that an answer to this question could be that collective action is initiated by a critical mass creating a ‘bandwagon effect’. The question then becomes where the critical mass itself comes from, which has been investigated using formal modeling (e.g. Centola, 2013; Chiang, 2007; Chwe, 1999; Gould, 1993; Kim & Bearman, 1997; Macy, 1990; Marwell, Oliver, & Prahl, 1988; Oliver & Marwell, 1988; Oliver, Marwell, & Teixeira, 1985).

We will position our contribution with regard to this field later in the paper, and will here just remark that while this research clearly relates to collective action in social movements, this tradition of research comes from a rather separate field and has had rather limited impact on social movement theory. A central reason for this is most likely differences in ontology and underlying assumptions (Myers & Oliver, 2008). As Emirbayer and Goodwin (1994, p. 1446) discuss, this also reflects a more general tendency within the social sciences that the often abstruse terminology and mathematical sophistication of network analysis and computational modeling have ‘prevented outsiders from venturing anywhere near it’. This has led to a lack of dialog between these different approaches, and in Emirbayer and Goodwin’s words: ‘a consequent impoverishment of their respective domains of social inquiry’.

**Free social spaces from a diffusion perspective**

Our intention here is to contextualize and motivate the use of formal modeling in a social movement context by adopting a diffusion perspective. Based on our brief review of the literature on free social spaces above, we argue that from a diffusion perspective, we can extract two main mechanisms or functions ascribed to these free social spaces. These functions are emphasized in most of the various perspectives and concepts relating to free social spaces, but have traditionally been conflated in the literature, rather than treated separately. (i) These spaces are often, to a varying degree, different or deviant from mainstream society; they constitute spaces where members can discuss collectively shared grievances and cultivate cultural frames that contradict those of surrounding society or dominant
groups. (ii) They constitute densely structured clusters in social movements, characterized by strong ties and high interconnectivity among the participants.

Thus, by departing from a diffusion perspective, we open up possibilities for using network modeling to simulate the network structural effects of free social spaces on the diffusion of social movements and mass mobilization in society. We argue that, in network terminology, the two functions of free social spaces can be operationalized as bias and clusterness. Bias is here defined as a predisposition for an individual to take part in a mobilization, thus capturing an aspect of the radical political position that is argued to characterize free social spaces, while clusterness refers to the degree of interconnectivity within the cluster. This operationalization allows us to separate or unpack these two properties of free social spaces that have previously been lumped together in the literature, and look at their different roles as well as their interaction.

Model description

The approach suggested here is to use formal network analytical tools to investigate whether the network structural aspects of free social spaces can be relevant in relation to diffusion. As is often the case in network modeling, focus lies on investigating properties of the system in order to find emergent effects on a more abstract level. The model presented here is simple. This also implies that the resulting findings are general, and potentially applicable to a wider range of phenomena – in practice, everything from formal organizations and movement communities to particle interaction and electricity grids.

While such generality perhaps becomes more opaque – and therefore conspicuous – with the use of formal models, we should remember that this applies also for any other piece of theorizing. For instance, concepts such as political opportunities, resource mobilization, brokerage, and emotions are in no way uniquely suited to account for social movements. Furthermore, the simplicity of the model is motivated by the fact that simple models are often argued to be more useful: they are more intuitive; their results are more likely to apply in specific cases; and they are easier to integrate into a sociological narrative mode of understanding. Even a cursory glance at the track records of this type of models also tells us that abstract and simple ones have been most successful within social scientific research. For example, the strength of the highly influential Schelling (1971) model of segregation lies not in any complicated representation of human agency, the incorporation of symbolic communication between agents, or an accurate implementation of the complex processes producing social segregation, but rather in that the simple assumptions made allow a deep understanding of the discovered dynamics. With a more complicated model, the conclusion would not have been as general and interesting, and the underlying causes would have been more difficult, or even impossible, to analyze (see e.g. Andersson, Törnberg, & Törnberg, 2014).

A possible objection could be that while simple models may be more useful, it would then be preferable to also draw abstract, broad theoretical conclusions from them, biting over multiple research fields. But the preference for abstract conclusions is exactly the reason that, as Diani (2011) argues, formal models have been rather separate from the social movement literature: it is not that their contributions are not wide enough to be relevant, it is more that they have not been argued for and their assumptions and conclusions have not been connected to the specific literature. For example, many simple models from theoretical physics are doubtless relevant for dynamics also in social movements, but this knowledge transfer is not automatic. It is far from trivial to make the necessary connections between the research strands, in part because it would require contextualization and argumentation for the relevance and applicability in relation to the existing research. Such argumentation is necessary as emergent phenomena do not necessarily matter in all contexts where their assumptions ostensibly apply – in some occasions, an emergent dynamic ‘shines through’ and becomes central in the macro-dynamics of the system, while in other occasions, other dynamics may dominate.

Based on the two main functions of free social spaces previously extracted from the literature, the model developed here focuses on whether, how, and to what extent the network structures of free social spaces affect the diffusion of social movements. The research most strongly related to this type
of modeling can be found in the literature on collective action. This field was initiated by Granovetter (1978), suggesting a central aim to be ‘to predict, from the initial distribution of thresholds, the ultimate number or proportion making each of two decisions’ (p. 1424). Central results show that the spread of collective action comes through bandwagon effects, and is related to (i) weak ties in the social network (see e.g. Chwe, 1999; Gould, 1993; Granovetter, 1973; Kim & Bearman, 1997; Macy, 1990; Watts & Strogatz, 1998) and (ii) moderate levels of homophily (e.g. Chiang, 2007; Krassa, 1988; Valente 1996).

In part due to (i), clusters are often seen as an impediment to cascades, as they reduce the number of the weak ties that are otherwise beneficial for the spread of social cooperation in the network. In part due to (ii), clusters are dense, reducing the number of weak ties, and are generally politically deviant, implying increased homophily. However, our model is to be grounded in assumptions and viewpoints established in the literature on social movements, rather than in the separate field of formal modeling within collective action. This also means that we focus on the spread of behavior and identity rather than of information, implying that we look at complex rather than simple contagions (Centola and Macy 2007; Watts 2002). But some models using complex contagion arrive at the same results. For example, Easley and Kleinberg (2010) show, using mathematical reasoning, that given that cascades start outside of the cluster, cascades are successful only if every node in the network is activated, and that if every node has the same number of neighbors, clusters will have negative effects for cascades.

That (ii), increased homophily, i.e. the probability that neighboring nodes have similar thresholds for activation, results in higher level of activation might imply that clusters with deviating threshold levels will have similar emergent effects. This can also be connected to the suggestion in critical mass theory that an initial group of activists can solve the large group problem by creating a bandwagon effect (Marwell & Oliver, 1993; Marwell et al., 1988; Oliver et al., 1985).

Free social spaces clearly relate to both (i) and (ii). They are dense clusters, reducing the number of weak ties, and are generally politically deviant, implying increased homophily. However, our model is to be grounded in assumptions and viewpoints established in the literature on social movements, rather than in the separate field of formal modeling within collective action. This also means that we focus on the spread of behavior and identity rather than of information, implying that we look at complex rather than simple contagions (Centola and Macy 2007). The reason for this is that free social spaces, as described above, are typically related to issues such as the formation of interpretive frames and collective identities rather than simply constituting a hub for information diffusion. Since the dynamics of complex contagions can be rather different from simple contagions, this may affect a central conclusion of the collective action literature: that weak ties are positive for the spread of collective action in social movements. Additionally, we want to loosen unrealistic assumptions such that every node needs to become activated for a contagion to be successful, or that each node has the same number of neighbors. We furthermore also explore the interaction between political deviance and clusterness, as these are arguably both characteristic of free social spaces.

The model presented here generally follows the established design decisions of existing models (see e.g. Centola, 2013; Centola and Macy 2007; Chiang, 2007), falling in the category of network cascade models. Such models analyze how activation can diffuse in a network, focusing on how different factors affect the probability for such spread to become dominant in the network. We refer to the probability that activation will spread to a majority of the network's nodes,\(^7\) given a certain distribution of network structures, as the network's 'cascade capacity'. Since the connections between the nodes represent the existence of mutual social contact, the model uses undirected ties.

The distribution of ties in the network follows the so-called Erdős–Rényi structure, commonly used in this context (see e.g. Watts & Strogatz, 1998), where each tie is distributed uniformly randomly between the nodes. This means that the distribution of ties follows a binomial distribution. To form the clusters, which here represent free social spaces, a fraction of the ties that span from inside the cluster to outside the cluster is removed and replaced with ties inside the clusters, resulting in more internal cluster ties than external (see Figure 1 for an illustrating example). The fraction between internal and external ties represents the previously introduced notion of clusterness. For example, with a clusterness value of .85, 85% of the cluster's external connections are removed and relocated, while a clusterness value of 0 represents a standard Erdős–Rényi structure. This definition of clusterness
captures the idea that free social spaces can prioritize either being densely connected internally, or prefer weak external ties.

A cascade is initiated by the activation of a randomly selected node and its neighbors. In a social movement context, this could for example be when a certain activist decides to take action together with a number of her friends. In each following time step, nodes that have more than a certain fraction – called their threshold – of their neighborhood activated themselves become activated – what is referred to as a complex contested contagion, which is standard in the collective action context (see e.g. Centola and Macy 2007). This continues until a steady state is reached. If a majority of the network has become activated at the end of the run, the cascade is classified as global.

Since the focus is on political spaces, we assume that the cascade is initiated inside the cluster and spreads from there. Furthermore, we here implement the previously introduced notion of bias, as a parameter lowering the activation threshold for the cluster nodes relative to the average threshold, meaning that the nodes inside the cluster are more easily activated. This, however, trivially has an effect of lowering the threshold needed for a cascade and may also have secondary effects that are difficult to intuitively predict. Because of this, we also implement a control case, in which the reduction of threshold is assigned to random nodes in the entire network.

The primary question that we aim to investigate with this model is whether political clusters have any effect on the diffusion of complex contagions, and to furthermore analyze the source of this potential effect and the interaction between the bias of clusters and their level of external connections. The design of the model implies an explorative approach: identifying relevant and unexpected dynamics, followed by a deeper analysis into its underlying mechanisms.

Technical description

We here include a more technical description of the model run. We define: $P_c$ as the clusterness parameter, $P_d$ as the average degree of the nodes, $P_b$ as bias parameter, $P_t$ is the activation threshold, $P_n$ is the number of nodes, and $P_s$ is the fraction of nodes belonging to the subcluster. $E$ is the set of edges, $N$ are the set of nodes. The model is run 1000 times for each step in parameter values, with a new network structure generated for each run to compensate for network structure heterogeneity and allow for higher generality and robustness. The parameter values $P_c, P_b,$ and $P_t$ are systematically varied over an interval to find how these parameters affect the model (other parameters were tested for sensitivity in separate runs). With 100 steps in bias and 200 steps in activation threshold and 12 steps in clusterness, the model is run a total of 240 million times. This constitutes a rather thorough exploration of parameter space, allowing an in-depth assessment of the model dynamics.4 Furthermore, these runs were evaluated for a number of network sizes to further validate their robustness and relevance.

In each such run, an Erdős–Renyi network structure is constructed in the following way. $P_d$ nodes are created, $P_n P_c$ of which are specified as belonging to the cluster. The specified mean degree is divided by two (since the network is undirected and any edge has two sides) and multiplied with the number of

---

4. The design of the model implies an explorative approach: identifying relevant and unexpected dynamics, followed by a deeper analysis into its underlying mechanisms.
node edges, i.e. $|E| = P_n P_C / 2$. From this set, $|E| P_C$ edges are selected where exactly one of the connected nodes belongs to the cluster. These are removed, and replaced by edges where both nodes belong to the cluster. Following this, nodes outside the cluster are set to have activation threshold $P_t$ and nodes in the cluster are set to have threshold $P_t - P_C$. A random node in the cluster is then selected, and set as activated to act as activation seed. All nodes connected to this node are also set as activated. After this, the following is repeated until no more changes occur: for each node, if the fraction of activated connected nodes is larger than threshold, the node is set as active in the next step. If the fraction of active nodes at the end of such a run is larger than .5, the cascade is set as global. This procedure is repeated 1000 times with new networks for each parameter step. Cascade capacity is defined as the fraction of times in these runs that a majority of the nodes are activated. The output from each such run is stored in lists – the processing of which results in the graphs presented below (Table 1).

**Model results**

We will now analyze the results of the model in a narrative manner, through a step-by-step process illustrating and describing the model output through different graphs. For reasons of clarity, we here take an explorative approach and iteratively describe and analyze the different graphs. We start by looking at how the presence of a cluster affects the cascade capacity – i.e. the likelihood that the complex contagions diffuse to a majority of the network nodes – without taking bias into account.

Looking at Figure 2, it is clear that the cluster has a network structural effect, as the cascade capacity varies strongly depending on the presence of a cluster. For example, at average threshold level of .270, a completely random network without clusters has an around 65% chance of a global cascade, while for the optimal value on clusterness (in relation to cascade capacity), it is more than 85%. This is intuitively unexpected, as it implies that a cluster itself – rather than its bias – impacts through emergent structural effects. Translating this to the notion of free social spaces, it implies that it is not necessary for the individuals associated to free social spaces to be in any relevant way different from the rest of the population; the simple fact that there is a cluster is adequate to make the space an important factor in the diffusion, something that would be difficult to derive using unaided intuition. This is clearly yet another dynamic of complex contagions that differs from that of simple contagions (Centola and Macy 2007).

At low clusterness levels, there are no discernible effects on the cascade capacity. Around .4, however, the cascade capacity increases significantly, as the cluster becomes more able to achieve internal activation. Past .7, the cascade capacity falls quickly, as the number of external connections from the cluster becomes too low for the cascade to spread successfully. This means that the clusters need to find a level of openness that both achieve internal coherence and external impact. As can be seen in the figure, the effect of having a cluster depends strongly on the threshold level: for threshold levels where no cascade is possible, or where cascades are almost certain to occur, the presence of clusters has little effect. The threshold levels that are most interesting are thus the ones where the effect of the cluster is as large as possible since these are in a sense balancing on the edge of a transition. In the following analysis, we will thus focus on the threshold levels in which the presence of clusters has the largest effects. As can be seen in Figure 2, the effect of varying the threshold value is far from linear and has a large impact on the cascade capacity. To further investigate the effects of changing

**Table 1.** These parameter values were used to generate all graphs, unless otherwise specified in the figure text. The step sizes constitute the resolution of the parameter values. The iteration count describes how many times the model was run for each parameter step, with new random network structures for each run to compensate for network heterogeneity.

<table>
<thead>
<tr>
<th>Network size</th>
<th>100 nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average degree</td>
<td>8</td>
</tr>
<tr>
<td>Step size for bias</td>
<td>.03</td>
</tr>
<tr>
<td>Step size for clusterness</td>
<td>.075</td>
</tr>
<tr>
<td>Step size for threshold</td>
<td>.0015</td>
</tr>
<tr>
<td>Iteration count</td>
<td>1000</td>
</tr>
</tbody>
</table>
threshold levels, we begin by looking at how varying the average threshold affects the cascade capacity for different parameter settings.

As can be seen in Figure 3, the transition from a very low probability to almost certainty for global cascades takes place relatively rapidly as the threshold value decreases (seeing the graph from right to left). The slope seems to be equally steep for the different settings, the difference between which seems to be mainly expressed in an offsetting of the transition to lower threshold values. The graph thus shows how relatively small changes in threshold values can result in increased probability for rapid global cascades. In a social movements context, this corresponds to the effect of a shift in opinion of the general population, e.g. an external event affecting the popular opinion. An illustrating example of this would be how the effects on public opinion caused by a single case of police brutality can lead to large-scale public mobilization. Thus, the results of the model illustrate that small causes can have big effects.

That increased bias results in higher cascade capacity, both when bias is assigned to the nodes inside the cluster or randomly follows trivially from the definition (since bias is operationalized as a decreased threshold level for certain nodes in the network). However, there are some interesting and unexpected differences between the case when the bias is assigned to the cluster nodes and when it is randomly assigned in the entire network. For example, the distance between the lines is smaller at higher bias levels with a randomly assigned bias, meaning that the effects of the presence of a cluster may be smaller in these cases. In other words, this indicates that the effect of the presence of a cluster on cascade capacity is larger when the bias is in the cluster compared to when the bias is randomly assigned in the entire network: political clusters are more likely to spark global cascades than what is expected.

This clearly warrants further investigation. We thus continue by returning to Figure 2, exploring in more detail how cascade capacity depends on clusterness, but this time also including different levels of bias.

Figure 4 illustrates how bias affects the likelihood of the cluster sparking global cascades. In other words, we investigate the effects of sharing political or cultural deviance. The bottom graph shows when the bias has been randomly assigned in the network. There is a striking and intuitively unexpected
difference between the case when the bias is assigned to the cluster nodes and when it is randomly assigned. Firstly, the graph clearly shows that the effect of the presence of a cluster is much larger when the bias is assigned to the cluster compared to the control case. Furthermore, the level of clusterness that leads to the optimal cascade capacity varies as a function of bias in the first case, but is constant in the latter. While the lines representing different bias cannot be compared with regard to the level of cascade capacity, since the threshold for each bias level is set differently, the shapes of the lines can however be compared, illustrating that the specific peak level for increasing level of bias is gradually shifted to the left in the graph. This indicates that increasing level of bias means that a lower level of clusterness is optimal for cascade capacity. In other words, there is an interaction between the clusters.
and the bias. Translating this to a social movement context, this would mean that radical free social spaces require more external connections, i.e. they need to be more open toward the surrounding society to spark global cascades.

We have seen that both bias and clusters have a positive effect on the cascade capacity, but it is not clear whether the effect is due simply to the combination between these factors, or whether the effect

Figure 4. In these graphs, we look at the threshold level at which the difference between the Erdős–Rényi structure and the optimal clusterness is largest, for each level of bias (note that this means that the lines in this graph cannot be compared with regard to the level of cascade capacity). The lines designate different levels of bias in the cluster. The upper graph shows the results when the cluster is biased, and the lower the control case, where the bias has been randomly assigned. As can be seen, when the bias is in the cluster, the level of bias affects the level of clusterness for which cascade capacity peaks. In the control case, there is no such effect.

Cascade capacity dependence on presence of network cluster for different bias levels

– with bias in cluster

– Bias randomly assigned

Figure 4. In these graphs, we look at the threshold level at which the difference between the Erdős–Rényi structure and the optimal clusterness is largest, for each level of bias (note that this means that the lines in this graph cannot be compared with regard to the level of cascade capacity). The lines designate different levels of bias in the cluster. The upper graph shows the results when the cluster is biased, and the lower the control case, where the bias has been randomly assigned. As can be seen, when the bias is in the cluster, the level of bias affects the level of clusterness for which cascade capacity peaks. In the control case, there is no such effect.

Cascade capacity dependence on presence of network cluster for different bias levels

– with bias in cluster

– Bias randomly assigned

Figure 4. In these graphs, we look at the threshold level at which the difference between the Erdős–Rényi structure and the optimal clusterness is largest, for each level of bias (note that this means that the lines in this graph cannot be compared with regard to the level of cascade capacity). The lines designate different levels of bias in the cluster. The upper graph shows the results when the cluster is biased, and the lower the control case, where the bias has been randomly assigned. As can be seen, when the bias is in the cluster, the level of bias affects the level of clusterness for which cascade capacity peaks. In the control case, there is no such effect.

Cascade capacity dependence on presence of network cluster for different bias levels

– with bias in cluster

– Bias randomly assigned

Figure 4. In these graphs, we look at the threshold level at which the difference between the Erdős–Rényi structure and the optimal clusterness is largest, for each level of bias (note that this means that the lines in this graph cannot be compared with regard to the level of cascade capacity). The lines designate different levels of bias in the cluster. The upper graph shows the results when the cluster is biased, and the lower the control case, where the bias has been randomly assigned. As can be seen, when the bias is in the cluster, the level of bias affects the level of clusterness for which cascade capacity peaks. In the control case, there is no such effect.

Cascade capacity dependence on presence of network cluster for different bias levels

– with bias in cluster

– Bias randomly assigned

Figure 4. In these graphs, we look at the threshold level at which the difference between the Erdős–Rényi structure and the optimal clusterness is largest, for each level of bias (note that this means that the lines in this graph cannot be compared with regard to the level of cascade capacity). The lines designate different levels of bias in the cluster. The upper graph shows the results when the cluster is biased, and the lower the control case, where the bias has been randomly assigned. As can be seen, when the bias is in the cluster, the level of bias affects the level of clusterness for which cascade capacity peaks. In the control case, there is no such effect.

Cascade capacity dependence on presence of network cluster for different bias levels

– with bias in cluster

– Bias randomly assigned

Figure 4. In these graphs, we look at the threshold level at which the difference between the Erdős–Rényi structure and the optimal clusterness is largest, for each level of bias (note that this means that the lines in this graph cannot be compared with regard to the level of cascade capacity). The lines designate different levels of bias in the cluster. The upper graph shows the results when the cluster is biased, and the lower the control case, where the bias has been randomly assigned. As can be seen, when the bias is in the cluster, the level of bias affects the level of clusterness for which cascade capacity peaks. In the control case, there is no such effect.
is in fact more than the sum of its parts. To explore this, we look at how the magnitude of the cluster's effect on the cascade capacity depends on the level of bias of the cluster.

Figure 5 illustrates the effect of the presence of an optimal cluster on the cascade capacity, i.e. a cluster with a level of clusterness that maximizes the likelihood of sparking global cascades. In other words, increasing value on the y-axis indicates that the presence of a cluster has an increasing effect on cascade capacity. What we can see from the graph is that the presence of a cluster becomes more important when there is a bias in the cluster, until a certain level where the effect pans out and then decreases. Put somewhat differently, up to a certain level, the more biased the cluster is, the more effect there is to the presence of a cluster on the cascade capacity. After this level, the impact of the cluster again falls. In the control case, the positive effect is minimal and wears off almost directly. At high bias levels in the control case, the presence of a cluster has almost no effects.

This implies a positive synergistic relation between bias and the presence of a cluster for lower levels of bias. That the effect is reversed at higher levels of bias may be explained by Figure 4: the peaks move left as the level of bias increases, meaning that as the level of bias increases, it seems as if the optimal clusterness decreases. In other words, with a high bias in the cluster, the optimal level of clusterness becomes lower. This also means that the difference between having an optimal cluster and not having any cluster becomes smaller, which results in that not having a cluster ultimately becomes more efficient at very high level of bias.

We can conclude that the benefit of having a cluster is to create a niche that can function as a base for the activation, from which it can then spread globally. However, if a group of nodes have very low activation, they are likely to be activated even if they are only weakly connected. The bottleneck is then whether they are able to spread their activation globally, which is facilitated by having many external connections, i.e. by not being strongly clustered. Relating this to social movements, the model indicates that the diffusion of social mobilization in a society is facilitated by the presence of clusters. The more radical the free social space is (as in deviant from surrounding society), the more important it becomes to have a cluster. However, this only goes up to a certain point; when the free social space becomes too radical/deviant in relation to the rest of society, the benefits of having a cluster in relation to the purpose of diffusion decrease.

We now explore this hypothesis by investigating the relation between the optimal level of clusterness at different levels of bias (see Graph 6).

Graph 6 shows which level of clusterness a cluster should have to be as efficient as possible at a given level of bias. Thus, the main difference from Figure 5 is that Figure 6 focuses on how the optimal value on clusterness is varying in relation to cluster bias. The results indicate that biased clusters are optimal at a higher level of external connections. As can be seen in the graph, clusters with a high bias should have more external connections than clusters with less bias to be likely to spark cascade. In other words, the less bias the cluster has, the higher degree of internal connections is needed to achieve optimal clusterness in relation to maximum cascade capacity. The reason for this is that biased clusters need fewer internal ties to activate internal nodes, allowing for more ties to be external – thus increasing the probability for global activation.

In a social movement context, Figure 5 shows that the benefits of having a cluster increase for radical free social spaces up to a certain point. What Figure 6 adds to this is that the more biased the free social space is, in relation to surrounding society, the lower becomes the optimal clusterness value (in relation to cascade capacity). In other words, the more radical the free space, the less dense it needs to be to become internally activated and, equivalently, the less radical the outside nodes compared to the free space, the more external connections are needed to enable global cascades.

Finally, we investigate the effects of the network size and density in relation to these dynamics to further validate the robustness of the model. This is illustrated in Figure 7, which compares the results for different number of network nodes. As can be seen in these graphs, while the cascade capacity is overall lower for larger networks, the effects on cascade capacity of having a cluster actually increase with larger network sizes. The reason for this is that a densely connected cluster obtains a more
important role as a connector and a starting point of cascades in larger and more weakly connected networks. This implies that the model dynamics described above are quite robust, which increases the likelihood of their empirical relevance.
Discussion

Let us begin by examining the dynamics of diffusion in the model networks. While it has been observed that long ties are always beneficial for the spread of simple contagions, complex contagions on the other hand first have to spread out locally before they can take advantage of long-range ties. Because of this, Ghasemiesfeh, Ebrahimi, and Gao (2013) observe that one can distinguish two phases of diffusion: the first phase is characterized by the spread of mobilization on a local level via strong, short-range ties, gathering a critical mass allowing for the transition to a second phase, in which spread can occur also

Figure 7. Graphs showing the effect of varying network sizes. The upper graph is similar to Figure 2, but shows threshold value .27 for varying node counts, with no bias. (Runs performed with bias showed the same result.) The lower graph corresponds to Figure 5, showing the size of the effect of the presence of the cluster for the threshold at which it has the greatest impact, for varying bias levels. As can be seen in these graphs, while the cascade capacity is lower for larger networks, the effects of having a cluster present increases with larger network sizes. These effects are likely to level out for larger networks (this could however not be formally shown, as it proved computationally unfeasible.) Results in both graphs were averaged over 500 iterations, average degree was 6, and 20% of nodes were assigned to be part of cluster. (Because of the increase in computational complexity with network size, the lower graph is shown with fewer node count levels.)
via long-range ties, allowing activation to spread quickly to more distant clusters. Having passed this threshold, complex contagion will start to spread in a manner more similar to simple contagions, i.e. taking advantage of shortcuts to other regions of the network, and quickly spreading globally. From the model dynamics, it seems that this describes the function of the free social space: it facilitates the initial phase of local activation, creating a foundation from which diffusion can occur via long ties. Looking at Figure 8, we can see that the initial phase of local diffusion occurs almost exclusively inside of the cluster, and is followed by a global spread through the cluster’s long-range ties that take the form of a separate wave of activation.

The model results indicate that there are highly general structural effects of clusters, facilitating the spread of complex contagions. This stands in sharp contrast to previous results on the spread of simple contagions, where clusters have been found to reduce cascade capacity (e.g. Watts & Strogatz, 1998). This indicates that the network structure characterizing free social spaces indeed seems relevant to take into account when studying what role such spaces can play in mass mobilization.

Figure 8. Illustrates the order in which the node activation occurs in successful cascades, by showing the fraction of cluster and non-cluster nodes to be activated at each time step. The upper graph has no difference in activation threshold between inside and outside the cluster, while the lower has a bias of .2. As can be seen, the cluster is activated the first few steps, while the bulk of the non-cluster nodes is activated later, peaking around step 6 or 7. The runs are averaged over 1000 iterations, clusterness = [.7, .75, .85], with activation threshold set to .27.
First of all, the mere presence of more densely structured relations within a cluster facilitates social practices to diffuse throughout the network, even when there are no further differences in bias. This phenomenon can be related to what Polletta (1999) calls indigenous structures, which are defined as free social spaces that are not explicitly political, but that still may have an important effect on uprisings. The effects of the presence of a cluster come from that it provides a niche from which dissent can then spread globally. The diffusion has two phases: first, it spreads internally in the cluster and second, it diffuses globally using the cluster as a base. The results from the model thus indicate that part of the reason that these indigenous structures are important for mobilization might lie in their network structure, providing a function as an incubator for the spread of dissent.

Secondly, in political free social spaces – what Polletta (1999) calls prefigurative spaces – the positive effects of the cluster structure become synergistically more important, up to a certain level of political deviance. This can be explained by that political spaces can be significantly more open (i.e. having more external connections), without risking the loss of internal coherence. At a very high level of political deviance, the effect of the presence of a cluster for diffusion wears off, as the cluster, in order to be optimal, must become so open that it is almost nonexistent. In other words, we have two seemingly contradictory results: (i) the effects of the presence of a cluster increase the more biased the cluster is, but (ii) the more biased the cluster is, the less clustered should it be to maximize the cascade capacity. These results are however both valid, even though the second factor counteracts the former. At a certain level of bias, the second factor outweighs the first, as the difference between having an optimal cluster and having none becomes smaller.

These factors, i.e. the presence of clusters and a high level of bias, allow uprisings to become possible at higher average thresholds than they otherwise would. The probability of global cascades to occur goes through a rather rapid transition with decreasing threshold level, which is illustrated in Figure 3. In the transition phase, small changes in public opinion can have a large impact in the probability of global cascades to occur.

The results from the model indicate that the efficacy of diffusion depends on a suitable relation between achieving internal and external activation. If the free social spaces are too open (have too many external connections), they risk losing internal coherence, but if they are too closed, they will be unable to spread their views. This relates in interesting ways to the discussion in Gamson (1996) on the relation between publicness and privateness/safety in free social space. In other words, this represents the difficult balance between being open enough to make new allies, attract potential sympathizers and to influence surrounding society, but at the same time remain sufficiently insulated from the hostile and the powerful to make free expression and participation in high-risk activism less risky. The results from this article thus provide guidance to Gamson’s central question whether radical groups should focus on seclusion and on building up strong internal coherence, or on openness and spreading the message to attract more sympathizers.

It should be noted that while these results are interesting and relate to previously largely unanswered issues in the literature, we do not consider the model as having proven that these factors are dominant in the real world. While formal modeling constitutes a useful way of abstracting effects of certain network configurations and subjecting them to detailed analysis, it is important not to fall victim to abstractions: multiple aspects of spaces could have been operationalized in very different ways, possibly with different outcomes (Lazonick 1993). These results are merely indications suggesting that the network effects of clusters are relevant to take into account when analyzing when, why, and how free social spaces influence the diffusion of social movements.

**Conclusion**

We have shown that the network structures and properties characterizing free social spaces, operationalized as a cluster with densely structured relations, indeed seem to have a positive effect on the diffusion of social practices that require complex contagion. This might contribute to explaining the effects of free social spaces observed in the literature. The main results from the model have some
interesting implications when it comes to understanding the relation between free social spaces and the diffusion of collective mobilization and social movements.

Up to a certain point, clusterness facilitates the diffusion of complex contagions, regardless of whether the cluster has a bias or not. Thus, this indicates that the network structure of free social spaces may facilitate the diffusion of social mobilization, regardless if the free social space is more radical than the surrounding society or not. When there is a bias in the cluster, the effects of having a cluster for diffusion are even larger. This implies a positive synergistic relation between bias and the presence of a cluster. However, this positive effect decreases when the bias in the cluster passes a certain point. This means that if the free social space becomes too radical, these positive effects decrease.

Furthermore, clusters with a high bias need more external connections (less clusterness) to be efficient in spreading activation. Thus, in order to diffuse more effectively, radical free social spaces might gain in trying to increase the number of (external) connections connecting the group with the rest of the society. The same goes the other way around; free social spaces that are not radically deviant from overall society might gain in increasing the ratio of internal ties and to build strong ties within the group in order to facilitate the global diffusion of mobilization. This is highly reminiscent of the analysis in Stark (1996) on the rise of Christianity. Stark argues that early Christian groups were effective in spreading their faith in part because they were able to remain open, while most new religious movements quickly become closed or semi-closed networks.

These results open up new, interesting perspectives that go beyond the predominant focus in the literature on the internal dynamics within free social spaces. The intention here has been to investigate and highlight how network structures might affect whether mobilizations succeed or fail to diffuse. In other words, how and when local disturbances do or do not snowball into mass mobilization, and what role free social spaces may play in these processes from a diffusion perspective. This also marks a step away from the often prevailing focus on ‘triggering events’ and external perturbation as the central point of analysis for understanding social mobilizations, emphasizing rather the culmination of underlying factors undermining the resilience of the system, such as network structural conditions and the alignment of collective action frames.

With this approach, we have also tried to illustrate the potentials of applying formal modeling for deepening our understanding of processes of diffusion within the field of social movement theory. This marks a move both in the direction requested by Diani (2003, 2011) concerning more research on the effect of the overall configuration of networks on collective action, but is also an attempt to more explicitly integrate formal modeling within the field of social movements, which has been requested by several scholars. In other words, we have tried to provide a tie, bridging these so far relatively isolated fields and thus render the methods more accessible to scholars interested in social movements and diffusion. We believe that different kinds of theories and methods play different roles and that formal modeling is one element in a diverse repertoire of methods, each capable of illuminating different aspects of social movements. In this regard, formal modeling can allow us to explore conditions that yet do not exist and help us get past impasses where reliance on intuition will get us no further.

Notes
1. This also relates to an adjacent issue that has been raised by scholars such as Vinthagen and Johansson (2013) and Simi and Futrell (2009) arguing that there is a theoretical gap in the literature concerning the link between low-profile, disguised, and non-organized everyday resistance on the one hand, and public, declared, and organized collective resistance on the other.
2. For example, Greenhalgh et al. (2005) have identified 13 different parallel diffusion literatures, ranging from, e.g. marketing, rural sociology, and organization studies to clinical epidemiology.
3. Nodes here represent individuals. For sake of simplicity, the network structure is assumed to be constant through time, meaning that we disregarded that mobilization can lead to the establishment of new social ties.
4. While methods such as network heterogeneity, thorough investigations of the parameter space, and averaging over large numbers of runs do contribute to making modeling conclusions more robust, it should be noted that they may still be unstable in the algorithmic space: small changes in the algorithm of the model may have large effects on its outcome. The severity of this issue is arguably under-appreciated in the modeling literature, perhaps in part due to the difficulty of systematically exploring the algorithmic space.
5. We have chosen to activate all seed neighbors primarily for two reasons: (i) it is reasonable that more connected nodes also have a larger initial impact when initiating a rebellion, and (ii) to simplify the model and reduce the number of variables. However, we recognize that this effect could play into the result in indirect ways, as the average number of seeded nodes will increase with cluster size due to higher average connectivity in the cluster. We have therefore performed validation runs where the number of seeded nodes was limited to a fixed count of random neighbors to control for this effect. These runs showed similar results, with two important differences: (i) the effect of having a cluster increased by almost a factor two, and (ii) the peak effect of the cluster occurred at lower bias values. The reasons for this is not that the cascade capacity is higher for the best cluster (it is more or less the same), but that it is significantly lower without cluster. This difference is reduced for higher levels of bias. In other words, when only a limited number of nodes are activated outside the cluster, this reduces the chance of by chance activating a highly connected node, whose neighbors can act as a form of informal cluster. When the bias is high, the difference is reduced since the cluster nodes are very likely to become activated in the first few steps in either case. All in all, these validation runs showcase the robustness of the model dynamics, and show that the effects shown with the main parameters are likely a lower bound for the importance of the cluster.

6. Looking at processes of radicalization and changes of general opinion in society, this can either take the form of radical groups and society, where both become more radical, or of society becoming more similar to the radical clusters. The latter would be illustrated by jumping between lines in the graph, while simultaneously moving left on the x-axis. A general pattern here is that the level of societal radicalism is, unsurprisingly, more important than the cluster bias level for the diffusion of dissent.

Acknowledgments
The authors thank Carl Cassegård and Justus Uitermark for stimulating discussions and valuable feedback. Thank you also to the editor and the two anonymous reviewers for many helpful comments.

Disclosure statement
No potential conflict of interest was reported by the authors.

Notes on contributors
Anton Törnberg is a PhD-candidate at the Department of Sociology and Work Science at Gothenburg University in Sweden. His research interests draw on complexity science and combining approaches and methodologies from computer science and sociology, focusing particularly on the field of social movements and collective action. He is currently involved in a research project that addresses discursive shifts in social media and the interplay between online discourses and offline action.

Petter Törnberg is a PhD-candidate within the Complex Systems Group at Physical Resource Theory at Chalmers University of Technology in Sweden. His work lays on the boundary between complexity science and the social sciences, focusing on developing new ways of understanding the evolution of societal systems.

References


Paper V
Combining CDA and topic modeling: Analyzing discursive connections between Islamophobia and anti-feminism on an online forum

Anton Törnberg
University of Gothenburg, Sweden

Petter Törnberg
Chalmers University of Technology, Sweden

Abstract
In this article we present an analysis of the discursive connections between Islamophobia and anti-feminism on a large Internet forum. We argue that the incipient shift from traditional media toward user-driven social media brings with it new media dynamics, relocating the (re)production of societal discourses and power structures and thus bringing about new ways in which discursive power is exercised. This clearly motivates the need to critically engage this field. Our research is based on the analysis of a corpus consisting of over 50 million posts, collected from the forum using custom web crawlers. In order to approach this vast material of unstructured text, we suggest a novel methodological synergy combining critical discourse analysis (CDA) and topic modeling – a type of statistical model for the automated categorization of large quantities of texts developed in computer science. By rendering an overview or ‘content map’ of the corpus, topic modeling provides an enriching complement to CDA, aiding discovery and adding analytical rigor.

Keywords
Anti-feminism, automated text analysis, critical discourse analysis, Internet forum, Islamophobia, methodology, social media, topic modeling

Corresponding author:
Anton Törnberg, Department of Sociology and Work Science, University of Gothenburg, Skanstorget 18, Box 720, 405 30 Gothenburg, Sweden.
Email: anton.tornberg@gu.se; tornberganton@gmail.com
Introduction

In a Scandinavian context, Islamophobia and anti-feminism have in recent years been central within the public debate. Feminists and Muslims have been the main targets of net-hatred in blogs and in comments fields, and these two groups are often characterized as two sides of the same ideological coin. This tendency is also reflected in the manifesto of the Norwegian extreme-right terrorist Anders Behring Breivik (2011), claiming that contained in the concepts ‘political correctness’ and ‘multi-culture’ lie cultural Marxism, pro-Islamism, and feminism, forming the ‘ideology’ which ‘now looms over western European society like a colossus’ (p. 21).

The focus in this article is to investigate and analyze the discursive connections between Islamophobia and anti-feminism on Flashback, a large Internet forum with a reputation for right-wing views. More specifically, how are Islam and feminism discursively connected on the forum? Are they cast as belonging to the same ideology? The data for the analysis are collected from the forum using custom web crawlers and consist of over 50 million posts. The reason for focusing on an Internet forum is the incipient shift toward social media as an increasingly important source for the (re)production of discursive power in society, but also because it is a unique source for studying everyday discourses outside the scope of mass media. In this sense, this large Internet forum is argued to have a function equivalent to that of traditional newspapers when it comes to producing and spreading societal discourses.

In order to approach this vast material, we suggest a novel methodological synergy between critical discourse analysis (CDA) and topic modeling: a new type of statistical model using hierarchical probabilistic modeling developed in computer science (Blei et al., 2003). By providing an overview, or ‘content map’, of the corpus, topic modeling constitutes an enriching complement to CDA, aiding discovery and adding analytical rigor. In addition, we complement this analysis with tools from social network analysis to illustrate how different discursive fields are connected through the engaged users.

The article begins by discussing the implication of the growth of social media for studying the (re)production of discursive power, arguing that we need to focus on social media in our analysis of societal discourses to a higher extent. Following this, we introduce topic modeling and illustrate how this method can be useful to approach and inductively structure large quantities of text. Then, we discuss some implications of combining CDA with topic modeling, relating to, for example, selection of data and the use of theory. The subsequent empirical analysis focuses on six different topics and topic categories that are manifesting various discursive connections between feminism and Islam. We show that the discursive connections between these issues are perhaps more complex than first expected and are mainly expressed through topics focusing on sexual assault, veils, and discrimination. We also show how gender equality seems to be used as a discursive strategy in order to criticize Islam.

From traditional to social media: A shift in the (re)construction of discursive power?

Traditionally, discourse analysts have often studied societal discourses by focusing on mass media, seen as playing a key role in the reproduction of dominant knowledge and
ideologies in society and the main channel through which the elite exercise their power (Van Dijk, 1993a, 2005). Another central reason for this focus is the lack of alternatives: it has traditionally been difficult to penetrate the realm of everyday discourses outside the scope of mass media.

While traditional media undeniably still constitutes a highly significant source of news, its influence is declining as people globally and across the age groups increasingly use Internet and social media as a substitute (Nielsen, 2012). This makes the almost exclusive focus on traditional media increasingly problematic. We here follow Kaplan and Haenlein’s (2010) broad definition of social media as ‘a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content’ (p. 61). Thus, the term obviously includes many rather different types of media and technologies, ranging from micro-blogs and online magazines to crowdsourcing, media sharing, and Internet forums.

New social media clearly shares many traits and functions often ascribed to traditional mass media, not least by framing issues and events and thus shaping people’s perceptions of reality and of social and political issues (Moscovici and Duveen, 2000). Following Van Dijk (1993b), we might say that they involve similar forms of social cognition. Social cognition mediates between the micro and macro levels of society by constituting the link between discourse and action, thus explaining how discourses transform social practice (see also Van Leeuwen, 2009). In this sense, social cognition links dominance and discourse by affecting how individuals and groups interpret the world and act upon this interpretation. This is also the focus of CDA, aiming to critically analyze these forms of texts and talks.

But social media is also in important ways different from traditional media. Social media builds upon interaction, communication, and networked individuals collaboratively sharing their narratives by creating and managing content. In this sense, the increase in social media usage thus also marks a shift from media consumers and passive observers to content creators. Through this shift toward user-driven, participatory information exchange, there are reasons to assume that the growth of new social media may bring with it new media dynamics, one example being online hate or online radicalization – a growing phenomenon that has received much attention over the last few years (Correa and Sureka, 2013). Online hate is generally characterized as content attempting to inflame public opinion against certain groups of people, generally based on race, religion, ethnicity, gender, or sexual orientation.

This apparent shift in the (re)production of societal discourses poses an implicit challenge to CDA: if CDA is indeed – as it claims to be – interested in describing, explaining, and criticizing the ways dominant discourses influence socially shared knowledge, attitudes, and ideologies, we must naturally continuously follow the source of these discourses as they shift in society. This indicates that we need to pay attention not only to top-down relations of dominance, but also acknowledge that power and dominance are not only imposed from the elite using mass media as channels through which they exercise their discursive power. Instead, as Van Dijk (1993b) notes, we need to pay more attention to how power and dominance can be jointly produced, and the aforementioned shift toward social media as an increasingly important societal actor indicates that this tendency will most likely become even more relevant in the future.
While many would agree that more focus needs to be placed on studying online social media, interested qualitatively oriented scholars have so far often been limited by multiple methodological challenges, perhaps foremost being the sheer amount of unstructured textual data often characterizing social media. Even relatively small data sets can be difficult to approach as it is hard to delineate, select, and confine materials of millions of texts, posts, or tweets. Making the matter worse, these texts are often short, lack discursive context, and are spread in complex and highly non-linear ways, making them difficult to extract and to study using established qualitative methodological and analytical approaches.

While these issues are indeed relevant for any discipline studying social media, they are perhaps particularly salient within discourse analysis where two main and interrelated criticisms are often raised: the arbitrary selection of texts (e.g. Koller and Mautner, 2004) and the small number of texts (e.g. Stubbs, 1994). The first criticism refers to the risk of ‘cherry-picking’, that is, that the author ‘picks a text to prove a point’, leading to obvious problems relating to representativeness and generalizability (Baker et al., 2008a; Stubbs, 1994, 1997). The second criticism concerns the often small data sets in discourse analytical studies, implying the risk of neglecting linguistic patterns that are less frequent or only cumulatively frequent. As Stubbs (1994) has observed, patterns of language use are often ‘not directly observable, because they are realized across thousands or millions of words of running text, and because they are not categorical but probabilistic’ (p. 204). In other words, many documents often contain only bits and pieces of ideologies, arguments, and discourses – small but systematic patterns and tendencies that may not be visible to the naked eye when restricted to small-\textit{n} studies (Stubbs, 1994).

In the following section, we will suggest a way of attacking social media data by combining CDA and topic modeling, thus forming a methodological synergy that can address these issues and that can be mutually beneficial to both fields. Topic modeling is similar to some methods applied within corpus linguistics (CL) and constitutes a valuable complement to discourse analysis by providing an overview or ‘content map’ of immense sets of documents, revealing small but systematic patterns and tendencies in the data. CDA can contribute by enabling a more thorough and systematic qualitative analysis that goes beyond superficial explorative findings and increases the ambition by reaching into the realm of understanding and explanation.

**Corpus-assisted discourse studies**

Combining CL and discourse analysis goes back quite some time, and studies have used various forms of computer-based techniques for handling large volumes of data (Mautner, 2009). This field has recently received increased attention under the name of corpus-assisted discourse studies (CADS), which is basically an umbrella term for approaches that integrate discourse analysis and techniques for corpus enquiries from CL (Cheng, 2013; Hardt-Mautner, 1995; Partington, 2006; Stubbs, 1996; Wodak and Meyer, 2009), either in the form of a methodological synthesis (Baker et al., 2008a) or as separate components combined as a way of triangulation (Baker and Levon, 2015). CL consists of various empirical methods of linguistic analysis using corpora as the primary data and starting point, with the aim of finding ‘probabilities, trends, patterns, co-occurrences of elements, features or groupings of features’ (Teubert and Krishnamurthy, 2007: 6).
While many of these methods are rather simplistic, focusing on, for example, computing frequencies and related statistical significance of certain words, others are more advanced, enabling qualitative examination of the collocational environment of certain words and description of salient semantic patterns (see e.g. Baker, 2006; Baker et al., 2008a). Although applying methods associated with CL is perhaps not yet central within mainstream CDA research, awareness of the potential of these methods seems to be growing and there have been a range of recent CDA studies using methods from CL (Baker et al., 2008b; Baker and McEnery, 2005; Hellsten et al., 2010; Nelson, 2006; Orpin, 2005).

This study takes a novel approach to the CADS perspective by, for the first time (to the best of our knowledge), combining CDA and topic modeling, a technique that was recently developed in computer science but which shares common traits with some of the methods conventionally applied within CADS. Topic modeling inductively finds a number of topics describing the text corpus – recurring clusters of co-occurring words. An important difference from most other techniques within CL is that unsupervised topic models inductively structure the data without using any pre-set keywords. This means that it is corpus-driven, that is, the analysis is driven by whatever patterns are salient in the data itself (Tognini-Bonelli, 2001). While there exists other unsupervised, inductive CL techniques (e.g. cluster analysis, keyness analysis, and word frequency lists), it is more common within CL to search for certain keywords and study them using, for example, collocation analysis (Pollach, 2012). Before we further discuss how we combine topic modeling and CDA in this particular study, we will first elaborate more on how topic modeling works.

**Topic modeling**

Topic modeling is basically a catchall term for a collection of methods and algorithms that uncover the hidden thematic structure in document collections by revealing recurring clusters of co-occurring words. While there are several different algorithms for performing topic modeling, the most common and also the one we use in this article is Latent Dirichlet allocation (LDA) (Blei et al., 2003).

LDA views each document as a bag-of-words. A topic is defined as a list of words with different assigned probabilities. The goals of topic modeling are that (1) the words from each document occur in as few topics as possible, while (2) each topic has as few words as possible. LDA tries to find an optimal solution for satisfying these two contradictory goals. The output is a pre-set number of topics, where each topic has a list of words with different probabilities, and each document is linked to a list of topics with different probabilities. It is of course also possible to look at this probability distribution in the other direction: the documents that are most strongly linked to a topic are the ones that it best describes.

Let us take a simplifying example – without assigning varying probabilities – to see how this works in practice. Say that we have the three following documents: (1) ‘John eats sausage’, (2) ‘Eating sausage’, and (3) ‘John eats’. We want to find two topics from these. Remember that the solution we want to find is one where each document occurs in as few topics as possible and where each topic has as few words as possible. In this simple case, no complicated algorithms are required as it is trivial to see that the solution are the topics: (1) ‘[John]’ and (2) ‘[sausage]’. ‘Eat’ is part of all documents and will
therefore be disregarded, given that we use stemming.) Topic 1 will be connected to documents 1 and 3, and topic 2 will be connected to documents 1 and 2. In more realistic cases, more sophisticated algorithms are required to approximate the optimal topics, each topic has different probabilities for each of its words, and each document has different probabilities for each of its assigned topics — but the basic principle is the same.

The algorithms that LDA uses for these calculations are based on Bayesian statistical theory (Gelman et al., 2014), where the topics and the per-document topic proportions are seen as latent variables in a hierarchical probabilistic model. The conditional distribution of those variables is approximated given an observed collection of documents. When applied to the documents in a corpus, inference produces a set of topics, and for each document an estimate of its topic proportions and to which topic each observed word is assigned. Since the underlying mathematics and algorithms applied in LDA are too technical to further elaborate in detail here and have already been covered in multiple publications, we refer the reader to Blei (2012) and Blei et al. (2003) for a more thorough description.

Generating and interpreting topics

As we have seen, the recurring clusters of co-occurring words that are generated using topic modeling are called topics. While these often resemble what is conventionally referred to as ‘themes’ and ‘topics’ within qualitative text analysis and discourse analyses (Chafe, 2001), these concepts should not be confused since the characteristics of the topics generated in topic modeling vary in relation to the data being analyzed.1

By automatically searching for word co-occurrences within textual units and identifying semantically coherent or internally homogeneous topics, topic modeling inductively discovers a structure of the corpus, largely unaffected by the researcher’s prior conceptualizations. While this does not eliminate the role of the researcher, it is often claimed that it turns the analytical work on its head by moving the work burden from identifying patterns internal to the text data, to the interpretation and theoretical conceptualization of patterns, and their relation to their social context (Krippendorff, 2004). For instance, as Mohr and Bogdanov (2013) argue, topic modeling

shift[s] the locus of subjectivity within the methodological program — interpretation is still required, but from the perspective of the actual modeling of the data, the more subjective moment of the procedure has been shifted over to the post-modeling phase of the analysis. (p. 560)

While we partly agree with this, it is nonetheless important to be aware that using topic modeling is not a neutral and rigorous process and the resulting topics do not reflect the one and only ‘true’ content of the corpus. Subjectivity not only leaks in through the theoretical assumptions underlying the model, but is also central throughout the entire research process. Topic modeling is often sensitive to the various settings and parameters used in the model, and changing these often yields rather different results. For example, one of the most central parameters is the number of topics specified by the user. While the suitable number of topics may in some cases be calculated statistically (Grimmer and Stewart, 2013), this is difficult when topics are used to identify themes for interpretation,
which is usually the case in social scientific studies. In these cases, the parameters are generally evaluated using more qualitative methods, according to whether they generate meaningful and analytically useful topics (Blei and Lafferty, 2007). Thus, similar to the standard procedure when applying most CL tools, this is a matter of experimenting and evaluating different numbers of topics and settings, combined with careful close readings, in order to estimate the quality of the topics. In general, one can say that a large number of topics leads to ‘higher resolution’, in other words more detailed topics, while fewer topics aggregate these into coarser overviews.

Consequently, using and calibrating topic models is, in our view, not an objective, strictly scientific process of optimization. Rather, following DiMaggio et al. (2013: 582), we think of topics rather as lenses for viewing a corpus of documents. As they argue, ‘[f]inding the right lens is different than evaluating a statistical model based on a population sample. The point is not to estimate population parameters correctly, but to identify the lens through which one can see the data most clearly’. Thus, just as different lenses are appropriate for different purposes, the ‘correct’ resolution and level of detail depend on the purpose of the study and the type of data. In the words of the statistician Box (1979), ‘All [models] are wrong; some are useful’ (p. 202).

Topic modeling in the social sciences

Topic modeling has been increasingly applied to various problems in the digital humanities, literature studies, and by historians, but also in a few cases in political science and within sociology (see e.g. Grimmer and Stewart, 2013; Roberts et al., 2014). It has also been the subject of a well-cited special issue in Poetics (2013). Many of these studies use LDA or related Bayesian approaches to infer latent topics in news articles, scientific journals, abstracts, and blog posts, but also in poetry and fiction. While topic modeling was initially intended primarily as a way of indexing and automatically categorizing large data sets, the applications in the humanities and social sciences are generally more theoretically driven. Topics have been claimed to capture various theoretical concepts such as media framing (DiMaggio et al., 2013; Hopkins, 2013), themes or thematic categories (Mohr and Bogdanov, 2013), dramatistic scene (Mohr and Bogdanov, 2013), and discourses, priming, and the relationality of meaning (see e.g. Bail, 2014).

In this study, however, we choose not to collapse topic modeling with any theoretical concept, but rather to use it as it was initially intended: as a tool for inductive empirical categorization. It is important to keep in mind that the technique was not particularly developed to address a certain theoretical issue or with certain theoretical concepts in mind. There is always a risk of overinterpretation when we are dealing with new, powerful methods of which the theoretical implications are yet to be fully understood. This is further exacerbated in this case due to the technical sophistication of the algorithms used and the resulting risks of blackboxing the method.

Following this, the current study uses topic modeling to render a valuable overview of the content, which can help us to categorize very large sets of documents and explore discursive changes over time, revealing small but systematic patterns and tendencies in the data. We will argue that such a methodological combination between CDA and topic
modeling can be mutually beneficial for both fields. Before we head over to the analysis, we will first discuss how topic modeling and CDA can be combined.

Combining CDA and topic modeling

We here follow Van Dijk’s (2008) definition of discourses as (re)contextualization of practices. In other words, material events and social practices happening in the ‘material world’ are reformulated in texts and talks. Generally, this means that most CDA studies are theoretically driven, in the sense that they often start with a macro-level focus on institutional power relations, dominance, and hegemonic ideologies, then investigate how these are exercised in the recontextualization of various events and social practices (Van Dijk, 1993b). In other words, focus often first lies on broader hegemonic ideologies or a certain ideological shift, then investigation takes place into how this ideology shapes and is shaped through the reconstruction or recontextualization of specific social practices. This is also apparent in the conventional procedure in CDA to use data selection through representative sampling in order to find ‘typical text’ where these ideological shifts are assumed to occur (Fowler, 1996; Lin, 2013). However, combining CDA with CL methods such as topic modeling turns this procedure around by instead enabling the analysis to start more inductively by focusing directly on the recontextualizations of various social practices and then exploring and identifying common discourses, underlying ideologies and tensions, and modifications within them.

For example, in this particular case, we first use topic modeling to automatically and inductively structure the text to find recurrent patterns, then we analyze the resulting topics more in-depth to find out what they are about, focusing both on the terms most closely associated with each topic and the related documents. In this way, we try to identify the broader discursive fields with the help of the topics. Discursive fields are here defined as the ‘dynamic terrain in which meaning contests occur’ (Steinberg, 1999: 748) and comprise various frames and discourses varying on a continuum ranging from consensus to heated contestation (Snow, 2004, 2008, 2013; Spillman, 1995; Steinberg, 1999). The boundaries of these discursive fields are never entirely fixed or clear, but are rather in constant flux, as the dynamic fields ‘emerge and evolve in the course of discussions and debates about contested issues and events’ and encompass protagonists, antagonists, and bystanders (Snow, 2004). In this sense, these discursive fields represent ‘public battlegrounds’ where different actors compete to ascribe meaning to a certain issue – a battleground that at the same time defines the limits of discussion on a particular issue.

In the next step, we code and analyze the ‘top documents’ relating to each of the relevant topics in order to (1) identify and analyze the dominant discourse(s) within this specific discursive field and (2) focus on the common discourses permeating through all or several of the topics, thus (potentially) representing an underlying ideology. By top documents, we refer to the documents that are most central to each topic, in the sense that they most prominently exhibit the word distribution of the topic (see the ‘Topic modeling’ section earlier). This is in fact also a key benefit of topic modeling: both finding central topics in a corpus and listing the documents that are most strongly associated with these topics. Focusing the analysis on these top documents enables a more systematic approach to the data and marks an important step away from the tendency of
'cherry-picking'. This also means that the approach suggested here is not to merge topic modeling and CDA into a single synthesis, but rather to use them as complementing steps in the analysis.

Data and procedure

The corpus for this study is extracted from Flashback, which is currently one of the largest web forums in the world. At the time of writing, there are 948,950 registered users and 49,491,019 posts, contained in different subforums and threads treating subjects ranging from computers, drugs, and family to culture, politics, and current crime. The forum is fully open to the public and is moderated by a set of privileged users. The forum is expanding with about 15,000–20,000 new posts per day and has around 2,300,000 unique visitors per week. While the forum includes discussions on many different topics, it has a reputation for leaning toward extreme-right opinions and it is often mentioned in Swedish media as a hub for online hatred and xenophobia.

Due to the size and reach of the forum, Flashback can be considered a highly relevant platform with impact and influence comparable to other media actors. Thus, with the background of the above discussion on the pending shift toward social media as an increasingly important actor, we argue that this forum serves as a channel for discursive power and has an equivalent function in producing and spreading societal discourses just as any traditional newspaper. This clearly motivates the need to critically examine how the discursive connections between Islamophobia and anti-feminism are manifested.

Using web crawlers, we downloaded and anonymized the entire forum in the time period between May 2000 and May 2013 into a local database, comprising over 50 million posts and about 968,289 users.

The analysis consists of two steps. In the first step, focus lies on analyzing the connections between anti-feminist and Islamophobic discourses. In order to do so, we created a subcorpus by selecting all posts containing at least one keyword associated with Islam ('muslim', 'islam', and 'arab'), together with a keyword associated with feminism ('femini'). We then ran an LDA set to 20 topics on this subcorpus, in total comprising 12,796 posts. This enables us to investigate more closely how anti-feminist and Islamophobic discourses are systematically discussed in relation to each other, in other words to reveal eventual discursive connections.

In the second step, we focus on whether these issues are directly connected through engaged users, and if so, which are the topics that connect them? Here, we created a new subcorpus by selecting all posts in all political subforums over all years without using any keywords and ran an LDA set to 80 topics. In total, this subcorpus comprises 6,038,773 posts, aggregated into 576,801 documents. Based on this, we used a network analytical software called Gephi to generate a discursive network, illustrating how the topics are connected through the engaged users (see Graph 2). In other words, this graph shows the interconnections between topics with shared/overlapping users as the connection. The overlap is calculated by finding how many of the N most frequent posters are in common between the topics, divided by what would be statistically expected. In this way, we can investigate whether the same users tend to be active in more than one topic.
These two steps concerning how the corpus is selected should not be confused with the following step of empirical categorization using the LDA model, which is inductive. Furthermore, using the keywords to create the specific subcorpus in the first step, there is obviously a risk of accidentally excluding a number of potentially relevant posts that either discuss the specific topic more indirectly or refer to it through relatively common derogatory terms such as ‘feminazi’, ‘sandnigger’, or ‘culture-enricher’. However, we consider this is a minor problem as most posts cite a previous submission that uses the specific term and therefore will be included in the analysis. A close reading of a selection sample indicates that our selection criteria and keywords seem to capture the intended texts.

In both steps we set the $\alpha$-parameter of the LDA model to 1.0. This defined the prior Dirichlet distribution the model assumes: the lower the $\alpha$-parameter, the more concentrated topic distribution the model will assume and generate, and a higher $\alpha$-parameter means more uniform topic distribution across each article. We have experimented with various topic numbers and $\alpha$-parameters and found the results to be robust.

For technical reasons, standard LDA generally works best for documents with a size of at least 1000 words. We therefore aggregated all posts in the subcorpus from each individual user in a specific thread within the same time period into chunks of 1000-word documents (for a similar approach, see e.g. Hong and Davison, 2010; Weng et al., 2010). Posts that are significantly longer than 1000 words are split into smaller chunks. In fact, as Zhao et al. (2011) have suggested, this approach can be regarded as an application of the author-topic model (Rosen-Zvi et al., 2010), where each document has a single author. An alternative would be to instead compile a number of subsequent posts from various users in the same thread. However, we do not see any advantages of this. First, to merge various opinions and discourses from different users in the same document would make it more difficult to read. Second, this would impede the following analysis on connections between users, which requires user-specific documents.

Furthermore, as conventional in LDA models, we excluded a number of stopwords from the analysis, that is, very common words like conjunctions and articles, using off-the-shelf lists for Swedish, Norwegian, Danish, and English, further extended with slang and abbreviations often used on the forum. Finally, while lemmatizing the corpus can often be useful to produce better topics, this proved difficult in our case due to the often informal language and the amount of, for example, slang, abbreviations, and misspellings.

In the analysis of the resulting topics, we distinguish between topics and topic categories. While the former refer to single topics (referred to as T1, T2 ...), the latter are a group of topics that are interpreted to belong to a common subject area, that is, all topics discussing an arguably similar issue or event, either in the same year or over time. These heuristic topic categories are hence manually constructed through close reading of the constituting topics and their top documents, with the purpose of enabling an overview of the results and to facilitate the analysis. This means that the decision as to how wide and inclusive these topic categories are set to be depends on the level and resolution of the analysis. Thus, a wide topic category consists of a number of topics that in turn can be thematized within smaller, more specific topic categories.
To perform the LDA models, we used Big Text Tool, an online-based application that includes various tools for automated text analysis and graphical illustrations. This application is free, easy-to-use, and customized particularly for social scientific studies using large corpuses.

**Result and analysis**

We now move to the presentation of the results and the analysis. As we can see in Graph 1, the model presents a number of separate topics focusing explicitly on either feminism or Islam – such as the Assange case (T7), net-hatred against women (T5), and extreme-right parties (T19). Since we are particularly interested in the overlapping topics, we chose to exclude these topics from the analysis. Analyzing the remaining overlapping topics enabled us to study the various discursive fields within which both feminism and Islam are jointly discussed. Focusing on the posts within these topics enabled us to more closely dissect and investigate the character of these intersections and the discursive connections between them. Thus, the main thrust of analysis will be on identifying underlying discourses connecting Islamophobia and anti-feminism, rather than any detailed analysis of each topic.

In general, much of the focus in the discussions of these topics lies on (1) the perception of women within Islam, (2) the claimed inherent tension between Islam and feminism, and (3) the alleged hypocrisy among feminists that is claimed to propagate for Islamic immigration. These issues seem to prevail as an underlying discourse penetrating most of the topics, although the topics often focus on more specific issues. We will here briefly discuss some of the fields within which the discursive connections are played out, analyzing quotes from the top documents in the topics. We use topic categories to facilitate the analysis, although we also focus on the separate topics within each category for a more nuanced and detailed analysis.
In the analysis, we identify two main topic categories:

1. inconsistency between feminism and Islam (T1, T3, T4, T17, and T18);
2. inconsistency within the political left (T10, T11, and T16).

Besides these, we also identify four separate topics:

1. children, traditions, family, and gender roles (T12);
2. race, ethnicity, and sexuality (T2);
3. religious private schools (T6);
4. Jews (T9).

First, there is a broad topic category focusing explicitly on the claimed inherent female oppression within Islam, consisting of particularly T1, T3, T4, T17, and T18 (see Graph 1). Posts in which these topics’ terms dominate word assignment thus deal with discussions on the inconsistency between feminism and Islam. They also discuss why feminists are perceived not to protest against male chauvinism and female oppression among immigrants in general and among Muslims in particular. This general tendency is clearly manifested in these posts below which are central in this topic category:

[…] you can’t even begin to see that islam and feminism are inconsistent/opposed? that the increased portion of muslims in Sweden should be a Swedish feminists worst fear?
I have probably never seen feminists demonstrating against the patriarchy of Muslims (where you really can speak of a patriarchy and gradual [sic] female oppression). It seems almost only important to destroy the Swedish traditions, not the others. A little unfair, I think, as a Swede.

Why can Ahmed have his traditions preserved, while mine are destroyed?

What is apparent here is an idea of an immanent essence of female oppression within Islam. Muslim countries are presented as strong patriarchates characterized by gender inequality, which is seen as a foundational pillar of Islam, and thus in direct opposition to feminism and gender equality. In the second quote, the user also expresses a criticism that while feminists ‘destroy’ Swedish traditions due to their claimed patriarchal tendencies, they are at the same time ignoring similar, or even worse, tendencies within Islam. This indicates that Muslims are argued to unfairly receive special treatment and enjoy privileges that so-called ethnic Swedes do not, that is, that there is a ‘reverse discrimination’ against ethnic Swedes – an argument which is recurrent in many of the posts. This statement also indicates that there is a claimed inconsistency within feminism by criticizing patriarchal structures in one context while ignoring it in another, something which we will come back to later.

Furthermore, this claimed essence of female oppression in Islam is manifested in different ways, which we can see in the various topics within this topic category. For example, T3 focuses specifically on discussions on whether immigrants are statistically overrepresented in violence against women, sexual assaults, and issues relating to ‘honor violence’. In these posts, Muslims are portrayed as violent and focus often lies on the alleged reasons for this, for example, due to genetic/biological causes or whether it is ‘inherent in their culture/religion’. T17 focuses particularly on discussions on veils and burqas and whether these should be considered as a free, individual choice or as an expression of compulsion. This topic also focuses on female rights according to the Quran. T4 focuses somewhat more explicitly on the essence of Islam: the Quran and Sharia laws. These discussions focus on various parts of the Quran that are claimed to be hostile toward women. Connections are frequently drawn between Islam, female oppression, pedophilia, and child marriage as well as to circumcision and genital mutilation.

Coming from this idea of an inherent and inevitable female oppression within Islam, the claim that many Swedish feminists nonetheless support Islam, or at least do not openly oppose it, is thus argued to illustrate a widespread hypocrisy among feminists and an inconsistency within feminism. This is a recurrent pattern in many posts in this topic category, for instance:

This is the interesting thing with Swedish feminists shameless hypocrisy. It’s hard to imagine anything more antifeminist than Islam, so it would be as you say logic if feminists spent their time criticizing Islam and their anti-equality perspective on women, particularly since this is a growing threat against the Swedish gender equality.

According to all logic feminists should hate Islam, but you never hear any criticism. Why? Because feminists know that Islam and feminism will never fit together. They choose to cover up and pretend as if it’s raining.
As we can see, the claimed inherent contradiction between Islam and feminism is thus again raised here by explicitly emphasizing the hypocrisy among feminists to avoid criticizing Islam and Muslims. Islam is portrayed as a ‘threat’ against gender equality, and feminists are claimed to ‘cover up’ this fact. Before analyzing this more in detail, we note that we find a similar argument also in the second topic category, consisting of T10, T11, and T16 and focusing particularly on the claimed lack of consistency within the political left. Similar to feminists, left-wing activists are argued to support Islam or at least avoid criticizing it:

You feminists and hbtq in the left why are you demonstrating together with men with African or Arabic origin? The men that you stand together with and scream think that hbtq persons are sick/suffers from an illness. Most of them think that you don’t have a right to live and should be shot or stoned to death, still you stand their side by side.

There is a burnout is the brain of the pc-ist (‘pkitens’) when two pc-topics contradict each other resulting in a tacit status quo.

Thus, as we have seen, there is a dominating tendency in the discussions to focus on Islam in relation to gender inequality, and this is frequently raised as an argument against immigration in general and against Muslims in particular. For instance, in one of the earlier quotes, Islam is argued to constitute a threat against Swedish gender equality. This reveals an interesting ostensible contradiction. Based on a broad overview of the central discussions on the forum, discrimination against women appears generally to be considered a minor or non-existent problem. At the same time, it is raised as a substantial issue when it comes to Muslims and immigrants. This suggests that gender equality is used deliberately as a discursive strategy among certain groups in order to criticize immigration, but also to ‘reveal’ what is seen as hypocrisy and contradiction within what is referred to as the ‘politically correct establishment’ that is alleged to both embrace feminism and immigration in general and Islam and Muslims in particular. In other words, this indicates that feminism and equality in these cases are used for the sake of the attack, rather than for any deeper, actual sympathies with feminism or concerns for gender equality.

Apart from these two topic categories, there are a number of separate topics treating other issues that are discursively and explicitly connecting feminism and Islam. One of these is T12, focusing on children, traditions, family, and traditional gender roles. Posts in which this topic dominates word assignment express criticism against the intention in feminism to ‘dissolve traditional gender roles’. For instance, as some users put it,

It’s not about whether gays are nice or not. Questioning the hetero-normativity (with the related pushing for ‘faggotry’) is a method to destroy the traditional man’s role.

Since the gender-pedagogue is pursuing a highly dubious operation, it appears to be better to let believing muslims take care of the children, since they at least don’t intend to break down the gender roles and sexual identity.

I am talking about Sweden today. And about that the choice is – hypothetically – between muslim values aiming to conserve/protect traditional gender roles, and feminists that want to get rid of them. It’s a question about how different cultures/religions relate to gender roles.
I’d choose feminism all days a week instead of the islamic oppression, women are our allies in the fight against the green pest.

Interestingly, as illustrated in these quotes, we may discern two distinct camps discussing these issues. In the first camp, we find the afore mentioned tendency where certain users claiming to advocate gender equality in Sweden see women as ‘their allies’ against the Islamic oppression and use this approach to criticize Islam. In the second camp, there is widespread criticism against radical feminism and how this is perceived to constitute a threat to ‘the family’ and traditional gender roles as well as ‘sexual identities’. Here, Islam and Muslims are seen as potential collaborators in the battle against ‘faggotry’, gender pedagogue, and feminism in general.

Furthermore, T2 focuses on racial issues and sex and discusses sexual relations between different ethnic groups and different ‘races’:

While in one moment, the swedish woman raves about feminism and tolerance in their fight against male chauvinism, in the other they jump to bed with misogynist arabs and other dominant and sexists men. Just in the same way, they intend to oppose misogyny and oppression, but at the time they consistently and loyalty rave to allow Islam political elements to emerge in society.

Again, the claimed inconsistency among feminists is brought to the fore, but this time focusing on sexual relations and ethnicity. Here, this inconsistency is illustrated in that Swedish women, on the one hand, fight male chauvinism, but at the same time go to bed with ‘misogynist Arabs’. The argument that Swedish (feminist) women prefer sexual subordination and dominant men in bed is a recurrent theme in the discussions and clearly intends to both illustrate the hypocrisy of feminists and consolidate traditional gender roles as well as to illustrate some deeper characteristics of ‘women’s nature’ that feminism is claimed to try to cover up.

T6 focuses on religious private schools, with particular focus on Islamic schools and the Pentecost. Here, criticism particularly toward Islam is thus articulated in that these religious schools are perceived as constituting a threat to Swedish values and to scientific perspectives. The connection to feminism is that religious schools are compared to the implementation of gender science in the education system, and these are claimed to constitute similar political regulations of scientific perspectives.

Finally, there is also a topic on Jews (T9). A central reason that this topic appears in the posts containing both feminism and Islam is that naturally Jews are claimed by some to be the root behind both feminism and ‘multiculturalism’. As one user quite straightforwardly puts it in a central post in the topic, ‘the problem is that since the entire PC [politically correct] multicult is a jewish invention it’s hard not to bring up the jews’.

In other words, the causality expressed in the majority of the posts within this topic is that the existing political elite (the ‘PC establishment’) – which is both feminist and pro-Islam – is in fact fundamentally Jewish. Or more correctly, it is connected to cultural Marxism, which in turn is seen as a Jewish invention. In fact, some users argue that cultural Marxism is the actual reason explaining the afore mentioned internal inconsistency among feminists supporting Islam and Muslim immigration.
**Topic connections through engaged users**

In the analysis so far, we have identified what seems to be an underlying ideology characterizing Islam as inherently misogynistic and feminists as inconsistent since they – despite this – are argued to support Islam and Muslim immigration. This ideology is recontextualized in several different fields and areas, including sexual relations and family and gender roles.

The next step is to investigate whether the same users in the forum are active in both discussions: in other words, whether they explicitly discuss both Islam and feminism as separate issues. This would enable us to see whether these issues are also directly connected through engaged users, and if so, which are the topics that connect them? This would indicate that these recontextualizations are explicitly related to each other.

In order to do this, we used the second subcorpus consisting of all posts in all political subforums over all years without using any keywords and ran an unsupervised LDA set to generate 80 topics. In total, the corpus comprises 6,038,773 posts, aggregated into 576,801 documents. We then investigated the overlap of users between these topics in order to see whether the same users tend to be active in more than one topic.

Since the discussions on Islam, Muslims, and feminism are relatively peripheral compared to all other topics and issues debated on the forum, a general network with the most connected topics would not include these topics in a meaningful way. Therefore, we created a form of ego network focusing explicitly on four topics that explicitly relate to either feminism (T9: women, men, feminism, and feminists) or to discussions that we previously have seen are central in the intersection between Islamophobia and anti-feminism (T30: children, child, parents, and abortion; T41: women, sex, girls, and rape; and T53: law, veil, prohibit, and discrimination). The results are illustrated in Graph 2. To facilitate the analysis, these central topics are larger and colored in red in the graph (i.e. the size of the nodes is not connected to the size of the topics) and the strength of the edges represents proportion of shared users. The alters in the network represent the topics that are directly connected to these focal topics. Noisy background topics are excluded.

As we can see in Graph 2, T9 that explicitly focuses on feminism is strongly connected with T30 (children, child, parents, and abortion) and T41 (women, sex, girls, and rape), but there are no direct connections to Islam, Muslims, or to any other topics. However, T41, T30, and particularly T53 – topics relating to discussions and issues that we previously have seen are central in the discursive connections between Islam and feminism – are indeed strongly connected to topics on immigrants and Islam.

This indicates that users who were explicitly discussing feminism were, in general, not also inclined to discuss Islam, Muslims, or immigrants. However, users discussing sexual assault and rape (T41), children, family issues, and abortion (T30), and veil/burqa and discrimination (T53) are in fact to some extent also discussing both feminism and immigrants, Muslims and Islam. This indicates that these topics serve as indirect connections, linking feminism and Islam. These results confirm the results from the analysis earlier, highlighting these specific fields as central in the recontextualizations of the Islamophobic and anti-feminist discourses. This also strengthens the previous hypothesis that there seems to be a tendency to use gender equality as a discursive strategy to criticize Islam and Muslims.
Discussion

By focusing on the topics, we have identified certain discursive fields where the recontextualizations of certain ideologies are ‘played out’: they are exercised and negotiated in these contexts. While these recontextualizations indeed concede to the reality of these events and practices, they are nonetheless formulated from a certain point of view and in line with certain interests. We have focused on a number of topics connecting anti-feminist and Islamophobic discussions, through, for example, discussions focusing on the claimed overrepresentation of Muslims in the statistics of violence against women, in the question of religious private schools, the phenomenon of burqas/veils, sexual relations between different ethnic groups, and family and gender roles.

By analyzing these topics and their associated documents, we have identified a common and pervading discourse, declaring a claimed immanent oppression of women within Islam and an alleged contradiction among feminists and the political left to be both in favor of gender equality and at the same time pro-Islam and positive toward Muslim immigration. This is also claimed to constitute some form of reverse discrimination, whereby Swedish (men) are criticized while Muslims are not. Furthermore, we have identified a tendency to use gender equality as a discursive strategy in order to criticize Islam and Muslims. In this sense, feminism seems to be used for the sake of the attack, rather than any deeper conviction. Finally, by analyzing the overlaps between the topics through engaged user, we further strengthen these results, showing that while feminism and Islam seem not to be explicitly and directly connected through users, they are indeed indirectly connected through topics focusing on sexual assault, veil and discrimination, and children and gender roles, serving as recontextualizing fields for the underlying ideologies.

Through this approach, we have also illustrated how topic modeling can serve as a powerful supplement to CDA, allowing the inductive exploration of large quantities of unstructured text data. The corpus analyzed here comprises in total about 50 million posts, with around 4.7 billion words, corresponding to about 10 million normal-spaced pages of text. Just reading this manually would take approximately 3–3.5 years of continuous reading. We have shown how topic modeling provides an alternative to this by rendering a valuable overview or map of the content, thus helping to find interesting discursive patterns that can be more closely investigated and analyzed using discourse analysis.

Such methodological synergy can be mutually beneficial and help in addressing some of the open issues in both fields by combining a valuable overview and structure, which are hard to extract with the naked eye, with sensitivity for linguistic nuances and implicit and symbolic meanings, which may not be visible for the automatic eye. This approach also harvests the strength of topic modeling by being close to what it was designed to do: exploring and categorizing large amounts of text rather than being used as a rigorous, stand-alone scientific method, as has so far often been the case in the social sciences.

In combination with CDA, topic modeling reduces the researcher bias and increases the credibility of the analysis, not least by reducing tendencies of ‘cherry-picking’ the data and guarding against over- and underinterpretations (Baker et al., 2008a; Stubbs, 1994). But it is also important not to get carried away by the bewildering surplus of accessible text and to avoid the temptation to ravenously devour data indiscriminately.
This approach should not, and cannot, replace careful consideration of what kinds of texts are necessary in order to answer one’s research questions and a critical awareness of what the corpus can actually be argued to represent. Furthermore, one should not overemphasize the automatic part of automated text analysis or downplay the effort involved in performing such analysis. It requires both time and effort to learn how to use these tools, but also persistent manual labor to make them work with the actual material, since data in these cases, as Baker (2006) puts it, often need to be ‘subtly massaged’ in order to produce the desired results (p. 179). This includes, for instance, various manual calibrations in order to produce coherent and interpretable topics. It is important not to conceal such subjective choices and to be aware of how this affects interpretation and analysis. Here, CDA can contribute with a more elaborated approach to hermeneutic interpretation processes and analytical methods for grasping and interpreting meanings in the selected texts, not least by making these processes more transparent. CDA also enables going beyond what is explicitly written and focusing on what could have been written but was not, or more subtle, coded strategies and symbolic meanings in the text.

In this particular study, we supplemented the analysis with meta-data of the users to produce a network with connections between topics through user activity, enabling us to connect and relate discourses to certain users. This is also an interesting aspect of Internet and social media data since it often includes relational data, thus enabling the study of how discourses and social groups are mutually constituted through interactions. This opens the possibility for future studies to investigate questions such as the following: ‘How do certain discourses form and diffuse within a network?’ ‘What influences do certain key actors and groups have on these processes?’ By focusing on the relation among various actors, we can also investigate how certain symbolically charged concepts and meaning constellations spread throughout and between social groups and the impact of what network analysts call network topologies. This indeed marks a potentially fruitful avenue for future research.

Finally, topic modeling has been criticized multiple times, on the one hand for not being rigorous enough to be used as a quantitative tool, and on the other for not being capable of moving beyond a mere descriptive exploration of the corpus when employed on its own. As we hope to have shown in this article, topic modeling becomes truly useful first when employed together with other methods that can grant both theoretical depth and more elaborated analytical techniques for interpretation. We think that such methodological synergy may contribute to transcending the boundaries between quantitative and qualitative approaches within text analysis.

Declaration of conflicting interests
The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.
Notes

1. Topics captured by Latent Dirichlet allocation (LDA) often vary greatly and directly depend on the type of text being analyzed. For instance, as Rhody (2012) has shown, topic models of more abstract and figurative text such as poetry tend to result in topics with less, or no, semantic coherence that can hardly be understood as ‘themes’.

2. For example, statistical tests for validation can be problematic and misleading in these cases since noisy data are often pushed into uninterpretable topics, which tend to increase the coherence of the other topics (DiMaggio et al., 2013).

3. We should note that we agree with the broadening of the concept of ‘social practice’ in Bennett (2013) to include not only the empirical sphere as the ‘real phenomena’ that are recontextualized, but also categories corresponding to structural or relational social forces such as class, race, and ideology.

4. ‘Arab’ is generally used as a synonym to ‘Muslim’ in the forum, and we therefore include it as a keyword in order to increase the size of the corpus. If ‘Arab’ had been used with a different meaning, this would likely have appeared as a separate topic in the model.

5. Gephi is an open-source network analysis and visualization software package written in Java. This application can be accessed free on: http://gephi.github.io/

6. Big Text Tool is accessible free on: http://www.toernberg.com

7. That is, all topics that are connected with other topics through shared users with a weight strength above a certain threshold.

8. However, we should note that this result is based on an LDA model of the corpus over all years. This means that there might be recent changes in the forum toward an increased connection between these topics. This, however, is not very likely since the corpus grows exponentially and significant changes in the last years would therefore most likely be reflected in the results.

References


Breivik A (2011) A European declaration of independence. Available at: https://info.publicintelligence.net/AndersBehringBreivikManifesto.pdf


**Author biographies**

**Anton Törnberg** is a PhD candidate in an interdisciplinary project between the Department of Sociology and Work Science at University of Gothenburg and the Division of Physical Resource Theory at Chalmers University of Technology. His research interests draw on Complexity Science and he is currently involved in a project focusing on combining approaches and methodologies from computer science and sociology, including various forms of automated text analysis and discourse analysis.

**Petter Törnberg** is a PhD candidate within the Complex Systems Group at Physical Resource Theory at Chalmers University of Technology. His work lays on the boundary between complexity science and the social sciences, focusing on developing new ways of understanding the evolution of societal systems. The basis for his work is the realization that innovation is usefully describable on the highly abstract level of ‘innovation in complex adaptive systems’, which implies theoretical and methodological integration between disciplines.
This thesis engages with questions on, and about, the boundary between the social and natural. The introductory essay observes a reinvigoration of naturalism in the increasing use of formal and computational methods for the study of social systems - in particular under the banner of Complexity Science. This tendency is put under scrutiny by revisiting the age-old philosophical question of naturalism and connecting ideas in complexity science with the work of e.g. Roy Bhaskar, Mario Bunge, William Wimsatt, and David Lane. This allows us to sketch a philosophical foundation for a complexity science of social systems, in the form of an integrative and methodologically pluralist complex realism. The five included papers exemplify and deepen this novel approach.